The fitness challenge of studying molecular adaptation.

David Coombes,1,2 James W.B. Moir,3 Anthony M. Poole,4 Tim F. Cooper,2,5 Renwick C.J. Dobson1,2,6\*

1 School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

2 Biomolecular Interaction Centre, University of Canterbury, Christchurch, New Zealand

3 Department of Biology, University of York, Heslington, York, YO10 5DD, United Kingdom

4 School of Biological Sciences, University of Auckland, Auckland, New Zealand

5 School of Natural and Computational Sciences, Massey University, Auckland, New Zealand

6 Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, VIC 3010, Australia.

\* To whom correspondence should be addressed: Prof. Renwick Dobson, Biomolecular Interaction Centre and School of Biological Sciences, University of Canterbury, Christchurch, New Zealand. Telephone (+64) 3 369 5145. E-mail: [renwick.dobson@canterbury.ac.nz](mailto:renwick.dobson@canterbury.ac.nz)

**Abstract**

Advances in bioinformatics and high-throughput genetic analysis increasingly allow us to predict the genetic basis of adaptive traits. These predictions can be tested and confirmed, but the molecular-level changes—*i.e.* the molecular adaptation—that link genetic differences to organism fitness remain generally unknown. In recent years, a series of studies have started to unpick the mechanisms of adaptation at the molecular level. In particular, this work has examined how changes in protein function, activity, and regulation cause improved organismal fitness. Key to addressing molecular adaptations is identifying systems and designing experiments that integrate changes in the genome, protein chemistry (molecular phenotype), and fitness. Knowledge of the molecular changes underpinning adaptations allow new insight into the constraints on, and repeatability of adaptations, and of the basis of non-additive interactions between adaptive mutations. Here we critically discuss a series of studies that examine the molecular-level adaptations that connect genetic changes and fitness.

(150 words)

**Introduction**

It is common to assume natural selection as the ultimate driver of molecular structure and function, but a causal relationship is difficult to demonstrate (1). Establishing which sequence changes are impactful and which have negligible effect is non-trivial and needs to encompass the entire organismal system (genotype, phenotype, fitness, environment) to be convincing (2). Indeed, high sequence-level variation between enzymes of identical function is a clear indication that many genetic changes have no phenotypic effect (3). Moreover, scepticism of a dominant role of natural selection in adaptive explanations of phenotypes (2, 4) is well-founded, particularly in species with small effective population sizes (5). Here random genetic drift speeds neutral processes that can lead to greater complexity, for example, biasing toward the evolution of protein complexes and division of labour through subfunctionalisation (where duplicate copies each perform part of the function that an ancestral non-duplicated gene product performed) (5-9).

Despite these challenges, modern genome sequencing and bioinformatics allow researchers to identify the specific genetic changes underlying adaptations. This is most easily done in experimentally evolved populations, where there are relatively few candidate genetic changes separating known ancestral and evolved populations, replicates aid in distinguishing selection from background variance (10), and complicating factors of migration and recombination are controlled (11). However, even when the genetic changes underlying adaptations are identified, establishing how those changes alter the molecular phenotypes (*i.e.* a molecular adaptation) and organismal phenotype, and how those changes affect fitness, is difficult to predict. Some studies of molecular adaptation have begun to untangle these problems by genetic manipulation (11, 12); proteins with and without the adaptation can be expressed and characterised to assess at least some aspects of molecular-level phenotypes (13), and isogenic clones with and without the adaptation can be competed to assess fitness (14). One way in which this kind of direct experimentation is important is to disentangle possible bases of selection when genetic changes have pleiotropic effects. For example, plants use colour pigments to attract pollinators for reproductive benefit, however the pigments can also have antimicrobial or antioxidant properties (15). Altering a biochemical pathway to produce similar compounds with only one property can in principle test the fitness effect of each phenotype. However, it is an open challenge whether individual properties can generally be separated. In the case of proteins, moonlighting functions (where a protein has more than one function) are thought to be commonplace (16) making it difficult to separately study the impact of individual functions on fitness.

In contrast to laboratory evolved populations, natural populations evolve in a complex dynamic environment with multiple selection pressures and no record of the outcompeted (ancestral) organism remains. This offers a major challenge when studying natural adaptations, as characterisation requires fitness testing in an appropriately modelled environment and comparison with non-adapted populations. This challenge has led to the common practice of comparing *in vitro* characteristics of adapted macromolecules (*e.g.* the Michaelis-Menten parameters) with those from distantly related extant species or ancestral reconstructions inferred using sequence analysis (17). These studies offer insight into how evolution optimises function and have value even when they do not contain direct fitness characterisations (18-21), but interpreting biophysical parameters in isolation can be misleading. For example, increasing enzyme catalytic function may be associated with a decrease in structural stability that may, in turn, be maladaptive (22), and cellular requirements to control metabolic flux mean that optimising a single enzyme is unlikely to affect most pathways (23, 24). In this context, Bar-Even *et al.* show that most enzymes are very far from the ‘ideal’ diffusion limited efficiency, and they suggest that optimisation only occurs when there is “a strong selective pressure, resulting from strong control over the metabolic flux and from the cost of producing many enzyme copies”. Therefore, it is difficult to link optimisation of a biophysical parameter (*e.g.* the Michaelis-Menten constant, or protein stability) to adaptation and caution should be used when interpreting results from these studies in the absence of comparative fitness data.

Although it is difficult to identify the molecular basis of adaptations, there is nevertheless good reason for optimism that we can make meaningful progress integrating genotype, molecular phenotype and fitness to understand the mechanisms of adaptation (2). This has been presented as a goal for a functional synthesis, the application of the tools of molecular biology to examine questions motivated by evolutionary biology (25). Below we present a series of example studies that contribute to this synthesis.

**Examining molecular-level adaptation through molecular characterisation and comparative fitness analysis**

Adaptive phenotypes provide an advantage over otherwise identical non-adapted organisms. This poses a challenge when characterising organismal adaptations, as organisms having the non-adapted phenotype (*e.g.*, an ancestor) are by definition at a fitness disadvantage, and are subject to being outcompeted (26) (though see below). In a molecular context, this can be overcome by comparative analysis of proteins from a wide range of organisms (27). In this section, we consider several studies that synthesise comparative approaches that identify candidate adaptive mutations with biochemical approaches that examine the function of those mutations. The studies focus on proposed molecular-level adaptive changes in haemoglobin, a protein used for the uptake and distribution of oxygen in most vertebrates (28).

Perutz hypothesised that bar-headed geese are adapted to high altitude through a substitution in haemoglobin (Pro119ɑ->Ala) that moves the equilibrium from a low O2 affinity T-state toward a high O2 affinity R-state (29). He reasoned that this substitution would cause a loss of van der Waals interactions between ɑ and β subunits and lower the energy required to transition into the high affinity R-state (30). In support of the Pro119ɑ->Ala substitution having this effect, exchanging alanine for proline in human haemoglobin increased oxygen affinity (P50), from 6.14 mm Hg (Pro119, low affinity) to 4.17 mm Hg (Ala119, high affinity) (measured under conditions typical of respiratory intake). This affinity change was associated with a ten-fold shift in the T- to R-state equilibrium—that is, there is a preference for the R-state (31). These results were consistent with the prediction of Perutz’s hypothesis and seemed to confirm a link between the Pro119ɑ->Ala substitution and high-altitude fitness, especially given that that substitution is one of just four separating bar-headed geese haemoglobin from lowland greylag geese haemoglobin (32, 33). However, a more recent study using geese haemoglobin isolated directly from blood has contradicted this link (34). It notes there is only a small increase in oxygen affinity (P50), changing from 4.24 mm Hg (greylag geese) to 3.72 mm Hg (bar-headed geese), which is consistent with a recent recombinant analysis (35), but more importantly reports there is no difference in T/R-state equilibrium between the species (34). Therefore, Ala119 does not appear to lower the energy barrier to enter the high affinity R-state in bar-headed geese and the hypothesised mechanism for high-altitude O2 uptake, proposed by Perutz, is not supported.

A further example underlining the difficulty of mapping putative genetic changes to adaptive molecular phenotypes is work on deer mice haemoglobin. Here, multiple mutations act non-additively to cause an adaptive phenotypic effect, a phenomenon known as epistasis (36). Different haemoglobin variants for ɑ-exon 2, ɑ-exon 3, and the β-exon exist in deer mice sub-populations depending on whether they live in a high-altitude (H) or low-altitude (L) habitat (37). The high-altitude variant combination (HH-H, where the first two letters indicate ɑ-exon variants and the third the β-exon variant) confers significantly higher haemoglobin efficiency than does the low-altitude combination (LL-L). Studies recombining exon variants show that exchanging any one lowland exon variant for a highland exon variant (i.e., HL-L, LH-L, or LL-H,) actually decreases O2 haemoglobin efficiency relative to the lowland wild-type (LL-L) (Fig. 1A) (36). This finding indicates that multiple epistatic interactions are required to increase haemoglobin efficiency (36). Crystal structures of haemoglobin with different variant combinations give a potential basis for these interactions. The lowland and highland exon variants have different hydrogen-bonding patterns to ɑ-helix 2 from the ɑ-subunit (residues 19-35) (Fig. 1B-D). Highland haemoglobin has one hydrogen bond from ɑHis113 (ɑ-exon 3) to ɑTyr24 on ɑ-helix 2 (shown in blue Fig. 1B). In contrast, lowland haemoglobin has two hydrogen-bonds, ɑHis50 (ɑ-exon 2) to ɑGlu30 on ɑ-helix 2, and βSer128 (β-exon) to ɑCys34 on ɑ-helix 2, which are not present in highland haemoglobin (shown in red Fig. 1C) (36). An exon substitution to the lowland wild-type will only add or remove one bond. For example, LH-L adds the bond from the highland variant but leaves the two lowland bonds in place (shown in Fig. 1D), which presumably increases molecular rigidity and the unfavourable dynamics provide less efficiency.

[insert Figure 1]

The two study systems discussed above highlight the difficulties in defining the connection between molecular-level phenotypes and fitness. In the first study, the differences between the haemoglobin sequence from the bar-headed (ɑAla119) and greylag geese (ɑPro119) lead to only small increases in O2 affinity, despite the same change causing a much greater effect when introduced into human haemoglobin. The demonstrates the pitfalls of interpreting adaptive features from an inferred ancestral model. The second haemoglobin reveals the importance of epistasis in determining the effect of a small number of molecular changes, such that potentially adaptive effects only occur in the context of a combination of multiple changes (38, 39). We note that a key limitation to both studies is that phenotypes are assessed at the level of O2 affinity, and the relationship of affinity to fitness is *assumed* from species presence at high or low altitude locations.

**Mapping enzyme phenotype to fitness**

A focus of functional synthetic studies has been to map the effect of mutations in enzymes first to functional phenotypes and then to organism fitness. These studies allow researchers to identify key mutations and functions underlying adaptation, and to identify functional responses that trade-off with one another, perhaps identifying alternative evolutionary possibilities. By considering genotype-phenotype maps in the context of the cell environment, it can also be possible to predict which evolutionary outcomes are more likely to be followed. We present below a series of studies examining dehydrogenase coenzyme preference.

Dehydrogenases can be separated into two evolutionary groups: those that use the Rossman fold and those that use the α/β fold (40, 41). Nucleotide binding is not separated by this grouping and members of either group can be specific for either NAD+ or NADP+ (42, 43). Both β-isopropylmalate dehydrogenase and isocitrate dehydrogenase are α/β fold dehydrogenases (44). Whereas β-isopropylmalate dehydrogenases only reduce NAD+ and are employed in the leucine biosynthesis pathway (45), isocitrate dehydrogenases reduce either NAD+ or NADP+, dependent on the organism of origin, and are employed in the citric acid cycle (46). Comparative studies between the NAD+ binding *Thermus thermophilus* β-isopropylmalate dehydrogenase and NADP+ binding *Escherichia coli* isocitrate dehydrogenase predict that six amino acids are responsible for differentiating between the nucleotides. This prediction was tested by constructing *T. thermophilus* β-isopropylmalate dehydrogenase variants that had different combinations of the six amino acids that specify nucleotide preference, characterising their molecular function and then determining fitness to produce a map showing how the amino acid variants relate to fitness benefit and nucleotide preference (44).

The *T. thermophilus* β-isopropylmalate dehydrogenase variants were tested *in vitro* for nucleotide specificity (*k*cat/*K*M) and were expressed in *E. coli* for fitness measurements (44). In general, amino acid combinations that allowed the use of either NAD+ or NADP+ (*i.e.* promiscuous enzymes) had a low substrate specificity (*k*cat/*K*M) and conferred low fitness (Fig. 2A). There were no variants with high enzyme specificity for both NAD+ and NADP+, but a small number of variants had high specificity for either NAD+ or NADP+ (44). In both cases, high specificity correlated with high fitness, but there was a general trend for higher fitness with NAD+, which matches the evolved wild-type preference of β-isopropylmalate dehydrogenases (47) (Fig. 2B). As the enzyme was expressed in a different organism for fitness testing, interaction with co-evolved macromolecules will not be a factor and intermolecular epistasis can be ruled out. Fitness benefit with NAD+ is expected to be linked by differences in cytosolic ratios of nucleotides, where NAD+ to NADH ratio is generally higher than NADP+ to NADPH ratio. This means that (when other factors are equal) cellular nucleotide concentrations are more thermodynamically favourable to complete the forward reaction with NAD+ and highlights that use of the substrate in β-isopropylmalate dehydrogenases is an adaptation (47).

[insert Figure 2]

The functional synthesis emphasizes the value of combining genetic and biochemical characterisation of alleles and correlating these data with fitness. In the β-isopropylmalate dehydrogenase experiment outlined above, the fitness effect of different variants involved in determining NAD+ and NADP+ preference can easily be determined. However, it is only when this information is paired with measurements of the specificity and activity of resulting enzymes, and of cellular substrate biochemistry, that mechanistic hypotheses connecting genotype and fitness can be made.

**Evolution experiments map mutations to their fitness effects**

Laboratory evolution experiments observe evolution as it occurs in a controlled environment. They usually begin with a single defined genotype and control processes such as migration and recombination that can complicate subsequent analysis of the genetic changes thought to underlie adaptation. Specific mutations can be added to or reverted from reference strains to test for their effect on fitness. Although such laboratory-based evolution experiments cannot directly investigate adaptation in the natural environment, they are nevertheless a useful proxy. Here, we will discuss three laboratory and one natural evolution experiments that synthesize genetic and fitness information, especially applied to implications for characterising the range of potential evolutionary outcomes available to a population in a given environment.

*Bacteriophage ΦX174*

Adaptive evolution can take a long time as it requires a population to acquire beneficial mutations and for selection to occur. A key advantage of studying adaptation in viral systems is their very short generation time and potentially large population sizes, which allow mutations to quickly arise and be selected. One bacteriophage that has been used in several evolutionary studies is ΦX174, a relatively simple phage with a genome of 5,386 bp encoding 11 proteins (48). This small size allows the genome to be easily sequenced and evolved changes to be tracked.

In an experiment by Wichman *et al.* (1999), two independent populations were continually supplied with a non-typical host (*Salmonella typhimurium*)and propagated at a novel temperature (43 °C), providing a strong selection pressure. The phage infection rate was tested each day. The first population had a mutation on day one in base 2167, corresponding to gene F (the major capsid protein), which increased in frequency throughout the population. This mutation conferred an approximately 4,000-fold increase in fitness, but was not found in the second population. Instead the second population had a large jump in fitness at day 4, after seven different mutations had occurred. These mutations were all different from the mutations in the first population, demonstrating that adaptation can occur through different trajectories (49). Further analysis indicates that most mutations occurring in the experiment caused an increase in fitness, suggesting that in this high-selection system, adaptive changes dominate evolution (49).

*Escherichia coli long-term evolution experiment*

Starting in 1988, the ongoing *E. coli* long-term evolution experiment (LTEE) follows 12 populations (started from a single clone) that has been propagated for 70,000 generations in a defined minimal medium supplemented with glucose (50). This environment selects for adaptations that increase nutrient acquisition and growth, both directly and indirectly through a decrease in resources spent on an unused capacity to respond to alternative environments. Parallel evolution, a signature of adaptation, has been uncovered in several genes (51), including: 1) *pykF* (pyruvate kinase), which regulates the glycolysis pathway (52); 2) *mrdA* (peptidoglycan D,D-transpeptidase), which is involved in cell wall synthesis (53); 3) *topA* (topoisomerase), which is used in DNA replication (54); and 4) *spoT* ((p)ppGpp synthase/hydrolase), which regulates the stringent response, reducing protein synthesis during environmental stress, including starvation (55, 56). Overall, the populations were ~70% fitter (determined through direct competition with the ancestor) after 50,000 generations and whilst not as prevalent as first thought (57), decreasing expression of proteins used for alternative environments was an important adaptive strategy for success in a simplified environment (58).

Nine distinct mutations in *pykF*, which encodes the enzyme pyruvate kinase, occurred early during the *LTEE* and were studied to determine whether interactions with later mutations within the *E. coli* genome affect fitness (14). In the first experiment, each mutation was added in isolation to the ancestral strain, which resulted in a fitness benefit of approximately 10% for non-synonymous mutations and 5% for a deletion. This demonstrates that each mutation is adaptive and confers an effect distinct from a simple loss of function (14). To investigate whether early *pykF* mutations worked in conjunction with later mutations occurring in other genes, evolved *pykF* alleles present in clones evolved for 20,000 generations were reverted to the ancestral allele. The difference in fitness benefit from this change ranged from 0-25%. In six cases, the evolved *pykF* allele was more beneficial in its own genetic background, indicating that subsequent beneficial mutations depended on the earlier *pykF* mutation for their fitness effect. In two cases, evolved *pykF* alleles were less beneficial in their evolved backgrounds, indicating that subsequent mutations had, in part, made redundant whatever physiological changes were caused by the pyruvate kinase substitutions. These experiments demonstrate that early adaptations can significantly shape the trajectory of future adaptations (14); however, the mechanism behind the fitness effect, *i.e.* the physiological change, is still to be determined.

Two of the most exciting findings made as part of the LTEE involve the evolution of complexity (59). In one case, a population evolved two separate cohabiting sub-populations. This happened when a cell lineage evolved to optimise glucose usage, but simultaneously lost the ability to metabolise a waste product, acetate (38). Glucose optimisation conferred an advantage following population transfer to fresh glucose replete media, but once glucose was depleted through cell growth, the population that could now use acetate flourished (60). Diversification of a monoclonal population into diversified sub-populations is an event observed in other evolution experiments. For example, Rosenzweig *et al.* (61) observed that *E. coli* populations differentiated into those that use glycerol, and those that use acetate. Experiments by Rainey & Travisano (62) demonstrate that *Pseudomonas fluorescens* quickly develops morphically diversified mutant populations, but only when the environment is not spatially resisted. The second example of complexity evolution in the LTEE, is the use citrate as a nutrient source. In the media, citrate is used as a chelating agent as *E. coli* doesn’t normally import citrate or use it as a carbon source in the presence of oxygen. Genome sequencing revealed that a key event in the evolution of citrate use was a tandem duplication of the citrate transport gene, which then diverged in regulatory mechanism to allow import of citrate in the presence of oxygen (59). The innovation was only noted in one of twelve systems in the LTEE but has been found in another evolution experiment without tandem duplication (63) indicating that adaptation to use citrate as a nutrient source is possible through different mechanisms. Together the above examples demonstrate how an organism can evolve complexity when exposed to underexploited niches in an environment, a situation that is typical of natural environments, where a diversity of compounds and neighbouring habitats exist.

*Modelling antibiotic resistance*

Laboratory-based evolution experiments are limited to investigating adaptations that occur within the system, however they can recreate some aspects of natural systems where adaptations of interest occur. In this final example, adaptation of antibiotic resistance was investigated by experiments using hypermutator cells that are prevalent in clinical settings (64).

Hypermutators have high mutation rates due to compromised DNA replication fidelity and/or repair mechanisms. High mutation rates are generally disadvantageous (65), but are often present in strains isolated from clinical settings (66). A possible explanation is that increased mutation rates lead to adaptive antibiotic resistance that provide an overall increase in fitness in the presence of antibiotics (65). For example, cefotaxime resistance can result from mutations in the β-lactamase TEM-1 gene (64) that involve G>A and C>T transitions, which are a signature of hypermutators deficient in the mismatch repair system protein (MutH) (67). Evolution experiments were undertaken in an environment that had increasing cefotaxime concentrations and started with wild-type, Δ*mut*T*,* or Δ*mut*H *E. coli* strains. Wild-type *E. coli* populations did not become resistant to high-levels of cefotaxime. The Δ*mut*T strains gained high-level resistance, but unlike clinical isolates this occurred through adaptation of penicillin binding protein (PBP3) (64). The Δ*mut*H *E. coli* adapted through G>A and C>T transition of the TEM-1 gene as predicted (64). Therefore, the evolution experiment recreates the adaptation of antibiotic resistance in a closed laboratory and links the results to environmental fitness in the natural (clinical) setting.

*Pre-adaptation to insecticide in a multicellular eukaryote*

Hartley *et al.* (68) examined the evolution of blowfly resistance to insecticides. They analysed genomic data from samples obtained in the field during a period when resistance became common, and historical pinned specimens collected before insecticide use. Their efforts provide a direct characterisation of the process of resistance adaptation that occurred in the natural environment.

The rapid adaptation of insecticide resistance in the blowfly led to the hypothesis that a preadapted allele variant is present in general populations. Australian blowfly have been observed to have two possible variations of esterase 3 (an enzyme in lipid metabolism) that provide organophosphate resistance (68). Leu251 variants (or other variants with a medium to small sidechain amino acid at position 251) provide malathion resistance (69) and function by repositioning His471 to prevent dealkylation of an active site serine residue (70). Investigating the preadaptation hypothesis, Leu/Ser/Thr251 variants were found in areas where selection pressure is low and in the pinned specimens predating insecticide use (71). This supports that blowfly are preadapted to malathion and provides a possible explanation for high-levels of worldwide malathion resistance, even in areas without widespread use. The esterase 3 Asp137 variation provides resistance to diazinon, where the aspartate is used to deprotonate water for nucleophilic attack on insecticides, changing the overall function of the enzyme to an organophosphate hydrolase (71). The investigation found that Asp137 variants were not found in areas of low insecticide selection pressure or the pinned-specimens, suggesting that blowfly are not preadapted for diazinon and potentially explains why resistance is less well spread. This observation is not surprising given that diazinon resistance is maladaptive due to increased winter mortality (62), which is potentially due to the loss of native lipid esterase activity in Asp137 variants (71).

In summary, evolution experiments provide a method to observe populations as they adapt to an environment, allowing the adaptive molecular features to be characterised and compared with the ancestral non-adapted features. Fitness is then measured within the experiment through competition and adaptive features also can be studied using isogenic clones for direct comparison. Combined, molecular characterisation alongside evolution experiment appear to provide the perfect method to analyse adaptation. However, as environment is an important part of fitness laboratory-based experiments may provide an inadequate model and nature-based experiments may have additional selection pressures that are more important to the characterised adaptation. Therefore, care still needs be taken to ensure that the investigated biological questions do not exceed the limitations of the system.

**Conclusion and perspectives**

*Importance to the field.* Adaptation shapes populations over time through the natural selection of individuals with beneficial features that increase fitness. Molecular adaptation studies provide empirical evidence to understand why organismal features exist, to map the evolutionary trajectories that shape populations, and to help predict future adaptations.

*Current thinking.* Characterising the fitness benefit of a given molecular adaptation (*e.g.* a substitution in a protein) is a major challenge when studying natural selection, as it requires the generational fitness to be measured in a replica of the environment. This review critically discusses the efforts to define fitness in molecular adaptation studies; these include mapping fitness landscapes with recombinant organisms, observing adaptations generated within a laboratory evolution experiment, modelling natural evolution in laboratory experiments, and following the adaptation of a species in nature from historic specimens.

*Future directions.* Over time adaptation studies have established the foundations to examine increasingly complex systems. Moving forward this trend will continue and experiments using larger organisms under dynamic selection pressures (*e.g.*, mice in a field (73)) will explore adaptation in more complex natural settings and investigate wholesale changes in metabolic pathways or macromolecular interactions at the organ level.

**References**

1. Allison JR, Lechner M, Hoeppner MP, Poole AM. Positive selection or free to vary? Assessing the functional significance of sequence change using molecular dynamics. PLoS One. 2016;11(2):e0147619.

2. Yi X, Dean AM. Adaptive landscapes in the age of synthetic biology. Molecular Biology and Evolution. 2019;36(5):890-907.

3. McShea DW, Brandon RN. Biology's first law: the tendency for diversity and complexity to increase in evolutionary systems: University of Chicago Press; 2010.

4. Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proceedings of the Royal Society B: Biological Sciences. 1979;205(1161):581-98.

5. Lynch M. The frailty of adaptive hypotheses for the origins of organismal complexity. Proceedings of the National Academy of Sciences. 2007;104 Suppl 1:8597-604.

6. Gray MW, Lukes J, Archibald JM, Keeling PJ, Doolittle WF. Cell biology. Irremediable complexity? Science. 2010;330(6006):920-1.

7. Stoltzfus A. On the possibility of constructive neutral evolution. Journal of Molecular Evolution. 1999;49(2):169-81.

8. Covello PS, Gray MW. On the evolution of RNA editing. Trends in Genetics. 1993;9(8):265-8.

9. Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW. Evolution of increased complexity in a molecular machine. Nature. 2012;481(7381):360.

10. Barrick JE, Lenski RE. Genome dynamics during experimental evolution. Nature Reviews Genetics. 2013;14(12):827-39.

11. Elena SF, Lenski RE. Microbial genetics: evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature Reviews Genetics. 2003;4(6):457.

12. Nowoshilow S, Schloissnig S, Fei J-F, Dahl A, Pang AWC, Pippel M, et al. The axolotl genome and the evolution of key tissue formation regulators. Nature. 2018;554:50.

13. Walkiewicz K, Benitez Cardenas AS, Sun C, Bacorn C, Saxer G, Shamoo Y. Small changes in enzyme function can lead to surprisingly large fitness effects during adaptive evolution of antibiotic resistance. Proceedings of the National Academy of Sciences. 2012;109(52):21408-13.

14. Peng F, Widmann S, Wunsche A, Duan K, Donovan KA, Dobson RCJ, et al. Effects of beneficial mutations in *pyk*F gene vary over time and across replicate populations in a long-term experiment with bacteria. Molecular Biology and Evolution. 2018;35(1):202-10.

15. Miller R, Owens SJ, Rørslett B. Plants and colour: Flowers and pollination. Optics & Laser Technology. 2011;43(2):282-94.

16. Copley SD. Moonlighting is mainstream: paradigm adjustment required. BioEssays. 2012;34(7):578-88.

17. Wilding M, Peat TS, Kalyaanamoorthy S, Newman J, Scott C, Jermiin LS. Reverse engineering: transaminase biocatalyst development using ancestral sequence reconstruction. Green Chemistry. 2017;19(22):5375-80.

18. Jackson CJ, Foo JL, Tokuriki N, Afriat L, Carr PD, Kim HK, et al. Conformational sampling, catalysis, and evolution of the bacterial phosphotriesterase. Proceedings of the National Academy of Sciences. 2009;106(51):21631-6.

19. Kaltenbach M, Burke JR, Dindo M, Pabis A, Munsberg FS, Rabin A, et al. Evolution of chalcone isomerase from a noncatalytic ancestor. Nature Chemical Biology. 2018;14(6):548.

20. Todd AE, Orengo CA, Thornton JM. Evolution of function in protein superfamilies, from a structural perspective. Journal of Molecular Biology. 2001;307(4):1113-43.

21. Wu Z, Kan SBJ, Lewis RD, Wittmann BJ, Arnold FH. Machine learning-assisted directed protein evolution with combinatorial libraries. Proceedings of the National Academy of Sciences. 2019;116(18):8852-8.

22. Wang X, Minasov G, Shoichet BK. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. Journal of Molecular Biology. 2002;320(1):85-95.

23. Fell DA, Thomas S. Physiological control of metabolic flux: the requirement for multisite modulation. Biochemical Journal. 1995;311(1):35-9.

24. Fell DA. Increasing the flux in metabolic pathways: A metabolic control analysis perspective. Biotechnology and Bioengineering. 1998;58(2‐3):121-4.

25. Dean AM, Thornton JW. Mechanistic approaches to the study of evolution: the functional synthesis. Nature Reviews Genetics. 2007;8(9):675.

26. Wiens JJ. Incomplete taxa, incomplete characters, and phylogenetic accuracy: Is there a missing data problem? Journal of Vertebrate Paleontology. 2003;23(2):297-310.

27. Joy JB, Liang RH, McCloskey RM, Nguyen T, Poon AFY. Ancestral reconstruction. PLoS Computational Biology. 2016;12(7):e1004763.

28. Wells RM. Evolution of haemoglobin function: Molecular adaptation to environment. Clinical and Experimental Pharmacology and Physiology. 1999;26(8):591-5.

29. Perutz MF. Species adaptation in a protein molecule. Molecular Biology and Evolution. 1983;1(1):1-28.

30. Fermi G, Perutz MF. Haemoglobin and myoglobin: Clarendon Press; 1981.

31. Weber RE, Jessen TH, Malte H, Tame J. Mutant hemoglobins (alpha 119-Ala and beta 55-Ser): functions related to high-altitude respiration in geese. Journal of Applied Physiology. 1993;75(6):2646-55.

32. Kleinschmidt T, Sgouros JG. Hemoglobin sequences. Biological Chemistry Hoppe-Seyler. 1987;368(6):579-615.

33. Oberthür W, Braunitzer G, Würdinger I. Hämoglobine, XLVII. Das hämoglobin der atreifengans (*Anser indicus*). Primärstruktur und physiologie der atmung, systematik und evolution. Hoppe-Seyler´s Zeitschrift für Physiologische Chemie. 1982;363(1):581-90.

34. Jendroszek A, Malte H, Overgaard CB, Beedholm K, Natarajan C, Weber RE, et al. Allosteric mechanisms underlying the adaptive increase in hemoglobin-oxygen affinity of the bar-headed goose. The Journal of Experimental Biology. 2018;221(Pt 18).

35. Natarajan C, Jendroszek A, Kumar A, Weber RE, Tame JRH, Fago A, et al. Molecular basis of hemoglobin adaptation in the high-flying bar-headed goose. PLoS Genetics. 2018;14(4):e1007331-e.

36. Natarajan C, Inoguchi N, Weber RE, Fago A, Moriyama H, Storz JF. Epistasis among adaptive mutations in deer mouse hemoglobin. Science. 2013;340(6138):1324-7.

37. Storz JF, Kelly JK. Effects of spatially varying selection on nucleotide diversity and linkage disequilibrium: insights from deer mouse globin genes. Genetics. 2008;180(1):367-79.

38. Plucain J, Hindré T, Le Gac M, Tenaillon O, Cruveiller S, Médigue C, et al. Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. Science. 2014;343(6177):1366-9.

39. Knies JL, Cai F, Weinreich DM. Enzyme efficiency but not thermostability drives cefotaxime resistance evolution in TEM-1 β-lactamase. Molecular Biology and Evolution. 2017;34(5):1040-54.

40. Rossmann MG, Moras D, Olsen KW. Chemical and biological evolution of a nucleotide-binding protein. Nature. 1974;250(5463):194-9.

41. Hurley JH, Dean AM, Koshland DE, Stroud RM. Catalytic mechanism of NADP+-dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and NADP+ complexes. Biochemistry. 1991;30(35):8671-8.

42. Carothers DJ, Pons G, Patel MS. Dihydrolipoamide dehydrogenase: Functional similarities and divergent evolution of the pyridine nucleotide-disulfide oxidoreductases. Archives of Biochemistry and Biophysics. 1989;268(2):409-25.

43. Zhang L, Ahvazi B, Szittner R, Vrielink A, Meighen E. Change of nucleotide specificity and enhancement of catalytic efficiency in single point mutants of *Vibrio harveyi* aldehyde dehydrogenase. Biochemistry. 1999;38(35):11440-7.

44. Chen R, Greer A, Dean AM. Redesigning secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase. Proceedings of the National Academy of Sciences. 1996;93(22):12171-6.

45. He Y, Mawhinney TP, Preuss ML, Schroeder AC, Chen B, Abraham L, et al. A redox-active isopropylmalate dehydrogenase functions in the biosynthesis of glucosinolates and leucine in *Arabidopsis*. The Plant Journal. 2009;60(4):679-90.

46. Akram M. Citric acid cycle and role of its intermediates in metabolism. Cell Biochemistry and Biophysics. 2014;68(3):475-8.

47. Miller SP, Lunzer M, Dean AM. Direct demonstration of an adaptive constraint. Science. 2006;314(5798):458-61.

48. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, et al. Nucleotide sequence of bacteriophage φX174 DNA. Nature. 1977;265(5596):687-95.

49. Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ. Different trajectories of parallel evolution during viral adaptation. Science. 1999;285(5426):422-4.

50. 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(6):1315-41.

51. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, et al. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature. 2009;461(7268):1243-7.

52. Donovan KA, Zhu S, Liuni P, Peng F, Kessans SA, Wilson DJ, et al. Conformational dynamics and allostery in pyruvate kinase. Journal of Biological Chemistry. 2016;291(17):9244-56.

53. Begg KJ, Donachie WD. Division planes alternate in spherical cells of *Escherichia coli*. Journal of Bacteriology. 1998;180(9):2564.

54. Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. Genetics. 2005;169(2):523-32.

55. Cooper TF, Rozen DE, Lenski RE. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proceedings of the National Academy of Sciences. 2003;100(3):1072-7.

56. Traxler MF, Summers SM, Nguyen H-T, Zacharia VM, Hightower GA, Smith JT, et al. The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. Molecular Microbiology. 2008;68(5):1128-48.

57. Leiby N, Marx CJ. Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. PLoS Biology. 2014;12(2):e1001789.

58. Maddamsetti R, Hatcher PJ, Green AG, Williams BL, Marks DS, Lenski RE. Core genes evolve rapidly in the long-term evolution experiment with *Escherichia coli*. Genome Biology and Evolution. 2017;9(4):1072-83.

59. Lenski RE. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. The ISME Journal. 2017;11(10):2181-94.

60. Rozen DE, Lenski RE. Long-term experimental evolution in *Escherichia coli.* VIII. Dynamics of a balanced polymorphism. The American Naturalist. 2000;155(1):24-35.

61. Rosenzweig RF, Sharp RR, Treves DS, Adams J. Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. Genetics. 1994;137(4):903-17.

62. McKenzie JA, Clarke GM. Diazinon resistance, fluctuating asymmetry and fitness in the Australian sheep blowfly, *Lucilia cuprina*. Genetics. 1988;120(1):213-20.

63. Hall BG. Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. Journal of Bacteriology. 1982;151(1):269-73.

64. Couce A, Rodríguez-Rojas A, Blázquez J. Bypass of genetic constraints during mutator evolution to antibiotic resistance. Proceedings of the Royal Society B: Biological Sciences. 2015;282(1804):20142698.

65. Couce A, Caudwell LV, Feinauer C, Hindré T, Feugeas J-P, Weigt M, et al. Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. Proceedings of the National Academy of Sciences. 2017;114(43):E9026-E35.

66. Komp Lindgren P, Higgins PG, Seifert H, Cars O. Prevalence of hypermutators among clinical *Acinetobacter baumannii* isolates. Journal of Antimicrobial Chemotherapy. 2015;71(3):661-5.

67. Denamur E, Matic I. Evolution of mutation rates in bacteria. Molecular Microbiology. 2006;60(4):820-7.

68. Hartley CJ, Newcomb RD, Russell RJ, Yong CG, Stevens JR, Yeates DK, et al. Amplification of DNA from preserved specimens shows blowflies were preadapted for the rapid evolution of insecticide resistance. Proceedings of the National Academy of Sciences. 2006;103(23):8757-62.

69. Heidari R, Devonshire AL, Campbell BE, Bell KL, Dorrian SJ, Oakeshott JG, et al. Hydrolysis of organophosphorus insecticides by in vitro modified carboxylesterase E3 from *Lucilia cuprina*. Insect Biochemistry and Molecular Biology. 2004;34(4):353-63.

70. Jackson CJ, Liu JW, Carr PD, Younus F, Coppin C, Meirelles T, et al. Structure and function of an insect alpha-carboxylesterase (alphaEsterase7) associated with insecticide resistance. Proceedings of the National Academy of Sciences. 2013;110(25):10177-82.

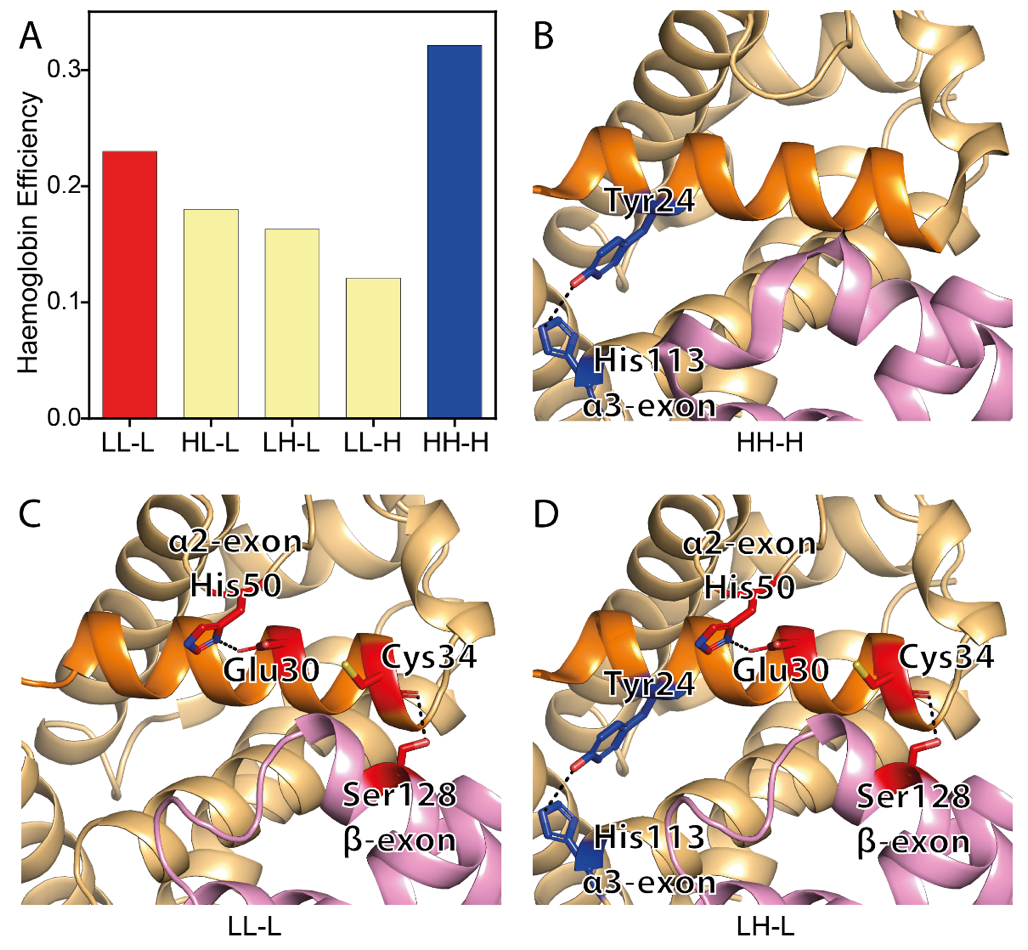
71. Newcomb RD, Campbell PM, Ollis DL, Cheah E, Russell RJ, Oakeshott JG. A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. Proceedings of the National Academy of Sciences. 1997;94(14):7464-8.

72. Dobzhansky T. Nothing in biology makes sense except in the light of evolution. The American Biology Teacher. 1973;35(3):125-29.

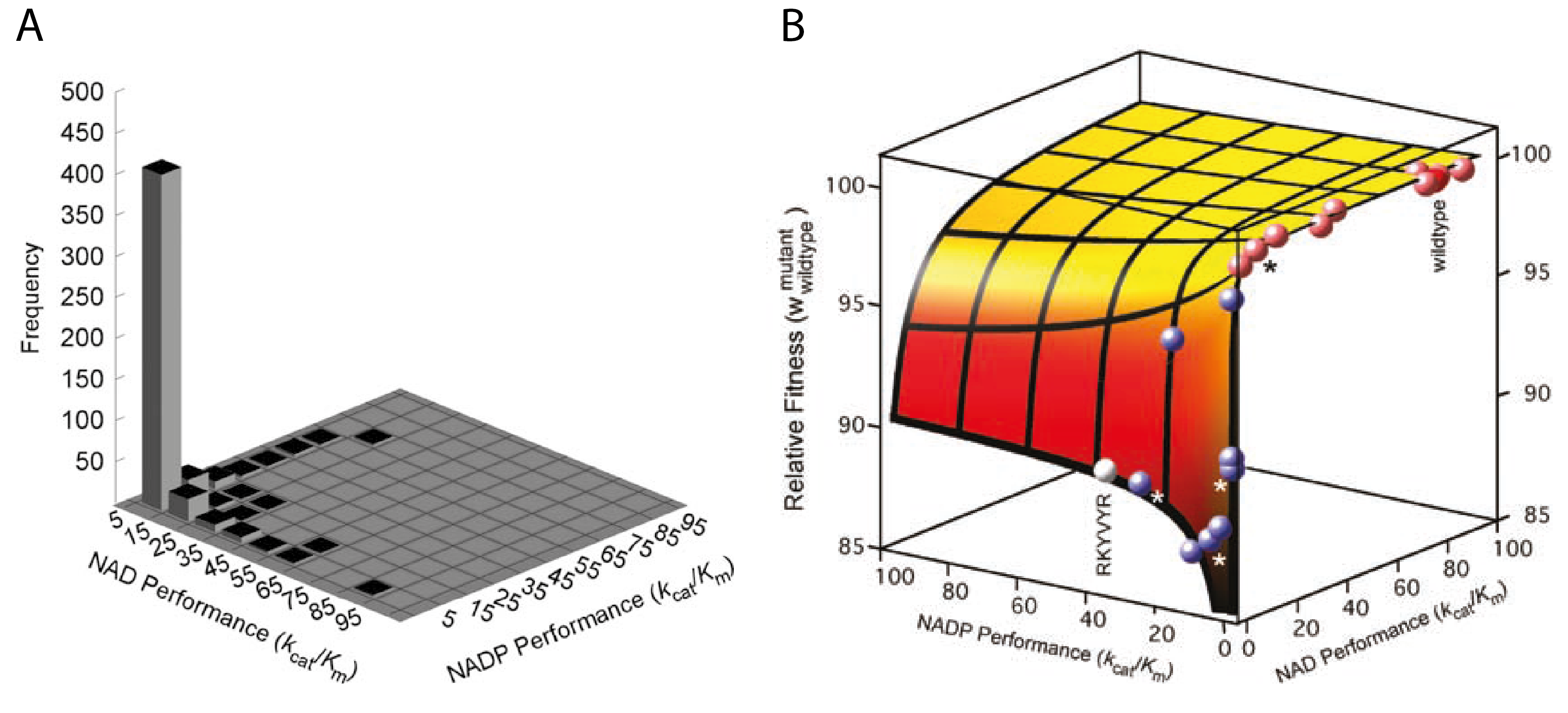
73. Barrett RDH, Laurent S, Mallarino R, Pfeifer SP, Xu CCY, Foll M, et al. Linking a mutation to survival in wild mice. Science. 2019;363(6426):499.

74. Cabrera A, Program CNIS, Golding G, Program CNIS, Campbell J, Program CNIS, et al. Characterization of clinical Methicillin-Resistant *Staphylococcus aureus* (MRSA) isolates from Canadian hospitals, 2010–2015. Open Forum Infectious Diseases. 2016;3(suppl\_1):1746.

**Figure 1.** Epistatic interaction between haemoglobin subunit variants from deer mouse (36). **A)** Highland (H) and lowland (L) variants are present for ɑ exon 2, exon 3, and the β subunit. Combination of possible variations constructed by mutagenesis and tested for haemoglobin efficiency, which is the difference between Log P50 value in conditions representing O2 uptake (stripped) and O2 offloading (with diphosphoglycerate). Interestingly, the normal lowland variant (red bar) had higher efficiency than combinations with a single highland variant (yellow bars) demonstrating that adaptation can occur through multiple non-additive mutation (epistasis). **B-D)** Hydrogen bonding differences between haemoglobin variants at helix 2 of the α-subunit (orange). **B)** The structure of highland variant haemoglobin (PDB: 5KER) has a hydrogen bond at the N-terminal of the helix (blue) from a difference in the α3 exon. **C)** The structure of lowland variant haemoglobin (PDB: 4H2L) has two hydrogen bonds near the C-terminal end of the helix (red) from differences in the α2 and β exons. **D)** Composite structure depicting lowland haemoglobin changed to have the α3 highland unit. This adds the highland hydrogen bond, but as the lowland bonds remain the arrangement has less efficiency than the original lowland variant.



**Figure 2.** Enzyme specificity and fitness of β-isopropylmalate dehydrogenase variants. A) Most variants of β-isopropylmalate dehydrogenase had low enzyme specificity (*k*cat/*K*M) meaning that these variants were poor enzymes with either substrate, which correlates with fitness, demonstrating that most non-synonymous mutations are maladaptive. Few variants had high specificity with NAD or NADP, but none had high-specificity with both substrates. B) In the experiment with high β-isopropylmalate dehydrogenase expression, small increases in NAD and NADP specificity caused large fitness increases, demonstrating the non-linearity involved in connecting molecular phenotypes to fitness. However, the importance of high substrate specificity to fitness is indicated from experiments at lower expression, when glucose is in excess and growth is high (47). Overall, the experimental trend indicates that NAD use for β-isopropylmalate dehydrogenase is preferable, as the product NADH is at lower levels and the reaction is more favourable. Figures adapted from “The Biochemical Architecture of an Ancient Adaptive Landscape” by Miller *et al.* (47), *Science* Vol. 310, Issue 5747, pp. 499-501. Reprinted with permission from American Association for the Advancement of Science (AAAS), copyright © 2013.

****