Finding General Patterns in Fitness Landscapes

by

Daniel Moisés Lyons

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2020

Doctoral Committee:

Associate Professor Adam Lauring, Co-Chair
Professor Jianzhi Zhang, Co-Chair
Professor Alex Kondrashov
Professor Patricia Wittkopp
Assistant Professor Robert Woods

Daniel Moisés Lyons

lyonsdm@umich.edu

ORCID iD: 0000-0002-5241-9335

Dedicación

Para Mami Dina y Papa Manuel. Me dieron la fuerza para estudiar.

Acknowledgements

Thank you to Adam Lauring and George Zhang for being exceptional mentors. Thank you Adam for the care you take in mentoring, including your concern for your students' well-being aside from our careers. Among the many ways you have made me a better scientist, thank you especially for always pushing me to think more carefully about how I will test my wild speculations. George, thank you for the intellectual environment you have created. I always feel welcome to ask new questions, to debate, and to speculate without fear of being wrong. You have taught me to think more critically and deeply.

Thank you to both George and Zhengting Zou for the many hours we've spent trying to figure out paradoxes in epistasis. They were some of the most fun I've ever had in science.

I am grateful to my dissertation committee, Alex Kondrashov, Patricia Wittkopp, and Robert Woods, for pushing my science and thinking forward. Thank you Trisha for opening my eyes to evolutionary genetics during my undergrad; I might not have gone into EEB if not for your wonderful teaching. Thank you Alex for your deep questions and for teaching me your vision of evolutionary biology in and out of your class. Thank you Bob for your critical questions that strike to the center of a problem. Thank you also to Aaron King and Luis Zaman for interesting discussions as well as beers at Biobev.

Thank you to the Lauring Lab for their constant support. I have immense gratitude to Will Fitzsimmons for his unending wealth of knowledge and skill in wet lab, for always helping out, and for making the lab fun. I must also thank Matt Pauly, for teaching me the ropes and getting me excited about viral evolution. Thanks also to Andrew Valesano, Yuan Li, Kayla Peck, and JT Mccrone for stimulating discussions and for always being there to field a question.

Thank you to the Zhang Lab for their wonderful intellectual community and for the jokes and conversation. Zhengting, you are the best collaborator I could ask for – a creative, critical thinker, and amazingly productive. Our interests and skills seem to complement each other and I hope this is the start of a long collaboration. Haiqing, thank you for the great times we've had being in the same cohort. I am always impressed by your critical thoughts. Thank you to all other Zhang lab members, including Rex, Xukang, and Siliang who are in the cohorts just below me and are always available for interesting discussion.

Thank you also to the MSTP department, including Ron, whose calm presence and wisdom has calmed many nerves. Thank you to Justine and Ellen, for their unending patience with my lack of promptness with email and for enriching the program in so many ways.

Finally, I am immensely grateful to my family and friends. I think the EEB department is the best department at UM because of its graduate students, who I can always look to for support, deep discussion, and enjoyment outside of work. I will miss all of our jokes, trips, and late nights and tight community. Thank you to my friends from high school who are also always there for support, discussion, and fun.

Thank you to my family for making me who I am.

Thank you uncle David; my critical thinking skills were developed by strategy games with you as a child. Uncle Russell, thank you for teaching me the wonders of math. Papa Don, thank you for telling me stories of your physics. Thank you to my brother, Michael, for always supporting me, for growing up with me, and for making me a better person. Thank you to my girlfriend, Elizabeth, for your unending patience with me, your emotional wisdom, and for your imaginativeness and the joy you give me in life.

Mom, some of my earliest memories are of the joy and surprise I got from the science experiments we did – like how snow melts slower under a blanket. Because of you, I am still looking at the world with wonder. Dad, I remember first learning about evolution from you in the kitchen when I asked you why some things taste good and other things taste bad. Because of you, I am still that curious child asking why.

Table of Contents

Dedicación ii

Acknowledgements iii

List of Tables ix

List of Figures x

Abstract xiii

Chapter 1 Introduction 1

- Overview 1
- Mutational Effects and Epistasis 3
  - Defining and measuring epistasis 4
  - Theories of epistasis and fitness landscapes 6
  - Emerging patterns of epistasis across phyla 11
- Mutational Effects and Epistasis in Influenza virus 14
  - Effects of Single Mutations 15
  - Epistasis in influenza viruses 21
- From idiosyncrasy to general patterns in fitness landscapes 26
- References 28

Chapter 2 Idiosyncratic Epistasis Creates Universals in Mutational Effects and Evolutionary Trajectories 42

- Abstract 42
- Introduction 42
- Results 44
  - Why epistasis could be highly idiosyncratic 44

Epistasis is highly idiosyncratic 46
Expected consequences of idiosyncratic epistasis 47
Idiosyncratic epistasis causes the trends of diminishing returns and increasing costs 48
Idiosyncratic epistasis causes slowing fitness drops in mutation accumulation 50
Idiosyncratic epistasis causes slowing fitness gains in adaptation 51
Discussion 53
Methods 54
Number of terms of phenotypic effects altered by a mutation 54
Differentially altered terms of effects in two genotypes caused by the same mutation 55
The fitness effect of a given mutation is negatively correlated with background fitness 55
Mutational effect is generally negatively correlated with background fitness 57
Mutational supply and evolutionary trajectories 61
Empirical phenotype landscapes 61
Simulating idiosyncratic fitness landscapes 62
Estimating idiosyncrasy index 63
Examining correlation between background fitness and mutational effect 64
Simulating evolutionary trajectories in mutation accumulation (MA) 65
Simulating adaptive trajectories 65
Data and code availability 66
Acknowledgements 66
References 67
Supplementary Materials 69
Chapter 3 Evidence for the Selective Basis of Transition-to-Transversion Substitution Bias in Two RNA Viruses 77
Abstract 77
Introduction 78
Results 80
Transitions are less detrimental in influenza A virus 80

An alternative statistical approach better captures Ts-Tv fitness differences 83
Transversions are more detrimental in larger influenza and HIV datasets 86
Differences in Ts and Tv fitness effects within radical and conservative substitution classes 88
Discussion 93
Materials and Methods 99
Data 99
AUC Analysis 99
Empirical CDF and Odds Ratios 100
Availability of Computer Code and Data 101
Acknowledgments 101
References 102
Supplemental Figures and Tables 107
Chapter 4 The Best Substitutions are also the Worst and Hardest to Make 110
Abstract 110
Introduction 111
Results 113
Mutational fitness effects in RNA viruses 113
Radical amino acid substitutions are more frequently and more highly detrimental than conservative ones 114
Radical amino acid substitutions are more highly beneficial than conservative ones 118
Non-synonymous multi-nucleotide mutations recapitulate findings for radical amino acid substitutions 118
Discussion 121
Methods 125
Data 125
Statistical Analyses 125
References 127
Supplemental Figures & Tables 132

Chapter 5 Discussion 139

Overview 139
Idiosyncratic epistasis as a null theory 141
Implications of Radical/Conservative Fitness Differences 144
Conclusion 146
References 148

List of Tables

Table 4-1 114

Supplemental Table 2-1 Description of empirical landscapes, ordered by idiosyncrasy indices shown in Figure 2-1C 76

Supplemental Table 3-1 Sample size of datasets 108

Supplemental Table 3-2 Amino Acid classifications 109

Supplemental Table 4-1 Numbers of radical/conservative detrimental and beneficial substitutions 133

Supplemental Table 4-2 Numbers of SN and MTN detrimental and beneficial substitutions ... 134

Supplemental Table 4-3 Average detrimental effects of radical/conservative substitutions 135

Supplemental Table 4-4 Average Detrimental Effects of SN and MTN Substitutions 136

Supplemental Table 4-5 Average Beneficial Effects of Radical/Conservative Substitutions 137

Supplemental Table 4-6 Average Beneficial Effects of SN and MTN Substitutions 138

List of Figures

Figure 1-1. Schematic of fitness landscapes 2

Figure 1-2. How mutational effects scale across levels of biological organization..... 4

Figure 1-3. Identifying epistasis. 5

Figure 1-4. Patterns of epistasis. 9

Figure 1-5. Epistasis can constrain adaptation. 24

Figure 2-1 Idiosyncratic index of a wide variety of phenotype landscapes. 46

Figure 2-2 Negative correlation between mutational effect and background fitness as a result of idiosyncratic epistasis. 49

Figure 2-3 Fitness declines decelerate during mutation accumulation as a result of idiosyncratic epistasis. 51

Figure 2-4 Adaptation slows as a result of idiosyncratic epistasis. 53

Figure 3-1 Differences in Ts and Tv Fitness Effects as measured by AUC 82

Figure 3-2 Comparisons of the Distribution of Ts and Tv Fitness Effects 84

Figure 3-3 Distribution of Ts and Tv Fitness Effects in Deep Mutational Scanning Datasets..... 88

Figure 3-4 Distribution of Radical and Conservative Amino Acid Changes 90

Figure 3-5 Distribution of Radical Ts and Radical Tv 91

Figure 3-6 Distribution of Conservative Ts and Conservative Tv 92

Figure 4-1. Difference in percent detrimental between radical and conservative substitutions. 115

Figure 4-2 Differences in mutational effects between radical and conservative substitutions... 117

Figure 4-3 Difference in percent detrimental between MTN and SN substitutions. 119

Figure 4-4 Differences in mutational effects between MTN and SN substitutions. 120

Figure 5-1 The role of the idiosyncratic epistasis theory and conservative versus radical mutations in the scaling of mutational affects across levels of biology. 140

Supplemental Figure 2-1 The high idiosyncrasy indices (Iid) observed are not due to phenotype measurement errors or the use of standard deviation (SD) instead of range of mutational effects. 69

Supplemental Figure 2-2 Negative correlation between mutational effect and background phenotype in GFP and RNA-folding landscapes. 70

Supplemental Figure 2-3 Fig. S3. Patterns of correlation between mutational effect and background fitness/phenotype for individual beneficial or deleterious mutations in various landscapes. 71

Supplemental Figure 2-4 Average fitness trajectories of mutation accumulation simulated in various n-order additive landscapes (k = 1) with different numbers of sites (n). 72

Supplemental Figure 2-5 Fitness declines decelerate during mutation accumulation as a result of idiosyncratic epistasis. 73

Supplemental Figure 2-6 Idiosyncratic epistasis is necessary but not sufficient to cause decelerating adaptations. 74

Supplemental Figure 2-7 Adaptation slows in empirical phenotype landscapes. 75

Supplemental Figure 3-1 Differences in Ts and Tv Fitness Effects as measured by AUC using Decreasing Thresholds 107

Supplemental Figure 4-1 Colors corresponding to strains in main figures 132

Abstract

Biological phenomena can be examined at multiple levels of organization. For example, the role of individual amino acids in protein function representing a low level, the interactions between genes underlying complex traits as an intermediate level, and evolutionary processes across taxa at the highest level. Phenomena at each level must emerge from processes at lower levels. Thus, understanding this emergence is crucial for a full understanding of any biological phenomenon, including the genetic basis of disease and the course of evolution. My thesis focuses on how mutational effects propagate across these levels of organization.

My second chapter develops a null theory for how mutational effects scale to complex phenotypes, from which universal evolutionary patterns emerge. Although mutational effects have predictable, deterministic effects on lower level phenotypes, the complexity of interactions that determine phenotypes like fitness results in a seeming randomness of mutational effects, or idiosyncrasy, at this higher level. Universal evolutionary patterns in adaptation, mutation accumulation, and the effects of mutations across fitness levels then emerge as statistical laws from this randomness. This null theory of "idiosyncratic interactions" resolves a paradox in the common, but mistaken, interpretation of these universal evolutionary patterns. My third and fourth chapters capture a deviation from this null theory, showing that a feature at the lowest level, the radical- or conservative-ness of amino acid substitutions, leads to predictable fitness differences and may bias phylogenetic patterns. Across most phyla, there is a substitution bias towards transitions versus transversions; there has been a long-standing debate over whether this

bias is due to a mutational bias or selection. In my third chapter I show that transversions are more detrimental to fitness in two RNA viruses, influenza virus and HIV. This can be partly explained by their greater likelihood to cause radical amino acid changes, which are more detrimental. Thus, selection is likely a major contributor to the transition-transversion substation bias. My fourth chapter examines further consequences of radical versus conservative amino acid changes. Uncovering the molecular constraints on proteins and the changes responsible for adaptations are major goals in molecular evolution. I find that while proteins are constrained in accepting radical amino acid changes in influenza virus, HIV, and additionally in Zika virus, such radical changes are also the most beneficial, conditional on being beneficial. Multi-nucleotide mutations, which are more likely to cause radical changes due to the genetic code, show the same pattern as radical changes. These findings have implications for viral evolution, phylogenetic inference of selection, and the evolution of the genetic code. Our null theory of idiosyncratic epistasis may help guide future work on how deviations from our predictions at the level of evolutionary patterns reflect the organization of interactions at the level of complex phenotypes and the biological properties of individual genes or mutations at the lowest level.

Chapter 1 Introduction

Note: This chapter is a modified version of the published article:

Lyons DM, Lauring AS. 2018. Mutation and Epistasis in Influenza Virus Evolution. Viruses. 10(8):407. doi:10.3390/v10080407.

Overview

The diversity of life is amazing, from the wings of a ruby-throated hummingbird flapping at hundreds of times per second to the intricate geometric shapes of viral capsids. Perhaps even just as beautiful is the fact that all of life's diversity is generated by the same evolutionary processes. To understand how life's diversity is produced through evolution, understanding the effects of mutations is crucial. Because all life originated from a common ancestor, this diversity must be generated by the accumulation of mutations through billions of years. However, it is not only the accumulation of single mutations that contribute to diversity, but the interactions between mutations. The same mutation can have varying effects depending on the other mutations present.

Epistasis refers to the genetic interactions between two or more mutations in a genome. It is useful to think about epistasis using the image of a mountainous landscape (Phillips 2008). Here, each point on the landscape is a genotype, and the height represents the fitness of the genotype (Figure 1-1). If there were no epistasis, then this fitness landscape would have a single peak, and populations might adapt, or climb up that peak, linearly with each mutation (Figure 1-1A). However, because of epistasis, these landscapes are more rugged, with different peaks and valleys representing unexpected fitness effects of mutational combinations (Figure 1-1B). The

shape of this landscape determines how a population will evolve (Phillips 2008). For example, how quickly it can climb a peak or which peaks are accessible without crossing a low fitness valley. The combinatorics of mutations vastly increases the potential diversity of life. The space of possible genotypes is enormous. Even the collection of all genomes of all organisms that have ever lived, represents a tiny point in a vast space of possible genotypes. For a genome the size of the influenza virus, the number of possible genotypes is larger than the number of particles in the observable universe.

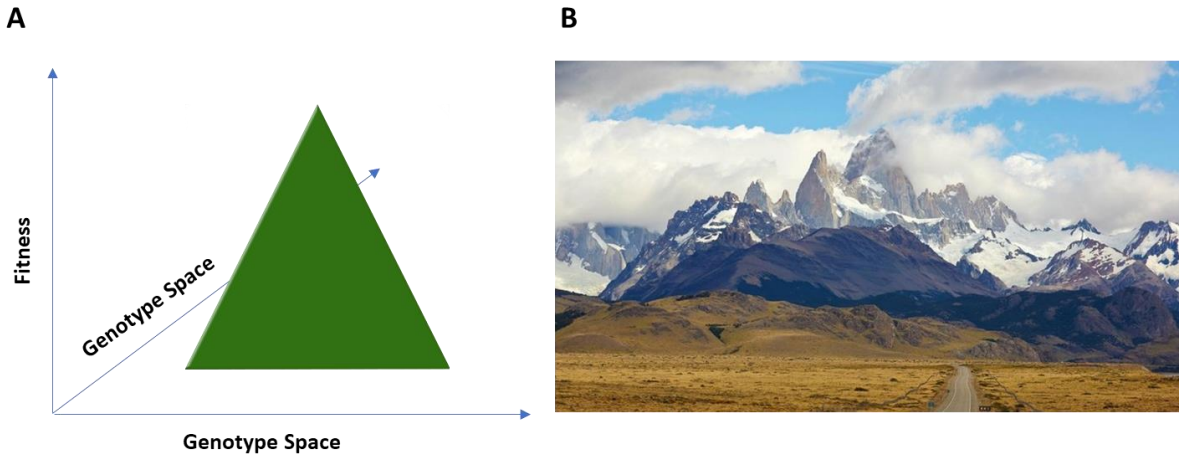


Figure 1-1. Schematic of fitness landscapes

A smooth fitness landscape without epistasis (A) and a rugged one with epistasis (B). In both, height represents fitness and the orthogonal axes represent genetic distance.

How can we understand this vast space in which evolution occurs? The projects in this dissertation advance our understanding from two perspectives. The first nearly dispenses with biological details, showing how universal patterns in evolutionary trajectories and mutational effects uncovered over the last decade can emerge from underlying randomness in mutational effects. The second turns to the biological details of mutational effects in RNA viruses, showing that the structure of the genetic code and properties of amino acids likely contribute to

widespread patterns in molecular evolution. Below, I review the literature on mutational effects and epistasis in general and as it pertains to the influenza virus, an RNA virus.

Mutational Effects and Epistasis

Enormous successes in molecular biology and biomedical science have been made by probing the effects of individual genes or mutations in model systems, from our basic understanding of bacterial gene regulation to the hormonal control of metabolism (Stent 1964; Pelleymounter et al. 1995). Life-saving cures have been developed by probing the structure and function of individual molecules with single mutations. A classic example is cystic fibrosis, which is commonly caused by a single amino acid deletion (Boyle and De Boeck 2013). Understanding how this mutation causes a misfolded protein led to the development of effective drug treatments.

Despite its historical success, the limits of research into the impact of single mutations and the function of individual molecules have become clear. No single gene or mutation can account for the development of complex traits or diseases like diabetes (Manolio et al. 2009). Often, a polymorphism linked to a trait in one population is unimportant in other populations, likely due to interactions with other polymorphisms (Moore 2003; Carlborg and Haley 2004). Findings in model systems like mice often do not translate to humans (Perel et al. 2007), partly due to epistasis with human-specific substitutions. Disease-causing polymorphisms in humans are frequently the wild-type in other species (Xu and Zhang 2014). While we can often understand the effect of individual mutations or interactions (Figure 1-2, bottom), it is extremely difficult to predict how such effects scale to higher levels of biological organization that involve many interactions – like the metabolism of an organism, complex phenotypes, or fitness (Figure 1-2, middle). If we cannot understand how the effects of individual mutations scale to intermediate

phenotypes or fitness, their effects on evolutionary processes are even more opaque (Figure 1-2, top).

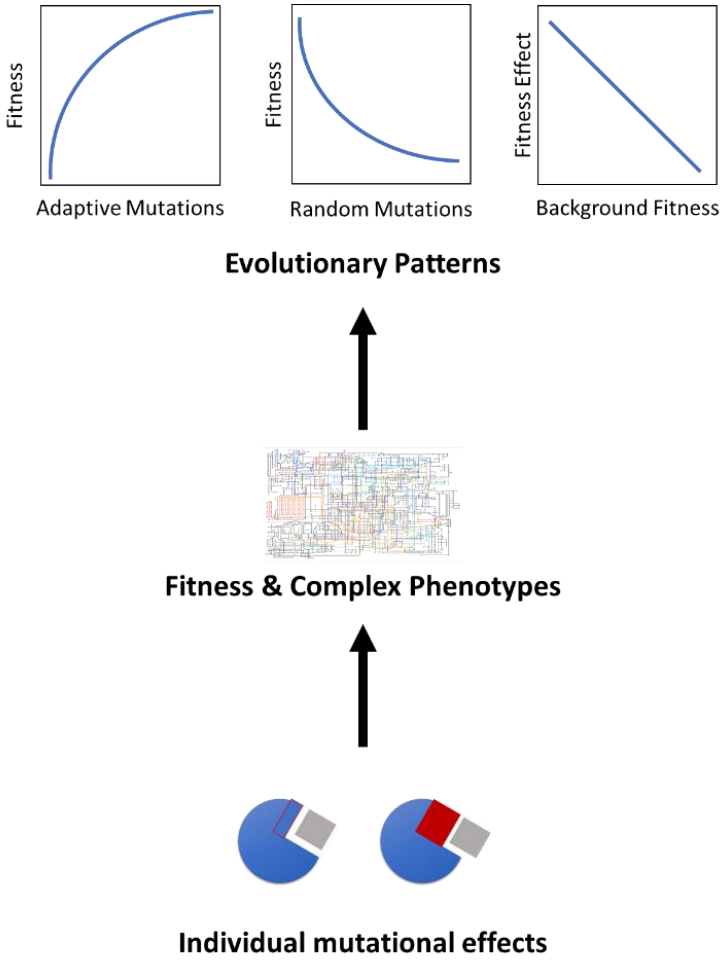


Figure 1-2. How mutational effects scale across levels of biological organization.

At the bottom level, the function of individual genes or effects of individual mutations are largely predictable given sufficient biophysical knowledge. Here, the red mutation towards a larger amino acid diminishes the function of the blue enzyme. Despite their deterministic actions, mutational effects at the level of fitness or complex phenotypes are highly unpredictable (middle). Somehow, unpredictability in the effect of mutations at the intermediate level leads to universal patterns during evolution. (top).

Defining and measuring epistasis

Epistasis (ϵ) refers to the genetic interactions between two or more mutations in a genome. Epistasis is most commonly defined as the difference between the observed fitness of the genome with both mutations i and j (w_{ij}) and the expected fitness given independent multiplicative effects of each single mutation (Figure 1-3A) (Mani et al. 2008). Thus, $\epsilon = w_{ij} - w_i \times w_j$, where $\epsilon = 0$ indicates no genetic interaction. Negative epistasis ($\epsilon < 0$) occurs when the fitness of a double mutant is less than expected. Positive epistasis ($\epsilon > 0$) occurs when the fitness of the double mutant is greater than expected. As we mainly discuss epistasis as either positive or

negative below, we refer the reader to (Mani et al. 2008) for a more detailed review of epistasis terminology.

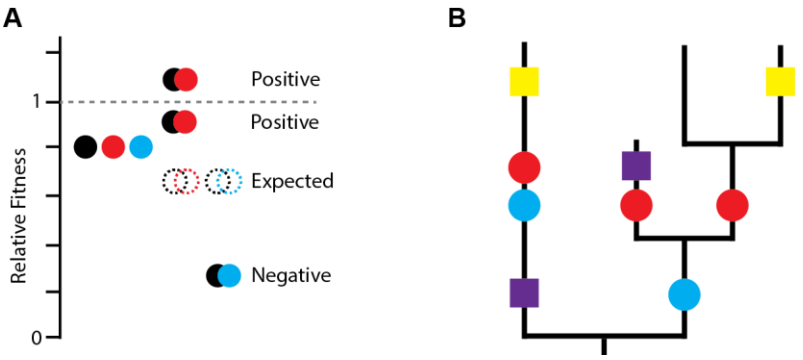


Figure 1-3. Identifying epistasis.

(A) Epistasis is defined based on deviation from the expected fitness assuming independent multiplicative effects. Black, red, and blue mutations are each deleterious (filled individual circles) and the expected fitness of pairwise combinations (empty circles) is shown. Black and red mutations exhibit positive epistasis because combining them is not as deleterious as expected and may even increase fitness. Black and blue mutations exhibit negative epistasis because combining them is more deleterious than expected. (B) Positive epistasis can be inferred phylogenetically. Red and blue mutations interact positively as they are always present together and occur closely in time. In contrast, yellow and purple mutations do not occur closely in time and are not necessarily present together.

Mutational fitness effects can be measured in a variety of ways. Site-directed mutagenesis allows for precise control of the identity and number of mutations created. This is usually combined with a competitive fitness assay that provides a precise measure of the fitness of a given mutant relative to the wild-type (Sanjuán 2010; Visher et al. 2016), an approach that is reliable, but labor-intensive. Deep mutational scanning (DMS) combines large-scale mutagenesis with bulk fitness measurements of a mutagenized library through next generation sequencing (Fowler and Fields 2014). This method allows for a nearly complete sampling of single mutations across a gene. To detect epistasis, DMS results can be compared across different genetic backgrounds (Doud et al. 2015; Haddox et al. 2018). Alternatively, a short region that can be covered by a

single sequencing read can be used to detect linkage between multiple mutations. However, DMS is less sensitive for lethal and low fitness mutations. A third strategy that can uncover mutational effects in natural environments is to examine patterns of polymorphism in a population, such as during viral infections by deep sequencing, or to examine substitution patterns phylogenetically (Zhang 2000; Gojobori et al. 2007; Keightley and Halligan 2011; Chen et al. 2019). These approaches interrogate the effect of mutations in a realistic environment (Nielsen and Yang 2003; Tamuri et al. 2012; Poon et al. 2016; McCrone et al. 2018). However, they often cannot assign fitness effects to individual mutations or determine the strength of interactions. Thus, these three approaches therefore provide complementary insights into the mutational fitness effects and epistasis.

Theories of epistasis and fitness landscapes

Theoretical work on the structure of fitness landscapes falls on a continuum from general mathematical models of interactions between elements of complex systems to theories informed by biological observations.

Mathematical models

The most general models include the NK-landscape, the House-of-Cards model, and the Rough Mount Fuji model, among others. In the NK-landscape, each of N loci (which could be genomic sites, genes, or even molecules in a magnet) interacts with K other randomly chosen loci (Kauffman and Weinberger 1989). A fitness contribution of each of the N loci is chosen randomly for each combination of the allelic states of each site's K interacting loci. The fitness of a genotype is the average of the fitness contributions of all N loci. When K is less than N, a single mutation does not change all the fitness contributions and thus fitness is correlated among related genotypes. The House-of-Cards model is a special case of the NK landscape, where K =

N-1 and the fitness of each genotype is chosen randomly, with no correlation among related genotypes (Fragata et al. 2019). Fitness correlation among genotypes is also tunable in the Rough Mount Fuji Model, but without an explicit model of interactions (Neidhart et al. 2014). This model begins with non-epistatic landscape and an optimal genotype and linearly decreasing fitness of each genotype with genetic distance from the optimum (like in Figure 1-1A). To introduce epistasis, a random variable akin to "noise" is added to the fitness of each genotype, creating a landscape more like Figure 1-1B. This noise variable can be tuned to introduce varying levels of epistasis.

The utility of these models has been to guide our intuition of how epistatic interactions influence the structure of the fitness landscape (Kryazhimskiy et al. 2009; Fragata et al. 2019). For example, how does fitness of peaks and the number of different peaks and their accessibility, or the ruggedness of the landscape, scale with epistasis (K or noise) (Kauffman and Levin 1987; Kauffman 1993; Neidhart et al. 2014)? The fitness of the best genotype in the landscape increases with epistasis and the number of loci. However, the mean fitness of peaks is the highest when there is epistasis, but only a small amount (e.g. $0 < K \ll N$). Moreover, the number of peaks increases exponentially with more epistasis, leading to lower accessibility of each individual peak as adaptive walks get stuck at a peak more quickly. Thus, although increasing N and K leads to higher fitness of the optimum genotype, adaptation will terminate at lower and lower fitness peaks. Interestingly, this can be thought of as a cost to complexity, as increasing the number of interacting parts (N and K) makes adaptation more difficult. The NK model has also been modified to investigate how genetic architecture influences landscape structure by choosing the interacting K sites non-randomly, say from neighboring sites. Some studies have shown that

a modular architecture can circumvent this cost to complexity (Earl and Deem 2004; Nowak and Krug 2015; Hwang et al. 2018). These models are also ideal for studying evolutionary processes that depend on epistasis and which require more knowledge of the fitness landscape than can be obtained from empirical data, such as the benefits of sexual recombination or the ability of drift to promote the crossing of fitness valleys (Kryazhimskiy et al. 2009; Nahum et al. 2015; Zagorski et al. 2016; Agarwala and Fisher 2019; Fragata et al. 2019; Reia and Campos). Another hope for these models has been to extrapolate global properties from small empirical landscapes (Kauffman and Weinberger 1989). Some studies have gauged the ruggedness of real landscapes by fitting these models to real data, but without much success beyond the unsurprising result that real landscapes lie somewhere between additive and the House-of-Cards model in terms of ruggedness (Fragata et al. 2019). A complementary use of these models is to gauge the reliability and properties of various measures of ruggedness and adaptability that can then be used to analyze empirical data (Siliang Song, personal correspondence). While the theoretical implications of these models have been important, their use in uncovering the parameters of empirical landscapes remains a challenge. Additionally, none of these models make predictions about the general sign of epistasis.

Biologically-grounded theories

More biologically informed theories make predictions about the general sign of epistasis and are more easy to connect to specific biological cases. However, some suffer from a lack of generality and they cannot be used as readily as mathematical models to investigate how epistasis influences global properties of the fitness landscape such as ruggedness.

The multiple hit hypothesis holds that epistasis depends on genome complexity and the extent of functional redundancy (Figure 1-4A) (Sanjuán and Elena 2006; Sanjuán and Nebot 2008). At one extreme are simple viral genomes encoding a small number of mostly essential, multifunctional proteins. Here, the first mutation may have a large effect, but the impact of additional mutations is smaller, since they cannot further break functions already broken by the first mutation. This is positive epistasis. In contrast, eukaryotes or organisms with larger genomes may have redundant pathways and regulatory mechanisms such as feedback inhibition, which tend to buffer the impact of single mutations but less so for multiple mutations, leading to negative epistasis.

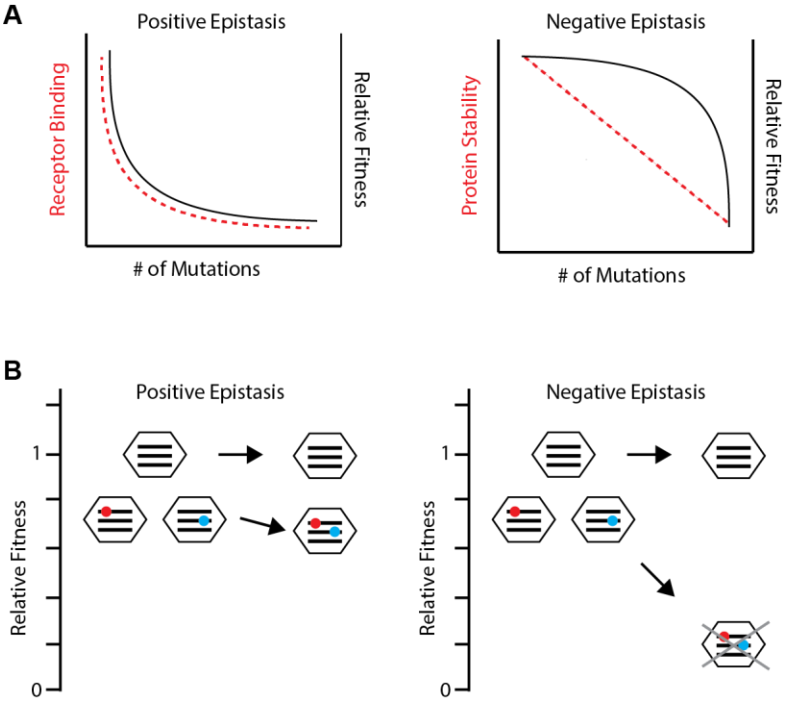


Figure 1-4. Patterns of epistasis.

(A) Distinct epistatic patterns are predicted based on how phenotypic effects (red dashed lines) correspond to fitness effects (black solid lines). Small genomes may encode a single mechanism for a particular function, such as receptor binding in HA (left). Without a backup mechanism, initial mutations have a large impact on receptor binding and fitness. Additional mutations have little further impact, as the function has already been destroyed, resulting in positive epistasis. Other phenotypes, like protein stability, can be reduced without affecting fitness until a threshold is reached (right). Thus, each additional mutation impacts protein stability similarly but increasingly impacts fitness, resulting in negative epistasis. (B) A viral population consisting of an unmutated genotype and two variants each with a slightly deleterious mutation (red and blue

circles) on different segments (black lines). Reassortment between the two variants can combine the two deleterious mutations. If epistasis is positive (left), the reassortant will have higher fitness than expected and the deleterious mutations may persist in the population, lowering the average fitness of the population. If epistasis is negative (right), the reassortant will be quickly purged from the population, leaving the unmutated genotype and raising the average fitness of the population.

In contrast, theories of selection for robustness predict that that high mutation rates can select for distinct mechanisms that buffer the impact of single mutations and lead to negative epistasis, even in simple genomes (Figure 1-4A) (Wilke and Christoph 2001; Wilke et al. 2003; Bershtein et al. 2006; Gros et al. 2009). For example, if fitness is reduced only when an underlying phenotype reaches a threshold, then the full deleterious impact of mutations affecting that phenotype will only be revealed when enough mutations accumulate to cross the threshold, resulting in negative epistasis (Bershtein et al. 2006).

The modular life model focuses on a different feature of complex organisms, modularity (Wei and Zhang 2019). This theory posits that an organism is composed of functionally and genetically distinct modules, for example, how well it metabolizes sugar or its resistance to environmental toxins. These modules each contribute to fitness separately but their contribution has an upper limit. Negative epistasis among beneficial mutations arises: A beneficial mutation improving functionality of a particular module will increase fitness greatly if the module is far from maximum functionality, but the same beneficial mutation will be less advantageous if the module's functionality is already near the maximum. This model has not been extended to detrimental mutations.

Fisher's geometric model (FGM) focuses on yet another property of organismal complexity – that mutations are highly pleiotropic. FGM posits that phenotypes are under stabilizing selection for an optimum. The predicted distribution of epistasis for an organism at optimal fitness is

approximated by a symmetric, normal distribution with mean 0 (Martin et al. 2007). FGM originated as a model for adaptation by small steps, which has found much support (Burch and Chao 1999).

Other biologically-based theories of epistasis only apply to particular functions of an organism, such as its metabolism, or only to epistasis within genes. For example, metabolic control theory applies to pathways in which the function depends on the functionality at each sequential step. Here, positive epistasis is predicted for mutations in serial steps of a pathway, negative epistasis for mutations in related but different (somewhat redundant) pathways, and no epistasis for unrelated pathways (Kacser and Burns 1981; Segrè et al. 2005; Maclean 2010). The reasoning is similar to that of the multiple hit hypothesis, with mutations in single pathway corresponding to those in a simple organism with no redundancy and mutations in related, redundant pathways corresponding to those in a complex organism. The final major theory of epistasis focuses on intragenic epistasis and is based on the thermodynamic threshold known to be important for protein folding stability (Bloom et al. 2005; Sarkisyan et al. 2016) (Figure 1-4A). Here, single mutations may not lower stability below the required threshold and will only have small effect on protein function and fitness. However, multiple mutations will push stability below the threshold, leading to a malformed protein and thus exhibiting high fitness effects and negative epistasis.

Emerging patterns of epistasis across phyla

The measurement of a complete fitness landscape for any organism and most genes is impossible due to their size. Furthermore, limited discriminatory power of most fitness assays and the feasibility of generating and testing large numbers of mutation pairs makes any quantitative studies of epistasis difficult. Thus, most studies have probed the fitness landscape either by

inference from evolutionary patterns or by focusing on a small number of variants or a gene of manageable size.

Epistasis across the genome

Over the last decade, some general patterns concerning adaptive mutations have been observed across nearly all phyla, from viruses to fungi (Couce and Tenaillon 2015). In the famous long-term evolution experiment with E. coli, fitness increase generally slows with time or with the number of fixed mutations (Barrick et al. 2009) (Figure 1-2, top-left plot). In addition, high fitness backgrounds increase in fitness to a lesser degree than do lower fitness backgrounds, given the same amount of time or number of mutations (Sanjuán et al. 2005; Barrick et al. 2009; MacLean et al. 2010; Wiser et al. 2013; Perfeito et al. 2014; Wünsche et al. 2017) (Figure 1-2, top-right plot). Other experiments have put beneficial mutations on backgrounds of varying fitness and found that beneficial fitness effect negatively correlates with background fitness; this is termed diminishing returns epistasis (Bull et al. 2000; MacLean et al. 2010; Chou et al. 2011; Khan et al. 2011; Kvitek and Sherlock 2011; Rokyta et al. 2011; Pearson et al. 2012; Flynn et al. 2013; Schenk et al. 2013; Wang, Yinhua et al. 2013; Caudle et al. 2014; Chou et al. 2014: 20; Kryazhimskiy et al. 2014; Schoustra et al. 2016; Wang et al. 2016; Wünsche et al. 2017). These observations are widely interpreted as a bias towards negative epistasis (concaved fitness landscape), as combining multiple beneficial mutations is less beneficial than expected.

Nearly universal patterns concerning detrimental mutations have also emerged, but they are thought to indicate a bias towards positive, not negative, epistasis in the fitness landscape. During mutation accumulation, which fixes successive detrimental mutations by subjecting an organism to extreme drift, fitness decrease generally slows with time or number of mutations

(deVisser et al. 1997; Burch and Chao 1999; Lenski et al. 1999; Poon and Otto 2000; Crotty et al. 2001; Wilke and Christoph 2001; You and Yin 2002; Wilke et al. 2003; Bonhoeffer et al. 2004; Burch and Chao 2004; Maisnier-Patin et al. 2005; Iglesia and Elena 2007; Perfeito et al. 2014) (Figure 1-2, top-middle plot). The corollary to diminishing returns has also been found, increasing costs, in which detrimental mutations are more detrimental in higher fitness backgrounds (Johnson et al. 2019) (Figure 1-2, top-right plot). These patterns are thought to indicate a bias towards positive epistasis in the fitness landscape and a convex shape.

Thus, these nearly universal patterns seem contradictory. Theories of epistasis can find both support and refutation for their predictions as to the sign of epistasis. No theory has been able to reconcile the opposite signs of epistasis inferred from beneficial and detrimental mutations. In fact, despite a decade or more of these observations, this apparent paradox was only recognized in the literature this past year (Miller 2019).

Epistasis within genes

In contrast to methods which track evolutionary change, DMS is more useful for identifying the interacting mutations and the strength of interaction and can infer the sign of epistasis directly. However, it is largely limited to intragenic epistasis. General findings are that epistasis is ubiquitous and is usually the strongest for pairwise interactions, although strong higher-order interactions are common as well.

Most studies attempt to explain the interactions they find from biological principles. Some mutational interactions can be explained by steric or charge interactions after close examination of protein structure (Melamed et al. 2013; Wu et al. 2017). A slight bias towards negative

epistasis seems common and can be predicted by the folding energy of the gene product (RNA or protein), supporting the thermodynamic threshold hypothesis (Bershtein et al. 2006; Araya et al. 2012; Olson et al. 2014; Li et al. 2016; Puchta et al. 2016; Sarkisyan et al. 2016; Bendixsen et al. 2017; Bendixsen et al. 2017). However, prediction is not straightforward - a novel biophysical model seems to be constructed for each new protein landscape – and how well each model explains the data is highly variable. Certainly the most common finding is that mutational effects and their interactions are very difficult to predict, or idiosyncratic (Domingo et al. 2018).

Is this unpredictability due to limited knowledge of the biophysical and chemical processes governing protein function? Or are biological interactions so numerous and complex that predicting their outcome is like predicting the outcome of die roll by tracking the collisions of all the air molecules in a room? Indeed, just as the sure knowledge that a die will come up a six one-sixth of the time, a statistical approach to epistasis can sometimes be more fruitful. One study used the statistical co-occurrence of amino acid variants in functional sequences to predict phenotypes with the same accuracy as using all experimentally-determined second-order epistasis terms (Poelwijk et al. 2019).

Mutational Effects and Epistasis in Influenza virus

Influenza viruses infect a large number of hosts, have high mutation rates, and frequently reassort. As a result, they have a tremendous capacity to explore a large number of potential sequences. Indeed, the ability of influenza populations to adapt to new hosts and to escape the immune system seems unlimited. However, mutations are often deleterious, which presents a barrier to viral adaptation. Furthermore, epistasis determines the mutational paths available and can make some adaptations inaccessible. Understanding how mutation and epistasis present both

barriers and opportunities for influenza virus evolution is essential in predicting viral evolution and designing better vaccines and antivirals.

Influenza A has a negative-sense, single stranded RNA genome of 14kb consisting of 8 sexually reassorting segments encoding a total of 11-12 proteins (Medina and García-Sastre 2011; Fields et al. 2013). The influenza envelope contains two interacting glycoproteins encoded by the haemagglutinin (HA) and neuraminidase (NA) segments. HA binds to sialic acids of cellular receptors, while NA facilitates the release of nascent virions by cleaving HA from bound sialic acids. Three proteins encoded by the PB2, PB1, and PA segments form the RNA-dependent RNA polymerase protein complex (RdRp). Nucleoprotein, encoded by the (NP) segment, packages the RNA and interacts with the RdRp complex. The matrix (M) segment encodes the M2 ion channel important for fusion of the virus with the cell and uncoating of RNA-NP complexes. The M1 protein, encoded by the M segment, is involved in assembly of virus particles. The non-structural segment NS encodes NS1, an antagonist of anti-viral responses, and the nuclear export protein NEP.

Effects of Single Mutations

The distribution of mutational fitness effects (DMFE) reveals the extent of genetic constraint on the influenza virus genome, how constraints vary between and within influenza proteins, and the structural and functional impacts of mutations.

The genome-wide DMFE in influenza virus

Site-directed mutagenesis has been used to characterize the genome-wide DMFE of single-mutations in a variety of viruses (Sanjuán 2010). Our laboratory applied this technique to an H1N1 influenza strain (Visher et al. 2016). We generated a library of 95 randomly selected point

mutations distributed across the influenza genome. We also generated 33 additional mutations in the segments encoding the hemagglutinin (HA) and neuraminidase (NA) proteins to compare the DMFE of these surface-exposed antigenic proteins (n=57) to the internal proteins encoded by the remaining 6 segments (n=71).

We measured fitness relative to the wild-type in a pairwise competition assay and used repeated transfection to distinguish true lethal mutations (fitness = 0) from failed viral rescue. In our dataset, 31.6% of all mutations were lethal. Approximately 40% of all viable mutations were highly detrimental (< 0.85), 50% mildly detrimental or neutral (0.85-1.05), and only seven were beneficial (> 1.05). The fitness among all viable mutations ranged from 0.26-1.13 with a mean of 0.82. Non-synonymous mutations were more deleterious than synonymous mutations, consistent with reduced genetic constraint at the level of RNA relative to protein. Two of the three non-coding mutations were lethal, consistent with the conserved roles of these regions in RNA packaging and genome replication (Watanabe et al. 2003; Dawson et al.).

In general, the DMFE for influenza virus is similar to those documented for other viruses with a variety of genomic structures, from single-stranded RNA viruses to DNA phages (Sanjuán et al. 2004; Carrasco et al. 2007; Domingo-Calap et al. 2009; Peris et al. 2010; Sanjuán 2010; Visher et al. 2016). The lethal fraction for influenza virus falls within the 20-40% range observed for other viruses. When scaled to exponential growth rate – the fitness surrogate in many studies – the average fitness effect in influenza virus is 12%, squarely within the 10-13% range found in other viruses (Sanjuán 2010; Visher et al. 2016). These large effects stand in marked contrast to those seen in cellular organisms (Eyre-Walker and Keightley 2007) and may reflect shared

genetic constraints across viruses, possibly related to the small size of their genomes (Sanjuán and Elena 2006).

While genome-wide patterns in the DMFE are similar across viruses, the DMFE varies between and within individual influenza genes. The antigenic proteins, HA and NA, evolve much more rapidly than other influenza proteins. This rapid evolution could be due to a history of stronger positive selection and/or an inherently greater mutational tolerance (Plotkin and Dushoff 2003). In our comparative study, we found that the antigenic proteins were generally more tolerant of mutation. The mean fitness of mutations in the surface proteins (0.88) was higher than for the internal proteins (0.78). Furthermore, the head region of HA, which is immunodominant and exhibits the greatest sequence diversity, had a higher mean fitness (0.77) than the stalk region (0.56), which exhibits lower sequence diversity. Other groups have also documented the relative mutational tolerance of HA. For example, Heaton et. al. mutagenized the entire influenza genome with 15-nucleotide insertions (Heaton et al. 2013). A disproportionate number of recovered variants had insertions in the head region of HA (7/20 recovered variants). DMS studies have further confirmed the mutational tolerance of rapidly-evolving HA regions, revealing high tolerance in antigenic domains and very low tolerance in the slower-evolving HA receptor binding pocket and stalk domain (Thyagarajan and Bloom 2014; Lee et al. 2018 Apr 10). These data suggest that mutational tolerance in HA and NA, particularly in the HA head, contributes to their greater evolutionary potential. It will be interesting to compare these patterns to those in the major antigenic proteins of other viruses (Fulton et al. 2015).

In influenza virus, other protein regions with immunodominant epitopes do not always recapitulate the trends in HA. For example, the solvent-exposed region of the nonstructural protein 1 (NS1), which likely interacts with host proteins to modulate immune responses, also has greater tolerance to insertions (Heaton et al. 2013). However, immune-targeted sites in the nucleoprotein (NP) do not show unusually high mutational tolerance (Bloom 2014; Thyagarajan and Bloom 2014). Perhaps these NP regions are less tolerant because they experience lower diversifying selection from the immune system. Alternatively, these sites could be inherently more constrained. Understanding the causes and consequences of varied mutational tolerance is highly relevant to vaccine design, as the lower mutational tolerance of the HA receptor-binding pocket and stalk make them attractive targets for a universal vaccine (Erbelding et al.).

Overall, the vast majority of mutations in influenza virus are lethal or deleterious. Given the virus's high mutation rate of 2-3 per genome replicated, a large proportion of newly replicated genomes will contain a lethal mutation, and many more will harbor one or more deleterious mutations (Visher et al. 2016; Pauly, Procario, et al. 2017). Within hosts, the constraints of deleterious mutations are manifest as high levels of purifying selection and limited genetic variation (Iqbal et al. 2009; Murcia et al. 2010; Dinis et al. 2016; Debbink et al. 2017; Leonard et al. 2017; Xue et al. 2017; McCrone et al. 2018; Xue and Bloom 2018 Jul 8). Deleterious mutations also impact influenza evolution at the global scale, because purifying selection does not always efficiently purge them from the population. Deleterious mutations may reach fixation by drift (e.g. during transmission bottlenecks) or by hitchhiking with adaptive mutations. Influenza virus phylogenies show a high deleterious mutation load (Pybus et al. 2007), and

models suggest that this load can slow antigenic evolution (Koelle and Rasmussen 2015; Raghwani et al. 2017).

Deep mutational scanning of influenza proteins

While site-directed mutagenesis has provided an overview of the genome-wide DMFE, DMS can interrogate nearly all amino acid substitutions in a single protein. In DMS studies, fitness is usually calculated as the change in frequency of a mutation in a pool of variants before and after passage or selection, relative to wild-type. This method is analogous to pairwise competition assays, and fitness measurements across studies are well correlated (Visher et al. 2016; Lyons and Lauring 2017). DMS studies by Bloom and colleagues include saturation mutagenesis of the HA and NP proteins from H1N1 and H3N2 strains (Bloom 2014; Thyagarajan and Bloom 2014; Doud et al. 2015; Doud and Bloom 2016; Lee et al. 2018 Apr 10). Studies by Sun and colleagues investigated many substitutions in nearly all sites in the six other influenza proteins (Wu et al. 2013; Wu, Young, Al-Mawsawi, Olson, Feng, Qi, Luan, et al. 2014; Wu, Young, Al-Mawsawi, Olson, Feng, Qi, Chen, et al. 2014; Wu et al. 2015; Du et al. 2016; Wu et al. 2016; Du et al. 2018). We are now close to a complete map of the fitness effects of all possible amino acid substitutions for an entire influenza virus genome.

Both sets of DMS studies clearly show that mutational tolerance varies widely across sites within a protein; some sites strongly prefer a single amino acid and others accept many different amino acids. The effect of any particular amino acid substitution is also highly site-specific. Wu et. al. investigated the link between protein stability and mutational effects in PA to shed light on why constraints may vary across sites (Wu et al. 2015). They found two categories of amino acid residues; those in which substitutions affected overall protein stability and those in which

substitutions were detrimental but did not affect stability. The latter were termed “functional” residues as they likely affected enzymatic functions of a protein (e.g. polymerase activity) or important protein-protein interactions (e.g. solvent-exposed sites).

More recent studies have used DMS in innovative ways. DMS-informed site-specific and parameter-free evolutionary models dramatically improve the fit of phylogenies (Bloom 2014; Doud et al. 2015; Hilton and Bloom 2018 Apr 17) and the inference of sites under positive selection (Bloom 2017; Hilton et al. 2017). Another promising avenue is the application of DMS to phenotypes other than fitness (Wu et al. 2013; Wu, Young, Al-Mawsawi, Olson, Feng, Qi, Luan, et al. 2014). Du et al. used DMS to identify mutations that increase IFN sensitivity while preserving replicative fitness and immunogenicity, leading to a potentially safe and effective vaccine strain (Du et al. 2018). Bloom and colleagues have used DMS to study the potential mutational pathways of antibody escape in HA and avian to human host-adaptation in PB2 (Doud et al. 2017; Doud et al. 2018; Soh et al. 2019).

There is now extensive data on the effects of single mutations in influenza virus. The vast majority of mutations are deleterious, with similar effects as in other viruses. Greater mutational tolerance in some antigenic sites may enable their rapid evolution whereas lower mutational tolerance makes other sites promising vaccine targets. The challenge now is to better understand the biological basis of mutational effects. Mutational fitness effects in the laboratory correlate with mutational frequency in nature (Visher et al. 2016; Lee et al. 2018); thus, these data could be used to improve predictive models of influenza evolution (Łuksza and Lässig 2014; Morris et al. 2017; Lee et al. 2018).

Epistasis in influenza viruses

Studies have employed site-directed mutagenesis to study interactions among small numbers of mutations, usually those implicated in adaptation to immune pressure or antiviral drugs (Bloom et al. 2010; Gong et al. 2013; Pauly, Lyons, et al. 2017). DMS been used for studying a small region in HA. DMS has also been used to compare mutational effects across different NP and HA genetic backgrounds (Bloom 2014; Lee et al. 2018 Apr 10). Shifts in mutational effects at a given site across different genetic backgrounds reflect epistatic interactions involving that site. However, comparative DMS studies can only detect epistatic interactions involving at least one divergent site and cannot precisely identify the interacting mutations.

Phylogenetic inference can be used to identify epistatic interactions in the virus's natural replication environment. One approach is to identify co-evolving sites (Shapiro et al. 2006; Akand and Downard 2018). If substitutions at one site are followed by second site substitutions more quickly than expected by chance, these substitutions likely enhance each other's beneficial effects (Figure 1-3B) (Kryazhimskiy et al. 2011). This approach can only detect positive epistasis and has limited power for rarer polymorphisms and weaker epistatic interactions. Furthermore, it has typically been applied to studies of within-gene epistasis, given the added complexity of reassortment and the computational costs of genome-wide scans (but see (Neverov et al. 2015)). Phylogenetic inference of between-gene epistasis in influenza relies on observed patterns of reassortment. Here, non-random patterns of reassortment among genome segments suggest incompatible interactions (Rambaut et al. 2008). These incompatibilities can also be detected as accelerated rates of evolution in reassortant lineages, as the newly combined

segments adapt to their new genetic environment (Neverov et al. 2014). While these studies identify gene-level epistasis, they typically do not identify the interacting sites.

General epistatic patterns in influenza viruses

Recent studies have elucidated patterns of within-gene epistasis. Comparative DMS of NP and HA in H3N2 and H1N1 backgrounds have found that both short-range physical interactions and long-range functional interactions within these proteins are common (Doud et al. 2015; Lee et al. 2018 Apr 10). Phylogenetic studies also find many long-range epistatic interactions (Shapiro et al. 2006; Nshogozabahizi et al. 2017). Additionally, sites exhibiting epistasis cluster with each other, which can be explained by structural changes affecting a particular region of the protein (Doud et al. 2015; Lee et al. 2018 Apr 10). Less is known about the type and magnitude of epistasis. A DMS study of 11 sites in the receptor binding region of HA found positive epistasis to be ubiquitous (Wu et al. 2017). This is in contrast to studies in other taxa showing that a folding stability threshold generally leads to negative epistasis across entire proteins (Figure 1-4A) (Bershtein et al. 2006; Lehner 2011; Li et al. 2016; Puchta et al. 2016; Sarkisyan et al. 2016).

The theoretical costs and benefits of reassortment largely depend on the type and magnitude of epistasis between mutations on different segments. Reassortment is advantageous in the setting of negative epistasis because combining deleterious mutations through reassortment will accelerate the rate at which they are purged from a population (Figure 1-3B) (Chao 1988; Kondrashov 1988; Kouyos et al. 2007). Conversely, positive epistasis slows the rate at which deleterious mutations are purged, making reassortment disadvantageous. Reassortment also underlies the process of antigenic shift and the associated spread of avian and swine viruses to

humans (Morens et al. 2009; Campbell et al. 2014; Danzy et al. 2014). However, segments do not reassort freely (Rambaut et al. 2008; Zeldovich et al. 2015), and differential pairwise epistasis among segments reflects their genetic incompatibilities. Here, epistasis imposes a fitness cost to reassortment, even between strains of the same subtype, and could limit host-range expansion (Ward et al. 2013; Neverov et al. 2014; Villa and Lässig 2017).

Epistasis in the adaptive evolution of influenza virus

Most studies of epistasis in influenza virus have focused on its role in antigenic evolution. HA evolution is characterized by a series of mutations with little apparent change in antigenicity, forming an antigenic cluster, followed by a mutation that leads to significant antigenic drift, called a cluster transition. Models show that epistatic interactions among individually neutral mutations can explain this pattern of evolution (Koelle et al. 2006; Taggi et al. 2013; Tria et al. 2013). The epistatic interactions in these antigenic clusters can lead to historical contingency. Mutations involved in a cluster transition also interact with mutations involved in the subsequent cluster transition, forming chains of interacting mutations (Nshogozabahizi et al. 2017). These chains suggest that the fixation of each substitution is contingent on the fixation of prior substitutions.

Studies employing site-directed mutagenesis and experimental evolution demonstrate how epistasis in HA leads to this historical contingency. First, the impact of a given mutation on antigenicity or receptor binding varies with genetic background (Nakajima et al. 2005; Das et al. 2013). This context dependence makes it harder to predict HA evolution and generalize molecular findings between strains. Second, mutations that mediate antigenic escape often have pleiotropic effects, and their success is contingent upon mutations that restore fitness. Antigenic

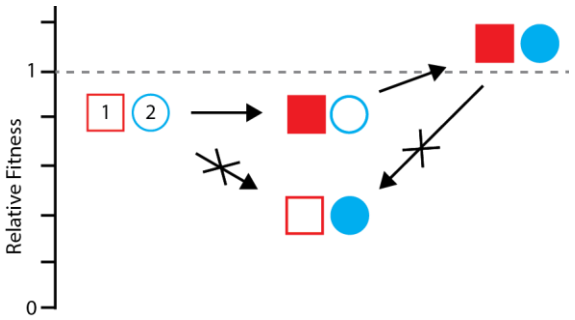


Figure 1-5. Epistasis can constrain adaptation.

The ancestral identity of loci (1) and (2) are shown as unfilled red and blue shapes. A mutation at locus 2 mediates immune escape (filled circle) but is detrimental if it occurs on the ancestral background. Thus, a compensatory mutation at locus 1 (filled square) is required before the escape mutation, limiting the accessibility of the higher fitness genotype (filled square and filled circle). The compensatory mutation also becomes entrenched. Once the antigenic mutation arises, reversion of the compensatory mutation to its ancestral state (unfilled square) would cause a fitness decrease, even though it was initially neutral. Such interactions can occur within or between genes and involve more than two loci.

escape variants in HA can decrease protein folding stability or alter sialic acid binding (Underwood et al. 1987; Mitnaul et al. 2000; Das et al. 2011; Myers et al. 2013; Wu et al. 2017; Kosik et al. 2018), and fitness can be restored by mutations in HA or NA with opposing effects (Mitnaul et al. 2000; Hensley et al. 2011; Kryazhimskiy et al. 2011; Myers et al. 2013; Wu et al. 2017; Kosik et al. 2018). In many cases, the deleterious side effects of an antigenic mutation are larger than its beneficial effects. This constrains adaptation, as the novel, but deleterious, antigenic mutation can only be selected if a compensatory mutation arises first (Figure 1-5) (Weinreich et al. 2005). Since many compensatory mutations are neutral, this often requires that the initial compensatory mutation arise by random drift, hitchhiking, or simultaneously with the novel antigenic mutation.

The adaptive evolution of HA is also constrained by entrenchment, whereby a substitution can no longer revert to its ancestral state without compromising fitness (Figure 1-5). For example, Wu et al. found that a substitution in the receptor binding site of H3, E190D, was reversible to the ancestral state within 10 years after the substitution arose, but not in more recent strains (Wu et al. 2018). Apparently, more recent mutations in the receptor binding site have altered its

structure such that E190 is no longer tolerated. Interestingly, all of the epistatically interacting mutations were located in antigenic sites and could explain why mutations that lead to antigenic changes in HA rarely revert.

Epistasis also influences the adaptive evolution of NA. Neuraminidase phylogenies reveal chains of interacting substitutions similar to those in HA cluster transitions, and resistance to oseltamivir and other neuraminidase inhibitors is constrained by epistasis. Resistance mutations reduce fitness by altering NA stability or enzymatic activity and are contingent on compensatory mutations in HA or NA (Bloom et al. 2010; Abed et al. 2011; Hensley et al. 2011; Ginting et al. 2012; Duan et al. 2014; Neverov et al. 2015). While oseltamivir was first introduced in 1999, the H274Y resistance mutation only arose 8 years later in a much different genetic background (Bloom et al. 2010; Abed et al. 2011; Duan et al. 2014). It then swept the population in a single year. This single epistatic interaction demonstrates how the starting genotype of a strain can have profound effects on whether it can adapt to a new selective pressure.

There are fewer examples of epistasis in other influenza proteins. Contingency has been identified in M2 and NP phylogenies and immune escape mutations in NP (Gong et al. 2013; Nshogozabahizi et al. 2017). We have found that mutations in PA and PB1 interact epistatically to mediate high-level resistance to mutagenic drugs in vitro (Pauly, Lyons, et al. 2017). Epistatic interactions in M2 may also mediate increased resistance to amantadine and/or increase virulence in amantadine-resistant strains (Abed et al. 2005; Dong et al. 2015; Durrant et al. 2015). For example, two mutations associated with amantadine resistance co-occur more frequently than predicted by chance (Durrant et al. 2015); the double mutant has become more prevalent in

recent years (Durrant et al. 2015) and has higher virulence in mice than either single mutant (Abed et al. 2005). Finally, recent work suggests that selection on non-antigenic phenotypes encoded by the remaining six segments can have profound effects on antigenic evolution of influenza virus (Koelle and Rasmussen 2015; Raghwani et al. 2017). Thus, defining epistasis across the genome is an important area for future study.

In contrast to single mutational effects, models and theoretical work on epistasis have outpaced empirical data in influenza virus. A handful of examples in several influenza genes demonstrate that epistasis is common and can lead to evolutionary flexibility via compensation, while at the same time constraining evolution through entrenchment and contingency. However, the general distribution of epistatic effects, including the sign and magnitude of epistasis, across the influenza genome is unknown. The general patterns of epistasis determine the likelihood of compensatory mutation and accessibility of adaptations, the consequences of reassortment, and thus the evolutionary fate of influenza populations. Novel methods are needed to investigate epistasis more extensively across the entire influenza virus genome.

From idiosyncrasy to general patterns in fitness landscapes

There are several puzzles in our picture of how mutational effects scale across levels of biology (Figure 1-2). How does the deterministic action of individual mutations lead to their highly variable effects in different genetic backgrounds at the level of complex phenotypes like fitness? The transition to the level of evolution presents a further problem. How can unpredictable and idiosyncratic individual effects give rise to the predictable and universal patterns in the collective effects of mutations and evolutionary trajectories? The universal patterns are themselves puzzling. According to common intuition, patterns involving beneficial versus detrimental

mutations suggest contradictory biases in epistasis. Current theories of epistasis do not resolve this contradiction and ignore the idiosyncrasy of mutational effects.

My next chapter resolves these puzzles, showing how the universal patterns can arise out of the idiosyncrasy, or near randomness, of mutational effects, without epistatic bias. Furthermore, it presents a null theory of epistasis that explains why deterministic mutational effects are idiosyncratic at higher phenotypic levels. The subsequent two chapters cut through the thicket of idiosyncrasy to find generalizations about mutational effects in RNA viruses. These generalizations help answer long-standing questions in molecular evolution and have implications for phylogenetic inferences and the evolution of the genetic code.

References

Abed Y, Goyette N, Boivin G. 2005. Generation and Characterization of Recombinant Influenza A (H1N1) Viruses Harboring Amantadine Resistance Mutations. Antimicrob Agents Chemother. 49(2):556–559. doi:10.1128/AAC.49.2.556-559.2005.

Abed Y, Pizzorno A, Bouhy X, Boivin G. 2011. Role of Permissive Neuraminidase Mutations in Influenza A/Brisbane/59/2007-like (H1N1) Viruses. PLOS Pathogens. 7(12):e1002431. doi:10.1371/journal.ppat.1002431.

Agarwala A, Fisher DS. 2019. Adaptive walks on high-dimensional fitness landscapes and seascapes with distance-dependent statistics. Theoretical Population Biology. 130:13–49. doi:10.1016/j.tpb.2019.09.011.

Akand EH, Downard KevinM. 2018. Identification of epistatic mutations and insights into the evolution of the influenza virus using a mass-based protein phylogenetic approach. Molecular Phylogenetics and Evolution. 121:132–138. doi:10.1016/j.ympev.2018.01.009.

Araya CL, Fowler DM, Chen W, Muniez I, Kelly JW, Fields S. 2012. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. PNAS. 109(42):16858–16863. doi:10.1073/pnas.1209751109.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature. 461(7268):1243–1247. doi:10.1038/nature08480.

Bendixsen DP, Østman B, Hayden EJ. 2017. Negative Epistasis in Experimental RNA Fitness Landscapes. J Mol Evol. 85(5):159–168. doi:10.1007/s00239-017-9817-5.

Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. 2006. Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. Nature. 444(7121):929–932. doi:10.1038/nature05385.

Bloom JD. 2014. An Experimentally Determined Evolutionary Model Dramatically Improves Phylogenetic Fit. Mol Biol Evol. 31(8):1956–1978. doi:10.1093/molbev/msu173.

Bloom JD. 2017. Identification of positive selection in genes is greatly improved by using experimentally informed site-specific models. Biology Direct. 12:1. doi:10.1186/s13062-016-0172-z.

Bloom JD, Gong LI, Baltimore D. 2010. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. Science. 328(5983):1272–1275. doi:10.1126/science.1187816.

Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, Arnold FH. 2005. Thermodynamic prediction of protein neutrality. PNAS. 102(3):606–611. doi:10.1073/pnas.0406744102.

Bonhoeffer S, Chappey C, Parkin NT, Whitcomb JM, Petropoulos CJ. 2004. Evidence for Positive Epistasis in HIV-1. Science. 306(5701):1547–1550. doi:10.1126/science.1101786.

Boyle MP, De Boeck K. 2013. A new era in the treatment of cystic fibrosis: correction of the underlying CFTR defect. The Lancet Respiratory Medicine. 1(2):158–163. doi:10.1016/S2213-2600(12)70057-7.

Bull JJ, Badgett MR, Wichman HA. 2000. Big-Benefit Mutations in a Bacteriophage Inhibited with Heat. Mol Biol Evol. 17(6):942–950.

Burch CL, Chao L. 1999. Evolution by Small Steps and Rugged Landscapes in the RNA Virus ϕ6. Genetics. 151(3):921–927.

Burch CL, Chao L. 2004. Epistasis and Its Relationship to Canalization in the RNA Virus φ6. Genetics. 167(2):559–567. doi:10.1534/genetics.103.021196.

Campbell PJ, Danzy S, Kyriakis CS, Deymier MJ, Lowen AC, Steel J. 2014. The M Segment of the 2009 Pandemic Influenza Virus Confers Increased Neuraminidase Activity, Filamentous Morphology, and Efficient Contact Transmissibility to A/Puerto Rico/8/1934-Based Reassortant Viruses. J Virol. 88(7):3802–3814. doi:10.1128/JVI.03607-13.

Carlborg Ö, Haley CS. 2004. Epistasis: too often neglected in complex trait studies? Nat Rev Genet. 5(8):618–625. doi:10.1038/nrg1407.

Carrasco P, Iglesia F de la, Elena SF. 2007. Distribution of Fitness and Virulence Effects Caused by Single-Nucleotide Substitutions in Tobacco Etch Virus. J Virol. 81(23):12979–12984. doi:10.1128/JVI.00524-07.

Caudle SB, Miller CR, Rokyta DR. 2014. Environment Determines Epistatic Patterns for a ssDNA Virus. Genetics. 196(1):267–279. doi:10.1534/genetics.113.158154.

Chao L. 1988. Evolution of sex in RNA viruses. Journal of Theoretical Biology. 133(1):99–112. doi:10.1016/S0022-5193(88)80027-4.

Chen Q, Lan A, Shen X, Wu C-I. 2019. Molecular Evolution in Small Steps under Prevailing Negative Selection: A Nearly Universal Rule of Codon Substitution. Genome Biol Evol. 11(10):2702–2712. doi:10.1093/gbe/evz192.

Chou H-H, Chiu H-C, Delaney NF, Segrè D, Marx CJ. 2011. Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. Science. 332(6034):1190–1192. doi:10.1126/science.1203799.

Chou H-H, Delaney NF, Draghi JA, Marx CJ. 2014. Mapping the Fitness Landscape of Gene Expression Uncovers the Cause of Antagonism and Sign Epistasis between Adaptive Mutations. PLOS Genetics. 10(2):e1004149. doi:10.1371/journal.pgen.1004149.

Couce A, Tenaillon OA. 2015. The rule of declining adaptability in microbial evolution experiments. Front Genet. 6. doi:10.3389/fgene.2015.00099. [accessed 2018 Dec 5]. https://www.frontiersin.org/articles/10.3389/fgene.2015.00099/full.

Crotty S, Cameron CE, Andino R. 2001. RNA virus error catastrophe: Direct molecular test by using ribavirin. PNAS. 98(12):6895–6900. doi:10.1073/pnas.111085598.

Danzy S, Studdard LR, Manicassamy B, Solorzano A, Marshall N, García-Sastre A, Steel J, Lowen AC. 2014. Mutations to PB2 and NP Proteins of an Avian Influenza Virus Combine To Confer Efficient Growth in Primary Human Respiratory Cells. J Virol. 88(22):13436–13446. doi:10.1128/JVI.01093-14.

Das SR, Hensley SE, David A, Schmidt L, Gibbs JS, Puigbò P, Ince WL, Bennink JR, Yewdell JW. 2011. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. PNAS. 108(51):E1417–E1422. doi:10.1073/pnas.1108754108.

Das SR, Hensley SE, Ince WL, Brooke CB, Subba A, Delboy MG, Russ G, Gibbs JS, Bennink JR, Yewdell JW. 2013. Defining Influenza A Virus Hemagglutinin Antigenic Drift by Sequential Monoclonal Antibody Selection. Cell Host & Microbe. 13(3):314–323. doi:10.1016/j.chom.2013.02.008.

Dawson WK, Lazniewski M, Plewczynski D. RNA structure interactions and ribonucleoprotein processes of the influenza A virus. Brief Funct Genomics. doi:10.1093/bfgp/elx028. [accessed 2018 May 25]. http://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elx028/4430348.

Debbink K, McCrone JT, Petrie JG, Truscon R, Johnson E, Mantlo EK, Monto AS, Lauring AS. 2017. Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. PLOS Pathogens. 13(1):e1006194. doi:10.1371/journal.ppat.1006194.

deVisser JA, Hoekstra RF, van den Ende H. 1997. Test of Interaction Between Genetic Markers That Affect Fitness in Aspergillus niger. Evolution. 51(5):1499–1505. doi:10.2307/2411202.

Dinis JM, Florek NW, Fatola OO, Moncla LH, Mutschler JP, Charlier OK, Meece JK, Belongia EA, Friedrich TC. 2016. Deep sequencing reveals potential antigenic variants at low frequency in influenza A-infected humans. J Virol. 90(7):3355–3365. doi:10.1128/jvi.03248-15.

Domingo J, Diss G, Lehner B. 2018. Pairwise and higher-order genetic interactions during the evolution of a tRNA. Nature. doi:10.1038/s41586-018-0170-7. [accessed 2018 Jun 5]. http://www.nature.com/articles/s41586-018-0170-7.

Domingo-Calap P, Cuevas JM, Sanjuá N R. 2009. The Fitness Effects of Random Mutations in Single-Stranded DNA and RNA Bacteriophages. doi:10.1371/journal.pgen.1000742.

Dong G, Peng C, Luo J, Wang C, Han L, Wu B, Ji G, He H. 2015. Adamantane-resistant influenza a viruses in the world (1902-2013): frequency and distribution of M2 gene mutations. PLoS ONE. 10(3):e0119115. doi:10.1371/journal.pone.0119115.

Doud MB, Ashenberg O, Bloom JD. 2015. Site-Specific Amino Acid Preferences Are Mostly Conserved in Two Closely Related Protein Homologs. Mol Biol Evol. 32(11):2944–2960. doi:10.1093/molbev/msv167.

Doud MB, Bloom JD. 2016. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. Viruses. 8(6):155. doi:10.3390/v8060155.

Doud MB, Hensley SE, Bloom JD. 2017. Complete mapping of viral escape from neutralizing antibodies. PLOS Pathogens. 13(3):e1006271. doi:10.1371/journal.ppat.1006271.

Doud MB, Lee JM, Bloom JD. 2018. How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin. Nature Communications. 9(1):1386. doi:10.1038/s41467-018-03665-3.

Du Y, Wu NC, Jiang L, Zhang T, Gong D, Shu S, Wu T-T, Sun R. 2016. Annotating Protein Functional Residues by Coupling High-Throughput Fitness Profile and Homologous-Structure Analysis. mBio. 7(6):e01801-16. doi:10.1128/mBio.01801-16.

Du Y, Xin L, Shi Y, Zhang T-H, Wu NC, Dai L, Gong D, Brar G, Shu S, Luo J, et al. 2018. Genome-wide identification of interferon-sensitive mutations enables influenza vaccine design. Science. 359(6373):290–296. doi:10.1126/science.aan8806.

Duan S, Govorkova EA, Bahl J, Zaraket H, Baranovich T, Seiler P, Prevost K, Webster RG, Webby RJ. 2014. Epistatic interactions between neuraminidase mutations facilitated the emergence of the oseltamivir-resistant H1N1 influenza viruses. Nature Communications. 5:5029. doi:10.1038/ncomms6029.

Durrant MG, Eggett DL, Busath DD. 2015. Investigation of a recent rise of dual amantadine-resistance mutations in the influenza A M2 sequence. BMC Genetics. 16(2):S3. doi:10.1186/1471-2156-16-S2-S3.

Earl DJ, Deem MW. 2004. Evolvability is a selectable trait. PNAS. 101(32):11531–11536. doi:10.1073/pnas.0404656101.

Erbelding EJ, Post DJ, Stemmy EJ, Roberts PC, Augustine AD, Ferguson S, Paules CI, Graham BS, Fauci AS. A Universal Influenza Vaccine: The Strategic Plan for the National Institute of Allergy and Infectious Diseases. J Infect Dis. doi:10.1093/infdis/jiy103. [accessed 2018 May 29]. http://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiy103/4904047.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat Rev Genet. 8(8):610–618. doi:10.1038/nrg2146.

Fields BN, Knipe DM, Howley PM, Ovid Technologies Inc. 2013. Fields virology. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins (Virology). http://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=n&CSC=Y&PAGE=booktext&D=books&AN=01735125$&XPATH=/PG(0).

Flynn KM, Cooper TF, Moore FB-G, Cooper VS. 2013. The Environment Affects Epistatic Interactions to Alter the Topology of an Empirical Fitness Landscape. PLOS Genetics. 9(4):e1003426. doi:10.1371/journal.pgen.1003426.

Fowler DM, Fields S. 2014. Deep mutational scanning: a new style of protein science. Nat Meth. 11(8):801–807. doi:10.1038/nmeth.3027.

Fragata I, Blanckaert A, Dias Louro MA, Liberles DA, Bank C. 2019. Evolution in the light of fitness landscape theory. Trends in Ecology & Evolution. 34(1):69–82. doi:10.1016/j.tree.2018.10.009.

Fulton BO, Sachs D, Beaty SM, Won ST, Lee B, Palese P, Heaton NS. 2015. Mutational Analysis of Measles Virus Suggests Constraints on Antigenic Variation of the Glycoproteins. Cell Reports. 11(9):1331–1338. doi:10.1016/j.celrep.2015.04.054.

Ginting TE, Shinya K, Kyan Y, Makino A, Matsumoto N, Kaneda S, Kawaoka Y. 2012. Amino Acid Changes in Hemagglutinin Contribute to the Replication of Oseltamivir-Resistant H1N1 Influenza Viruses. J Virol. 86(1):121–127. doi:10.1128/JVI.06085-11.

Gojobori J, Tang H, Akey JM, Wu C-I. 2007. Adaptive evolution in humans revealed by the negative correlation between the polymorphism and fixation phases of evolution. PNAS. 104(10):3907–3912. doi:10.1073/pnas.0605565104.

Gong LI, Suchard MA, Bloom JD. 2013. Stability-mediated epistasis constrains the evolution of an influenza protein. eLife. 2:e00631. doi:10.7554/eLife.00631.

Gros P-A, Nagard HL, Tenaillon O. 2009. The Evolution of Epistasis and Its Links With Genetic Robustness, Complexity and Drift in a Phenotypic Model of Adaptation. Genetics. 182(1):277–293. doi:10.1534/genetics.108.099127.

Haddox HK, Dingens AS, Hilton SK, Overbaugh J, Bloom JD. 2018. Mapping mutational effects along the evolutionary landscape of HIV envelope. Chakraborty AK, editor. eLife. 7:e34420. doi:10.7554/eLife.34420.

Heaton NS, Sachs D, Chen C-J, Hai R, Palese P. 2013. Genome-wide mutagenesis of influenza virus reveals unique plasticity of the hemagglutinin and NS1 proteins. PNAS. 110(50):20248–20253. doi:10.1073/pnas.1320524110.

Hensley SE, Das SR, Gibbs JS, Bailey AL, Schmidt LM, Bennink JR, Yewdell JW. 2011. Influenza A Virus Hemagglutinin Antibody Escape Promotes Neuraminidase Antigenic Variation and Drug Resistance. PLOS ONE. 6(2):e15190. doi:10.1371/journal.pone.0015190.

Hilton SK, Bloom JD. 2018 Apr 17. Modeling site-specific amino-acid preferences deepens phylogenetic estimates of viral divergence. bioRxiv.:302703. doi:10.1101/302703.

Hilton SK, Doud MB, Bloom JD. 2017. phydms: software for phylogenetic analyses informed by deep mutational scanning. PeerJ. 5:e3657. doi:10.7717/peerj.3657.

Hwang S, Schmiegelt B, Ferretti L, Krug J. 2018. Universality Classes of Interaction Structures for NK Fitness Landscapes. J Stat Phys. 172(1):226–278. doi:10.1007/s10955-018-1979-z.

Iglesia F de la, Elena SF. 2007. Fitness Declines in Tobacco Etch Virus upon Serial Bottleneck Transfers. J Virol. 81(10):4941–4947. doi:10.1128/JVI.02528-06.

Iqbal M, Xiao H, Baillie G, Warry A, Essen SC, Londt B, Brookes SM, Brown IH, McCauley JW. 2009. Within-host variation of avian influenza viruses. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 364(1530):2739–2747. doi:10.1098/rstb.2009.0088.

Johnson MS, Martsul A, Kryazhimskiy S, Desai MM. 2019. Higher-fitness yeast genotypes are less robust to deleterious mutations. Science. 366(6464):490–493. doi:10.1126/science.aay4199.

Kacser H, Burns JA. 1981. The Molecular Basis of Dominance. Genetics. 97(3–4):639–666.

Kauffman S, Levin S. 1987. Towards a general theory of adaptive walks on rugged landscapes. Journal of Theoretical Biology. 128(1):11–45. doi:10.1016/S0022-5193(87)80029-2.

Kauffman SA. 1993. The Origins of Order: Self-organization and Selection in Evolution. Oxford University Press.

Kauffman SA, Weinberger ED. 1989. The NK model of rugged fitness landscapes and its application to maturation of the immune response. Journal of Theoretical Biology. 141(2):211–245. doi:10.1016/S0022-5193(89)80019-0.

Keightley PD, Halligan DL. 2011. Inference of Site Frequency Spectra From High-Throughput Sequence Data: Quantification of Selection on Nonsynonymous and Synonymous Sites in Humans. Genetics. 188(4):931–940. doi:10.1534/genetics.111.128355.

Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. 2011. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. Science. 332(6034):1193–1196. doi:10.1126/science.1203801.

Koelle K, Cobey S, Grenfell B, Pascual M. 2006. Epochal Evolution Shapes the Phylodynamics of Interpandemic Influenza A (H3N2) in Humans. Science. 314(5807):1898–1903. doi:10.1126/science.1132745.

Koelle K, Rasmussen DA. 2015. The effects of a deleterious mutation load on patterns of influenza A/H3N2's antigenic evolution in humans. eLife. 4:e07361. doi:10.7554/eLife.07361.

Kondrashov AS. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature. 336(6198):435–440. doi:10.1038/336435a0.

Kosik I, Ince WL, Gentles LE, Oler AJ, Kosikova M, Angel M, Magadán JG, Xie H, Brooke CB, Yewdell JW. 2018. Influenza A virus hemagglutinin glycosylation compensates for antibody escape fitness costs. PLOS Pathogens. 14(1):e1006796. doi:10.1371/journal.ppat.1006796.

Kouyos RD, Silander OK, Bonhoeffer S. 2007. Epistasis between deleterious mutations and the evolution of recombination. Trends in Ecology & Evolution. 22(6):308–315. doi:10.1016/j.tree.2007.02.014.

Kryazhimskiy S, Dushoff J, Bazykin GA, Plotkin JB. 2011. Prevalence of Epistasis in the Evolution of Influenza A Surface Proteins. PLOS Genet. 7(2):e1001301. doi:10.1371/journal.pgen.1001301.

Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. 2014. Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science. 344(6191):1519–1522. doi:10.1126/science.1250939.

Kryazhimskiy S, Tkačik G, Plotkin JB. 2009. The dynamics of adaptation on correlated fitness landscapes. PNAS. 106(44):18638–18643. doi:10.1073/pnas.0905497106.

Kvitek DJ, Sherlock G. 2011. Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape. PLOS Genetics. 7(4):e1002056. doi:10.1371/journal.pgen.1002056.

Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, Bloom JD. 2018. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. PNAS. 115(35):E8276–E8285. doi:10.1073/pnas.1806133115.

Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, Bloom JD. 2018 Apr 10. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. bioRxiv.:298364. doi:10.1101/298364.

Lehner B. 2011. Molecular mechanisms of epistasis within and between genes. Trends in Genetics. 27(8):323–331. doi:10.1016/j.tig.2011.05.007.

Lenski RE, Ofria C, Collier TC, Adami C. 1999. Genome complexity, robustness and genetic interactions in digital organisms. Nature. 400(6745):661. doi:10.1038/23245.

Leonard AS, McClain MT, Smith GJD, Wentworth DE, Halpin RA, Lin X, Ransier A, Stockwell TB, Das SR, Gilbert AS, et al. 2017. The effective rate of influenza reassortment is limited during human infection. PLOS Pathogens. 13(2):e1006203. doi:10.1371/journal.ppat.1006203.

Li C, Qian W, Maclean CJ, Zhang J. 2016. The fitness landscape of a tRNA gene. Science. 352(6287):837–840. doi:10.1126/science.aae0568.

Łuksza M, Lässig M. 2014. A predictive fitness model for influenza. Nature. 507(7490):57–61. doi:10.1038/nature13087.

Lyons DM, Lauring AS. 2017. Evidence for the Selective Basis of Transition-to-Transversion Substitution Bias in Two RNA Viruses. Mol Biol Evol. 34(12):3205–3215. doi:10.1093/molbev/msx251.

Maclean RC. 2010. Predicting epistasis: an experimental test of metabolic control theory with bacterial transcription and translation. Journal of Evolutionary Biology. 23(3):488–493. doi:10.1111/j.1420-9101.2009.01888.x.

MacLean RC, Perron GG, Gardner A. 2010. Diminishing Returns From Beneficial Mutations and Pervasive Epistasis Shape the Fitness Landscape for Rifampicin Resistance in Pseudomonas aeruginosa. Genetics. 186(4):1345–1354. doi:10.1534/genetics.110.123083.

Maisnier-Patin S, Roth JR, Fredriksson Å, Nyström T, Berg OG, Andersson DI. 2005. Genomic buffering mitigates the effects of deleterious mutations in bacteria. Nat Genet. 37(12):1376–1379. doi:10.1038/ng1676.

Mani R, St.Onge RP, Hartman JL, Giaever G, Roth FP. 2008. Defining genetic interaction. PNAS. 105(9):3461–3466. doi:10.1073/pnas.0712255105.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. 2009. Finding the missing heritability of complex diseases. Nature. 461(7265):747–753. doi:10.1038/nature08494.

Martin G, Elena SF, Lenormand T. 2007. Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nat Genet. 39(4):555–560. doi:10.1038/ng1998.

McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. 2018. Stochastic processes constrain the within and between host evolution of influenza virus. eLife Sciences. 7:e35962. doi:10.7554/eLife.35962.

Medina RA, García-Sastre A. 2011. Influenza A viruses: new research developments. Nat Rev Micro. 9(8):590–603. doi:10.1038/nrmicro2613.

Melamed D, Young DL, Gamble CE, Miller CR, Fields S. 2013. Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. RNA. 19(11):1537–1551. doi:10.1261/rna.040709.113.

Miller CR. 2019. The treacheries of adaptation. Science. 366(6464):418–419. doi:10.1126/science.aaz5189.

Mitnaul LJ, Matrosovich MN, Castrucci MR, Tuzikov AB, Bovin NV, Kobasa D, Kawaoka Y. 2000. Balanced Hemagglutinin and Neuraminidase Activities Are Critical for Efficient Replication of Influenza A Virus. J Virol. 74(13):6015–6020.

Moore JH. 2003. The Ubiquitous Nature of Epistasis in Determining Susceptibility to Common Human Diseases. HHE. 56(1–3):73–82. doi:10.1159/000073735.

Morens DM, Taubenberger JK, Fauci AS. 2009. The Persistent Legacy of the 1918 Influenza Virus. New England Journal of Medicine. 361(3):225–229. doi:10.1056/NEJMp0904819.

Morris DH, Gostic KM, Pompei S, Bedford T, Łuksza M, Neher RA, Grenfell BT, Lässig M, McCauley JW. 2017. Predictive Modeling of Influenza Shows the Promise of Applied

Evolutionary Biology. Trends in Microbiology. 0(0). doi:10.1016/j.tim.2017.09.004. [accessed 2017 Nov 8]. http://www.cell.com/trends/microbiology/abstract/S0966-842X(17)30209-3.

Murcia PR, Baillie GJ, Daly J, Elton D, Jervis C, Mumford JA, Newton R, Parrish CR, Hoelzer K, Dougan G, et al. 2010. Intra- and Interhost Evolutionary Dynamics of Equine Influenza Virus. J Virol. 84(14):6943–6954. doi:10.1128/JVI.00112-10.

Myers JL, Wetzel KS, Linderman SL, Li Y, Sullivan CB, Hensley SE. 2013. Compensatory Hemagglutinin Mutations Alter Antigenic Properties of Influenza Viruses. J Virol. 87(20):11168–11172. doi:10.1128/JVI.01414-13.

Nahum JR, Godfrey-Smith P, Harding BN, Marcus JH, Carlson-Stevermer J, Kerr B. 2015. A tortoise-hare pattern seen in adapting structured and unstructured populations suggests a rugged fitness landscape in bacteria. Proc Natl Acad Sci USA. 112(24):7530–7535. doi:10.1073/pnas.1410631112.

Nakajima K, Nobusawa E, Nagy A, Nakajima S. 2005. Accumulation of Amino Acid Substitutions Promotes Irreversible Structural Changes in the Hemagglutinin of Human Influenza AH3 Virus during Evolution. J Virol. 79(10):6472–6477. doi:10.1128/JVI.79.10.6472-6477.2005.

Neidhart J, Szendro IG, Krug J. 2014. Adaptation in Tunably Rugged Fitness Landscapes: The Rough Mount Fuji Model. Genetics. 198(2):699–721. doi:10.1534/genetics.114.167668.

Neverov AD, Kryazhimskiy S, Plotkin JB, Bazykin GA. 2015. Coordinated Evolution of Influenza A Surface Proteins. PLOS Genetics. 11(8):e1005404. doi:10.1371/journal.pgen.1005404.

Neverov AD, Lezhnina KV, Kondrashov AS, Bazykin GA. 2014. Intrasubtype Reassortments Cause Adaptive Amino Acid Replacements in H3N2 Influenza Genes. PLOS Genetics. 10(1):e1004037. doi:10.1371/journal.pgen.1004037.

Nielsen R, Yang Z. 2003. Estimating the Distribution of Selection Coefficients from Phylogenetic Data with Applications to Mitochondrial and Viral DNA. Mol Biol Evol. 20(8):1231–1239. doi:10.1093/molbev/msg147.

Nowak S, Krug J. 2015. Analysis of adaptive walks on NK fitness landscapes with different interaction schemes. J Stat Mech. 2015(6):P06014. doi:10.1088/1742-5468/2015/06/P06014.

Nshogozabahizi JC, Dench J, Aris-Brosou S. 2017. Widespread Historical Contingency in Influenza Viruses. Genetics. 205(1):409–420. doi:10.1534/genetics.116.193979.

Olson CA, Wu NC, Sun R. 2014. A Comprehensive Biophysical Description of Pairwise Epistasis throughout an Entire Protein Domain. Current Biology. 24(22):2643–2651. doi:10.1016/j.cub.2014.09.072.

Pauly MD, Lyons DM, Fitzsimmons WJ, Lauring AS. 2017. Epistatic Interactions within the Influenza A Virus Polymerase Complex Mediate Mutagen Resistance and Replication Fidelity. mSphere. 2(4):e00323-17. doi:10.1128/mSphere.00323-17.

Pauly MD, Procario MC, Lauring AS. 2017. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. eLife Sciences. 6:e26437. doi:10.7554/eLife.26437.

Pearson VM, Miller CR, Rokyta DR. 2012. The Consistency of Beneficial Fitness Effects of Mutations across Diverse Genetic Backgrounds. PLOS ONE. 7(8):e43864. doi:10.1371/journal.pone.0043864.

Pelleymounter MA, Cullen MJ, Baker MB, Hecht R, Winters D, Boone T, Collins F. 1995. Effects of the obese gene product on body weight regulation in ob/ob mice. Science. 269(5223):540–543. doi:10.1126/science.7624776.

Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, Macleod M, Mignini LE, Jayaram P, Khan KS. 2007. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ. 334(7586):197. doi:10.1136/bmj.39048.407928.BE.

Perfeito L, Sousa A, Bataillon T, Gordo I. 2014. Rates of Fitness Decline and Rebound Suggest Pervasive Epistasis. Evolution. 68(1):150–162. doi:10.1111/evo.12234.

Peris JB, Davis P, Cuevas JM, Nebot MR, Sanjuán R. 2010. Distribution of Fitness Effects Caused by Single-Nucleotide Substitutions in Bacteriophage f1. Genetics. 185(2):603–609. doi:10.1534/genetics.110.115162.

Phillips PC. 2008. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 9(11):855–867. doi:10.1038/nrg2452.

Plotkin JB, Dushoff J. 2003. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. PNAS. 100(12):7152–7157. doi:10.1073/pnas.1132114100.

Poelwijk FJ, Socolich M, Ranganathan R. 2019. Learning the pattern of epistasis linking genotype and phenotype in a protein. Nat Commun. 10(1):1–11. doi:10.1038/s41467-019-12130-8.

Poon A, Otto SP. 2000. Compensating for Our Load of Mutations: Freezing the Meltdown of Small Populations. Evolution. 54(5):1467–1479. doi:10.1111/j.0014-3820.2000.tb00693.x.

Poon LLM, Song T, Rosenfeld R, Lin X, Rogers MB, Zhou B, Sebra R, Halpin RA, Guan Y, Twaddle A, et al. 2016. Quantifying influenza virus diversity and transmission in humans. Nature Genetics. 48(2):195–200. doi:10.1038/ng.3479.

Puchta O, Cseke B, Czaja H, Tollervey D, Sanguinetti G, Kudla G. 2016. Network of epistatic interactions within a yeast snoRNA. Science. 352(6287):840–844. doi:10.1126/science.aaf0965.

Pybus OG, Rambaut A, Belshaw R, Freckleton RP, Drummond AJ, Holmes EC. 2007. Phylogenetic Evidence for Deleterious Mutation Load in RNA Viruses and Its Contribution to Viral Evolution. Mol Biol Evol. 24(3):845–852. doi:10.1093/molbev/msm001.

Raghwani J, Thompson RN, Koelle K. 2017. Selection on non-antigenic gene segments of seasonal influenza A virus and its impact on adaptive evolution. Virus Evol. 3(2). doi:10.1093/ve/vex034. [accessed 2018 Mar 21]. https://academic.oup.com/ve/article/3/2/vex034/4614565.

Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. 2008. The genomic and epidemiological dynamics of human influenza A virus. Nature. 453(7195):615–619. doi:10.1038/nature06945.

Reia SM, Campos PRA. Analysis of statistical correlations between properties of adaptive walks in fitness landscapes. Royal Society Open Science. 7(1):192118. doi:10.1098/rsos.192118.

Rokyta DR, Joyce P, Caudle SB, Miller C, Beisel CJ, Wichman HA. 2011. Epistasis between Beneficial Mutations and the Phenotype-to-Fitness Map for a ssDNA Virus. PLOS Genetics. 7(6):e1002075. doi:10.1371/journal.pgen.1002075.

Sanjuán R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 365(1548):1975–82. doi:10.1098/rstb.2010.0063.

Sanjuán R, Cuevas JM, Moya A, Elena SF. 2005. Epistasis and the Adaptability of an RNA Virus. Genetics. 170(3):1001–1008. doi:10.1534/genetics.105.040741.

Sanjuán R, Elena SF. 2006. Epistasis correlates to genomic complexity. PNAS. 103(39):14402–14405. doi:10.1073/pnas.0604543103.

Sanjuán R, Moya A, Elena SF. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. PNAS. 101(22):8396–8401. doi:10.1073/pnas.0400146101.

Sanjuán R, Nebot MR. 2008. A Network Model for the Correlation between Epistasis and Genomic Complexity. PLOS ONE. 3(7):e2663. doi:10.1371/journal.pone.0002663.

Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, et al. 2016. Local fitness landscape of the green fluorescent protein. Nature. 533(7603):397–401. doi:10.1038/nature17995.

Schenk MF, Szendro IG, Salverda MLM, Krug J, Visser JAGM de. 2013. Patterns of Epistasis between Beneficial Mutations in an Antibiotic Resistance Gene. Mol Biol Evol. 30(8):1779–1787. doi:10.1093/molbev/mst096.

Schoustra S, Hwang S, Krug J, Visser JAGM de. 2016. Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. Proc R Soc B. 283(1837):20161376. doi:10.1098/rspb.2016.1376.

Segrè D, DeLuna A, Church GM, Kishony R. 2005. Modular epistasis in yeast metabolism. Nature Genetics. 37(1):77–83. doi:10.1038/ng1489.

Shapiro B, Rambaut A, Pybus OG, Holmes EC. 2006. A Phylogenetic Method for Detecting Positive Epistasis in Gene Sequences and Its Application to RNA Virus Evolution. Mol Biol Evol. 23(9):1724–1730. doi:10.1093/molbev/msl037.

Soh YS, Moncla LH, Eguia R, Bedford T, Bloom JD. 2019. Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans. Ferguson NM, Kawaoka Y, Fodor E, tenOever BR, editors. eLife. 8:e45079. doi:10.7554/eLife.45079.

Stent GS. 1964. The Operon: On Its Third Anniversary. Science. 144(3620):816–820.

Taggi L, Colaiori F, Loreto V, Tria F. 2013. Dynamical correlations in the escape strategy of Influenza A virus. EPL. 101(6):68003. doi:10.1209/0295-5075/101/68003.

Tamuri AU, Reis M dos, Goldstein RA. 2012. Estimating the Distribution of Selection Coefficients from Phylogenetic Data Using Sitewise Mutation-Selection Models. Genetics. 190(3):1101–1115. doi:10.1534/genetics.111.136432.

Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. eLife. 3:e03300. doi:10.7554/eLife.03300.

Tria F, Pompei S, Loreto V. 2013. Dynamically correlated mutations drive human Influenza A evolution. Scientific Reports. 3:2705. doi:10.1038/srep02705.

Underwood PA, Skehel JJ, Wiley DC. 1987. Receptor-binding characteristics of monoclonal antibody-selected antigenic variants of influenza virus. J Virol. 61(1):206–208.

Villa M, Lässig M. 2017. Fitness cost of reassortment in human influenza. PLOS Pathogens. 13(11):e1006685. doi:10.1371/journal.ppat.1006685.

Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS. 2016. The Mutational Robustness of Influenza A Virus. PLOS Pathog. 12(8):e1005856. doi:10.1371/journal.ppat.1005856.

Wang Y, Arenas CD, Stoebel DM, Flynn K, Knapp E, Dillon MM, Wünsche A, Hatcher PJ, Moore FB-G, Cooper VS, et al. 2016. Benefit of transferred mutations is better predicted by the fitness of recipients than by their ecological or genetic relatedness. PNAS. 113(18):5047–5052. doi:10.1073/pnas.1524988113.

Wang, Yinhua, Arenas, Carolina Díaz, Stoebel, Daniel M., Cooper, Tim F. 2013. Genetic background affects epistatic interactions between two beneficial mutations. Biology Letters. 9(1):20120328. doi:10.1098/rsbl.2012.0328.

Ward MJ, Lycett SJ, Avila D, Bollback JP, Leigh Brown AJ. 2013. Evolutionary interactions between haemagglutinin and neuraminidase in avian influenza. BMC Evolutionary Biology. 13:222. doi:10.1186/1471-2148-13-222.

Watanabe T, Watanabe S, Noda T, Fujii Y, Kawaoka Y. 2003. Exploitation of Nucleic Acid Packaging Signals To Generate a Novel Influenza Virus-Based Vector Stably Expressing Two Foreign Genes. J Virol. 77(19):10575–10583. doi:10.1128/JVI.77.19.10575-10583.2003.

Wei X, Zhang J. 2019. Patterns and Mechanisms of Diminishing Returns from Beneficial Mutations. Mol Biol Evol. 36(5):1008–1021. doi:10.1093/molbev/msz035.

Weinreich DM, Watson RA, Chao L, Harrison R. 2005. Perspective:sign epistasis and genetic constraint on evolutionary trajectories. Evolution. 59(6):1165–1174. doi:10.1554/04-272.

Wilke CO, Christoph A. 2001. Interaction between directional epistasis and average mutational effects. Proceedings of the Royal Society of London B: Biological Sciences. 268(1475):1469–1474. doi:10.1098/rspb.2001.1690.

Wilke CO, Lenski RE, Adami C. 2003. Compensatory mutations cause excess of antagonistic epistasis in RNA secondary structure folding. BMC Evolutionary Biology. 3:3. doi:10.1186/1471-2148-3-3.

Wiser MJ, Ribeck N, Lenski RE. 2013. Long-Term Dynamics of Adaptation in Asexual Populations. Science. 342(6164):1364–1367. doi:10.1126/science.1243357.

Wu NC, Du Y, Le S, Young AP, Zhang T-H, Wang Y, Zhou J, Yoshizawa JM, Dong L, Li X, et al. 2016. Coupling high-throughput genetics with phylogenetic information reveals an epistatic interaction on the influenza A virus M segment. BMC Genomics. 17:46. doi:10.1186/s12864-015-2358-7.

Wu NC, Olson CA, Du Y, Le S, Tran K, Remenyi R, Gong D, Al-Mawsawi LQ, Qi H, Wu T-T, et al. 2015. Functional Constraint Profiling of a Viral Protein Reveals Discordance of Evolutionary Conservation and Functionality. PLOS Genetics. 11(7):e1005310. doi:10.1371/journal.pgen.1005310.

Wu NC, Thompson AJ, Xie J, Lin C-W, Nycholat CM, Zhu X, Lerner RA, Paulson JC, Wilson IA. 2018. A complex epistatic network limits the mutational reversibility in the influenza hemagglutinin receptor-binding site. Nature Communications. 9(1):1264. doi:10.1038/s41467-018-03663-5.

Wu NC, Xie J, Zheng T, Nycholat CM, Grande G, Paulson JC, Lerner RA, Wilson IA. 2017. Diversity of Functionally Permissive Sequences in the Receptor-Binding Site of Influenza Hemagglutinin. Cell Host & Microbe. 21(6):742-753.e8. doi:10.1016/j.chom.2017.05.011.

Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, Chen S-H, Lu I-H, Lin C-Y, Chin RG, et al. 2014. High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. Scientific Reports. 4:4942. doi:10.1038/srep04942.

Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, Luan HH, Li X, Wu T-T, Sun R. 2014. High-Throughput Identification of Loss-of-Function Mutations for Anti-Interferon Activity in the Influenza A Virus NS Segment. J Virol. 88(17):10157–10164. doi:10.1128/JVI.01494-14.

Wu NC, Young AP, Dandekar S, Wijersuriya H, Al-Mawsawi LQ, Wu T-T, Sun R. 2013. Systematic identification of H274Y compensatory mutations in influenza A virus neuraminidase by high-throughput screening. Journal of virology. 87(2):1193–9. doi:10.1128/JVI.01658-12.

Wünsche A, Dinh DM, Satterwhite RS, Arenas CD, Stoebel DM, Cooper TF. 2017. Diminishing-returns epistasis decreases adaptability along an evolutionary trajectory. Nature Ecology & Evolution. 1(4):0061. doi:10.1038/s41559-016-0061.

Xu J, Zhang J. 2014. Why human disease-associated residues appear as the wild-type in other species: Genome-scale structural evidence for the compensation hypothesis. Molecular Biology and Evolution. 31(7):1787–1792. doi:10.1093/molbev/msu130.

Xue KS, Bloom JD. 2018 Jul 8. Reconciling disparate estimates of viral genetic diversity during human influenza infections. bioRxiv.:364430. doi:10.1101/364430.

Xue KS, Stevens-Ayers T, Campbell AP, Englund JA, Pergam SA, Boeckh M, Bloom JD. 2017. Parallel evolution of influenza across multiple spatiotemporal scales. eLife Sciences. 6:e26875. doi:10.7554/eLife.26875.

You L, Yin J. 2002. Dependence of Epistasis on Environment and Mutation Severity as Revealed by in Silico Mutagenesis of Phage T7. Genetics. 160(4):1273–1281.

Zagorski M, Burda Z, Waclaw B. 2016. Beyond the Hypercube: Evolutionary Accessibility of Fitness Landscapes with Realistic Mutational Networks. PLOS Computational Biology. 12(12):e1005218. doi:10.1371/journal.pcbi.1005218.

Zeldovich KB, Liu P, Renzette N, Foll M, Pham ST, Venev SV, Gallagher GR, Bolon DN, Kurt-Jones EA, Jensen JD, et al. 2015. Positive Selection Drives Preferred Segment Combinations during Influenza Virus Reassortment. Mol Biol Evol. 32(6):1519–1532. doi:10.1093/molbev/msv044.

Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J Mol Evol. 50(1):56–68.

Chapter 2 Idiosyncratic Epistasis Creates Universals in Mutational Effects and Evolutionary Trajectories

Note: This chapter is a modified version of the article under review:

Daniel M. Lyons*, Zhengting Zou*, Haiqing Xu, and Jianzhi Zhang. Idiosyncratic epistasis creates universals in mutational effects and evolutionary trajectories. *Under review* at Nature Ecology and Evolution.

Abstract

Patterns of epistasis and shapes of fitness landscapes are of wide interest because of their bearings on a number of evolutionary theories. The common phenomena of slowing fitness increases during adaptations and diminishing returns from beneficial mutations are believed to reflect a concave fitness landscape and a preponderance of negative epistasis. Paradoxically, fitness drops tend to decelerate and harm from deleterious mutations shrinks during accumulation of random mutations, patterns thought to indicate a convex fitness landscape and a predominance of positive epistasis. Current theories cannot resolve this apparent contradiction. Here we show that the phenotypic effect of a mutation varies substantially depending on the specific genetic background and that this idiosyncrasy in epistasis creates all of the above trends without requiring a biased distribution of epistasis. The idiosyncratic epistasis theory explains the universalities in mutational effects and evolutionary trajectories as emerging from randomness due to biological complexity.

Introduction

Epistasis, or genetic interaction among a set of mutations, impacts the phenotypic effects of mutations and shapes fundamental evolutionary processes[1]. Epistasis is said to be positive (or

negative) for a particular trait such as fitness if the trait value of the individual with multiple mutations is greater (or smaller) than the expectation from the corresponding single mutants under no epistasis[1]. A number of evolutionary theories such as the mutational deterministic hypothesis of the evolution of sexual reproduction[2] and the hypothesis of reduction in mutational load by truncation selection against deleterious mutations depend on assumptions of general trends of epistasis[3]. Universal patterns involving epistasis are emerging from decades of intense investigations[4,5]. For instance, many experimental evolution studies have shown that fitness increase slows during the organismal adaptation to a constant environment[6]. While the speed of fitness increase is typically measured per unit time[6], the same trend is observed when the speed is measured per mutation accrued[7]. This phenomenon of slowing adaptation is at least in part due to diminishing returns epistasis, a common observation that advantageous mutations are less beneficial on fitter genetic backgrounds[8-11]. Because diminishing returns epistasis is a form of negative epistasis, the above observations are thought to indicate a preponderance of negative epistasis between beneficial mutations and a concave fitness landscape[12]. If the process of adaptation is reversed by reverting the accepted beneficial mutations, one should see accelerating fitness drops and negative epistasis between deleterious mutations. Contrary to this expectation, mutation accumulation experiments in the near absence of selection have revealed decelerating fitness declines[13-15] and manipulative experiments have demonstrated that deleterious mutations tend to be less harmful in less fit genetic backgrounds (a.k.a. increasing costs epistasis because of the higher costs of deleterious mutations in fitter genotypes)[16]. These observations concerning deleterious mutations are thought to indicate a convex fitness landscape and a predominance of positive epistasis[12,13].

Apparently, the inferred shape of the fitness landscape and distribution of epistasis from climbing fitness peaks contrast the shape and distribution inferred from going down fitness peaks[12]. We term this contradiction the uphill-downhill paradox. Although several theoretical models have been proposed to explain the inferred prevalence of either negative or positive epistasis[9,10,13,16-20], these models cannot simultaneously explain both in the same species, leaving the uphill-downhill paradox unresolved. For instance, some authors have suggested that the negative epistasis among beneficial mutations accumulated during adaptations are not representative of the entire fitness landscape, due to the biased sampling of mutations[21,22]. But this explanation does not apply to mutations randomly accrued in mutation accumulation experiments so cannot fully resolve the uphill-downhill paradox. Below we propose and demonstrate that epistasis is generally idiosyncratic and that this idiosyncrasy is responsible for the general trends in both climbing and descending from fitness peaks.

Results

Why epistasis could be highly idiosyncratic

Let g be the population growth rate (a.k.a. Malthusian fitness, logarithm of Wrightian fitness, or fitness for short) of a genotype in an environment and let n be the number of nucleotide sites in the genome that impact g. In general, g can be expressed as the sum of 2^n-1 terms of fitness effects, including the additive effect of every site, the interactive effect of every pair of sites, the interactive effect of every triplet of sites, and so on (see Methods). We refer to this model of fitness landscape as the n-order model, because it includes all terms up to the n-order interaction. It can be shown that a mutation at a single site changes up to $2^{n-1}/(2^n-1) \approx 50\%$ of all terms of effects making up g. Under the assumption that the interactive terms are idiosyncratic (i.e., varying with the interacting nucleotides involved), a single mutation can differentially alter as

many as 2^{n-2} (or ~25% of) terms of effects in two genotypes that differ by only one nucleotide; this number can rise up to 2^{n-1} (or ~50% of) terms if the two genotypes are more different (see Methods). Given the potential of such a large fraction of differentially affected terms of *g*, it is not surprising that the same mutation could have vastly different effects in different genotypes. As long as the idiosyncrasy assumption holds, the same argument can be made for any phenotypic trait whose value is expressed as the sum of all additive and interactive terms of effects. Of course, not all 2^n-1 terms of effects are of the same magnitude, which would increase or decrease the effective fraction of terms differentially altered by a mutation. Regardless, the above consideration elucidates why mutational effects could be highly sensitive to the genetic background when biological interactions are complex.

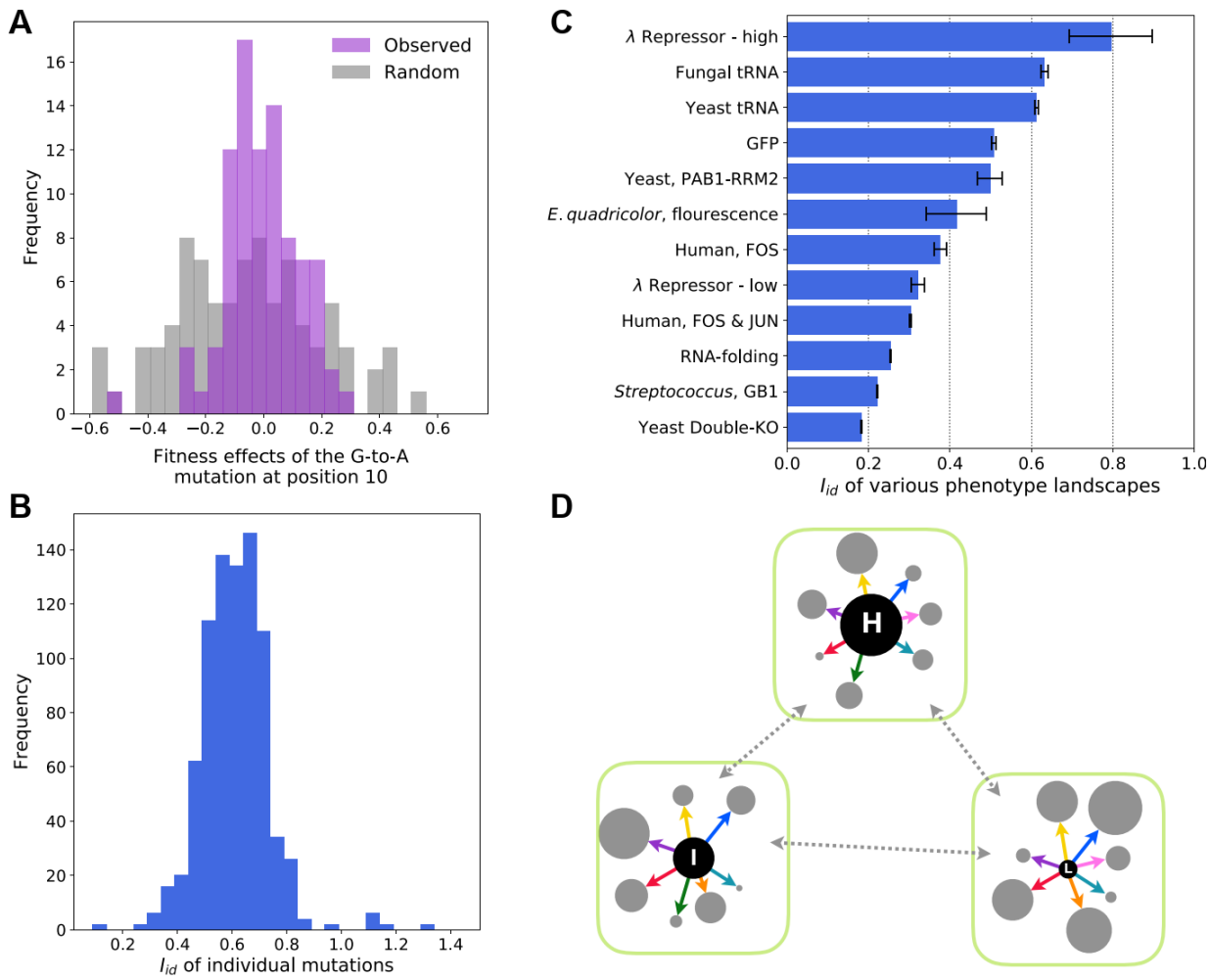


Figure 2-1 Idiosyncratic index of a wide variety of phenotype landscapes.

(A) Frequency distribution of the fitness effects of the mutation from G to A at position 10 across all available genetic backgrounds (purple) and the corresponding distribution of the fitness difference between two random genotypes for the same number of genotype pairs (grey) in the yeast tRNA fitness landscape. (B) Frequency distribution of the standard deviation (SD)-based idiosyncrasy index (Iid), which is the ratio of the SD of fitness effects of a particular mutation on different backgrounds to the SD of fitness differences between random genotype pairs, for all individual mutations in the yeast tRNA landscape. (C) SD-based Iid of various phenotype landscapes. Error bars show standard errors. Detailed information of each landscape is provided in Table S1. (D) Schematic of a highly idiosyncratic fitness landscape. Genotypes are represented by circles, and the fitness of a genotype is represented by the circle size. The three black circles labeled with H, I, L respectively indicate three focal genotypes with relatively high, intermediate, and low fitness values, whereas the grey circles represent one-mutation neighbors of the focal genotypes. Each light-green outlined area encompasses a focal genotype and some one-mutation neighbors. Solid arrows indicate single mutations, whereas dotted arrows indicate multiple mutations. Solid arrows of the same color indicate the same mutation.

Epistasis is highly idiosyncratic

To quantify the above sensitivity that originates from idiosyncratic epistasis, we define an idiosyncratic index (I_{id}) for a mutation as the variation in the fitness difference between genotypes that differ by the mutation, relative to the variation in the fitness difference between random genotypes for the same number of genotype pairs. Here, the variation may be measured by standard deviation (SD), range, or other statistics. We can further compute the I_{id} for a fitness landscape by averaging I_{id} of individual mutations considered. The I_{id} for a landscape varies from 0 to 1, corresponding to the minimal and maximal levels of idiosyncrasy, respectively. We first estimated I_{id} for the fitness landscape of a yeast tRNA gene that includes experimentally measured fitness of over 65,000 genotypes[23]. For example, the G-to-A mutation at site 10 has a fitness effect varying from -0.53 to 0.29 (SD = 0.13) on 88 different backgrounds. For comparison, the fitness difference between a randomly picked genotype and another randomly picked genotype varies from -0.59 to 0.74 (SD = 0.26) for 88 genotype pairs sampled (**Fig. 2-1A**). So, the ratio of the two SDs is 0.49. This analysis was repeated for 828 single mutations (considering reverse mutations) (**Fig. 2-1B**), and the average ratio of SD is $I_{id} = 0.612 \pm 0.005$ (SE). I_{id} can be similarly defined for non-fitness traits, and we estimated I_{id} for a variety of empirical phenotype landscapes that are experimentally determined[23-31] and one that is computationally predicted (RNA-folding) (**Table 2-S1**). Overall, I_{id} varies from 0.18 to 0.80

among the 12 landscapes examined, with a mean of 0.43 (**Fig. 2-1C**). Hence, in an average phenotype landscape, a particular mutation's effects across different backgrounds are 43% as variable as if they are randomly drawn from the effects of any number of mutations on any genetic background. To exclude the possibility that the observed idiosyncrasy is largely due to imprecise phenotyping, we computed I_{id} for the same 828 mutations in the tRNA landscape, but used fitness estimates from different numbers of experimental replicates, because the measurement error should decrease with the number of replicates. We found that I_{id} is insensitive to the number of replicates (**Fig. 2-S1A**), suggesting that the high idiosyncrasy is not explained by potentially imprecise phenotyping. Additionally, phenotypic values in the RNA-folding landscape were computed deterministically without measurement error, but mutational-effects are still a quarter as idiosyncratic as the maximum ($I_{id} = 0.25$). We similarly observed high idiosyncrasy when range instead of SD of effects was used in estimating I_{id} (**Fig. 2-S1B**).

Expected consequences of idiosyncratic epistasis

The substantial idiosyncrasy observed suggests that phenotype landscapes are quite rugged (**Fig. 2-1D**). Below we demonstrate its consequences with regard to fitness, but the same applies to other traits. In a maximally idiosyncratic fitness landscape such as the one described by the house-of-cards model[32], fitness values (circle sizes in **Fig. 2-1D**) of neighboring genotypes connected through single mutations are uncorrelated. The fitness of a neighboring genotype of a high- or a low-fitness focal genotype is expected to be the same. Hence, the fitness difference between a neighboring genotype (grey circle) and the focal genotype (black circle) is expected to be less positive or more negative as the fitness of the focal genotype rises. In other words, beneficial mutations are less beneficial and deleterious mutations are more deleterious on fitter genotypes, causing diminishing returns and increasing costs, respectively. These arguments

apply not only to the effects of the same mutation on different genetic backgrounds but also to the effects of different mutations on different backgrounds. That is, an arbitrary mutation on a relatively fit background is expected to be less beneficial or more detrimental than another arbitrary mutation on a relatively unfit background. Under the foregoing model of *g*, one can mathematically prove that, in the presence of idiosyncrasy of at least one interactive term, the correlation between the fitness effect of a mutation and the background fitness is negative, for both the same and different mutation(s) (see Methods). In the case of different mutations, among-site/state variation in the additive effect further contributes to the negative correlation (see Methods). Importantly, all of the above occurs even with no bias toward positive or negative epistasis in the fitness landscape and no fitness estimation error.

Idiosyncratic epistasis causes the trends of diminishing returns and increasing costs

To examine whether the extent of idiosyncratic epistasis in an actual fitness landscape is sufficient to explain the observed diminishing returns and increasing costs, we simulated a series of 16 fitness landscapes with $n = 16$ binary sites, under the *n*-order model of *g*. In the *k*th landscape in the series ($1 \leq k \leq 16$), we considered up to the *k*th order of interaction. That is, each term of effect from the first to the *k*th order interaction is a random variable independently drawn from the standard normal distribution whereas all other terms are set to 0. When *k* rises from 1 to 16, I_{id} increases from 0 to 0.69 (**Fig. 2-2A**), which is close to the theoretically predicted value (see Methods). In all simulated landscapes except the one with $I_{id} = 0$, most if not all mutations exhibit a negative Pearson's correlation between fitness effect and background fitness (boxes in **Fig. 2-2A**). In addition, the larger the *k* and I_{id}, the more negative the correlations (boxes in **Fig. 2-2A**), supporting the role of idiosyncratic epistasis in creating the negative correlations. For comparison, 87.8% of mutations from the yeast tRNA fitness landscape show a

negative correlation between fitness effect and background fitness (**Fig. 2-2B**). A similar trend is seen in other empirical phenotype landscapes (**Fig. 2-S2A**, **B**). Separating mutations that are beneficial or detrimental on the wild-type background or an arbitrary background reveals the familiar patterns of diminishing returns and increasing costs in both the simulated and empirical landscapes (**Fig. 2-S3**).

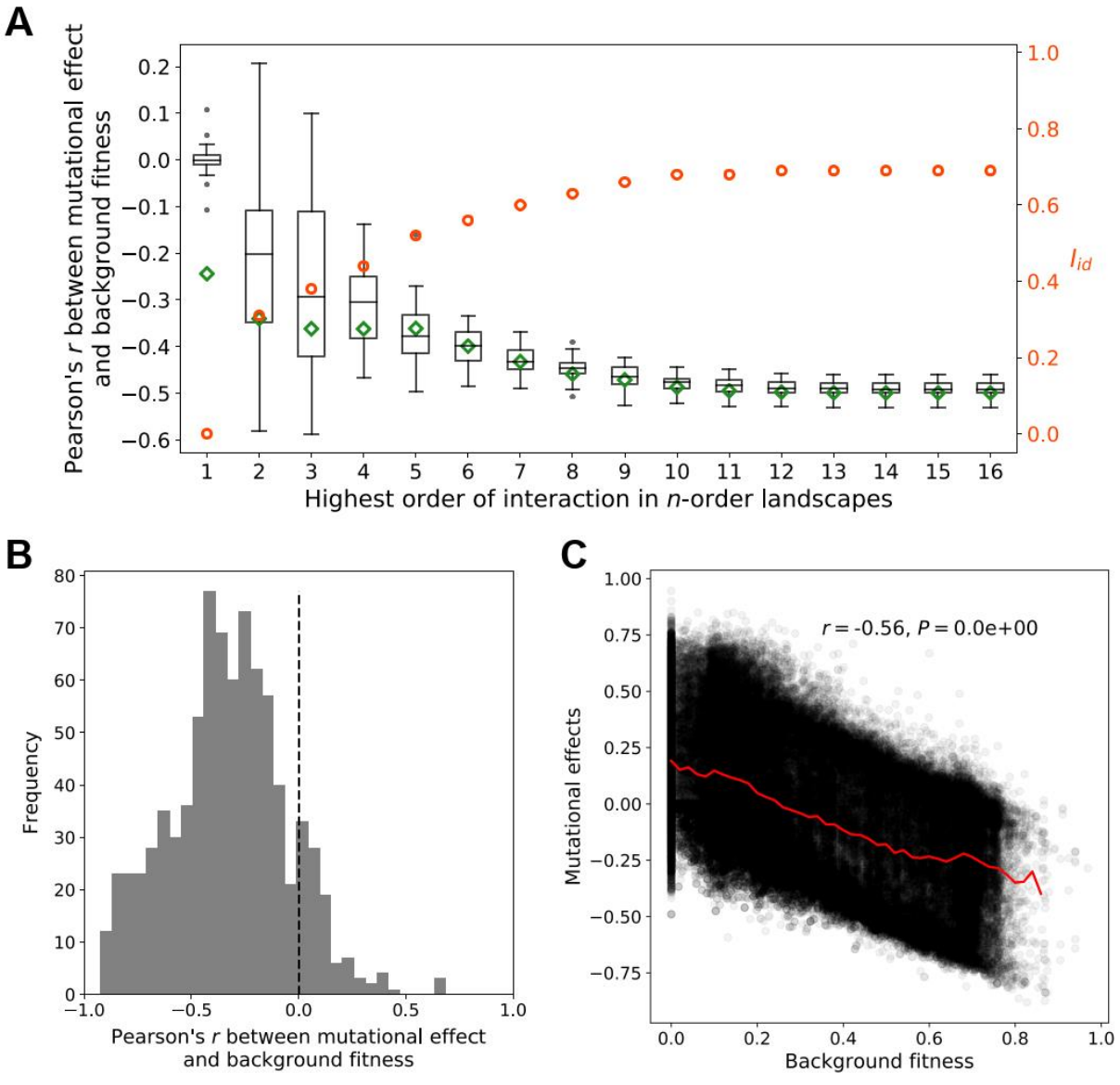


Figure 2-2 Negative correlation between mutational effect and background fitness as a result of idiosyncratic epistasis.

(A) Boxplot showing the distribution of Pearson's correlation coefficient (r) between mutational effect and background fitness for individual mutations in a series of n-order landscapes. The lower and upper edges of a box represent the first (qu1) and third

(qu3) quartiles, respectively, the horizontal line inside the box indicates the median (md), the whiskers extend to the most extreme values inside inner fences, md ± 1.5(qu3 − qu1), and the grey dots represent values outside the inner fences (outliers). Green diamonds represent r values for all mutations pooled in each landscape, while red circles indicate SD-based Iid of the landscapes. (B) Distribution of r for 828 individual mutations in the tRNA fitness landscape. (C) Relationship between background fitness and mutational effect for 414 mutations (reversions not considered) in the tRNA fitness landscape. The red line depicts the running mean in non-overlapping X-axis bins of width = 0.02, in all bins with more than 10 data points. To avoid a spurious correlation due to shared measurement error on the X- and Y-axis, we used three replicates of background fitness measures for the X-axis and three other replicates for the Y-axis in (B) and (C). For each mutation and its reversion, a randomly picked one is considered in (C).

Furthermore, the effects of different mutations also negatively correlate with background fitness in the simulated landscapes (green diamonds in **Fig. 2-2A**; see Methods), as well as in the tRNA fitness landscape (**Fig. 2-2C**) and other empirical phenotype landscapes (**Fig. 2-S2C**, **D**). As mentioned, the negative correlation in the simulated landscape with $I_{id} = 0$ (green diamond in **Fig. 2-2A**) is due to the contribution from the among-site/state variation in additive effect; in the absence of this variation, all genotypes are equally fit, so the correlation disappears.

Idiosyncratic epistasis causes slowing fitness drops in mutation accumulation

When random mutations accrue in a relatively fit population in the near absence of selection, population fitness is expected to decline. Because idiosyncratic epistasis renders random mutations on average less deleterious on relatively unfit genotypes than on relatively fit genotypes, the fitness drop of the population is expected to decelerate during mutation accumulation until it reaches the mean fitness of all genotypes in the landscape, around which the fitness should subsequently fluctuate. We confirmed this prediction in the simulated *n*-order landscapes: As *k* and I_{id} increase, the slowing curvature becomes more prominent (**Fig. 2-3A**). A change in mutational supply explains why the fitness decline is decelerating even when $I_{id} = 0$ (see Methods), and as expected, this trend diminishes as *n* rises (**Fig. 2-S4**). For comparison, decelerating fitness declines are apparent during simulated mutation accumulations in the tRNA fitness landscape (**Fig. 2-3B**) and other empirical phenotype landscapes (**Fig. 2-S5**).

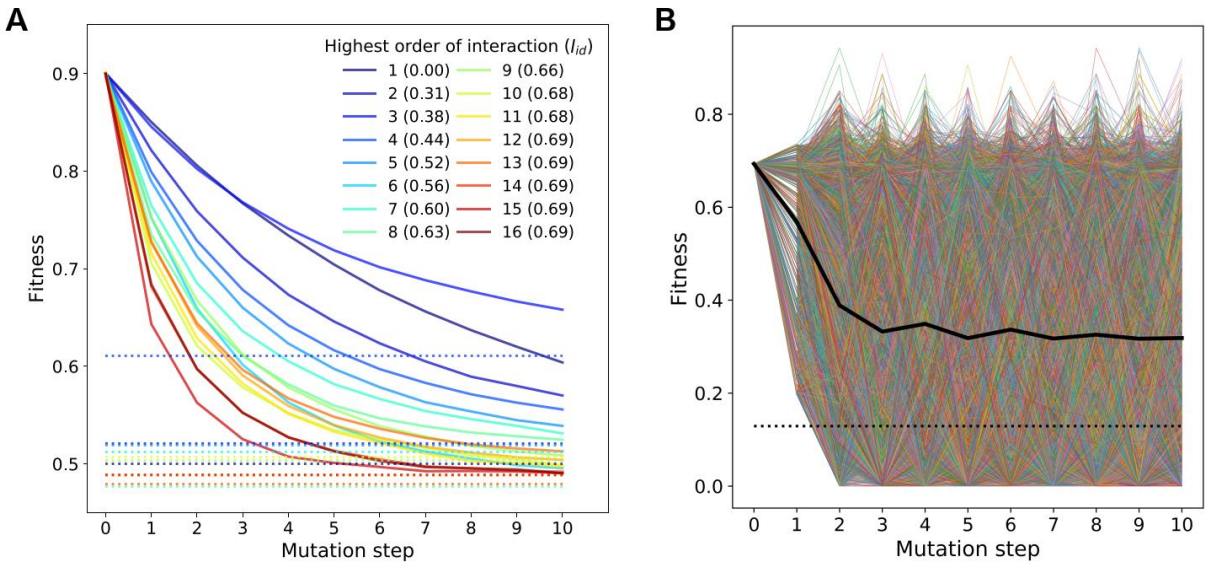


Figure 2-3 Fitness declines decelerate during mutation accumulation as a result of idiosyncratic epistasis.

(A) Average fitness trajectories simulated in a series of n-order landscapes. (B) Ten thousand fitness trajectories of mutation accumulation simulated in the tRNA fitness landscape, with the average fitness of all trajectories at each step shown in black. Dotted lines indicate the mean fitness of all genotypes in the corresponding landscape.

The drop in fitness to the mean of all genotypes during mutation accumulation is observed in the *n*-order landscapes (**Fig. 2-3A**) and RNA-folding landscape (**Fig. 2- S5B**), while the average mutation accumulation trajectory in tRNA (**Fig. 2-3B**) and GFP (**Fig. 2-S5A**) landscapes fluctuates above the mean of all genotypes. This latter phenomenon is due to preferential sampling of genotypes close to the wild-type in the experimental data, "trapping" many simulated mutation accumulation trajectories around the wild-type. The former two theoretically simulated/calculated landscapes do not have such biases.

Idiosyncratic epistasis causes slowing fitness gains in adaptation

Idiosyncratic epistasis, in combination with certain distributions of genotype fitness or interactive effects, creates the phenomenon of decelerating fitness gains during adaptation. In a solely additive landscape with $I_{id} = 0$, each adaptive trajectory is basically a random ordering of the beneficial mutations. Thus, the mean fitness increase of every step is the mean of the effect

of all beneficial mutations, leading to a linear average trajectory regardless of the fitness distribution (**Fig. 2-4A**). In an idiosyncratic landscape with $I_{id} > 0$, the mutation fixed at each step during adaptation is a random draw from beneficial mutations instead of all mutations. Because of this bias, the shape of the adaptive trajectory, unlike that of mutation accumulation, is dependent on the distribution of genotype fitness or interactive effects. For example, in a house-of-cards model (a special case of *n*-order landscapes with only the highest-order interaction term), when fitness of all genotypes is gamma distributed with shape parameter >1, =1, or <1 (**Fig. 2-S6A**), fitness rises sublinearly, linearly, and superlinearly with the number of mutations accumulated, respectively (**Fig. 2-S6B**) (see also [33]). Although not sufficient for creating a decelerating adaptive trajectory, idiosyncrasy causes a decelerating trajectory in a wide range of full *n*-order landscapes, including, for example, those with normal- (**Fig. 2-4A**), gamma-, and beta-distributed interaction effects (**Fig. 2-S6C**, **D**). As expected, adaptation slows more dramatically with greater I_{id} (**Fig. 2-4A**; **Fig. 2-S6C**, **D**). Simulated adaptation also decelerates in the tRNA fitness landscape (**Fig. 2-4B**) as well as in other empirical phenotype landscapes (**Fig. 2-S7**), suggesting that the empirical cases fulfill both the idiosyncrasy and fitness distribution requirements.

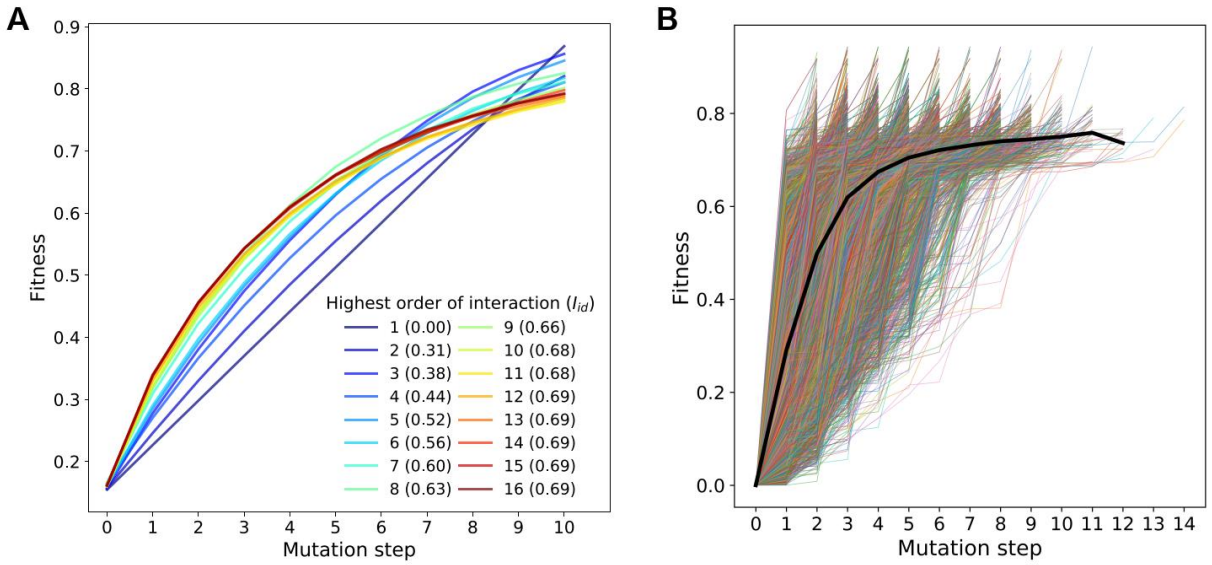


Figure 2-4 Adaptation slows as a result of idiosyncratic epistasis.

(A) Average adaptive trajectories simulated in a series of n-order landscapes. (B) A total of 15,878 adaptive trajectories simulated in the tRNA fitness landscape, with the average fitness of all trajectories shown in black, at each step when the trajectory number exceeds 10.

Discussion

In summary, we proposed a simple theory that uses the idiosyncrasy of epistasis to explain some of the most commonly observed patterns of mutational effects and evolutionary trajectories. Phenotype landscapes of a variety of genes and taxa confirm our assumption of idiosyncratic epistasis. Contrary to the common intuition, our work shows that diminishing returns and decelerating adaptations do not suggest a bias toward negative epistasis in the underlying fitness landscape or a concave landscape. Similarly, increasing costs and slowing fitness declines during mutation accumulation do not indicate a bias toward positive epistasis or a convex landscape. Thus, our theory resolves the uphill-downhill paradox.

Although the idiosyncrasy of epistasis is a major characteristic of empirical phenotype landscapes (**Fig. 2-1C**), biological interactions are not completely idiosyncratic. Rather, idiosyncratic epistasis should serve as a null model for the role of epistasis in mutational effects and evolution. For example, the relationship between the mutational robustness of a genotype (i.e., fitness insensitivity to mutation) and its adaptability/evolvability to environmental challenges is debated[34-36]. Our theory reveals an intrinsic positive correlation between robustness and adaptability due to idiosyncratic epistasis, because, as the fitness of a genotype rises, deleterious mutations are more detrimental (i.e., lower robustness) and advantageous mutations are less beneficial (i.e., lower adaptability). Deviations from this null expectation may reveal interesting forms of epistasis beyond idiosyncrasy. Similarly, because slowing fitness drops during mutation accumulation naturally emerge from idiosyncratic epistasis, such

observations need not be explained by selection for "genomic buffering against the fitness reduction caused by accumulated mutations"[13]. Rather, when this trend is absent or when the opposite trend is observed, selection for mutational robustness of the wild-type may be invoked[37]. Additionally, other processes such as clonal interference[38] and changes in mutational supply[39] may enhance some of the universals caused by idiosyncratic epistasis.

How does the idiosyncrasy of epistasis arise from the underlying deterministic biological interactions? The *n*-order model reveals that the number of interactive terms determining the phenotype of a genotype is potentially astronomical and that the same mutation differentially alters a substantial fraction of these terms in even slightly different genotypes. Consequently, it is difficult to predict the mutational effect in any particular genotype despite the underlying deterministic biological interactions, much like the apparently random outcome of a die roll that is deterministically shaped by myriad factors such as the movement of air molecules. That the universal trends of mutational effects and evolutionary trajectories emerge from this randomness due to idiosyncratic epistasis is no more surprising than the tendency of observing a smaller number in a second roll of a die when the first roll yields a five.

Methods

Number of terms of phenotypic effects altered by a mutation

Let g be the Malthusian fitness (fitness for short) of a genotype in an environment and let n be the number of nucleotide sites in a genome that are relevant to g. Here g equals the sum of the additive fitness effect of every site (i.e., first order interaction), the interactive effect of every pair of sites (i.e., second order interaction), the interactive effect of every triplet of sites (i.e., third order interaction), and so on. That is, g contains $\binom{n}{k}$ terms of effects of the k^{th} order of

interaction ($1 \leq k \leq n$), for a total of 2^n-1 terms. Among these terms, 2^{n-1} terms involve any particular site. Thus, a mutation at a single site potentially changes $2^{n-1}/(2^n-1) \approx 50\%$ of all terms making up g.

Differentially altered terms of effects in two genotypes caused by the same mutation

When the same mutation of allele P changing to Q at site i occurs on two different genotypes that differ at m sites (i is not one of the m sites), the differentially altered terms of effects in these genotypes must involve site i and at least one of the m sites, and may also involve site(s) identical between the two genotypes. The total number of differentially altered terms equals the number of terms involving i and at least one other site minus the number of terms involving i and at least one other site that is identical between the two genotypes. The resulting number is $(2^{n-1} - 1) - (2^{n-m-1} - 1) = (2^m - 1)2^{n-m-1}$. When $m = 1$, the above number is 2^{n-2}. That is, up to $2^{n-2}/(2^n-1) \approx 25\%$ of terms are differentially altered by the same mutation in two genotypes that differ at only one site. When $m = n$-1, the above number is $2^{n-1} - 1$. That is, up to $(2^{n-1}-1)/(2^n-1) \approx 50\%$ of terms are differentially altered by the same mutation in two maximally different genotypes.

The fitness effect of a given mutation is negatively correlated with background fitness

Let us consider the n-order landscape model and focus on the mutation from the P allele to the Q allele at site k of the genome. We examine the fitness effect of this mutation on different genetic backgrounds. Let X represent any genotype with the P allele at site k. Among them, x_i is the i^{th} genotype whose Malthusian fitness is R_{x_i}. R_{x_i} can be written as $R_{x_i} = A_{x_i} + I_{x_i} + I'_{x_i}$, where A_{x_i} is the sum of additive (i.e., 1st order interactive) effects, I_{x_i} is the sum of the 2nd to n^{th} order interactive effects involving the focal site k, and I'_{x_i} is the sum of the 2nd to n^{th} order interactive effects that do not involve site k.

Similarly, let *Y* represent any genotype with the *Q* allele at site *k*. For each genotype x_i, we have a corresponding genotype y_i that is identical to x_i except that site *k* now has the *Q* allele. R_{y_i}, the fitness of y_i, can be written as $R_{y_i} = A_{y_i} + I_{y_i} + I'_{y_i}$, where A_{y_i} is the sum of additive (i.e., 1st order interactive) effects, I_{y_i} is the sum of the 2nd to n^{th} order interactive effects involving site *k*, and I'_{y_i} is the sum of the 2nd to n^{th} order interactive effects that do not involve site *k*.

Note that the difference in the additive effect between alleles *P* and *Q* is a constant that is not influenced by sites other than *k*. That is, $A_{y_i} - A_{x_i} = C$. Therefore, we have

$$Cov(A_Y, A_X) = \frac{1}{N}\sum_{i=1}^{N}\Big(A_{y_i} - E(A_Y)\Big)\Big(A_{x_i} - E(A_X)\Big)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\Big(A_{x_i} + C - E(A_X + C)\Big)\Big(A_{x_i} - E(A_X)\Big)$$

$$= \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\Big(A_{x_i} - E(A_X)\Big)\Big(A_{x_i} - E(A_X)\Big) = Var(A_X).$$

Here, *N* is the total number of pairs of (x_i, y_i) and equals 4^{n-1} for a genome with *n* sites each with four states, *Cov* stands for covariance, and *Var* stands for variance.

Also note that, because x_i and y_i are the same except at site *k*, $I'_{x_i} = I'_{y_i}$. Let $Cor(I_X, I_Y)$ be the Pearson correlation between I_X and I_Y. We have $Cor(I_X, I_Y) = Cov(I_X, I_Y)/\sqrt{Var(I_X)Var(I_Y)} \leq 1$. Hence, $Cov(I_X, I_Y) \leq \sqrt{Var(I_X)Var(I_Y)}$. Under the reasonable assumption that the corresponding interactive terms of x_i and y_i are sampled from the same distribution, $Var(I_X)$ and $Var(I_Y)$ are expected to be equal. Hence, $Cov(I_X, I_Y) \leq \sqrt{Var(I_X)Var(I_X)} = Var(I_X)$. Thus, $Cov(I_X, I_Y) = Var(I_X)$ when I_X and I_Y have a correlation of 1; otherwise $Cov(I_X, I_Y) < Var(I_X)$.

When epistasis is to some extent idiosyncratic, I_Y does not correlate perfectly with I_X, resulting in $Cov(I_X, I_Y) < Var(I_X)$.

Under the assumption of independence among the interactive terms of a genotype, we have *Cov*(mutational effect, fitness of the background genotype) = $Cov(R_Y - R_X, R_X) = Cov\big((A_Y - A_X) + (I_Y - I_X) + (I'_Y - I'_X),\ A_X + I_X + I'_X\big) = Cov(C, A_X) + Cov(I_X, I_Y) - Var(I_X) + Cov(0, I'_X) = Cov(I_X, I_Y) - Var(I_X) < 0$. This mathematical result means that, when epistasis is to some extent idiosyncratic, for any given mutation, we expect a negative correlation between the background fitness and mutational effect, which is exactly what diminishing returns of beneficial mutations and increasing costs of deleterious mutations are. The above result holds when fitness is replaced with any phenotypic trait as long as the trait value of each genotype can be expressed as the sum of the 2^n-1 terms of effects.

Mutational effect is generally negatively correlated with background fitness

Below we show that the preceding result about a given mutation also applies to different mutations. That is, we expect a negative correlation between the mutational effect and background fitness even when different mutations are considered. R_t, the Malthusian fitness of genotype *t*, can be expressed by $R_t = I_t^1 + I_t^2 + \cdots + I_t^n + I_t^{1,2} + I_t^{1,3} + \cdots + I_t^{n,n-1} + \cdots + I_t^{1,2,\cdots,n}$. Here, the superscript indicates the site(s) involved in an additive or interactive term. For instance, I_t^2 stands for the additive effect of site 2 and $I_t^{1,2}$ stands for the interactive effect between sites 1 and 2.

Let *X* represent an arbitrary genotype and *Y* represent another genotype that differs from *X* by a particular mutation named *W* that occurs at site *k*. We have

$$R_X = I_X^1 + I_X^2 + \cdots + I_X^n + I_X^{1,2} + I_X^{1,3} + \cdots + I_X^{n-1,n} + \cdots + I_X^{1,2,\cdots,n}$$

$$R_Y = I_Y^1 + I_Y^2 + \cdots + I_Y^n + I_Y^{1,2} + I_Y^{1,3} + \cdots + I_Y^{n-1,n} + \cdots + I_Y^{1,2,\cdots,n}.$$

In the above, all corresponding terms between I_X and I_Y are equal except for the terms involving k. So, $R_Y - R_X = \left(I_Y^k - I_X^k\right) + \left(I_Y^{1,k} - I_X^{1,k}\right) + \cdots + \left(I_Y^{n,k} - I_X^{n,k}\right) + \cdots + \left(I_Y^{(1,2,\cdots,n),k} - I_X^{(1,2,\cdots,n),k}\right)$.

Under the assumption that all *I* terms in an *R* are independent from one another, *Cov*(mutational effect, background fitness)=$Cov(R_Y - R_X, R_X) = Cov\left(I_Y^k - I_X^k, I_X^k\right) + Cov\left(I_Y^{1,k} - I_X^{1,k}, I_X^{1,k}\right) + \cdots + Cov\left(I_Y^{n,k} - I_X^{n,k}, I_X^{n,k}\right) + \cdots + Cov\left(I_Y^{(1,2,\cdots,n),k} - I_X^{(1,2,\cdots,n),k}, I_X^{(1,2,\cdots,n),k}\right)$.

According to the law of total variance and the law of total covariance, we can expand each term in the above equation. Let us use the second order interaction between site *s* and site *k* as an example. $Cov\left(I_Y^{s,k} - I_X^{s,k}, I_X^{s,k}\right) = Cov\left(I_Y^{s,k}, I_X^{s,k}\right) - Var\left(I_X^{s,k}\right) = E\left(Cov\left(I_Y^{s,k}, I_X^{s,k}\middle|W\right)\right) + Cov\left(E\left(I_Y^{s,k}\middle|W\right), E\left(I_X^{s,k}\middle|W\right)\right) - E\left(Var\left(I_X^{s,k}\middle|W\right)\right) - Var\left(E\left(I_X^{s,k}\middle|W\right)\right) = E\left(Cov\left(I_Y^{s,k}, I_X^{s,k}\middle|W\right) - Var\left(I_X^{s,k}\middle|W\right)\right) + Cov\left(E\left(I_Y^{s,k} - I_X^{s,k}\middle|W\right), E\left(I_X^{s,k}\middle|W\right)\right)$.

As shown in the section about a given mutation, as long as there is some degree of idiosyncrasy, $Cov\left(I_Y^{s,k}, I_X^{s,k}\middle|W\right) < Var\left(I_X^{s,k}\middle|W\right)$. So, $E\left(Cov\left(I_Y^{s,k}, I_X^{s,k}\middle|W\right) - Var\left(I_X^{s,k}\middle|W\right)\right) < 0$. Further, $Cov\left(E\left(I_Y^{s,k} - I_X^{s,k}\middle|W\right), E\left(I_X^{s,k}\middle|W\right)\right) = 0$, because $E\left(I_Y^{s,k} - I_X^{s,k}\middle|W\right) = 0$ under the reasonable assumption that, given *W*, $I_X^{s,k}$ and $I_Y^{s,k}$ follow the same distribution. Hence, $Cov\left(I_Y^{s,k} - I_X^{s,k}, I_X^{s,k}\right) < 0$. The same conclusion applies to all terms except the first-order interactive

(additive) term, which is $Cov(I_Y^k - I_X^k, I_X^k) = E\left(Cov(I_Y^k, I_X^k|W) - Var(I_X^k|W)\right) + Cov\left(E(I_Y^k - I_X^k|W), E(I_X^k|W)\right)$. Because additive effects are independent of the genetic background, given *W*, I_Y^k and I_X^k are both fixed and are two randomly sampled values from the same distribution. Hence, $Cov(I_Y^k, I_X^k|W) = 0$ and $Var(I_X^k|W) = 0$. So $E\left(Cov(I_Y^k, I_X^k|W) - Var(I_X^k|W)\right) = 0$. $E(I_Y^k|W) = I_Y^k|W$ and $E(I_X^k|W) = I_X^k|W$. As *W* varies, $I_Y^k|W$ and $I_X^k|W$ are two random variables from the same distribution. They have the same variance and are not usually completely correlated. So, $Cov\left(E(I_Y^k - I_X^k|W), E(I_X^k|W)\right) = Cov(I_Y^k - I_X^k|W, I_X^k|W) = Cov(I_Y^k, I_X^k|W) - Var(I_X^k|W) < 0$. Under the special case when all additive terms are equal, $Cov\left(E(I_Y^k - I_X^k|W), E(I_X^k|W)\right) = 0$.

Thus, *Cov*(mutational effect, background fitness) = $E\left(Cov(I_Y^k, I_X^k|W) - Var(I_X^k|W)\right) + Cov\left(E(I_Y^k - I_X^k|W), E(I_X^k|W)\right) + E\left(Cov(I_Y^{1,k}, I_X^{1,k}|W) - Var(I_X^{1,k}|W)\right) + Cov\left(E(I_Y^{1,k} - I_X^{1,k}|W), E(I_X^{1,k}|W)\right) + \cdots + E\left(Cov(I_Y^{n,k}, I_X^{n,k}|W) - Var(I_X^{n,k}|W)\right) + Cov\left(E(I_Y^{n,k} - I_X^{n,k}|W), E(I_X^{n,k}|W)\right) + \cdots + E\left(Cov\left(I_Y^{(1,2,\cdots,n),k}, I_X^{(1,2,\cdots,n),k}\middle|W\right) - Var\left(I_X^{(1,2,\cdots,n),k}\middle|W\right)\right) + Cov\left(E\left(I_Y^{(1,2,\cdots,n),k} - I_X^{(1,2,\cdots,n),k}\middle|W\right), E\left(I_X^{(1,2,\cdots,n),k}\middle|W\right)\right) = Cov(I_Y^k - I_X^k|W, I_X^k|W) + E\left(Cov(I_Y^{1,k}, I_X^{1,k}|W) - Var(I_X^{1,k}|W)\right) + \cdots + E\left(Cov(I_Y^{n,k}, I_X^{n,k}|W) - Var(I_X^{n,k}|W)\right) + \cdots + E\left(Cov\left(I_Y^{(1,2,\cdots,n),k}, I_X^{(1,2,\cdots,n),k}\middle|W\right) - Var\left(I_X^{(1,2,\cdots,n),k}\middle|W\right)\right) < 0.$

Therefore, in the *n*-order model, mutational effect is negatively correlated with background fitness even for different mutations. As shown in the above mathematical derivation, this

negative correlation has two sources: unequal additive effects and idiosyncratic epistasis. Given the same additive effects, increasing the idiosyncrasy in epistasis strengthens the negative correlation. As in the preceding section, the result here applies to any phenotypic trait as long as the trait value of a genotype can be expressed as the sum of the 2^n-1 terms of effects.

Expected idiosyncrasy index under the *n*-order landscape model

The variance of the effect of a particular mutation across all genetic backgrounds can be calculated as follows. Let *X* represent an arbitrary genotype and *Y* represent another genotype that differs from *X* at site *k* only. We have shown earlier that

$$R_X = I_X^1 + I_X^2 + \cdots + I_X^n + I_X^{1,2} + I_X^{1,3} + \cdots + I_X^{n-1,n} + \cdots + I_X^{1,2,\cdots,n}$$

$$R_Y = I_Y^1 + I_Y^2 + \cdots + I_Y^n + I_Y^{1,2} + I_Y^{1,3} + \cdots + I_Y^{n-1,n} + \cdots + I_Y^{1,2,\cdots,n}.$$

In the above, all corresponding terms between I_X and I_Y are equal except for the terms involving *k*. So, $R_Y - R_X = \left(I_Y^k - I_X^k\right) + \left(I_Y^{1,k} - I_X^{1,k}\right) + \cdots + \left(I_Y^{n,k} - I_X^{n,k}\right) + \cdots + \left(I_Y^{(1,2,\cdots,n),k} - I_X^{(1,2,\cdots,n),k}\right)$ and $Var\left(R_y - R_x\right) = Var\left(I_Y^k - I_X^k\right) + Var\left(I_Y^{1,k} - I_X^{1,k}\right) + \cdots + Var\left(I_Y^{n,k} - I_X^{n,k}\right) + \cdots + Var\left(I_Y^{(1,2,\cdots,n),k} - I_X^{(1,2,\cdots,n),k}\right)$. If we assume that all interactive terms for *X* and *Y* are independent with the same variance σ^2, $Var(R_Y - R_X) = 2^n\sigma^2$.

If there are *M* states at each site, among the k^{th} order interactive terms, $\binom{n}{k}(1/M)^k$ terms are expected to be the same between two random genotypes. One can show that $\sum_{k=1}^{n}\binom{n}{k}(1/M)^k = \sum_{k=1}^{n}\frac{n!}{k!(n-k)!}\cdot(1/M)^{\mathrm{k}} \approx \sum_{k=1}^{n}\frac{n^k}{k!}\cdot(1/M)^k \approx e^{\frac{n}{M}}$. Thus, two random genotypes are expected to differ by approximately $2^n - 1 - e^{n/M}$ terms. Hence, the variance of the fitness difference between two random genotypes is $Var(R_Y - R_X) = 2(2^n - 1 - e^{\frac{n}{M}})\sigma^2$. Because $M \geq 2$, $e^{\frac{n}{M}} <$

2^n. So, when n is large, $Var(R_Y - R_X)$ is approximately $2^{n+1}\sigma^2$. Therefore, the idiosyncrasy index becomes $\frac{\sqrt{2^n}\sigma}{\sqrt{2^{n+1}}\sigma} = \frac{1}{\sqrt{2}}$ = ~0.71. Our numerical finding (the most right red dot in Fig. 2A) confirms this result.

Mutational supply and evolutionary trajectories

During adaptation, if the supply of beneficial mutations diminishes as the fitness of a population rises, the speed of population fitness increase per unit time will decline. However, if the speed of fitness increase is measured per beneficial mutation accrued as in the present study, the reducing supply of beneficial mutations will not reduce the speed of fitness increase.

During mutation accumulation in the near absence of selection, as the population fitness declines, the supply of beneficial mutations should increase and the supply of deleterious mutations should decrease. Thus, even under a purely additive model, the speed of population fitness drop slows. When only the first few mutations accrued are examined, however, this phenomenon of slowing fitness drops should be minimal under the purely additive model unless the number of possible mutations is very limited

Empirical phenotype landscapes

An unbiased search for phenotype landscape data published between 2000 and 2019 was preformed using Google Scholar with words such as “epistasis”, “fitness landscape”, or “genetic interaction”. A total of 18 datasets were found for which quantitative phenotype values were published or could be calculated without extensive analysis (e.g., studies reporting only sequencing reads were excluded) and which included genotypes with at least two mutations in comparison with the reference genotype. Measured phenotypes included protein function such

as log(fluorescence), log(Wrightian fitness) or growth rate, and colony size. For landscapes reporting genotypes with nucleotide mutations, all 12 classes of single mutations were considered. For landscapes reporting genotypes with amino acid mutations, all 380 mutations between any two amino acids were considered as single mutations. Genotypes with fitness at the minimum detection limit (e.g., non-fluorescent GFP genotypes) or that were lethal or non-growing (e.g., tRNA genotypes with Wrightian fitness relative to the wild-type = 0.5) were excluded. A final set of 12 studies with at least 10 single mutations and at least an average of 10 fitness effects measured per mutation were used for further analysis. **Table S1** lists the basic information of these phenotype landscapes. The original study of the tRNA fitness landscape reported Wrightian fitness relative to the wild-type; we computed Malthusian fitness = log(Wrightian fitness) in the present study.

To map the RNA-folding landscape, we studied a sequence of 72 nucleotides, the length of the tRNA gene used in the tRNA fitness landscape. The phenotype studied was the absolute value of the minimum free energy (MFE) of a sequence, calculated using ViennaRNA (https://www.tbi.univie.ac.at/RNA/). Two single mutants and the corresponding double mutant were randomly created for each of 2 million random background genotypes, and this set of genotypes was used for subsequent analyses.

Simulating idiosyncratic fitness landscapes

We simulated a series of 16-site fitness landscapes under *n*-order models with two states (A/T) per site, including all 65,536 genotypes. The fitness of a genotype is determined by additive effects (referred to as first order interactions) and interactive effects. For the k^{th} order interaction ($1 \leq k \leq 16$), there are $\frac{16!}{(16-k)!k!}$ interactive terms. For each of these terms, there are 2^k possible

state combinations. The fitness effect of each state combination of each interaction term for each order of interaction is drawn independently from the standard normal distribution, and the fitness of the genotype concerned is the sum of all these terms. Sixteen landscapes were made by including successively increasing orders of interactions. For instance, the first landscape contains only 1st order interactions (purely additive), the second landscape contains only 1st and 2nd order interactions, and the sixteenth landscape contains all orders of interactions. As expected, I_{id} increased with the number of orders of interactions included (orange circles in Fig. 2A). In each landscape, fitness values are linearly scaled to the interval of [0, 1]. In each of these landscapes, epistasis between mutations is symmetrically distributed with the mean equal to 0. We also simulated additive landscapes with larger *n* values to examine the linearity of fitness drops during mutation accumulation.

Estimating idiosyncrasy index

For each single mutation in a fitness landscape, we calculated its fitness effects on all genetic backgrounds available. For each mutation, we also derived a control set of fitness effects by randomly sampling (with replacement) the same number of pairs of genotypes from the landscape as used for the mutation and computing the fitness difference for each pair. We then calculated the range of fitness effects and standard deviation (SD) of fitness effects for each mutation and its control dataset. For each mutation, we calculated the ratio in the SD (or range) between the actual data and the control data. The average ratio across all single mutations is the I_{id} of the landscape, and the error bars in Fig. 1C are the standard error of the mean (SE). The same method is used to estimate I_{id} of other phenotype landscapes. Although empirical phenotype landscape data typically include only a small fraction of nonrandomly sampled genotypes and their phenotypes, this nonrandom sampling is not expected to substantially affect

I_{id} estimation, because both the variation of the effect of a mutation and the variation in the control data are estimated using the available landscape data.

Examining correlation between background fitness and mutational effect

For the simulated *n*-order landscapes and empirical landscapes (tRNA fitness, GFP activity, and RNA-folding), Pearson's correlation coefficient was calculated between mutational effect on a particular trait and background trait value for each single mutation. Mutations appearing on less than four backgrounds were excluded. Pearson's correlation coefficient was also calculated between all mutational effects and background trait values for each landscape.

In the tRNA fitness landscape, the fitness of each genotype was measured in six replicates. To exclude artificial correlation due to measurement error, the background fitness of each case of a single mutation is calculated using the mean fitness value from replicates 1-3, while the mutational effect is computed using mean fitness from replicates 4-6. Additionally, two mutations which are the reverse of each other on the same backgrounds can automatically create a negative correlation between all mutational effects and background fitness. Hence, in each landscape where this could occur we randomly chose a mutation or its reversion when pooling all mutations together (green diamonds in Fig. 2A; Fig. 2C; Fig. S2C-D).

For analysis of diminishing returns and increasing costs, mutations were deemed beneficial or detrimental depending on their effect on the wild-type genotype in GFP and tRNA, or a random arbitrary genotype in RNA-folding, or on a genotype with fitness value closest to the average fitness in the *n*-order landscapes.

Simulating evolutionary trajectories in mutation accumulation (MA)

For each empirical landscape, MA from an initial genotype was simulated by randomly choosing single mutations until the resulting genotype was non-functional (GFP) or for a maximum of 10 mutational steps (tRNA) or 50 mutational steps (RNA-folding). For all plots concerning MA, the mean phenotype value of the landscape was calculated from all genotypes.

For the GFP landscape, genotypes were not allowed to be revisited within a trajectory. If an MA trajectory was part of another simulated trajectory, the shorter trajectory was discarded. A total of 3,069 MA trajectories were simulated from each of 3,069 initial genotypes with activity equal to or greater than that of the wild-type. In the tRNA fitness landscape, 10,000 MA trajectories were simulated starting from the wild-type genotype. In the *n*-order fitness landscapes, 10,000 MA trajectories were simulated starting from the genotype with fitness closest to the 90^{th} percentile. For the RNA-folding landscape, a total of 350 MA trajectories were simulated starting with the final genotypes from the simulated adaptations.

Simulating adaptive trajectories

For each empirical landscape, adaptation from an initial genotype was simulated by randomly choosing a single beneficial mutation, which increased the value of the trait concerned, until no more single beneficial mutations were available. A total of 5,000 adaptive trajectories starting from 3,441 initial genotypes chosen from the bottom 15% of genotypes (activity $\leq$ -0.4) were simulated for the GFP landscape. A total of 350 adaptive trajectories starting from 350 initial genotypes chosen from the bottom 0.0175% of genotypes in the RNA-folding landscape were simulated. In the tRNA fitness landscape, we simulated five adaptive trajectories starting from each genotype with fitness = 0.5; trajectories longer than two steps were retained, totaling 15,878

trajectories. In the *n*-order fitness landscapes, adaptations start from all genotypes in the bottom 20% of fitness distribution; among 10 adaptation simulations starting from each genotype, trajectories equal to or longer than two steps were retained.

Data and code availability

Data analysis and simulations for all landscapes except the tRNA and model landscapes were performed using R version 3.5.2. Analysis and simulations for the tRNA and model landscapes were performed using Python version 3.6.9. All figures were made with matplotlib package in Python and Keynote. Code and new data are available at https://github.com/lyonsdm/idiosyncrasy.

Acknowledgements

We thank Alexey Kondrashov, Adam Lauring, and members of the Zhang lab for valuable comments. This work was supported by U.S. National Institutes of Health (NIH) research grant GM103232 to J.Z. D.M.L was supported by NIH 5F31AI140618-02.

References

1 Phillips, P. C. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**, 855-867 (2008).

2 Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435-440 (1988).

3 Crow, J. F. & Kimura, M. Efficiency of truncation selection. *Proc Natl Acad Sci U S A* **76**, 396-399 (1979).

4 Zhang, J. Epistasis analysis goes genome-wide. *PLoS Genet* **13**, e1006558 (2017).

5 Kemble, H., Nghe, P. & Tenaillon, O. Recent insights into the genotype-phenotype relationship from massively parallel genetic assays. *Evol Appl* **12**, 1721-1742 (2019).

6 Couce, A. & Tenaillon, O. A. The rule of declining adaptability in microbial evolution experiments. *Front Genet* **6**, 99 (2015).

7 Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* **461**, 1243-1247 (2009).

8 MacLean, R., Perron, G. G. & Gardner, A. Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in Pseudomonas aeruginosa. *Genetics* **186**, 1345-1354 (2010).

9 Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D. & Marx, C. J. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190-1192 (2011).

10 Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519-1522 (2014).

11 Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E. & Cooper, T. F. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193-1196 (2011).

12 Miller, C. R. The treacheries of adaptation. *Science* **366**, 418-419 (2019).

13 Maisnier-Patin, S. *et al.* Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nat Genet* **37**, 1376-1379 (2005).

14 Perfeito, L., Sousa, A., Bataillon, T. & Gordo, I. Rates of fitness decline and rebound suggest pervasive epistasis. *Evolution* **68**, 150-162 (2014).

15 de la Iglesia, F. & Elena, S. F. Fitness declines in Tobacco etch virus upon serial bottleneck transfers. *J Virol* **81**, 4941-4947 (2007).

16 Johnson, M. S., Martsul, A., Kryazhimskiy, S. & Desai, M. M. Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science* **366**, 490-493 (2019).

17 Wei, X. & Zhang, J. Patterns and mechanisms of diminishing returns from beneficial mutations. *Mol Biol Evol* **36**, 1008-1021 (2019).

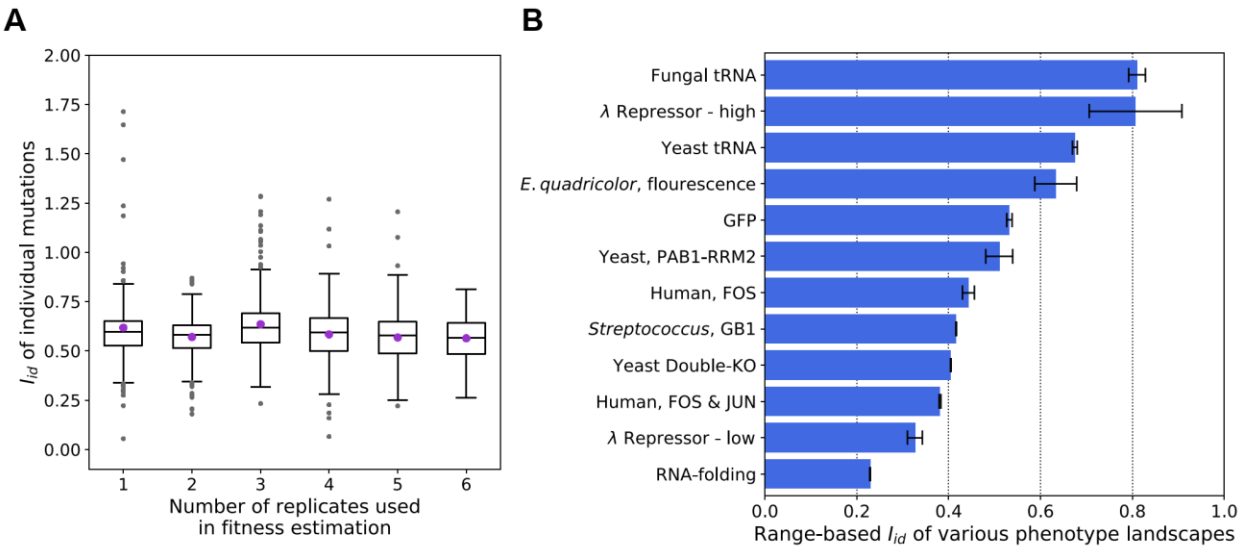
18 Sanjuan, R. & Elena, S. F. Epistasis correlates to genomic complexity. *Proc Natl Acad Sci U S A* **103**, 14402-14405 (2006).

19 Schoustra, S., Hwang, S., Krug, J. & de Visser, J. A. Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. *Proc Biol Sci* **283** (2016).

20 Blanquart, F., Achaz, G., Bataillon, T. & Tenaillon, O. Properties of selected mutations and genotypic landscapes under Fisher's geometric model. *Evolution* **68**, 3537-3554 (2014).

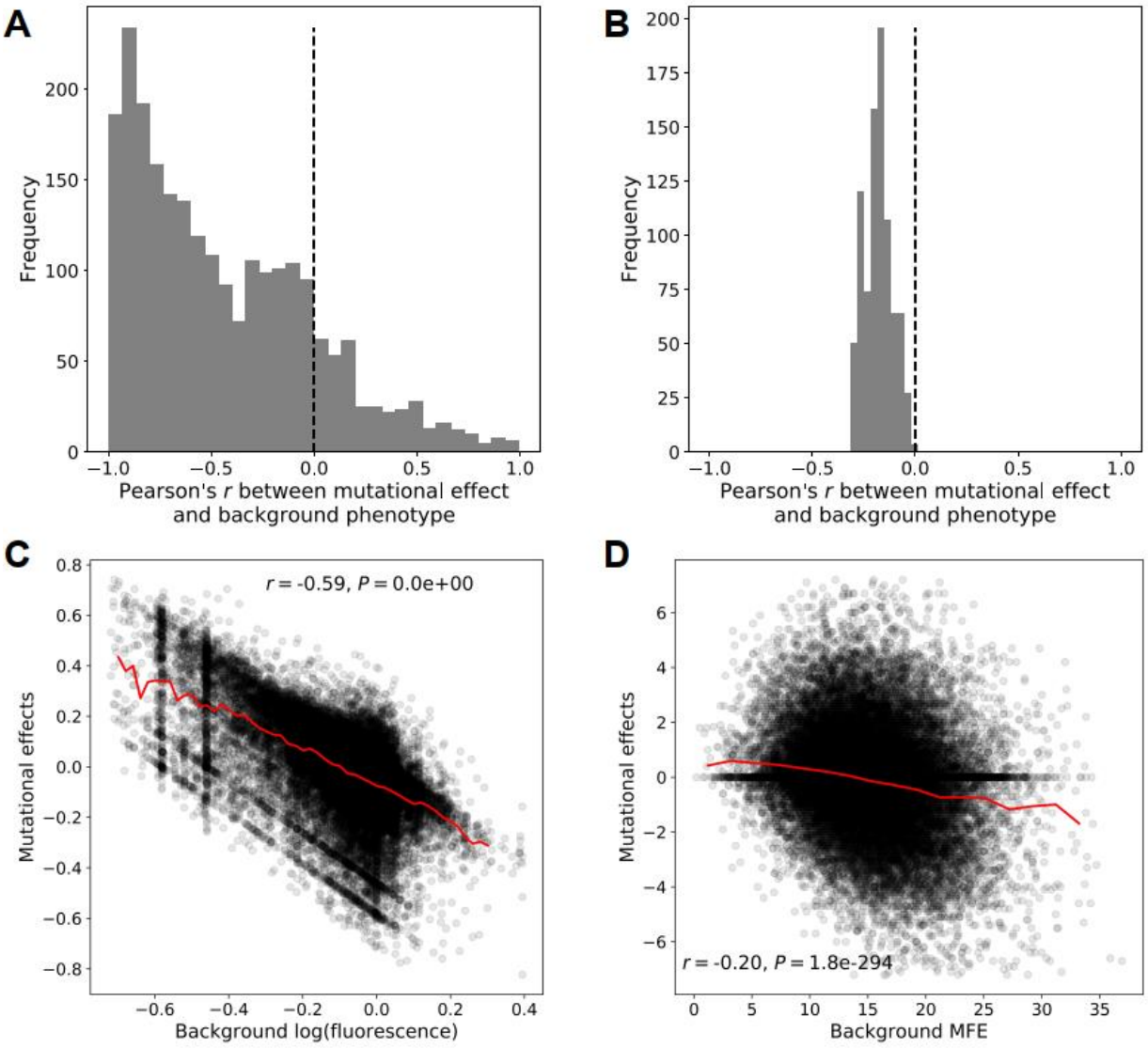
21 Draghi, J. A. & Plotkin, J. B. Selection biases the prevalence and type of epistasis along adaptive trajectories. *Evolution* **67**, 3120-3131 (2013).
22 Greene, D. & Crona, K. The changing geometry of a fitness landscape along an adaptive walk. *PLoS Comput Biol* **10**, e1003520 (2014).
23 Li, C., Qian, W., Maclean, M. & Zhang, J. The fitness landscape of a tRNA gene. *Science* **352**, 837-840 (2016).
24 Li, X., Lalic, J., Baeza-Centurion, P., Dhar, R. & Lehner, B. Changes in gene expression predictably shift and switch genetic interactions. *Nat Commun* **10**, 3886 (2019).
25 Domingo, J., Diss, G. & Lehner, B. Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature* **558**, 117-121 (2018).
26 Sarkisyan, K. S. *et al.* Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397-401 (2016).
27 Melamed, D., Young, D. L., Gamble, C. E., Miller, C. R. & Fields, S. Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. *RNA* **19**, 1537-1551 (2013).
28 Poelwijk, F. J., Socolich, M. & Ranganathan, R. Learning the pattern of epistasis linking genotype and phenotype in a protein. *Nat Commun* **10**, 4213 (2019).
29 Diss, G. & Lehner, B. The genetic landscape of a physical interaction. *eLife* **7**, e32472 (2018).
30 Olson, C. A., Wu, N. C. & Sun, R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr Biol* **24**, 2643-2651 (2014).
31 Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016).
32 de Visser, J. A. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet* **15**, 480-490 (2014).
33 Seetharaman, S. & Jain, K. Adaptive walks and distribution of beneficial fitness effects. *Evolution* **68**, 965-975 (2014).
34 Masel, J. & Trotter, M. V. Robustness and evolvability. *Trends Genet* **26**, 406-414 (2010).
35 Wagner, A. Robustness and evolvability: a paradox resolved. *Proc Biol Sci* **275**, 91-100 (2008).
36 Wei, X. & Zhang, J. Why phenotype robustness promotes phenotype evolvability. *Genome Biol Evol* **9**, 3509-3515 (2017).
37 Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D. S. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444**, 929-932 (2006).
38 Gerrish, P. J. & Lenski, R. E. The fate of competing beneficial mutations in an asexual population. *Genetica* **102-103**, 127-144 (1998).
39 Silander, O. K., Tenaillon, O. & Chao, L. Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS Biol* **5**, e94 (2007).

Supplementary Materials



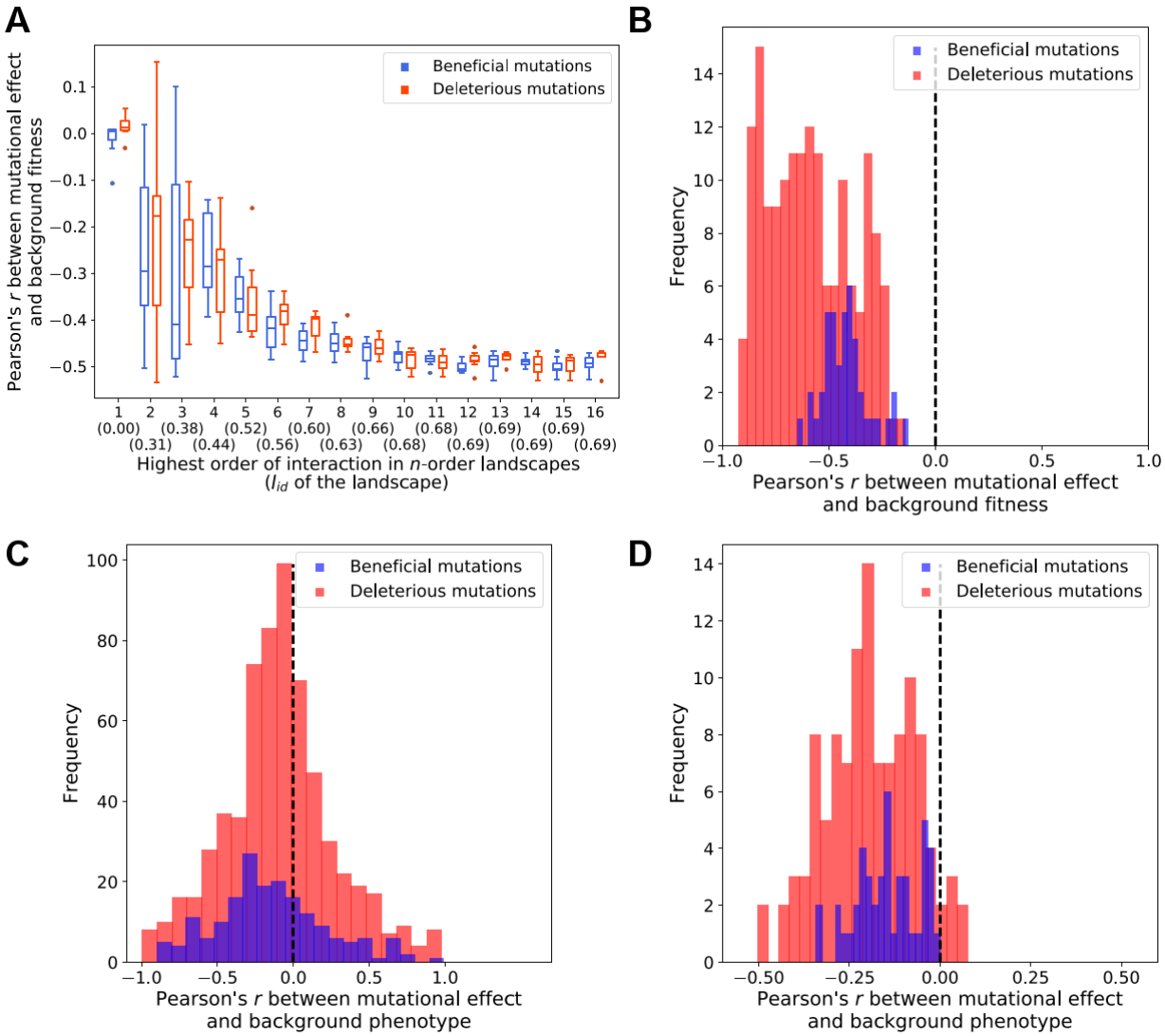
Supplemental Figure 2-1 The high idiosyncrasy indices (Iid) observed are not due to phenotype measurement errors or the use of standard deviation (SD) instead of range of mutational effects.

(A) SD-based Iid of the yeast tRNA fitness landscape is insensitive to the number of experimental replicates used in the fitness estimation. Boxplots show the distribution of Iid values of 828 single mutations in the tRNA landscape, calculated based on different numbers of replicates. The lower and upper edges of a box represent the first (qu1) and third (qu3) quartiles, respectively, the horizontal line inside the box indicates the median (md), the whiskers extend to the most extreme values inside inner fences, md ± 1.5(qu3 − qu1), and the grey dots represent values outside the inner fences (outliers). Violet dots show mean Iid of all mutations calculated based on respective numbers of replicates. (B) Range-based Iid for various phenotype landscapes. Error bars show standard errors. Detailed information of each landscape is provided in Table S1.



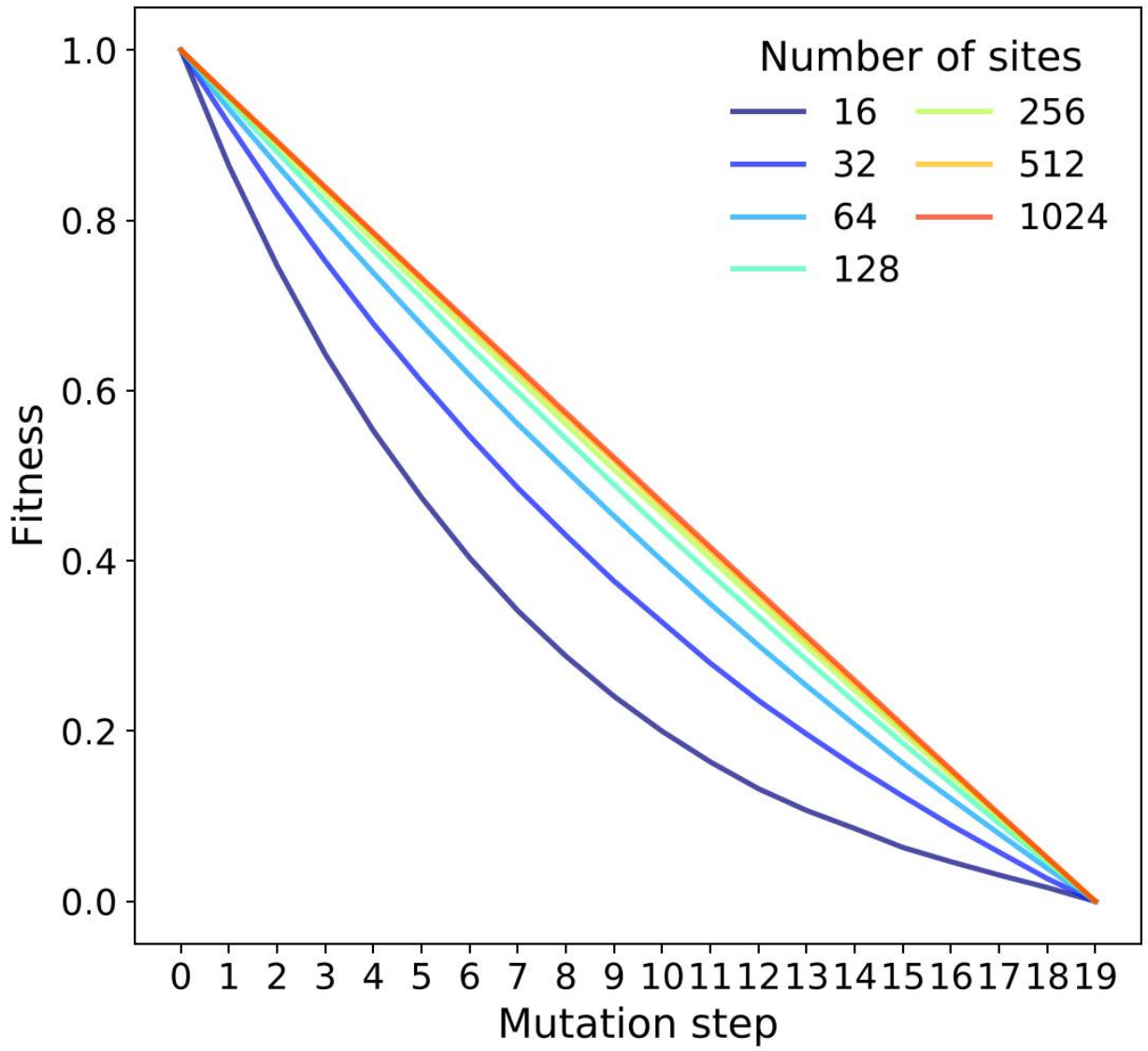
Supplemental Figure 2-2 Negative correlation between mutational effect and background phenotype in GFP and RNA-folding landscapes.

(A) Distribution of Pearson's correlation coefficient (r) between mutational effect and background phenotype for individual mutations in the GFP landscape. (B) Distribution of r for individual mutations in the RNA-folding landscape. (C) Relationship between background phenotype and mutational effect for all mutations in the GFP landscape. (D) Relationship between background phenotype and mutational effect for all mutations in the RNA-folding landscape. MFE, minimum free energy. The red line depicts the running mean in non-overlapping X-axis bins of width = 0.02 and 2 in (C) and (D), respectively, in all bins with more than 10 data points. There is no measurement error in the RNA-folding landscape. Shared measurement error between mutational effect and background fitness cannot be controlled for in GFP as replicate fitness measurements are not available. For each mutation and its reverse, we considered a random one of them in (C) and (D).



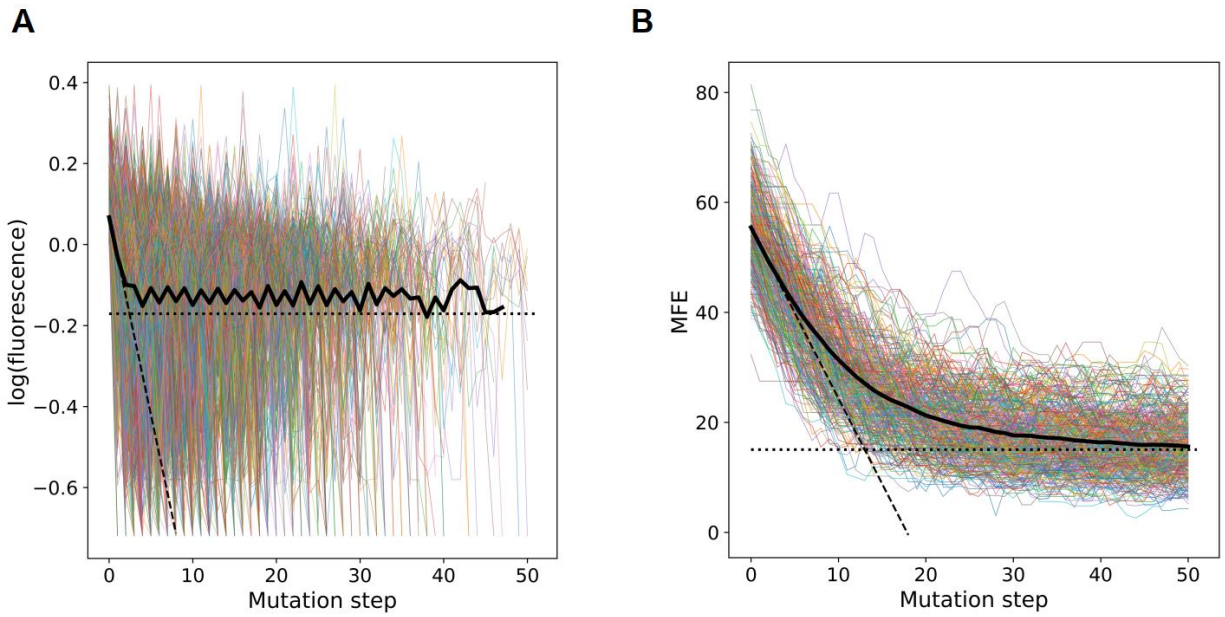
Supplemental Figure 2-3 Fig. S3. Patterns of correlation between mutational effect and background fitness/phenotype for individual beneficial or deleterious mutations in various landscapes.

(A) Boxplots showing distributions of correlations in a series of n-order landscapes of 16 sites (where the highest order of nonzero interaction is indicated on the X-axis) for beneficial (blue) and deleterious (red) mutations, respectively. The lower and upper edges of a box represent the first (qu1) and third (qu3) quartiles, respectively, the horizontal line inside the box indicates the median (md), the whiskers extend to the most extreme values inside inner fences, md ± 1.5(qu3 − qu1), and the dots represent values outside the inner fences (outliers). (B-D) Frequency distributions of correlations for individual beneficial mutations (blue) and deleterious mutations (red) in the tRNA (B), GFP (C), and RNA-folding (D) landscapes. Whether a mutation is beneficial or deleterious is determined in reference to the wild-type (tRNA and GFP) or an arbitrary reference genotype (n-order and RNA-folding). The wider distribution for deleterious than beneficial mutations is at least in part due to the larger number of deleterious than beneficial mutations.



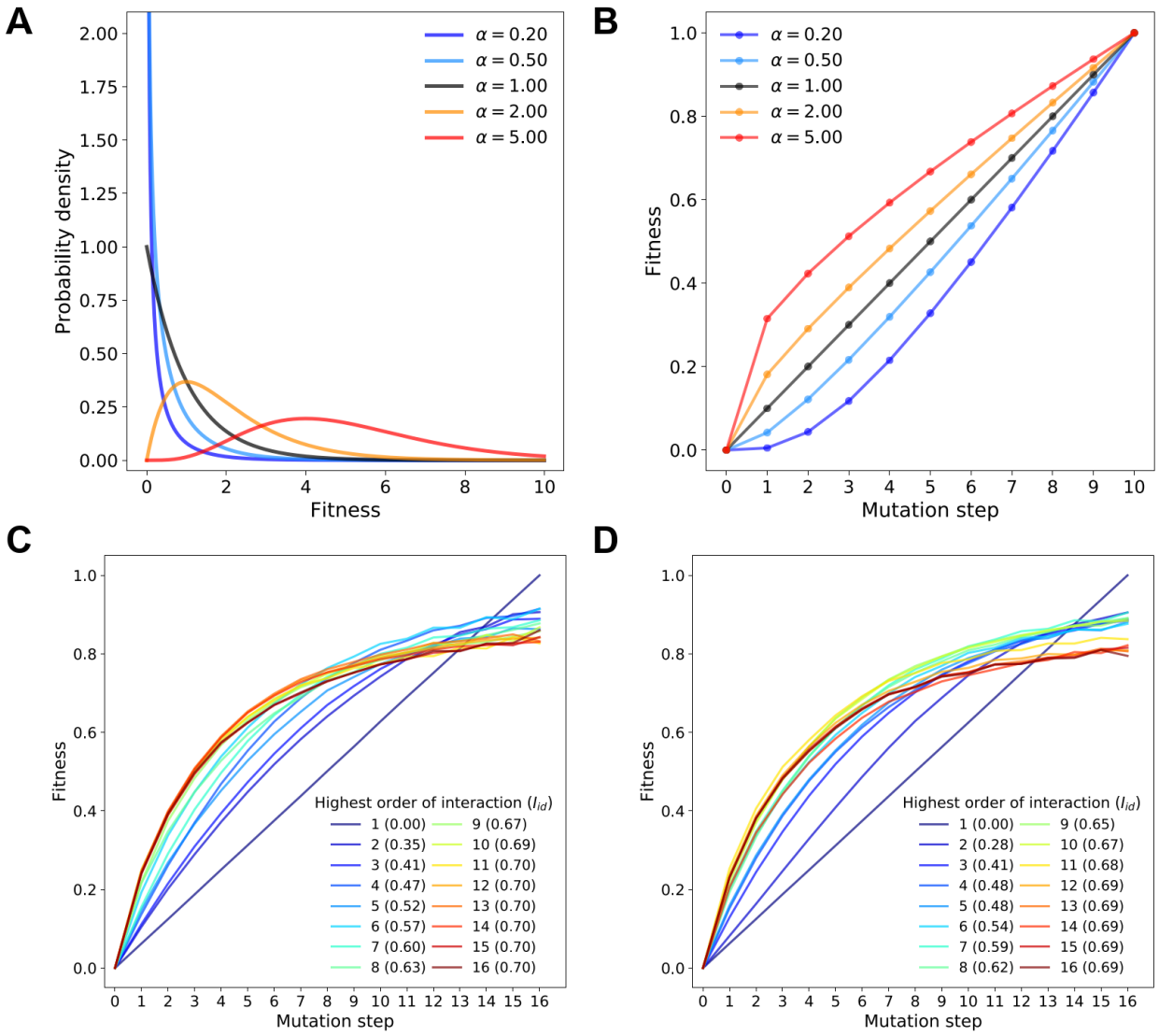
Supplemental Figure 2-4 Average fitness trajectories of mutation accumulation simulated in various n-order additive landscapes (k = 1) with different numbers of sites (n).

The mean trajectories are scaled so that the minimum fitness appearing in the trajectory is 0 and the maximum is 1 to allow direct comparison.



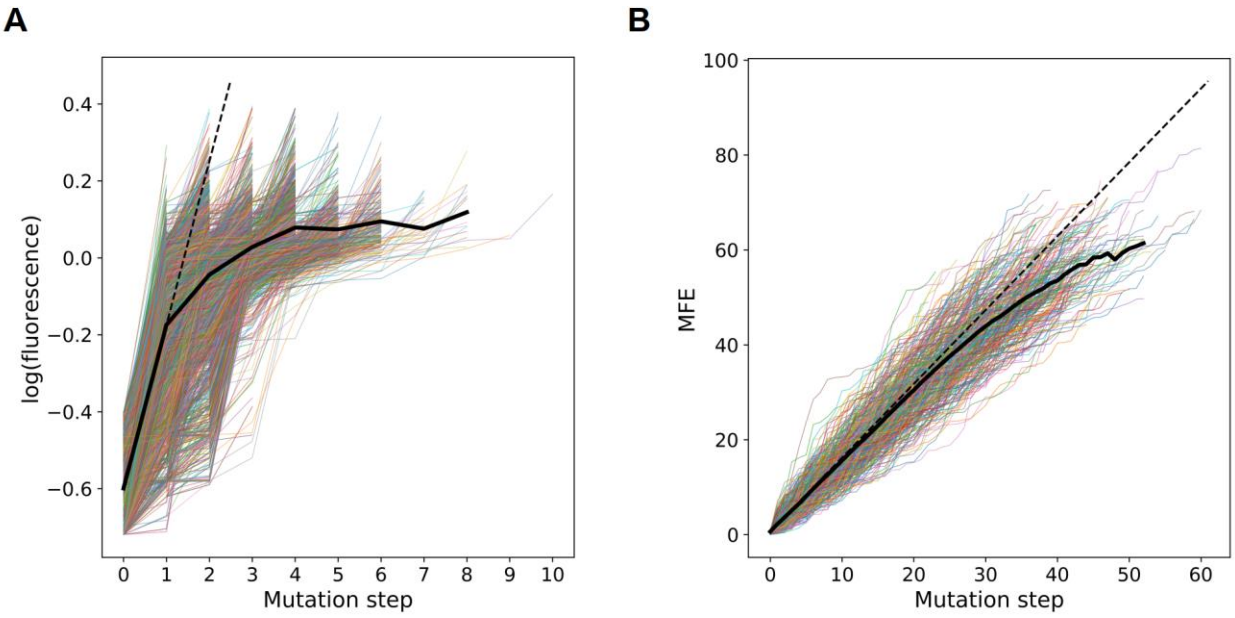
Supplemental Figure 2-5 Fitness declines decelerate during mutation accumulation as a result of idiosyncratic epistasis.

(A) A total of 5000 fitness trajectories of mutation accumulation simulated in the GFP landscape, with the average trajectory shown in black, at each step when the trajectory number exceeds 10. (B) A total of 350 fitness trajectories of mutation accumulation simulated in the RNA-folding landscape, with the average trajectory shown in black. The dotted lines indicate the mean phenotypic value of all genotypes in the landscape, excluding non-active genotypes in the GFP landscape. For comparison, the dashed line in (A) or (B) represents the predicted linear decline given the slope in the first mutational step.



Supplemental Figure 2-6 Idiosyncratic epistasis is necessary but not sufficient to cause decelerating adaptations.

(A) Gamma distributions of genotype fitness for house-of-cards landscapes, with different values of the gamma shape parameter \alpha. (B) Theoretically computed mean fitness trajectories of adaptation on landscapes in (A) with corresponding colors. (C) Average adaptive trajectories starting from the genotype with the lowest fitness (0), simulated in a series of n-order landscapes of 16 sites where each nonzero interaction term of each genotype is drawn from a gamma distribution of \alpha = 1. (D) Average adaptive trajectories starting from the genotype with the lowest fitness (0), simulated in a series of n-order landscapes of 16 sites where each nonzero interaction term of each genotype is drawn from a beta distribution with a=b=0.25. For each landscape in (C) and (D), the distribution of epistasis between mutations is symmetrical with mean equal to 0.



Supplemental Figure 2-7 Adaptation slows in empirical phenotype landscapes.

(A) A total of 5000 adaptive trajectories simulated in the GFP landscape, with the average trajectory shown in black, at each step when the trajectory number exceeds 10. (B) A total of 350 adaptive trajectories simulated in the RNA-folding landscape, with the average trajectory shown in black, at each step when the trajectory number exceeds 10. For comparison, the dashed line in (A) or (B) represents the predicted linear increase given the slope in the first mutational step.

Supplemental Table 2-1 Description of empirical landscapes, ordered by idiosyncrasy indices shown in Figure 2-1C

Landscape name	Description	Phenotype	# of genotypes	# of single mutations	References
λ Repressor - high	Phage lambda repressor (CI) at high expression	Repression of GFP expression	887	174	[24]
Fungal tRNA	tRNA-Arg(CCU) of post-whole-genome-duplication yeast species	log(fitness)	4,175	36	[25]
Yeast tRNA	*S. cerevisae* tRNA-Arg(CCU)	log(fitness)	65,537	828	[23]
GFP	*A.victoria* green fluorescent protein	log(fluorescence)	30,123	2,314	[26]
Yeast, PAB1-RRM2	*S. cerevisae* RNA-recognition motif 2 of poly(A)-binding protein 1	Fitness relative to wild-type (relative enrichment score)	40,722	10,668	[27]
E. quadricolor, flourescence	*Entacmaea quadricolor* fluorescent protein	log(fluorescence)	8,191	26	[28]
Human, FOS	Physical interaction between human FOS and JUN proteins, with mutations in FOS	Strength of FOS-JUN protein physical interaction	17,685	1,594	[29]
λ Repressor – low	Phage lambda repressor (CI) at low expression	Repression of GFP expression	887	174	[24]
Human, FOS & JUN	Physical interaction between human FOS and JUN proteins, with mutations in the two genes	Strength of FOS-JUN protein physical interaction	108,833	12,160	[29]
RNA-folding	72-nucleotide sequence, folding predicted by ViennaRNA	Minimum free energy	2,000,000	864	Present study
Streptococcus, GB1	IgG-binding domain of protein G	Biding affinity to IgG-FC	536,962	20,900	[30]
Yeast Double-KO	*S. cerevisae* double gene deletions, genome-wide	Fitness relative to wild-type (colony size)	14,578,495	10,564	[31]

Chapter 3 Evidence for the Selective Basis of Transition-to-Transversion Substitution Bias in Two RNA Viruses

Note: This chapter is a modified version of the published article:

Lyons DM, Lauring AS. 2017. Evidence for the Selective Basis of Transition-to-Transversion Substitution Bias in Two RNA Viruses. Mol Biol Evol. 34(12):3205–3215. doi:10.1093/molbev/msx251.

Abstract

The substitution rates of transitions are higher than expected by chance relative to those of transversions. Many have argued that selection disfavors transversions, as nonsynonymous transversions are less likely to conserve biochemical properties of the original amino acid. Only recently has it become feasible to directly test this selective hypothesis by comparing the fitness effects of a large number of transition and transversion mutations. For example, a recent study of six viruses and one beta-lactamase gene did not find evidence supporting the selective hypothesis. Here we analyze the relative fitness effects of transition and transversion mutations from our recently published genome-wide study of mutational fitness effects in influenza virus. In contrast to prior work, we find that transversions are significantly more detrimental than transitions. Using what we believe to be an improved statistical framework, we also identify a similar trend in two HIV datasets. We further demonstrate a fitness difference in transition and transversion mutations using four deep mutational scanning datasets of influenza virus and HIV, which provided adequate statistical power. We find that three of the most commonly cited radical/conservative amino acid categories are predictive of fitness, supporting their utility in studies of positive selection and codon usage bias. We conclude that selection is a major contributor to the transition:transversion substitution bias in viruses and that this effect is only

partially explained by the greater likelihood of transversion mutations to cause radical as opposed to conservative amino acid changes.

Introduction

Fifty years ago, Walter Fitch noted that the nucleotide substitution pattern in cytochrome c is non-random (Fitch 1967). If random, transversions (purine-pyrimidine changes) should be observed twice as often as transitions (purine to purine or pyrimidine to pyrimidine changes) solely due to the accessible mutations. However, Fitch observed that transitions are more common than transversions. In fact, this transition-transversion (Ts:Tv) substitution bias has been noted across many proteins and phyla, and phylogenetic inferences account for this bias by weighting transversions more than transitions (Gojobori et al. 1982; Kumar 1996; Wakeley 1996; Petrov and Hartl 1999; Rosenberg et al. 2003; Lynch 2010; Duchêne et al. 2015).

The underlying reasons for this widespread Ts:Tv substitution bias are largely unknown. Two main hypotheses, which are not mutually exclusive, have emerged to explain this phenomenon: the mutational hypothesis and the selective hypothesis. The mutational hypothesis holds that the transition mutation rates of polymerases are higher than the transversion rates. This hypothesis is supported by the observation of a transitional bias in both coding and non-coding regions (Zhang and Gerstein 2003; Jiang and Zhao 2006) as well as mutation rate analyses showing higher transition mutation rates (Denver et al. 2004; Pauly, Procario, et al. 2017). The selective hypothesis posits that natural selection disfavors transversions. This hypothesis is based on the observation that, depending on codon usage, non-synonymous transitions are more likely to conserve important biochemical properties of the original amino acid (Vogel and Kopun 1977; Miyata et al. 1979; Zhang 2000). For example, a mutation that changes the charge of an amino

acid is a “radical” change, while one that does not is a “conservative” change. However, this provides only indirect evidence for the selective hypothesis and the extent to which radical/conservative distinctions are predictive of fitness is unclear. Radical changes do occur less often than conservative ones during protein evolution. Arguments based on this observation can be circular (Dagan et al. 2002; Yampolsky and Stoltzfus 2005). If the transition mutation rate is higher and transitions are more likely to be conservative, then conservative changes will occur more often simply due to the transitional mutation bias. Furthermore, the radical/conservative amino acid distinctions may be overly broad and arbitrary. For example, the hydrophobicity of amino acids may be more constrained for some proteins while their size may be more constrained for other proteins. One could arbitrarily choose a biochemical distinction that would suggest transversions are more likely to be conservative.

Only recently has it become experimentally tractable to test directly the selective hypothesis by comparing the fitness effects of a large number of transition and transversion mutations. A recent study (Stoltzfus and Norris 2016) compared the fitness effects of missense transitions and transversions reported in eight studies of mutational fitness effects. This meta-analysis included: a beta-lactamase gene (TEM1), two HIV genes (integrase and capsid), and five genome-wide studies of viruses (Sanjuán et al. 2004; Purificación Carrasco et al. 2007; P. Carrasco et al. 2007; Domingo-Calap et al. 2009; Peris et al. 2010; Jacquier et al. 2013; Rihn et al. 2013; Rihn et al. 2015). Stoltzfus and Norris did not identify a statistically significant difference in the fitness effects of transitions and transversions (Ts-Tv) in any of the viral datasets (2016). They did find a statistically significant difference after combining the data, which was deemed to be of questionable biological significance.

Here, we revisit this question using our recently published library of randomly distributed point mutations in influenza A virus (Visher et al. 2016). In contrast to other viral datasets, we find that transitions are significantly less detrimental than transversions. We apply what we believe is an improved statistical framework and identify a similar trend in the HIV integrase and capsid datasets (Rihn et al. 2013; Rihn et al. 2015). We expand our analysis to include deep mutational scanning studies of one HIV gene and two genes from two different strains of influenza (Bloom 2014; Doud et al. 2015; Doud and Bloom 2016; Haddox et al. 2016). The distribution of fitness effects of transversions is shifted toward more detrimental effects compared to transitions at some points along the fitness distribution in each gene, and transitions are never more detrimental. Three of the most commonly cited radical/conservative distinctions are predictive of mutational fitness effects. However, transversions are more detrimental than transitions even when controlling for their greater likelihood to be radical.

Results

Transitions are less detrimental in influenza A virus

We recently published a library of 128 point mutants in influenza A virus (Visher et al. 2016). As in other studies of viral mutational fitness effects, the substitution types were chosen at random and the fitness values were assessed individually. Our library contains 95 mutants distributed across the eight genomic RNA in proportion to the size of each segment and an additional 33 random mutations in the segments encoding the surface proteins hemagglutinin (HA) and neuraminidase (NA). Thus, a total of 57 mutations occur in HA and NA and 71 in the other six segments. For the present study, we excluded 27 synonymous mutations and 6

beneficial missense mutations and considered the remaining 95 missense transitions and transversions that had fitness values equal to or less than one. We performed our analyses on the total library (N = 95), the genes encoding the internal proteins only (N = 53), and the genes encoding the surface proteins only (N = 42) (Supplemental Table 3-1; Total, Internal, and Surface datasets, respectively). Each of these data subsets is larger than the analogous genome-wide datasets of other viruses.

As described by Stoltzfus and Norris (Stoltzfus and Norris 2016), we identified differences in the fitness effects of Ts and Tv by calculating the area under the curve (AUC) of a receiver operating characteristic (ROC) curve. An ROC curve plots the true positive rate against the false positive rate of a binary classifier system as the discrimination threshold varies. Consider a hypothetical example using a binary classifier system to predict whether a mutation is a transition or a transversion without prior knowledge of its identity. If the fitness of the mutation is above a fitness level threshold, it is categorized as a transition, and if it is below the threshold, it is categorized as a transversion. The AUC is equivalent to the probability that a randomly chosen transition is more fit than a randomly chosen transversion. The AUC can be calculated from the Mann-Whitney U test statistic (see Materials and Methods) (Hanley and McNeil 1982; Mason and Graham 2002). An AUC of 1 would indicate that all transitions are more fit than all transversions, and an AUC of 0 would indicate that all transversions are more fit than all transitions. The null expectation is an AUC of 0.50. To identify the points along the fitness distribution at which there is a Ts-Tv difference, we calculated the AUC among mutations at or above 10 successively higher fitness thresholds, starting at 0 and increasing each threshold by 0.10.

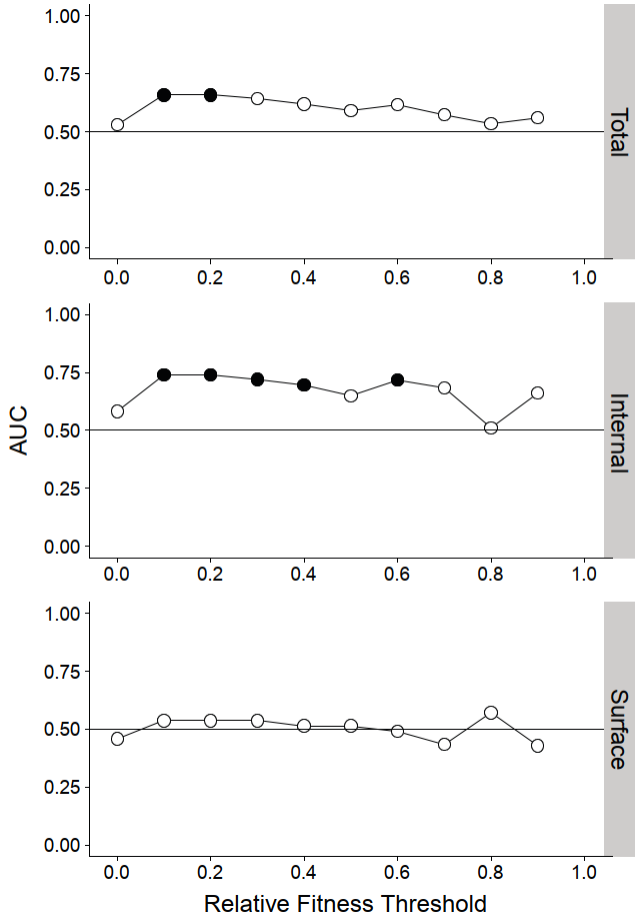


Figure 3-1 Differences in Ts and Tv Fitness Effects as measured by AUC

Transition-transversion (Ts-Tv) fitness differences as measured by area under the curve of an ROC curve (AUC) for the total, internal, and surface influenza datasets. To identify the points along the fitness distribution at which there is a Ts-Tv difference, the AUC was calculated among mutations at or above 10 successively higher fitness thresholds, starting at 0 and increasing by 0.10. Filled circles denote $p < 0.05$ for a one-sided Mann-Whitney U test where the alternative is transitions are more fit at that threshold. Lines shown to clarify trends only. Plotted data and raw p-values can be found at https://github.com/lauringlab/tstv_paper.

Using the AUC criterion, we did not find a statistically significant difference between the fitness effects of transition and transversion mutations across the total dataset, the internal dataset, or the surface (HA/NA) dataset (**Figure *3-1***, relative fitness threshold of 0). However, we detected a significant difference when we examined only non-lethal mutations (**Figure *3-1***, relative fitness threshold of 0.1; there were no mutations with a fitness value between 0 and 0.1). Among the viable fraction, transitions are significantly more fit than transversions (AUC = 0.65,

p = 0.03). This fitness difference is more pronounced in the internal dataset (AUC = 0.74, p = 0.02) but is not present among the surface dataset. As the thresholds approach 1, the Ts-Tv difference approaches 0.5 and loses significance (perhaps due to decreasing sample size) in both the genome-wide and internal datasets. In contrast, we never found transversions to be significantly more fit than transitions. Thus, viable transitions are more fit than viable transversions, and this fitness difference varies between the internal and surface datasets.

Our findings in influenza differ from those in other viruses (HIV integrase, HIV capsid, TEV, F1, VSV, Qβ, ϕX174) (Stoltzfus and Norris 2016). While selective constraints could potentially differ among these viruses, another factor could be the greater statistical power of our influenza dataset due to our larger sample size as compared to the other genome-wide viral datasets. Our influenza virus datasets were smaller, however, than those in the HIV integrase and capsid studies.

An alternative statistical approach better captures Ts-Tv fitness differences

The fact that we could only identify a significant difference in influenza by excluding lethal mutations suggests an inherent bias in the AUC threshold analysis and led us to reexamine our statistical framework. If transitions are more likely to be lethal in influenza, this would offset their advantage among viable mutations. More generally, as the fitness threshold increases, the AUC reflects only the Ts-Tv differences at the higher end of the fitness distribution. In fact, when we applied decreasing, as opposed to increasing, thresholds, many of our conclusions were opposite from those obtained with increasing thresholds (Supplemental Figure 3-1). In this case, the influenza total dataset is weighted by strongly detrimental transitions and no fitness differences between Ts and Tv were detected at any threshold. In the larger HIV capsid and

combined integrase and capsid datasets, we observed a previously unrecognized, and statistically significant, Ts-Tv difference at decreasing thresholds driven by the inclusion of strongly detrimental transversions (Stoltzfus and Norris 2016).

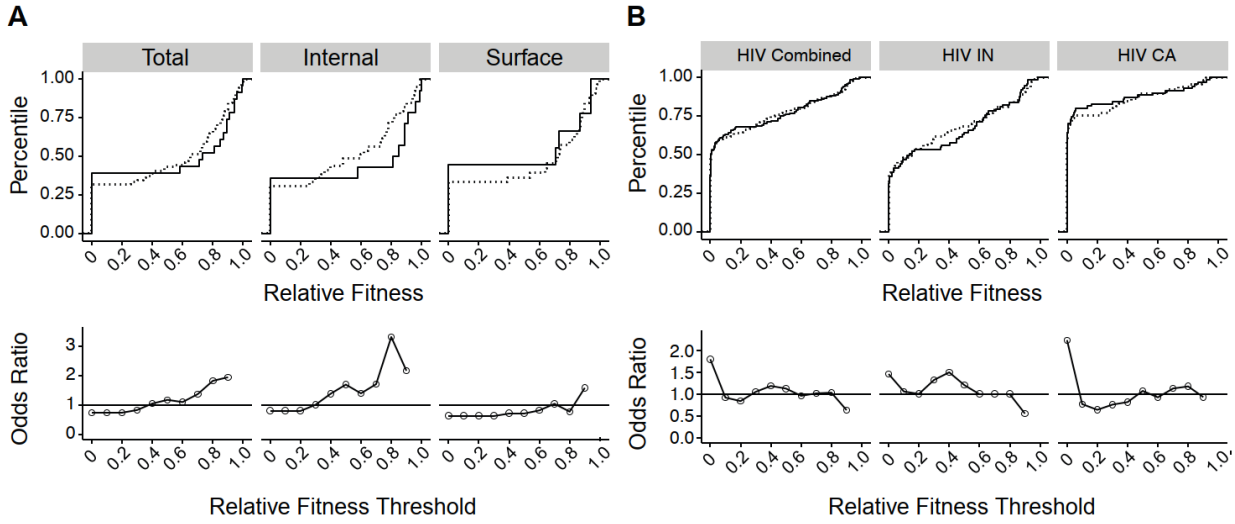


Figure 3-2 Comparisons of the Distribution of Ts and Tv Fitness Effects

Empirical cumulative distribution functions of transitions (solid line) and transversions (dotted line) in our influenza datasets (A, top) and the HIV combined, integrase (IN), and capsid (CA) datasets (B, top). Odds ratios indicate the odds of a transversion versus a transition to be at or below each of 10 relative fitness thresholds as estimated by a Fisher test for our influenza datasets (A, bottom) and HIV datasets (B, bottom). Filled circles denote $p < 0.05$ for a two-sided test with Holm-Bonferroni correction. Lines shown to clarify trends only.

To capture differences in the distribution of fitness effects between transitions and transversions more completely, we compared the empirical cumulative distribution functions (CDF) of each mutation type. The CDF reveal only subtle differences between transitions and transversions in the total influenza dataset, with more transitions at low fitness levels and more transversions at intermediate fitness levels (**Figure *3-2***A, top). The difference between transitions and transversions is greater for the internal dataset, where 75% of transversions have a lower fitness compared to only 50% of transitions at a fitness threshold of 0.8. However, transitions are proportionally over-represented below a fitness of 0.3. In contrast, the AUC threshold analysis for the internal influenza dataset seems to suggest a Ts-Tv fitness difference

starting at 0.1 and decreasing to no difference at 0.8 (**Figure *3-1***), precisely the opposite of the differences in the distributions (Figure 3-2A).

Recognizing these issues, we implemented a different statistical approach that is an explicit comparison of the Ts and Tv CDF and is therefore better able to resolve differences in the distributions of fitness effects. We used 10 fitness thresholds from 0 to 0.9 with a step of 0.1. At each fitness level, we performed a Fisher test to compare the proportion of transversions and transitions at or below the threshold. The estimated effect size is the odds ratio for transversions to be at or below the threshold compared to transitions. As this approach never excludes data, it is less biased by the fitness values at the tails of the distributions. Unlike the AUC (see Figure 3-1), the odds ratios closely follow the divergence in the corresponding CDF (compare top and bottom panels in Figure 3-2).

We used this approach to reanalyze data from our influenza datasets as well as those for HIV integrase (IN) and capsid (CA). Using a conservative Holm-Bonferroni correction for multiple comparisons, we did not find a statistically significant difference between transition and transversion mutations in influenza at any fitness level (Figure 3-2A, bottom). The CDF suggests that the impact of transitions relative to transversions at low fitness levels (0-0.3) is indeed offset by their impact at higher ones (0.6-1). There are no significant Ts-Tv differences in either of the two HIV datasets or in the combined dataset. However, there is a trend across the HIV datasets suggesting that transversions are more likely to be lethal as compared to transitions (Figure 3-2B, bottom, first threshold), an effect missed by the AUC analysis. We note that while the original HIV studies considered a fitness below 0.02 to be lethal (Rihn et al. 2013; Rihn et al. 2015); we

applied a strict and consistent criterion for lethality across the datasets and considered a fitness of 0 to be lethal.

Transversions are more detrimental in larger influenza and HIV datasets

While the influenza and HIV datasets are relatively large for studies of viral mutational fitness effects, they sample only a small fraction of the total number of possible point mutations. The influenza library contains a median of 11 missense mutations per gene and the HIV IN and CA datasets have 156 and 135 missense mutations, respectively. To increase our power, we analyzed available data from deep mutational scanning (DMS) studies of four viral proteins: the nucleoprotein (NP) of influenza A/Puerto Rico/8/1934 H1N1 and influenza A/Aichi/2/1968 H3N2, the HA protein from the same strain as our mutants (influenza A/WSN/1933 H1N1), and the HIV envelope (ENV) protein (Bloom 2014; Doud et al. 2015; Doud and Bloom 2016; Haddox et al. 2016). Importantly, these four studies all used the same approach and were performed in the same laboratory.

DMS uses high-throughput mutagenesis to introduce every single amino acid substitution in a given gene followed by deep sequencing to measure the change in frequency of each mutation after passage or selection. The effect of each mutation is often reported as a site preference, which represents the expected proportion of an amino acid at a site if all amino acids at that site were present at equal proportions prior to passaging. We derived a relative site preference from these data by dividing the site preference for each mutant by the site preference for the "wild type" amino acid. We found that relative site preference is a reasonable surrogate for relative fitness, as they are well correlated for mutations in the WSN33 HA gene (Spearman correlation 0.71, $p = 2.5 \times 10^{-5}$, Table 1). Our fitness values for WSN33 NP also exhibit a

statistically significant correlation with the DMS data from the closely related PR8 H1N1 strain, but not with the more distant H3N2 strain.

The DMS studies report changes at the amino acid level. Therefore, for each codon in the nucleotide sequence, we asked which amino acid substitutions could only be made by a single transition (Ts-only) and which could only be made by a single transversion (Tv-only). We excluded the amino acid substitutions that were accessible by both transitions and transversions as well as those that required more than one mutation per codon. We then compared the relative site preferences of Ts-only amino acid changes to those of Tv-only using the Fisher threshold strategy. We used an initial threshold of 0.05 rather than 0, since there are no site preferences of 0 in these datasets, and DMS studies are known to under-sample the lethal fraction. The large sample sizes of the DMS studies (Supplemental Table 3-1) allowed us to use more thresholds (increasing each by 0.05 instead of 0.01), thereby identifying Ts-Tv differences in the CDF with greater precision.

We found transversions to be significantly more detrimental than transitions at a subset of relative site preference levels in three of the four DMS datasets (Figure 3-3). In NP (H3N2), transversions tend to be more detrimental than transitions across most of the fitness distribution, but no threshold achieved statistical significance using a Holm-Bonferroni correction (Figure 3-3A). In NP (H1N1), transversions are significantly more likely to be highly detrimental than transitions (first threshold), and there is a trend for transversions to be more detrimental at higher relative site preferences as well (Figure 3-3A). Contrary to the trend in our smaller influenza study, the Ts-Tv fitness differences are larger and more broadly distributed in genes coding for

the two surface proteins, HA and HIV ENV (Figure 3-3B), as compared to the genes coding for the internal influenza NP proteins. In HA (H1N1), transversions are significantly more detrimental than transitions across most of the fitness distribution. The Ts-Tv difference is especially pronounced in HIV ENV, for which the odds of a transversion being highly detrimental (thresholds from 0.05-0.20) is 2-5 times greater than those of a transition. Across the four datasets, we never found transitions to be significantly more detrimental than transversions, and the odds ratio is rarely below 1 for any of the datasets. Thus, with an improved statistical approach and greater power, we found transitions to be less damaging than transversions in proteins from two viruses.

Differences in Ts and Tv fitness effects within radical and conservative substitution classes

We next asked why transversions are more detrimental than transitions. The genetic code constrains the type of amino acid substitutions accessible by mutation, and it has been proposed

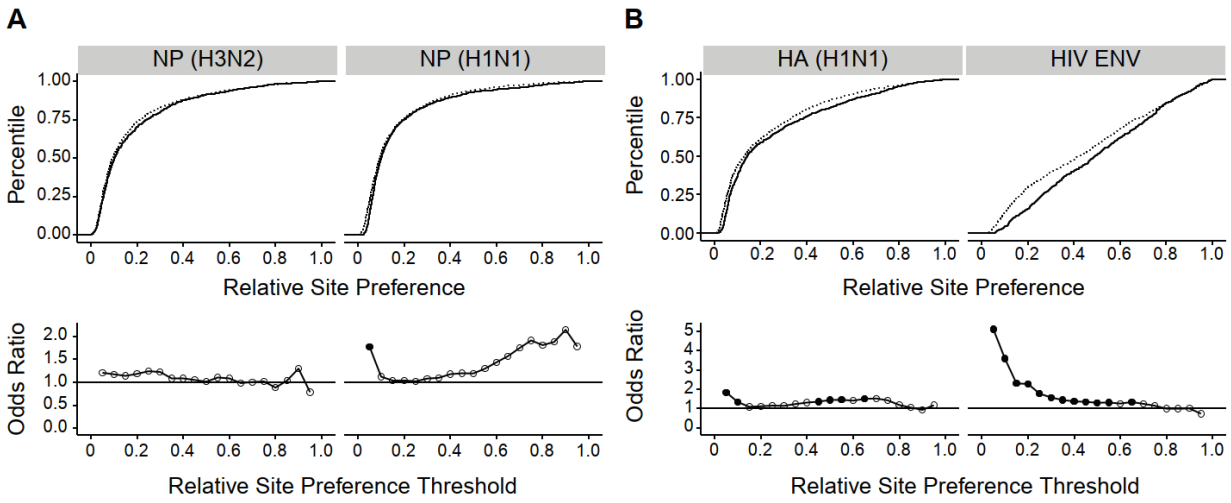


Figure 3-3 Distribution of Ts and Tv Fitness Effects in Deep Mutational Scanning Datasets

Empirical cumulative distribution functions of transitions (solid line) and transversions (dotted line) in two nucleoprotein (NP) proteins (A, top) and two antigenic surface proteins influenza hemagglutinin (HA) and HIV envelope (ENV) (B, top). Odds ratio estimated by a Fisher test comparing the odds of a transversion versus a transition to be at or below each of 19 fitness thresholds, beginning at a fitness of 0.05 and increasing by 0.05, for the same datasets (A and B, bottom). Filled circles denote $p < 0.05$ for a two-sided test with Holm-Bonferroni correction. Lines shown to clarify trends only. Plotted data and raw p-values can be found at https://github.com/lauringlab/tstv_paper.

that transversions are more detrimental because they are more likely to cause substitutions that radically alter biochemical properties of the original amino acid. We therefore examined whether the observed fitness differences could be explained by the differences in the accessibility of radical versus conservative amino acid changes by transitions and transversions.

We used the Fisher threshold strategy to test whether radical amino acid changes are more detrimental than conservative changes for three of the most commonly cited biochemical distinctions, which categorize amino acids based on charge, polarity, and polarity and size, (Miyata et al. 1979; Zhang 2000) (see Supplemental Table 3-2). Similar categories have been used in other studies of protein evolution (Epstein 1967; Grantham 1974). Radical amino acid substitutions of all three types are more detrimental than conservative changes in the two NP proteins across much of the fitness distribution (Figure 3-4). Radical changes of polarity (red) and polarity and size (blue) are also more detrimental than conservative changes in the two surface proteins HA (H1N1) and HIV ENV. Radical charge changes (black) have similar effects on fitness as compared to conservative changes in both HA (H1N1) and HIV ENV. Despite this variation in the impact of changes in charge, these simple categories are remarkably predictive of fitness effects across these four proteins.

Transversions may be more detrimental than transitions in these four proteins if they are more likely than transitions to cause a radical amino acid change. As above, we considered amino acid substitutions that could only be made by a single transition (Ts-only) or by a single transversion (Tv-only). Using a Fisher test, we compared the odds that a Tv-only amino acid is radical to the odds that a Ts-only amino acid change is radical as defined by each of the three

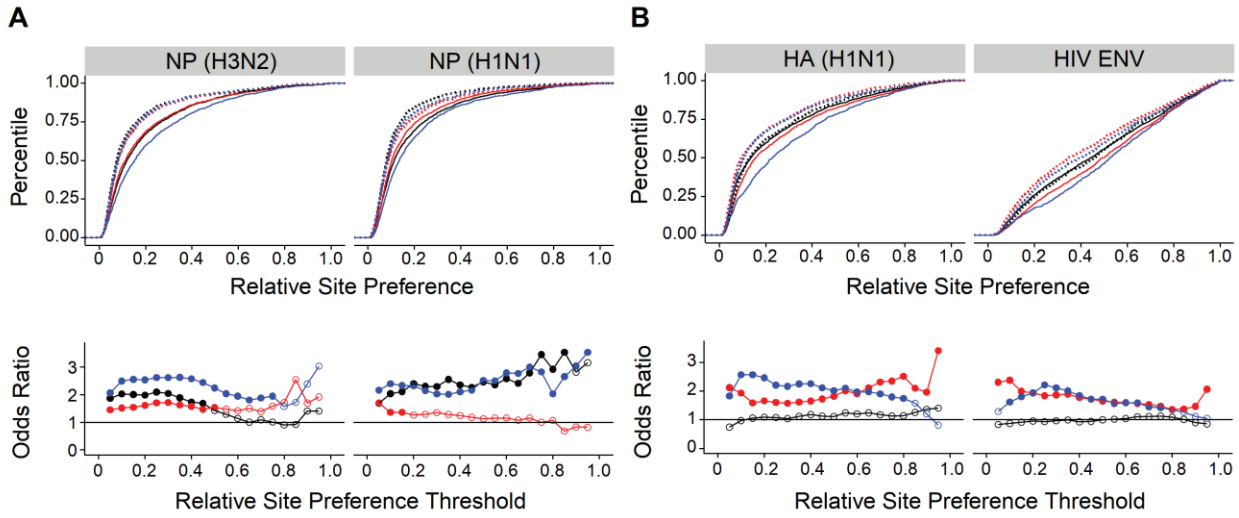


Figure 3-4 Distribution of Radical and Conservative Amino Acid Changes

Empirical cumulative distribution functions (CDF) of conservative (solid lines) and radical (dotted lines) amino acid changes in deep mutational scanning datasets of two NP proteins (A, top) and two antigenic surface proteins (B, top). Shown for amino acid changes classified by charge (black), polarity (red), and polarity and size (blue). Odds ratio estimated by a Fisher test comparing the odds of a radical versus a conservative amino acid change to be at or below each of 19 fitness thresholds, beginning at a fitness of 0.05 and increasing in steps of 0.05, for the same datasets (A and B, bottom). Color scheme is the same as in the CDFs. Filled circles denote $p < 0.05$ for a two-sided test with Holm-Bonferroni correction. Lines shown to clarify trends only. Plotted data and raw p-values can be found at https://github.com/lauringlab/tstv_paper.

categories above. We considered all possible Tv-only and Ts-only amino acid changes in these four proteins. For all four proteins, the odds of a transversion causing a radical change of any of these three types is significantly greater than the odds of a transition causing a radical change (Table 2). This difference is greatest for amino acid substitutions that affect polarity for all genes.

We next examined whether the fact that transversions are more likely to be radical explains all of the observed differences in Ts-Tv fitness effects in the DMS studies. If the fitness differences can be accounted for by this bias, the differences should be eliminated when comparing both radical Ts to radical Tv and conservative Ts to conservative Tv. Alternatively, if this bias does not account for the difference, one would see a difference in fitness between transitions and transversions among radical or among conservative changes of a given category.

We first compared radical transitions to radical transversions. In NP (H3N2), there are no significant Ts-Tv differences among radical charge (black) or radical polarity and size (blue) changes. However, among radical polarity (red) changes, transversions are more detrimental than transitions—an effect not seen in the overall dataset (Figure 3-3A). In NP (H1N1), radical transversions are more detrimental than radical transitions for all three amino acid classifications at the first threshold (Figure 3-5A). In both NP proteins, the Ts-Tv fitness difference is greater in magnitude among radical polarity changes as compared to the differences among all mutations (the odds ratio is > 3 in Figure 3-5A but is ≤ 2 in Figure 3-3A), indicating that controlling for radical transversions can increase rather than eliminate Ts-Tv differences. In HA (H1N1), there are no significant Ts-Tv differences among radical charge changes. However, among radical polarity and polarity and size changes, transversions are more detrimental than transitions.

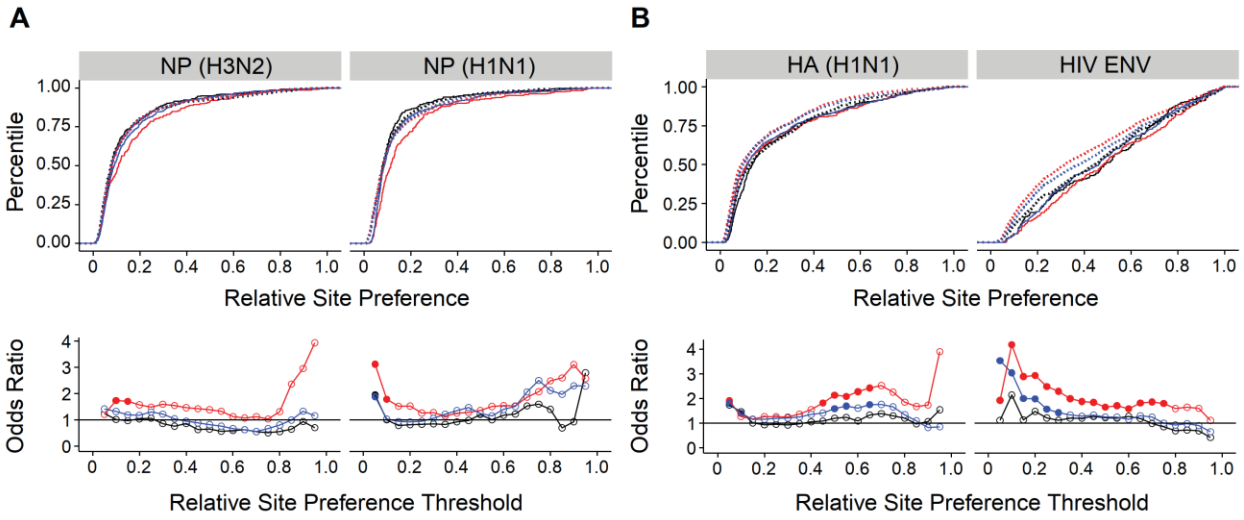


Figure 3-5 Distribution of Radical Ts and Radical Tv

Empirical cumulative distribution functions of radical transitions (solid lines) and radical transversions (dotted lines) in deep mutational scanning datasets of two NP proteins (A, top) and two antigenic surface proteins (B, top). Computed for radical amino acid changes classified by charge (black), polarity (red), and polarity and size (blue). Odds ratio estimated by a Fisher test comparing the odds of a radical transversion versus a radical transition to be at or below each of 19 fitness thresholds, beginning a t a fitness of 0.05 and increasing by 0.05, for the same datasets (A and B, bottom). Color scheme is the same as in the CDFs. Filled circles denote $p < 0.05$ for a two-sided test with Holm-Bonferroni correction. Lines shown to clarify trends only. Plotted data and raw p-values can be found at https://github.com/lauringlab/tstv_paper.

Similarly, in HIV ENV, there are no Ts-Tv differences among radical charge changes (Figure 3-5B). Transversions are more detrimental than transitions among radical polarity and polarity and size changes, although to a lesser degree as compared to overall in Figure 3-3. Thus, even among radical substitutions of three different amino acid categories, transitions tend to be less detrimental than transversions.

We then compared conservative transitions to conservative transversions (Figure 3-6). In both NP proteins, there are no significant Ts-Tv differences among conservative changes of all three amino acid classes (Figure 3-6A). In HA (H1N1), there are no significant Ts-Tv differences among conservative polarity or polarity and size changes. However, transversions are more detrimental than transitions among conservative charge changes. In HIV ENV, transversions are more detrimental than transitions among conservative changes of all three types (Figure 3-6B). Among conservative polarity and size changes, the odds ratio at the second threshold (>6) is

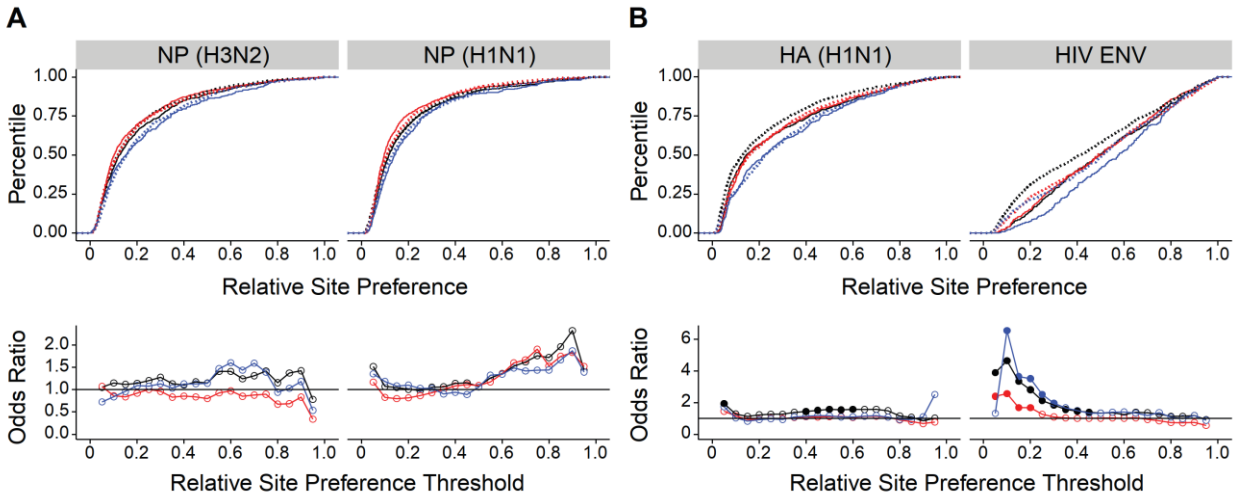


Figure 3-6 Distribution of Conservative Ts and Conservative Tv

Empirical cumulative distribution functions of conservative transitions (solid lines) and conservative transversions (dotted lines) in deep mutational scanning datasets of two NP proteins (A, top) and two antigenic surface proteins (B, top). Computed for conservative amino acid changes classified by charge (black), polarity (red), and polarity and size (blue). Odds ratio estimated by a Fisher test comparing the odds of a conservative transversion versus a conservative transition to be at or below each of 19 fitness thresholds, beginning at a fitness of 0.05 and increasing by 0.05, for the same datasets (A and B, bottom). Color scheme is the same as in the CDFs. Filled circles denote $p < .05$ for a two-sided test with Holm-Bonferroni correction. Lines shown to clarify trends only. Plotted data and raw p-values can be found at https://github.com/lauringlab/tstv_paper.

higher than the odds ratios when comparing all Ts and Tv mutations (Figure 3-3B, <4), indicating that controlling for the conservation of transitions can increase rather than eliminate Ts-Tv differences. However, the Ts-Tv differences are reduced among conservative polarity changes as compared to overall (compare the first thresholds—the odds ratio is ~ 2 in Figure 3-6B but ~ 5 in Figure 3-3B). Thus, conservative transversions are more detrimental than conservative transitions for these three categories in some of the datasets.

In sum, transversions are more detrimental than transitions either among radical or among conservative changes of all three amino acid classes in all four proteins (Table 3). In some cases, Ts-Tv fitness differences were increased when we constrained the analysis to just radical or conservative changes. For NP (H3N2), the constrained analysis revealed a Ts-Tv fitness difference among radical polarity changes that was not observed overall. In other cases, Ts-Tv fitness differences were eliminated or reduced, mostly when constraining the analysis to conservative changes. Thus, these three amino acid categories at best only partially explain the Ts-Tv fitness differences in these proteins, with conservative transitions and transversions generally being more similar in fitness than overall.

Discussion

We addressed a longstanding question in molecular evolution, whether the observed Ts:Tv substitution bias is due to a mutational bias or to selection disfavoring transversions. We found that missense transversions are more detrimental to fitness than transitions in two RNA viruses, influenza and HIV. Our study therefore provides direct support for the selective hypothesis. Furthermore, transversions are more detrimental even when controlling for their

greater likelihood of causing a radical amino acid change. These data demonstrate that commonly used classifications of amino acid changes may not adequately capture the varying selective constraints on different proteins.

The fitness differences between transitions and transversions can be measured in multiple ways and are not well described by a single summary statistic. In four analyzed DMS datasets, the distribution of fitness effects of transversions is shifted toward more deleterious effects. However, they differ at the fitness level at which the shift occurs. We suggest that one explanation for finding a null result is the use of AUC as a summary statistic, which can over-weight effects at the ends of the fitness distribution and obscure differences in other regions. In contrast, our use of a Fisher test and large DMS datasets allowed for explicit comparisons of Ts and Tv fitness effects along the distributions without sacrificing power.

Several observations support the idea that the small but significant Ts-Tv fitness differences we identify are biologically relevant and can plausibly explain the Ts:Tv substitution bias. First, despite variation in the Ts-Tv fitness differences, transitions are never more detrimental than transversions, and transversions are either similar to transitions or more detrimental (Table 3). This consistent trend suggests that we are identifying a biologically important generality in the effects of transitions and transversions and not simply subtle variations in effects in the highly powered DMS datasets. Second, the Ts-Tv fitness differences in many cases are similar to or even greater than those between radical and conservative amino acid changes. For example, at the low-fitness end of the HIV ENV distribution, the Ts-Tv fitness difference (Figure 3-3) is greater than that between radical and conservative changes of any type

at any fitness level for any gene studied (Figure 3-4). The radical/conservative distinction has widely accepted evolutionary consequences—conservative substitutions occur more often than radical ones in proteins under purifying selection (Epstein 1967; Clarke 1970; Miyata et al. 1979; Eyre-Walker et al. 2000; Zhang 2000; Miller and Kumar 2001; Duda et al. 2002; Popadin et al. 2007). If the fitness differences between radical and conservative changes have consequences for protein evolution, then the similar or greater fitness differences between transversions and transitions are likely to be consequential as well. Finally, the biological relevance of these effects is also supported by our own and other measurements of Ts and Tv mutation rates in several influenza strains (Bloom 2014; Pauly, Procario, et al. 2017). The Ts:Tv mutational bias is 2-3.6, significantly less than the average observed Ts:Tv substitution ratio of 5.24 in influenza (Duchêne et al. 2015). These measured mutational biases demonstrate that the selective and mutational hypotheses for the Ts:Tv substitution bias are not mutually exclusive. An important area for future work will be to determine the relative impact of the transitional mutational and selective biases on the overall Ts:Tv substitution bias, particularly in varying genomic contexts (e.g. coding vs. noncoding regions).

While transversions are more likely to be radical than transitions, this bias only partially accounts for the observed differences in fitness effects. This is perhaps not surprising, as the radical/conservative distinction may not capture the varying constraints on proteins of diverse structure and function. For example, the radical/conservative distinction did not always predict fitness in viral genes. We suggest that the Ts-Tv distinction might be able to better capture these differing functional constraints because transversions are more likely to be radical for a number of different amino acid categories, not just the three analyzed here (Stoltzfus and Norris 2016).

Dozens of amino acid categories and many other metrics, such those provided by Polyphen and SIFT, exist for predicting the fitness effects of amino acid substitutions (Kawashima et al. 2008; Kumar et al. 2009; Stoltzfus and Norris 2016). Any one of our simple categories cannot themselves explain the Ts-Tv fitness difference, but their combination, represented in the Ts-Tv distinction, can be quite generally predictive of fitness. Therefore, just as the radical/conservative substitution ratio has been used to detect relaxed selection or positive selection (Hughes et al. 1990; Eyre-Walker et al. 2000; Zhang et al. 2002; Pupko et al. 2003; Zhang and Webb 2004; Tennessen 2005; Shen et al. 2009; Wernegreen 2011), our data support the use of the Ts:Tv ratio as an independent, and perhaps more general, test of selection.

We focused on non-synonymous point mutations because the available evidence suggests that these have greater fitness impacts than synonymous or non-coding mutations (Cuevas et al. 2012). Significant fitness effects from synonymous substitutions are more often observed with large scale changes rather than individual mutations; for example, a complete change in the codon usage of a gene (Lauring et al. 2012). Other selective pressures on synonymous or non-coding sequences include regulation of replication and translation (Groeneveld et al. 1995; Klovins et al. 1998), targeting by host RNAses (Klovins et al. 1997), and G+C content and thermostability of RNA structures with various functions (Schultes et al. 1997; Smit et al. 2009; Watts et al. 2009). While it is possible that these selective pressures also contribute to the observed fitness disadvantage of transversions, the main factor is likely the amino acid change.

Our data suggest that the predictive value of the radical/conservative amino acid distinctions may vary due to differing functions of the structural and nonstructural proteins of

viruses. Genes encoding for the surface proteins often have a history of intense frequency-dependent selection and may exhibit tolerance to mutations that allow for immune escape while preserving their essential functions of binding and fusion (Stephens and Waelbroeck 1999; Plotkin and Dushoff 2003; Thyagarajan and Bloom 2014; Doud and Bloom 2016; Visher et al. 2016). We therefore expected radical amino acid changes, which may allow for immune escape, to exhibit a less pronounced fitness disadvantage in the surface proteins (HA and ENV) as compared to the internal NP proteins. This is true for charge changes, but not for polarity and polarity and size changes. This observation is also in agreement with two studies of codon usage bias in HA and HIV ENV (Stephens and Waelbroeck 1999; Plotkin and Dushoff 2003). These studies found that, as compared to non-antigenic regions or genes, the antigenic regions exhibit a bias toward codons that tend to mutate non-synonymously, but not toward codons that tend to mutate to radical polarity and size changes. Charge changes were not evaluated. Thus, these genes may be more tolerant of charge changes that allow for immune escape. If correct, one might expect a bias toward codons that preferentially mutate to radical charge changes and that charge changes, rather than polarity and/or size changes, more often lead to escape from host immune pressure.

The observed Ts-Tv fitness differences suggest an evolutionarily-informed approach to improving antiviral strategies. Mutagenic drugs have been used to cause extinction of a variety of viruses in cell culture, a strategy called lethal mutagenesis (Anderson et al. 2004; Bull et al. 2007). There has been little consideration regarding the choice of mutagenic drug, and most commonly employed mutagens cause transitions (Crotty et al. 2001; Ruiz-Jarabo et al. 2003; Graci and Cameron 2008; Dapp et al. 2009). We suggest that the most effective way to achieve

lethal mutagenesis may be by using drugs that increase the rate of the more deleterious transversion mutations. In fact, a previous report from our lab showed that the influenza RNA polymerase makes fewer transversions than transitions (Pauly and Lauring 2015). Additionally, 5-azacytidine, a mutagenic drug that causes transversions, is more effective at reducing viral infectivity than two drugs that cause transitions (Pauly and Lauring 2015). Given our results, we speculate that the same may be true for HIV.

Here we find that despite being broad mutational categories, transitions and transversions can capture functional constraints in very different proteins in two viruses. Although the underlying reason for the relative fitness advantage of transitions likely depends on the structure of the genetic code and the accessibility of different types of amino acids, we have shown that the reason is not as simple as the lower likelihood of a transition causing radical changes of certain broad categories. One possibility is that the codon usage in RNA viruses may have evolved in part to buffer a transitional mutation load (Sanjuán 2010; Lauring et al. 2012) due to their high mutation rates and underlying transitional mutation bias (Drake and Holland 1999; Pauly, Procario, et al. 2017). Identifying the combination of biochemical factors that lead to the fitness advantage of transitions, the relative effects of selection and mutational biases on the overall Ts:Tv substitution bias, and the degree to which these results extend beyond RNA viruses will be important areas of further research.

Materials and Methods

Data

All fitness and site preference data were obtained from supplementary material in the published articles or provided by the authors directly. Please see the original papers for details on measurements fitness and site preference (Rihn et al. 2013; Thyagarajan and Bloom 2014; Doud et al. 2015; Rihn et al. 2015; Doud and Bloom 2016; Haddox et al. 2016). To identify transition-only and transversion-only accessible amino acid substitutions for the mutational scanning data, we obtained the backbone nucleotide sequence of the genes in which the amino acid substitutions were made. These were provided in supplemental files in the published articles for HA (H1N1), NP (H3N2) and HIV ENV. The sequence for NP (H1N1) was obtained from Genbank (Accession number EF467822.1). All sequences can be found online at https://github.com/lauringlab/tstv_paper. For all our analyses, we excluded beneficial mutations, synonymous mutations, and amino acid substitutions accessible by both transitions and transversions or requiring more than one nucleotide mutation.

AUC Analysis

Our AUC analysis was performed exactly as in Stoltzfus and Norris 2016. An ROC curve plots the true positive rate against the false positive rate of a binary classifier system as the discrimination threshold varies. The AUC is the area under this curve. Consider a hypothetical example in which a fitness value between 0 and 1 serves as a discrimination threshold used to predict whether a mutation is a transition or a transversion. A mutation with a fitness value above the threshold level will be classified as a transition and below the threshold level as a transversion. Thus, the true positive rate is the proportion of transitions above the threshold and the false positive rate is the proportion of transversions above the threshold. If transitions and

transversions do not differ in their fitness effects, the true positive rate will be equal to the false positive rate at all threshold levels and an ROC curve would show a 1:1 line. The AUC in this null case is half of the total ROC plot area, or 0.5. If transitions generally have a higher fitness than transversions, the true positive rate will be higher than the false positive rate at most threshold levels. The corresponding ROC curve would have a steeper slope than a 1:1 line and have an AUC greater than 0.5. The greater the difference in fitness between transitions and transversions, the greater the difference between the true positive rates and the false positive rates, leading to a steeper ROC curve and a greater AUC. The AUC is mathematically equivalent to the chance that a randomly chosen positive instance of the classifier system is ranked higher than a randomly chosen negative instance (Hanley and McNeil 1982; Mason and Graham 2002). Thus, for our analysis, the AUC is the probability that a randomly chosen transition has a higher fitness value than a randomly chosen transversion. The AUC is calculated from the Mann-Whitney U test (Hanley and McNeil 1982; Mason and Graham 2002): $AUC = (pairs - statistic)/pairs$, where $pairs = number\ of\ transitions \times number\ of\ transversions$ and *statistic* is the Mann-Whitney U test statistic comparing the fitness values of transitions and transversions. Statistics were calculated using the wilcox.test() function in R. All p values are for a one-sided Mann-Whitney U test where the alternative hypothesis is that transitions are ranked higher than transversions.

Empirical CDF and Odds Ratios

Empirical CDF were computed using the ggplot2 stat_ecdf function in R. Odds ratios were estimated by Fisher's exact test using the fisher.test() function in R. When the odds ratio was calculated as infinite (e.g. when transversions fall below a relative fitness threshold but no transitions fall below the threshold), the estimated lower 95% confidence interval of the odds

ratio was plotted. All p values are for a two-sided test. Holm-Bonferroni correction was implemented when comparisons were performed at multiple fitness level thresholds for a given dataset.

Availability of Computer Code and Data

R version 3.3.2 was used for all data analysis and to create all figures. Scripts and data are available online at https://github.com/lauringlab/tstv_paper as are all plotted data along with unadjusted p values for all figures.

Acknowledgments

We thank George Zhang for helpful guidance and a critical reading of the manuscript. We thank Kayla Peck for helpful comments on the manuscript and review of R scripts. We thank Jesse Bloom for suggestions on appropriate use of the DMS data. We thank Rafael Sanjuan for providing the raw data for the VSV study and Suzannah Rihn and Paul Bieniasz for providing the raw data for the HIV IN and CA studies. This work was supported by a Clinician Scientist Development Award from the Doris Duke Charitable Foundation (CSDA 2013105) and R01 AI118886, both to ASL. DML was supported by University of Michigan Medical Scientist Training Program (T32GM007863).

References

Anderson JP, Daifuku R, Loeb LA. 2004. Viral error catastrophe by mutagenic nucleosides. Annu. Rev. Microbiol. 58:183–205. doi:10.1146/annurev.micro.58.030603.123649.

Bloom JD. 2014. An Experimentally Determined Evolutionary Model Dramatically Improves Phylogenetic Fit. Mol. Biol. Evol. 31:1956–1978. doi:10.1093/molbev/msu173.

Bull JJ, Sanjuán R, Wilke CO. 2007. Theory of Lethal Mutagenesis for Viruses. J. Virol. 81:2930–2939. doi:10.1128/JVI.01624-06.

Carrasco P., Daròs JA, Agudelo-Romero P, Elena SF. 2007. A real-time RT-PCR assay for quantifying the fitness of tobacco etch virus in competition experiments. J. Virol. Methods 139:181–188. doi:10.1016/j.jviromet.2006.09.020.

Carrasco Purificación, Iglesia F de la, Elena SF. 2007. Distribution of Fitness and Virulence Effects Caused by Single-Nucleotide Substitutions in Tobacco Etch Virus. J. Virol. 81:12979–12984. doi:10.1128/JVI.00524-07.

Clarke B. 1970. Selective Constraints on Amino-acid Substitutions during the Evolution of Proteins. Nature 228:159–160. doi:10.1038/228159a0.

Crotty S, Cameron CE, Andino R. 2001. RNA virus error catastrophe: Direct molecular test by using ribavirin. Proc. Natl. Acad. Sci. 98:6895–6900. doi:10.1073/pnas.111085598.

Cuevas JM, Domingo-Calap P, Sanjuán R. 2012. The Fitness Effects of Synonymous Mutations in DNA and RNA Viruses. Mol. Biol. Evol. 29:17–20. doi:10.1093/molbev/msr179.

Dagan T, Talmor Y, Graur D. 2002. Ratios of Radical to Conservative Amino Acid Replacement are Affected by Mutational and Compositional Factors and May Not Be Indicative of Positive Darwinian Selection. Mol. Biol. Evol. 19:1022–1025. doi:10.1093/oxfordjournals.molbev.a004161.

Dapp MJ, Clouser CL, Patterson S, Mansky LM. 2009. 5-Azacytidine Can Induce Lethal Mutagenesis in Human Immunodeficiency Virus Type 1. J. Virol. 83:11950–11958. doi:10.1128/JVI.01406-09.

Denver DR, Morris K, Lynch M, Thomas WK. 2004. High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome. Nature 430:679–682. doi:10.1038/nature02697.

Domingo-Calap P, Cuevas JM, Sanjuá N R. 2009. The Fitness Effects of Random Mutations in Single-Stranded DNA and RNA Bacteriophages. doi:10.1371/journal.pgen.1000742.

Doud MB, Ashenberg O, Bloom JD. 2015. Site-Specific Amino Acid Preferences Are Mostly Conserved in Two Closely Related Protein Homologs. Mol. Biol. Evol. 32:2944–2960. doi:10.1093/molbev/msv167.

Doud MB, Bloom JD. 2016. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. Viruses 8:155. doi:10.3390/v8060155.

Drake JW, Holland JJ. 1999. Mutation rates among RNA viruses. Proc. Natl. Acad. Sci. 96:13910–13913. doi:10.1073/pnas.96.24.13910.

Duchêne S, Ho SY, Holmes EC. 2015. Declining transition/transversion ratios through time reveal limitations to the accuracy of nucleotide substitution models. BMC Evol. Biol. 15:36. doi:10.1186/s12862-015-0312-6.

Duda TF, Vanhoye D, Nicolas P. 2002. Roles of diversifying selection and coordinated evolution in the evolution of amphibian antimicrobial peptides. Mol. Biol. Evol. 19:858–864.

Epstein CJ. 1967. Non-randomness of Ammo-acid Changes in the Evolution of Homologous Proteins. Nature 215:355–359. doi:10.1038/215355a0.

Eyre-Walker A, Keightley PD, Smith NGC, Gaffney D. 2000. Quantifying the Slightly Deleterious Mutation Model of Molecular Evolution. Mol. Biol. Evol. 19:2142–2149. doi:10.1093/oxfordjournals.molbev.a004039.

Fitch WM. 1967. Evidence suggesting a non-random character to nucleotide replacements in naturally occurring mutations. J. Mol. Biol. 26:499–507. doi:10.1016/0022-2836(67)90317-8.

Gojobori T, Li W-H, Graur D. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. J. Mol. Evol. 18:360–369. doi:10.1007/BF01733904.

Graci JD, Cameron CE. 2008. Therapeutically targeting RNA viruses via lethal mutagenesis. Future Virol. 3:553–566. doi:10.2217/17460794.3.6.553.

Grantham R. 1974. Amino Acid Difference Formula to Help Explain Protein Evolution. Science 185:862–864.

Groeneveld H, Thimon K, van Duin J. 1995. Translational control of maturation-protein synthesis in phage MS2: a role for the kinetics of RNA folding? RNA 1:79–88.

Haddox HK, Dingens AS, Bloom JD. 2016. Experimental Estimation of the Effects of All Amino-Acid Mutations to HIV's Envelope Protein on Viral Replication in Cell Culture. PLOS Pathog. 12:e1006114. doi:10.1371/journal.ppat.1006114.

Hanley JA, McNeil BJ. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36. doi:10.1148/radiology.143.1.7063747.

Hughes AL, Ota T, Nei M. 1990. Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol. Biol. Evol. 7:515–524. doi:10.1093/oxfordjournals.molbev.a040626.

Jacquier H, Birgy A, Nagard HL, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit E, Poulain J, Barnaud G, et al. 2013. Capturing the mutational landscape of the beta-lactamase TEM-1. Proc. Natl. Acad. Sci. 110:13067–13072. doi:10.1073/pnas.1215206110.

Jiang C, Zhao Z. 2006. Mutational spectrum in the recent human genome inferred by single nucleotide polymorphisms. Genomics 88:527–534. doi:10.1016/j.ygeno.2006.06.003.

Kawashima S, Pokarowski P, Pokarowska M, Kolinski A, Katayama T, Kanehisa M. 2008. AAindex: amino acid index database, progress report 2008. Nucleic Acids Res. 36:D202–D205. doi:10.1093/nar/gkm998.

Klovins J, Berzins V, Duin JV. 1998. A long-range interaction in Qβ RNA that bridges the thousand nucleotides between the M-site and the 3′ end is required for replication. RNA 4:948–957.

Klovins J, van Duin J, Olsthoorn RCL. 1997. Rescue of the RNA phage genome from RNase III cleavage. Nucleic Acids Res. 25:4201–4208. doi:10.1093/nar/25.21.4201.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4:1073–1081. doi:10.1038/nprot.2009.86.

Kumar S. 1996. Patterns of Nucleotide Substitution in Mitochondrial Protein Coding Genes of Vertebrates. Genetics 143:537–548.

Lauring AS, Acevedo A, Cooper SB, Andino R. 2012. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. Cell Host Microbe 12:623–632. doi:10.1016/j.chom.2012.10.008.

Lynch M. 2010. Rate, molecular spectrum, and consequences of human mutation. Proc. Natl. Acad. Sci. 107:961–968. doi:10.1073/pnas.0912629107.

Mason SJ, Graham NE. 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. Q. J. R. Meteorol. Soc. 128:2145–2166. doi:10.1256/003590002320603584.

Miller MP, Kumar S. 2001. Understanding human disease mutations through the use of interspecific genetic variation. Hum. Mol. Genet. 10:2319–2328. doi:10.1093/hmg/10.21.2319.

Miyata T, Miyazawa S, Yasunaga T. 1979. Two types of amino acid substitutions in protein evolution. J. Mol. Evol. 12:219–236. doi:10.1007/BF01732340.

Pauly MD, Lauring AS. 2015. Effective Lethal Mutagenesis of Influenza Virus by Three Nucleoside Analogs. J. Virol. 89:3584–3597. doi:10.1128/JVI.03483-14.

Pauly MD, Procario MC, Lauring AS. 2017. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. eLife 6:e26437. doi:10.7554/eLife.26437.

Peris JB, Davis P, Cuevas JM, Nebot MR, Sanjuán R. 2010. Distribution of Fitness Effects Caused by Single-Nucleotide Substitutions in Bacteriophage f1. Genetics 185:603–609. doi:10.1534/genetics.110.115162.

Petrov DA, Hartl DL. 1999. Patterns of nucleotide substitution in Drosophila and mammalian genomes. Proc. Natl. Acad. Sci. 96:1475–1479. doi:10.1073/pnas.96.4.1475.

Plotkin JB, Dushoff J. 2003. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. Proc. Natl. Acad. Sci. 100:7152–7157. doi:10.1073/pnas.1132114100.

Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. Proc. Natl. Acad. Sci. 104:13390–13395. doi:10.1073/pnas.0701256104.

Pupko T, Sharan R, Hasegawa M, Shamir R, Graur D. 2003. Detecting excess radical replacements in phylogenetic trees. Gene 319:127–135. doi:10.1016/S0378-1119(03)00802-3.

Rihn SJ, Hughes J, Wilson SJ, Bieniasz PD. 2015. Uneven Genetic Robustness of HIV-1 Integrase. J. Virol. 89:552–567. doi:10.1128/JVI.02451-14.

Rihn SJ, Wilson SJ, Loman NJ, Alim M, Bakker SE, Bhella D, Gifford RJ, Rixon FJ, Bieniasz PD. 2013. Extreme Genetic Fragility of the HIV-1 Capsid. PLOS Pathog 9:e1003461. doi:10.1371/journal.ppat.1003461.

Rosenberg MS, Subramanian S, Kumar S. 2003. Patterns of Transitional Mutation Biases Within and Among Mammalian Genomes. Mol. Biol. Evol. 20:988–993. doi:10.1093/molbev/msg113.

Ruiz-Jarabo CM, Ly C, Domingo E, de la Torre JC. 2003. Lethal mutagenesis of the prototypic arenavirus lymphocytic choriomeningitis virus (LCMV). Virology 308:37–47.

Sanjuán R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 365:1975–82. doi:10.1098/rstb.2010.0063.

Sanjuán R, Moya A, Elena SF. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc. Natl. Acad. Sci. U. S. A. 101:8396–8401. doi:10.1073/pnas.0400146101.

Schultes E, Hraber PT, LaBean TH. 1997. Global similarities in nucleotide base composition among disparate functional classes of single-stranded RNA imply adaptive evolutionary convergence. RNA 3:792–806.

Shen Y-Y, Shi P, Sun Y-B, Zhang Y-P. 2009. Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability. Genome Res. 19:1760–1765. doi:10.1101/gr.093138.109.

Smit S, Knight R, Heringa J. 2009. RNA structure prediction from evolutionary patterns of nucleotide composition. Nucleic Acids Res. 37:1378–1386. doi:10.1093/nar/gkn987.

Stephens CR, Waelbroeck H. 1999. Codon Bias and Mutability in HIV Sequences. J. Mol. Evol. 48:390–397. doi:10.1007/PL00006483.

Stoltzfus A, Norris RW. 2016. On the Causes of Evolutionary Transition:Transversion Bias. Mol. Biol. Evol. 33:595–602. doi:10.1093/molbev/msv274.

Tennessen JA. 2005. Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. J. Evol. Biol. 18:1387–1394. doi:10.1111/j.1420-9101.2005.00925.x.

Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. eLife 3:e03300. doi:10.7554/eLife.03300.

Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS. 2016. The Mutational Robustness of Influenza A Virus. PLOS Pathog 12:e1005856. doi:10.1371/journal.ppat.1005856.

Vogel F, Kopun M. 1977. Higher frequencies of transitions among point mutations. J. Mol. Evol. 9:159–180. doi:10.1007/BF01732746.

Wakeley J. 1996. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. Trends Ecol. Evol. 11:158–162. doi:10.1016/0169-5347(96)10009-4.

Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, Burch CL, Weeks KM. 2009. Architecture and Secondary Structure of an Entire HIV-1 RNA Genome. Nature 460:711–716. doi:10.1038/nature08237.

Wernegreen JJ. 2011. Reduced Selective Constraint in Endosymbionts: Elevation in Radical Amino Acid Replacements Occurs Genome-Wide. PLOS ONE 6:e28905. doi:10.1371/journal.pone.0028905.

Yampolsky LY, Stoltzfus A. 2005. The Exchangeability of Amino Acids in Proteins. Genetics 170:1459–1472. doi:10.1534/genetics.104.039107.

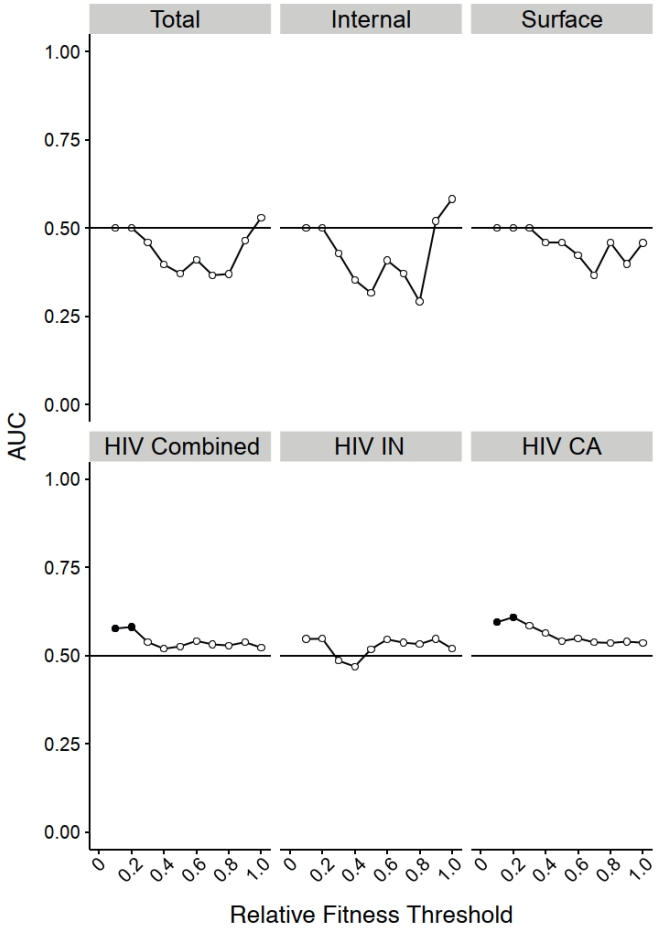
Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J. Mol. Evol. 50:56–68.

Zhang J, Webb DM. 2004. Rapid evolution of primate antiviral enzyme APOBEC3G. Hum. Mol. Genet. 13:1785–1791. doi:10.1093/hmg/ddh183.

Zhang J, Zhang Y, Rosenberg HF. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat. Genet. 30:411–415. doi:10.1038/ng852.

Zhang Z, Gerstein M. 2003. Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. Nucleic Acids Res. 31:5338–5348. doi:10.1093/nar/gkg745.

Supplemental Figures and Tables



Supplemental Figure 3-1 Differences in Ts and Tv Fitness Effects as measured by AUC using Decreasing Thresholds

Ts-Tv fitness differences as measured by AUC among mutations at or below successively lower fitness thresholds in our influenza datasets (top) and the HIV datasets (bottom). Thresholds decrease by a fitness of 0.1, starting at 1 and ending at 0.1. Filled circles denote $p < 0.05$ for a one-sided Mann-Whitney U test where the alternative is transitions are more fit. Lines shown to clarify trends only. Plotted data and raw p-values can be found at https://github.com/lauringlab/tstv_paper.

Supplemental Table 3-1 Sample size of datasets

Dataset	Ts[a]	Tv[b]	Total
HIV Combined	143	138	281
HIV IN	73	73	146
HIV CA	70	65	135
Influenza Total	23	72	95
Influenza Internal	14	39	53
TEV	20	27	47
F1	16	28	44
Influenza Surface	9	33	42
VSV	7	24	31
Qβ	2	27	29
ϕX174	3	18	21
HA (H1N1)	934	2064	2998
NP (H1N1)	878	1870	2748
NP (H3N2)	849	1833	2682
HIV ENV	684	1794	2478

[a] Number of transition mutations
[b] Number of transversion mutations

Supplemental Table 3-2 Amino Acid classifications

CHARGE	
negative	D,E
neutral	A,P,G,S,T,F,W,Y,I,L,M,V,N,Q,C
positive	R,H,K
POLARITY	
nonpolar	A,P,F,W,I,L,M,V
polar	G,S,T,Y,R,H,K,D,E,N,Q,C
POLARITY & SIZE	
neutral & small	A,P,G,S,T
nonpolar & large	F,W,Y
nonpolar & small	I,L,M,V
polar & large	R,H,K
polar & small	D,E,N,Q
special	C

Chapter 4 The Best Substitutions are also the Worst and Hardest to Make

Abstract

Understanding the evolutionary constraints on proteins and the properties of mutations and genes responsible for adaptive changes are major goals in molecular evolution. Many such studies have assumed that amino acid classifications based on simple biochemical properties can capture differences in fitness effects between substitutions that radically versus conservatively change these biochemical properties. For example, like the non-synonymous to synonymous substitution ratio, the ratio of radical to conservative substitutions has been used to detect positive selection. Similarly, studies examining whether adaptation proceeds by large or small steps have assumed that these radical/conservative distinctions reflect mutational effect size. However, there is some doubt as to the extent to which radical/conservative distinctions are predictive of fitness effects. Phylogenetic evidence that radical substitutions are more deleterious may instead reflect mutational biases or the structure of the genetic code. Here, using deep mutational scanning studies of several RNA viruses, we show that common radical/conservative classifications are in fact predictive of fitness differences. Interestingly, radical substitutions are both more detrimental and more beneficial, when beneficial. These same patterns hold for multi-nucleotide codon mutations, which are more likely to cause a radical substitutions, as compared to single-nucleotide mutations. Thus, our results validate the use of radical/conservative classifications in the study of protein evolution and support the hypothesis that the structure of the genetic code evolved to minimize errors.

Introduction

Decades of work in molecular evolution have assumed that radical versus conservative changes in the biochemical properties of amino acids correspond to their fitness effects. In the 1960's Woese and Crick proposed that the genetic code evolved to minimize the detrimental effect of single nucleotide mutations because more similar amino acids (in terms of polarity or other properties) have more similar codons (Woese 1965; Woese et al. 1966; Crick 1968). Studies on how protein evolution is shaped by purifying selection make the same assumption that substitutions between amino acids that are more similar in biochemical properties have less detrimental fitness effects (Epstein 1967; Clarke 1970; Miyata et al. 1979; Eyre-Walker et al. 2000; Zhang 2000; Miller and Kumar 2001; Duda et al. 2002; Popadin et al. 2007; Chen, Lan, et al. 2019).

Radical/conservative amino acid classifications have also been important in the study of adaptive evolution. For example, does adaptive evolution proceed through small or large steps? A number of conflicting conclusions on this question at the molecular level have arisen (Hughes et al. 2000; Rand et al. 2000; Bergman and Eyre-Walker 2019; Chen, He, et al. 2019; Chen, Lan, et al. 2019). Such studies have assumed that simple radical/conservative classifications capture true differences in the size of phenotypic and fitness effects. Like the ratio of non-synonymous to synonymous substitutions (ns/s), the ratio of radical to conservative substitutions has been used to detect positive selection (Hughes et al. 1990; Eyre-Walker et al. 2000; Hughes et al. 2000; Zhang et al. 2002; Pupko et al. 2003; Zhang and Webb 2004; Tennessen 2005; Gojobori et al. 2007; Shen et al. 2009; Wernegreen 2011). Possibly the first example is Nei and colleagues' detection of positive selection at MHC antigen-binding loci, helping explain the maintenance of

polymorphisms due to host-pathogen coevolution (Hughes et al. 1990). Additionally, the ns/s ratio is unreliable when synonymous substitutions are saturated, making the radical/conservative substitution ratio particularly useful for investigating selection in highly diverged sequences (Hanada et al. 2007).

Despite their widespread use, there is some doubt about the extent to which radical/conservative distinctions are predictive of fitness effects. Radical substitutions occur less often during protein evolution than radical ones, supporting the idea that they are more detrimental (Epstein 1967; Clarke 1970; Grantham 1974; Miyata et al. 1979; Zhang 2000). However, this bias may simply reflect the structure of the genetic code or mutational biases, rather than selection (Dagan et al. 2002; Yampolsky and Stoltzfus 2005). There is widespread evidence of a mutational bias towards transitions (Zhang and Gerstein 2003; Jiang and Zhao 2006; Baer et al. 2007; Denver et al. 2009; Pauly et al. 2017), and transitions are more likely to cause conservative substitutions (Vogel and Kopun 1977; Miyata et al. 1979; Zhang 2000). The assumption that radical substitutions are more highly beneficial also requires validation – evidence of positive selection for radical substitutions can be affected by mutational or codon usage biases (Dagan et al. 2002).

With the advent of high-throughput mutation screens, it now possible to directly test the assumptions underlying the use of radical/conservative distinctions. Recently, we showed that radical substitutions are more detrimental than conservative ones in several proteins of two RNA viruses. Here, using three of the most commonly used radical/conservative classifications, we extend this analysis to additional proteins, and viral strains and species, and we test the additional assumptions about the beneficial effects of radical substitutions. We find that radical

substitutions are more likely to be detrimental and more highly detrimental, yet are also more highly beneficial. Furthermore, we provide direct evidence that the structure of the genetic code minimizes detrimental effects, albeit with a tradeoff with the size of beneficial effects. Substitutions requiring two or more nucleotide mutations are both more detrimental and more highly beneficial.

Results

Mutational fitness effects in RNA viruses

To directly test whether radical/conservative amino acid distinctions are predictive of fitness differences, a large dataset of mutational fitness effects with no biases towards the type of mutations included is crucial. Recent deep mutational scanning (DMS) studies of RNA viruses from the Bloom lab meet this requirement. These studies introduce every single amino acid substitution in a given gene followed by deep sequencing to measure the change in frequency of each substitution after selection. We curated ten DMS datasets of various strains of three RNA viruses, Influenza A, Zika, and HIV-1 (Doud et al. 2015; Doud and Bloom 2016; Haddox et al. 2016; Haddox et al. 2018; Lee et al. 2018; Soh et al. 2019; Sourisseau et al. 2019: 19). Fitness was usually reported as amino acid preference which we transformed to a proxy for relative fitness, relative preference, by dividing by the wild-type amino acid preference (Bloom 2015; Lyons and Lauring 2017). Relative preference correlates well with relative fitness measured in lower-throughput experiments (Visher et al. 2016; Lyons and Lauring 2017). We then transformed relative preference to fitness effect by taking the log of the relative preference values as in (Soh et al. 2019). Basic information about the ten datasets are in Table 4-1 with more details in Supplemental Tables 4-1 thru 4-6.

Table 4-1 Viral DMS Datasets

Species	Strain	Gene	Total Mutations	Avg. Deleterious Effect	Avg. Beneficial Effect
HIV-1	BF520	ENV	12578	-1.546	0.304
	BG505	ENV	12711	-2.01	0.408
	LAI	ENV	13540	-1.828	0.904
Influenza A virus	A/PR8/1934 (H1N1)	NP	9462	-4.065	0.726
	A/Aichi/2/1968 (H3N2)	NP	9462	-3.278	0.732
	A/WSN/1933 (H1N1)	HA	10716	-3.191	0.55
	A/Perth/16/2009 (H3N2)	HA	10755	-2.474	0.525
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	14421	-2.705	0.467
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	14421	-2.432	0.47
Zika virus	MR766	ENV	9576	-4.31	0.829

Radical amino acid substitutions are more frequently and more highly detrimental than conservative ones

We classified amino acids according to their charge, polarity, or polarity and size (polarity-size) as in (Miyata et al. 1979; Zhang 2000; Lyons and Lauring 2017). Radical amino acid substitutions are substitutions between amino acids with different classes (e.g. from a negative to a positively charged amino acid) while conservative substitutions are those between amino acids in the same category. We hypothesized that radical substitutions are more detrimental than conservative ones, in terms of the likelihood they are detrimental and their effect size.

For each of the ten DMS datasets, we first calculated the percent of radical or conservative substitutions that are detrimental and plotted the difference in this percentage (radical minus conservative) (Figure 4-1). Radical substitutions were more likely to be detrimental than conservative ones in 27 of 30 comparisons (ten datasets across the three classifications) (Figure 4-1). Considering the average across datasets, the percentage that are detrimental for radical substitutions is significantly greater than that for conservative ones based on polarity and polarity-size ($p < .05$, t-test) and trends greater for charge.

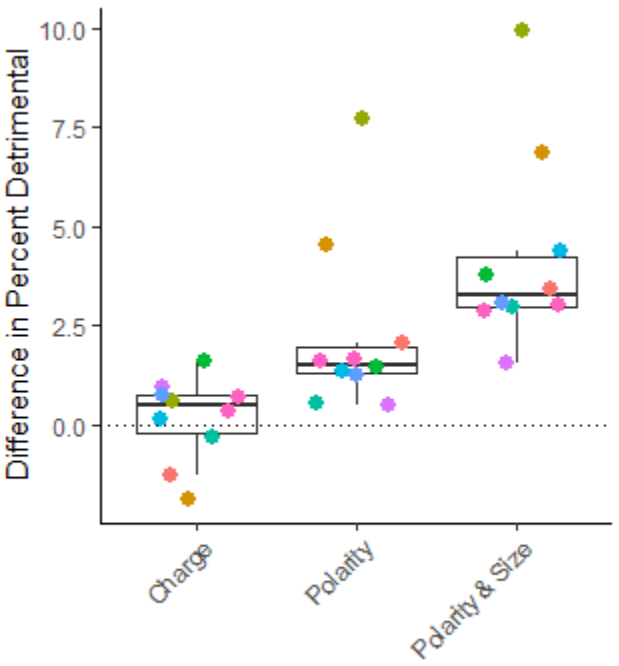


Figure 4-1. Difference in percent detrimental between radical and conservative substitutions.

Difference is calculated as the percentage of radical charge, polarity, or polarity and size changes that were detrimental minus that for conservative changes. Colors correspond to different datasets as in Figure S-1

We determined whether radical substitutions are more highly detrimental than conservative substitutions by comparing the detrimental mutational effects for each type of substitution in each dataset. Figure 4-2A shows the difference in average detrimental effect (Table 4-S3) between radical and conservative substitutions for all datasets for the three amino acid classifications. Radical substitutions were more detrimental for 28 out of 30 comparisons and significantly so for each of the ten datasets for polarity and polarity-size, and for six of the ten datasets for charge ($p < .05$, t-test). Additionally, radical substitutions were significantly less detrimental than conservative ones for only one dataset and only for charge.

To see whether the difference in detrimental effect sizes between radical and conservative substitutions are substantial biologically, we calculated this difference as a percentage of the overall average detrimental effect size. This was done for the 28 comparisons in which radical substitutions were more detrimental. The differences were quite large, representing on average from 10-25% of the overall average detrimental effect (Figure 4-2B). In other words, the *difference* in effect size between radical and conservative substitutions is 10-25% of the *entire* effect size of an average detrimental mutation.

Notably, the polarity-size classification was the most predictive of both the likelihood a radical substitution would be detrimental and its detrimental effect size, whereas charge was the least predictive.

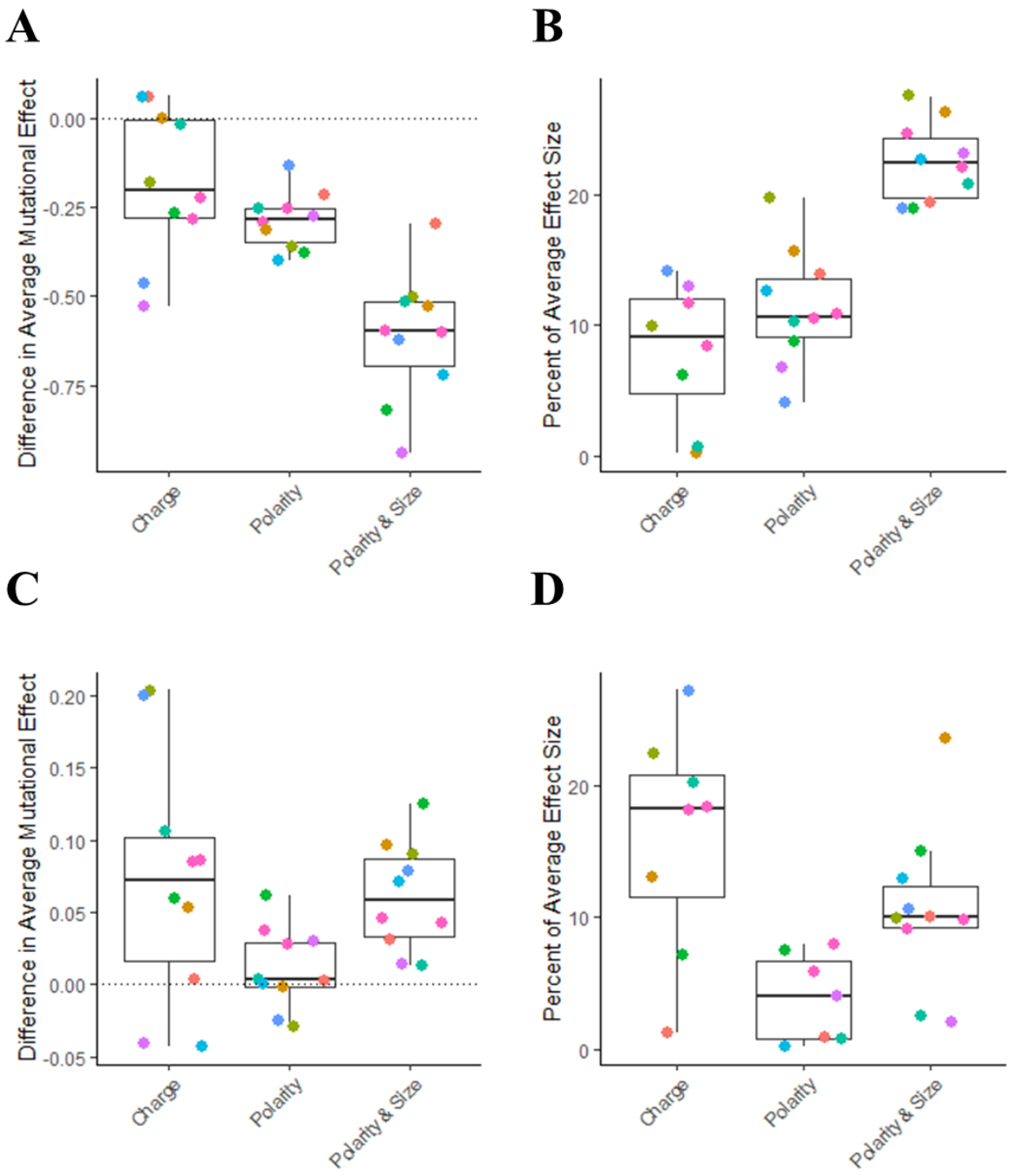


Figure 4-2 Differences in mutational effects between radical and conservative substitutions.

Differences in average mutational effects were calculated as the average effect of radical substitutions minus that for conservative substitutions for each dataset in each amino acid classification, conditional on being detrimental (A) or beneficial (C). The difference in effects between radical and conservative amino acids as a percentage of the overall average effect size for detrimental (B) or beneficial (B) substitutions.

Radical amino acid substitutions are more highly beneficial than conservative ones

To test the common assumption that radical substitutions are more beneficial, conditional on being beneficial, we compared the mutational effects of radical versus conservative beneficial mutations in each dataset. Figure 4-2C shows the difference in average beneficial effect (Table 4-S5) between radical and conservative substitutions for all datasets for the three amino acid classifications. Radical substitutions were more beneficial than conservative ones in 25 out 30 comparisons and significantly more beneficial for six and three out of ten datasets for charge and polarity-size, respectively ($p < .05$, t-test). The differences in beneficial effect sizes were quite large, representing on average from 5-20% of the effect size of an average beneficial mutation (Figure 4-2D). Finally, conservative substitutions were never significantly more beneficial than radical ones.

Non-synonymous multi-nucleotide mutations recapitulate findings for radical amino acid substitutions

Due to the structure of the genetic code, single nucleotide mutations are more likely to cause a conservative amino acid substitutions than a radical one. Given our results on radical/conservative mutational effects, we hypothesized that amino acid substitutions requiring two or three nucleotide mutations are more likely to be detrimental and more highly detrimental, but more beneficial when beneficial.

We classified amino acid substitutions as requiring only one, two, or three nucleotide mutations based on the standard genetic code. We then performed the same analyses as for radical/conservative substitutions, but comparing multi-nucleotide (MTN) to single nucleotide (SN) codon changes.

Both two- and three-MTN changes were more likely to be detrimental than SN changes for all but one dataset in the two-MTN class (Figure 4-3). Considering the average across datasets, the percentage that was detrimental was significantly greater for all classes of MTN changes than for SN changes ($p < .05$, t-test). MTN changes were also significantly more detrimental than SN changes for all datasets and all MTN classes ($p < .05$, t-test) (Figure 4-4A). The differences in detrimental effect sizes represented on average from 10-15% of the effect size of an average detrimental mutation (Figure 4-4B, Table 4-S4).

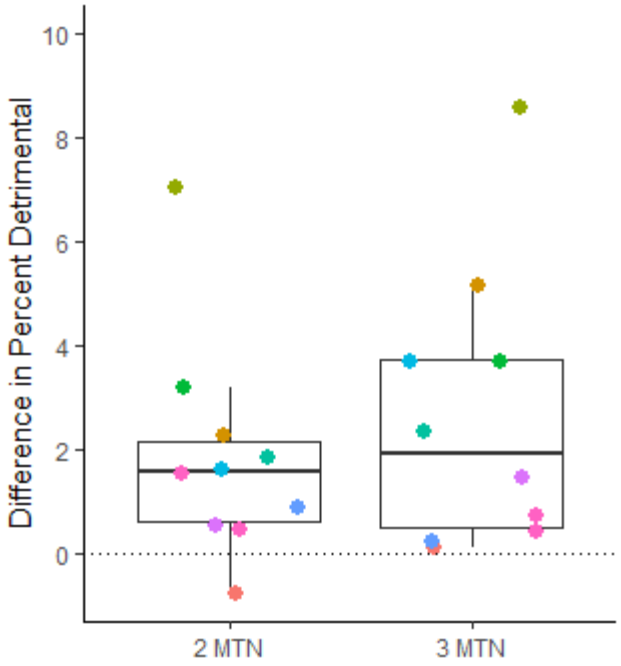


Figure 4-3 Difference in percent detrimental between MTN and SN substitutions.

Difference is calculated as the percentage of substitutions that were detrimental among those requiring at minimum 2 or 3 MTN mutations minus that for substitutions requiring only one nucleotide mutation.

MTN changes were also more beneficial than SN changes, conditional on being beneficial, for 15 of 20 comparisons (10 datasets across two or three MTN changes) (Figure 4-4C, Table 4-S6). This was significant for only one dataset for two- and three-MTN changes, but SN changes were never significantly more beneficial than MTN changes ($p < .05$, t-test). The differences in

beneficial effect sizes (when MTN was more beneficial than SN) represented on average from 5-10% of the effect size of an average beneficial mutation (Figure 4-4D).

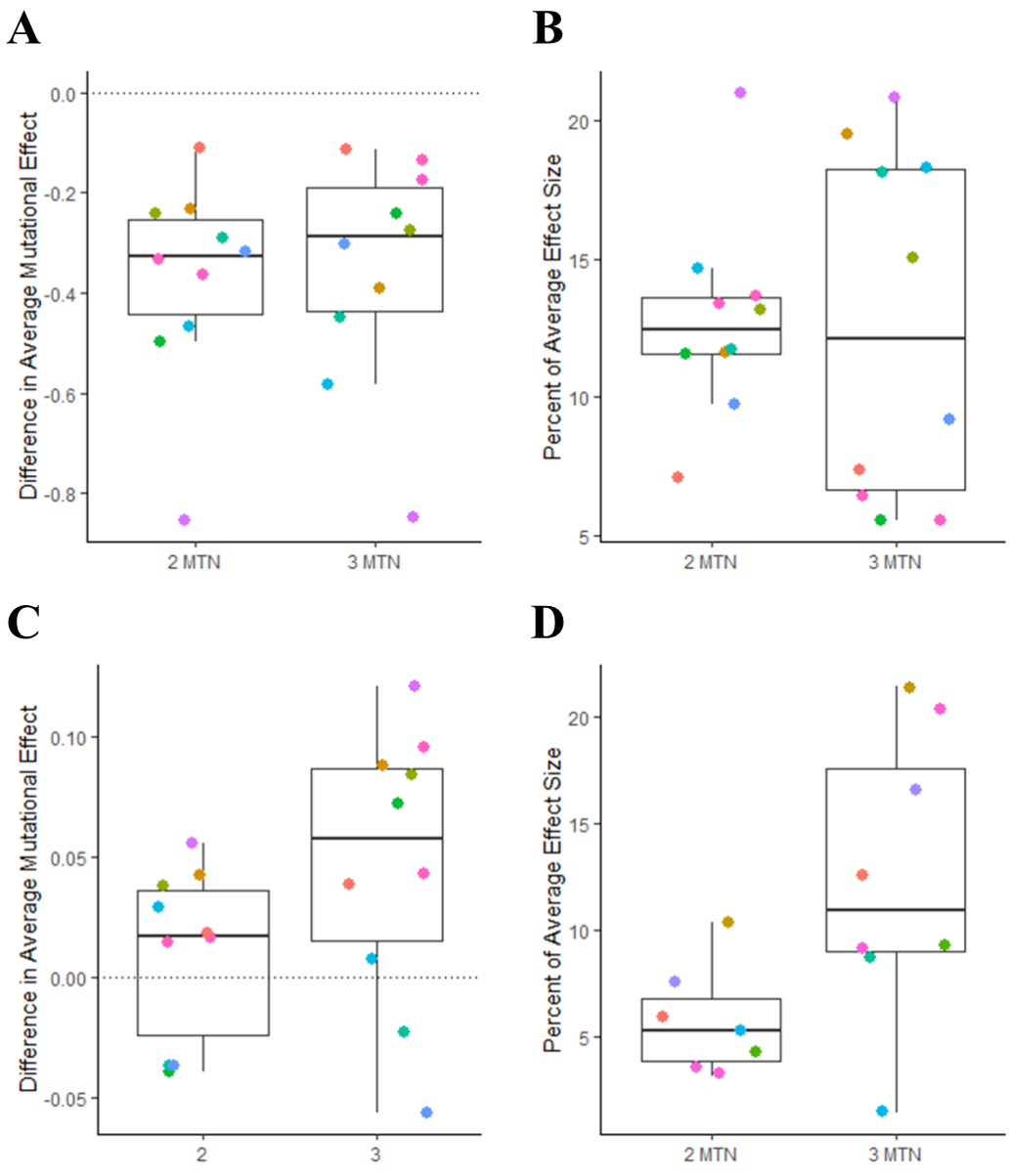


Figure 4-4 Differences in mutational effects between MTN and SN substitutions.

Differences in average mutational effects were calculated as the average effect of substitutions requiring at minimum 2 or 3 MTN mutations minus that for substitutions requiring only one nucleotide mutation, conditional on being detrimental (A) or beneficial (C). The difference in effects between MTN and SN substitutions as a percentage of the overall average effect size for detrimental (B) or beneficial (B) substitutions.

Discussion

Here, we used experimental data to verify the long assumed fitness differences between radical and conservative substitutions for three commonly used classifications across a variety of genes in three viral species. Radical substitutions are more likely to be detrimental and more detrimental, but also more beneficial, than conservative substitutions. Multi-nucleotide codon changes, which are more likely to be radical, show the same pattern as compared to single-nucleotide codon changes.

It is difficult to know whether the fitness differences we detect are large enough to impact substitutions rates, but one line of reasoning suggests so. The effective population size is estimated at 500 for influenza virus in the human population and 1000 for the intra-host HIV-1 population (Rouzine and Coffin 1999; Kouyos et al. 2006; Bedford et al. 2011). Thus, the threshold fitness effect for a mutation that is more governed by selection than drift (1/Ne) is approximately .002 (1/500), using the lower Ne estimate. Given an average mutational effect size of .2 for viable mutations across the influenza genome (Visher et. al., 2016), a 1% effect size difference would make the difference between a neutrally evolving conservative mutation and a negatively (or positively) selected radical mutation. On average, the differences we observe are 5-25 times greater than this difference. Of course, this reasoning considers only the situation when the conservative mutation is nearly neutral, but not the situation when they are both detrimental or beneficial. Simulations using the determined mutational biases in these viruses could be used to further predict the consequences of these fitness differences and compare to phylogenetic data.

These viral DMS datasets are ideal for studying radical versus conservative and MTN vs SN changes because they measure fitness effects for all 19 amino acid substitutions at each position. Expanding this analysis to genes from other taxa would be helpful; however, many datasets do not contain all possible amino acid substitutions. For example, a recent GFP dataset only contains about 6 amino acid substitutions per position and no substitutions that require three nucleotide changes (data not shown) (Sarkisyan et al. 2016). Whether the included substitutions are random or not is unclear and bias against some substitutions due to low fitness would make interpretation challenging. Thus, more empirical work is needed in other taxa.

Our results support the use of radical/conservative distinctions in positive selection (Hughes et al. 1990; Eyre-Walker et al. 2000; Hughes et al. 2000; Zhang et al. 2002; Pupko et al. 2003; Zhang and Webb 2004; Tennessen 2005; Gojobori et al. 2007; Shen et al. 2009; Wernegreen 2011). Future work could analyze more radical/conservative classifications to identify the best ones for detecting positive selection. The results of one of the most widely used tests for positive selection, the branch-site test (Zhang et al. 2005), have recently come under scrutiny because it assumes that MTN codon changes are due to successive single substitutions. Instead, MTN codon changes may be due to a bias in polymerases to cause simultaneous multiple mutations in nearby sites. This mutational bias has been observed in taxa from yeast to humans (Schrider et al. 2011). One study has found that the positive signals from this test can be almost completely explained by this mutational bias (Venkat et al. 2018). On the other hand, a recent phylogenetic analysis suggest that multiple substitutions within a codon are enriched compared to multiple substitutions straddling two adjacent codons (Zhengting Zou, personal correspondence). This suggests positive selection for MTN codon changes even in the setting of simultaneous multiple

nucleotide mutations. Our data that MTN codon changes are more beneficial than SN changes further support this hypothesis.

The long-standing debate between saltationism versus gradualism has a corollary at the molecular level in terms of whether adaptive evolution proceeds through small, i.e. conservative, or large, radical, steps (Bell 2009). Fisher's geometric theory (FGM) supposes that most beneficial mutations must be of small effect in order to minimize pleiotropic detrimental effects (Tenaillon 2014). Under FGM, more radical changes have pleiotropic tradeoffs which diminish both their chance of being beneficial and their beneficial effect size. We find that radical substitutions are indeed less likely to be beneficial, but contrary to FGM, they are also the most beneficial compared to conservative substitutions. Thus, the largest improvements in fitness might require large phenotypic changes, such as those from radical substitutions. Furthermore, adaptive evolution may be biased towards higher effect mutations because they have a higher probability of fixation (Kimura 1983). Conflicting conclusions have been made as to how these factors play out during adaptive evolution (Hughes et al. 2000; Rand et al. 2000; Bergman and Eyre-Walker 2019; Chen, He, et al. 2019; Chen, Lan, et al. 2019). A possible resolution to the debate is our finding that while radical substitutions are more beneficial, they are less frequently beneficial and more detrimental. This pattern may lead to different conclusions depending on how negative selection is controlled for and whether the question is about the relative frequency or importance of large versus small steps (Chen, He, et al. 2019).

In new or more stressful environments where wildtype fitness is lower, the supply and effect size of beneficial mutations are often higher (Bell 2009). It is hypothesized that radical phenotypic

changes may be more necessary for fitness improvement, as the fitness peak may lie in very different phenotypic region. Thus, adaptive evolution may proceed by larger steps in this circumstance. Indeed, radical amino acid substitutions are enriched in cases of environmental stress (Luo et al. 2017; Xu et al. 2017). DMS studies of viral strains under the stress of drugs, antibodies, or new cell types could be used to experimentally test the prediction that radical substitutions are less detrimental and even more beneficial under such environmental stresses. Moreover, such analyses could help identify the types of mutations important in the evolution of drug or immune resistance and host-switching, with implications for public health.

Interestingly, comparison between experimental data and previous phylogenetic results on radical/conservative differences may reveal biological differences in protein constraints or selection pressures across taxa and genes. For example, in mammals, radical versus conservative charge changes showed the largest differences in substitutions rates among the three types of radical changes (Zhang 2000). However, in the viral datasets, radical versus conservative changes in charge show the least differences in detrimental fitness differences. Insensitivity to radical charge changes seems to be most pronounced among the antigenic surface proteins HA and ENV compared to the internal NP protein. These proteins have a long history of immune selection and have evolved robustness to changes that allow for immune escape (Stephens and Waelbroeck 1999; Plotkin and Dushoff 2003; Thyagarajan and Bloom 2014; Doud and Bloom 2016; Visher et al. 2016). Thus, differences from mammalian genes may reflect the antigenic role of these viral genes. It is possible such differences are also reflected in phylogenetic patterns, and monitoring viral populations for charge substitutions may improve predictions of vaccine escape.

Selection for error minimization is a major hypothesis for the evolution of the genetic code, but has relied on the assumption that radical substitutions are more detrimental (Woese 1965; Woese et al. 1966; Crick 1968; Haig and Hurst 1991; Freeland and Hurst 1998; Geyer and Mamlouk 2018; Tripathi and Deem 2018). Our results provide direct evidence that the genetic code minimizes the detrimental effects of single-nucleotide changes. Interestingly, this comes with a greater likelihood of mutations being beneficial, but a tradeoff in terms of the size of beneficial effects. How might these factors enhance or impede adaptive evolution? Has the genetic code traded evolvability for robustness? Whatever the origin of the genetic code, its structure and the properties of amino acids lead to the pattern that the best substitutions are also the worst and hardest to make.

Methods

Data

All amino acid preference data were obtained from supplementary material in the published articles. Please see the original papers for details on measurements of amino acid preference. For all of our analyses, we excluded synonymous mutations.

Statistical Analyses

All *P* values are for a two-sided t-test using the t.test() function in R. All statistical tests compared radical/MTN or SN/conservative substitutions within each individual dataset, with the exception of the percent detrimental. The percent of radical/MTN or SN/conservative substitutions that were detrimental was calculated for each dataset and the results were then pooled for the statistical test.

References

Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat Rev Genet. 8(8):619–631. doi:10.1038/nrg2158.

Bedford T, Cobey S, Pascual M. 2011. Strength and tempo of selection revealed in viral gene genealogies. BMC Evolutionary Biology. 11(1):220. doi:10.1186/1471-2148-11-220.

Bell G. 2009. The Oligogenic View of Adaptation. Cold Spring Harb Symp Quant Biol. 74:139–144. doi:10.1101/sqb.2009.74.003.

Bergman J, Eyre-Walker A. 2019. Does Adaptive Protein Evolution Proceed by Large or Small Steps at the Amino Acid Level? Mol Biol Evol. 36(5):990–998. doi:10.1093/molbev/msz033.

Bloom JD. 2015. Software for the analysis and visualization of deep mutational scanning data. BMC Bioinformatics. 16:168. doi:10.1186/s12859-015-0590-4.

Chen Q, He Z, Lan A, Shen X, Wen H, Wu C-I. 2019. Molecular Evolution in Large Steps—Codon Substitutions under Positive Selection. Mol Biol Evol. 36(9):1862–1873. doi:10.1093/molbev/msz108.

Chen Q, Lan A, Shen X, Wu C-I. 2019. Molecular Evolution in Small Steps under Prevailing Negative Selection: A Nearly Universal Rule of Codon Substitution. Genome Biol Evol. 11(10):2702–2712. doi:10.1093/gbe/evz192.

Clarke B. 1970. Selective Constraints on Amino-acid Substitutions during the Evolution of Proteins. Nature. 228(5267):159–160. doi:10.1038/228159a0.

Crick FHC. 1968. The origin of the genetic code. Journal of Molecular Biology. 38(3):367–379. doi:10.1016/0022-2836(68)90392-6.

Dagan T, Talmor Y, Graur D. 2002. Ratios of Radical to Conservative Amino Acid Replacement are Affected by Mutational and Compositional Factors and May Not Be Indicative of Positive Darwinian Selection. Mol Biol Evol. 19(7):1022–1025. doi:10.1093/oxfordjournals.molbev.a004161.

Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK, Lewis SC, Okamoto K, Thomas WK, Lynch M, et al. 2009. A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. PNAS. 106(38):16310–16314. doi:10.1073/pnas.0904895106.

Doud MB, Ashenberg O, Bloom JD. 2015. Site-Specific Amino Acid Preferences Are Mostly Conserved in Two Closely Related Protein Homologs. Mol Biol Evol. 32(11):2944–2960. doi:10.1093/molbev/msv167.

Doud MB, Bloom JD. 2016. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. Viruses. 8(6):155. doi:10.3390/v8060155.

Duda TF, Vanhoye D, Nicolas P. 2002. Roles of diversifying selection and coordinated evolution in the evolution of amphibian antimicrobial peptides. Mol Biol Evol. 19(6):858–864.

Epstein CJ. 1967. Non-randomness of Ammo-acid Changes in the Evolution of Homologous Proteins. Nature. 215(5099):355–359. doi:10.1038/215355a0.

Eyre-Walker A, Keightley PD, Smith NGC, Gaffney D. 2000. Quantifying the Slightly Deleterious Mutation Model of Molecular Evolution. Mol Biol Evol. 19(12):2142–2149. doi:10.1093/oxfordjournals.molbev.a004039.

Freeland SJ, Hurst LD. 1998. The Genetic Code Is One in a Million. J Mol Evol. 47(3):238–248. doi:10.1007/PL00006381.

Geyer R, Mamlouk A. 2018. On the efficiency of the genetic code after frameshift mutations. PeerJ. 6:e4825–e4825. doi:10.7717/peerj.4825.

Gojobori J, Tang H, Akey JM, Wu C-I. 2007. Adaptive evolution in humans revealed by the negative correlation between the polymorphism and fixation phases of evolution. PNAS. 104(10):3907–3912. doi:10.1073/pnas.0605565104.

Grantham R. 1974. Amino Acid Difference Formula to Help Explain Protein Evolution. Science. 185(4154):862–864.

Haddox HK, Dingens AS, Bloom JD. 2016. Experimental Estimation of the Effects of All Amino-Acid Mutations to HIV's Envelope Protein on Viral Replication in Cell Culture. PLOS Pathogens. 12(12):e1006114. doi:10.1371/journal.ppat.1006114.

Haddox HK, Dingens AS, Hilton SK, Overbaugh J, Bloom JD. 2018. Mapping mutational effects along the evolutionary landscape of HIV envelope. Chakraborty AK, editor. eLife. 7:e34420. doi:10.7554/eLife.34420.

Haig D, Hurst LD. 1991. A quantitative measure of error minimization in the genetic code. J Mol Evol. 33(5):412–417. doi:10.1007/BF02103132.

Hanada K, Shiu S-H, Li W-H. 2007. The Nonsynonymous/Synonymous Substitution Rate Ratio versus the Radical/Conservative Replacement Rate Ratio in the Evolution of Mammalian Genes. Mol Biol Evol. 24(10):2235–2241. doi:10.1093/molbev/msm152.

Hughes AL, Green JA, Garbayo JM, Roberts RM. 2000. Adaptive diversification within a large family of recently duplicated, placentally expressed genes. PNAS. 97(7):3319–3323. doi:10.1073/pnas.97.7.3319.

Hughes AL, Ota T, Nei M. 1990. Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol Biol Evol. 7(6):515–524. doi:10.1093/oxfordjournals.molbev.a040626.

Jiang C, Zhao Z. 2006. Mutational spectrum in the recent human genome inferred by single nucleotide polymorphisms. Genomics. 88(5):527–534. doi:10.1016/j.ygeno.2006.06.003.

Kimura M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press.

Kouyos RD, Althaus CL, Bonhoeffer S. 2006. Stochastic or deterministic: what is the effective population size of HIV-1? Trends in Microbiology. 14(12):507–511. doi:10.1016/j.tim.2006.10.001.

Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, Bloom JD. 2018. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. PNAS. 115(35):E8276–E8285. doi:10.1073/pnas.1806133115.

Luo H, Huang Y, Stepanauskas R, Tang J. 2017. Excess of non-conservative amino acid changes in marine bacterioplankton lineages with reduced genomes. Nature Microbiology. 2(8):1–9. doi:10.1038/nmicrobiol.2017.91.

Lyons DM, Lauring AS. 2017. Evidence for the Selective Basis of Transition-to-Transversion Substitution Bias in Two RNA Viruses. Mol Biol Evol. 34(12):3205–3215. doi:10.1093/molbev/msx251.

Miller MP, Kumar S. 2001. Understanding human disease mutations through the use of interspecific genetic variation. Hum Mol Genet. 10(21):2319–2328. doi:10.1093/hmg/10.21.2319.

Miyata T, Miyazawa S, Yasunaga T. 1979. Two types of amino acid substitutions in protein evolution. J Mol Evol. 12(3):219–236. doi:10.1007/BF01732340.

Pauly MD, Procario MC, Lauring AS. 2017. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. eLife Sciences. 6:e26437. doi:10.7554/eLife.26437.

Plotkin JB, Dushoff J. 2003. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. PNAS. 100(12):7152–7157. doi:10.1073/pnas.1132114100.

Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. PNAS. 104(33):13390–13395. doi:10.1073/pnas.0701256104.

Pupko T, Sharan R, Hasegawa M, Shamir R, Graur D. 2003. Detecting excess radical replacements in phylogenetic trees. Gene. 319:127–135. doi:10.1016/S0378-1119(03)00802-3.

Rand DM, Weinreich DM, Cezairliyan BO. 2000. Neutrality tests of conservative-radical amino acid changes in nuclear- and mitochondrially-encoded proteins. Gene. 261(1):115–125. doi:10.1016/S0378-1119(00)00483-2.

Rouzine IM, Coffin JM. 1999. Linkage disequilibrium test implies a large effective population number for HIV in vivo. Proc Natl Acad Sci USA. 96(19):10758–10763. doi:10.1073/pnas.96.19.10758.

Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, et al. 2016. Local fitness landscape of the green fluorescent protein. Nature. 533(7603):397–401. doi:10.1038/nature17995.

Schrider DR, Hourmozdi JN, Hahn MW. 2011. Pervasive Multinucleotide Mutational Events in Eukaryotes. Current Biology. 21(12):1051–1054. doi:10.1016/j.cub.2011.05.013.

Shen Y-Y, Shi P, Sun Y-B, Zhang Y-P. 2009. Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability. Genome Res. 19(10):1760–1765. doi:10.1101/gr.093138.109.

Soh YS, Moncla LH, Eguia R, Bedford T, Bloom JD. 2019. Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans. Ferguson NM, Kawaoka Y, Fodor E, tenOever BR, editors. eLife. 8:e45079. doi:10.7554/eLife.45079.

Sourisseau M, Lawrence DJP, Schwarz MC, Storrs CH, Veit EC, Bloom JD, Evans MJ. 2019. Deep Mutational Scanning Comprehensively Maps How Zika Envelope Protein Mutations Affect Viral Growth and Antibody Escape. Journal of Virology. 93(23). doi:10.1128/JVI.01291-19. [accessed 2020 Feb 13]. https://jvi.asm.org/content/93/23/e01291-19.

Stephens CR, Waelbroeck H. 1999. Codon Bias and Mutability in HIV Sequences. J Mol Evol. 48(4):390–397. doi:10.1007/PL00006483.

Tenaillon O. 2014. The Utility of Fisher's Geometric Model in Evolutionary Genetics. Annual Review of Ecology, Evolution, and Systematics. 45(1):179–201. doi:10.1146/annurev-ecolsys-120213-091846.

Tennessen JA. 2005. Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. Journal of Evolutionary Biology. 18(6):1387–1394. doi:10.1111/j.1420-9101.2005.00925.x.

Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. eLife. 3:e03300. doi:10.7554/eLife.03300.

Tripathi S, Deem MW. 2018. The Standard Genetic Code Facilitates Exploration of the Space of Functional Nucleotide Sequences. J Mol Evol. 86(6):325–339. doi:10.1007/s00239-018-9852-x.

Venkat A, Hahn MW, Thornton JW. 2018. Multinucleotide mutations cause false inferences of lineage-specific positive selection. Nature Ecology & Evolution. 2(8):1280–1288. doi:10.1038/s41559-018-0584-5.

Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS. 2016. The Mutational Robustness of Influenza A Virus. PLOS Pathog. 12(8):e1005856. doi:10.1371/journal.ppat.1005856.

Vogel F, Kopun M. 1977. Higher frequencies of transitions among point mutations. J Mol Evol. 9(2):159–180. doi:10.1007/BF01732746.

Wernegreen JJ. 2011. Reduced Selective Constraint in Endosymbionts: Elevation in Radical Amino Acid Replacements Occurs Genome-Wide. PLOS ONE. 6(12):e28905. doi:10.1371/journal.pone.0028905.

Woese CR. 1965. Order in the genetic code. PNAS. 54(1):71–75. doi:10.1073/pnas.54.1.71.

Woese CR, Dugre DH, Dugre SA, Kondo M, Saxinger WC. 1966. On the Fundamental Nature and Evolution of the Genetic Code. Cold Spring Harb Symp Quant Biol. 31:723–736. doi:10.1101/SQB.1966.031.01.093.

Xu S, He Z, Zhang Z, Guo Z, Guo W, Lyu H, Li J, Yang M, Du Z, Huang Y, et al. 2017. The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. Natl Sci Rev. 4(5):721–734. doi:10.1093/nsr/nwx065.

Yampolsky LY, Stoltzfus A. 2005. The Exchangeability of Amino Acids in Proteins. Genetics. 170(4):1459–1472. doi:10.1534/genetics.104.039107.

Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J Mol Evol. 50(1):56–68.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. Mol Biol Evol. 22(12):2472–2479. doi:10.1093/molbev/msi237.

Zhang J, Webb DM. 2004. Rapid evolution of primate antiviral enzyme APOBEC3G. Hum Mol Genet. 13(16):1785–1791. doi:10.1093/hmg/ddh183.

Zhang J, Zhang Y, Rosenberg HF. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat Genet. 30(4):411–415. doi:10.1038/ng852.

Zhang Z, Gerstein M. 2003. Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. Nucleic Acids Res. 31(18):5338–5348. doi:10.1093/nar/gkg745.

Supplemental Figures & Tables

Supplemental Figure 4-1 Colors corresponding to strains in main figures

Supplemental Table 4-1 Numbers of radical/conservative detrimental and beneficial substitutions

Species	Strain	Gene	Overall	Con. Charge	Radical Charge	Con. Polarity	Radical Polarity	Con. Polarity & Size	Radi Polar & Si
HIV-1	BF520	ENV	12578	6811, 843	4319, 605	5438, 780	5692, 668	1700, 286	943 116
	BG505	ENV	12711	6599, 1115	4180, 817	5198, 1101	5581, 831	1591, 422	918 151
	LAI	ENV	13540	5757, 2105	4193, 1485	4415, 1948	5535, 1642	1332, 715	861 287
Influenza A virus	A/PR8/ 1934 (H1N1)	NP	9462	5170, 136	4088, 68	4616, 114	4642, 90	1472, 53	778 151
	A/Aichi /2/196 8 (H3N2)	NP	9462	5039, 241	4022, 160	4492, 230	4569, 171	1421, 104	764 297
	A/WSN /1933 (H1N1)	HA	10716	5948, 353	4174, 241	5042, 334	5080, 260	1536, 156	858 438
	A/Perth /16/20 09 (H3N2)	HA	10755	5850, 432	4151, 321	4987, 391	5014, 362	1523, 160	847 593
	A/Gree n- winged Teal/Oh io/175/ 1986 (Avian)	PB2	14421	7536, 579	5879, 427	6531, 554	6884, 452	2103, 221	1131 785
	A/Gree n- winged Teal/Oh io/175/ 1986 (Avian)	PB2	14421	7601, 514	5952, 354	6601, 484	6952, 384	2128, 196	1142 672
Zika virus	MR766	ENV	9576	5393, 244	3832, 107	4520, 208	4705, 143	1449, 106	777 245

Supplemental Table 4-2 Numbers of SN and MTN detrimental and beneficial substitutions

Species	Strain	Gene	Overall	SN	2 MTN	3 MTN
HIV-1	BF520	ENV	4619, 577	5912, 797	599, 74	4619, 577
	BG505	ENV	4404, 881	5775, 973	600, 78	4404, 881
	LAI	ENV	3935, 1742	5469, 1693	546, 155	3935, 1742
Influenza A virus	A/PR8/1934 (H1N1)	NP	3936, 101	4844, 98	478, 5	3936, 101
	A/Aichi/2/1968 (H3N2)	NP	3836, 189	4757, 190	468, 22	3836, 189
	A/WSN/1933 (H1N1)	HA	4183, 295	5400, 283	539, 16	4183, 295
	A/Perth/16/2009 (H3N2)	HA	4129, 364	5317, 355	556, 34	4129, 364
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	5664, 480	7085, 475	666, 51	5664, 480
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	5758, 386	7120, 440	675, 42	5758, 386
Zika virus	MR766	ENV	3764, 221	4985, 121	476, 9	3764, 221

Supplemental Table 4-3 Average detrimental effects of radical/conservative substitutions

Species	Strain	Gene	Overall	Con. Charge	Radical Charge	Con. Polarity	Radical Polarity	Con. Polarity & Size	Radic Polar & Siz
HIV-1	BF520	ENV	-1.546	-2.008	-1.846	-2.087	-2.162	-2.012	-1.5
	BG505	ENV	-2.01	-1.569	-1.435	-1.591	-1.651	-1.509	-1.29
	LAI	ENV	-1.828	-1.751	-1.627	-1.895	-1.989	-1.934	-1.39
Influenza A virus	A/PR8/1934 (H1N1)	NP	-4.065	-3.832	-3.926	-4.215	-4.204	-4.36	-3.27
	A/Aichi/2/1968 (H3N2)	NP	-3.278	-3.073	-3.21	-3.375	-3.345	-3.535	-2.75
	A/WSN/1933 (H1N1)	HA	-3.191	-3.215	-2.989	-3.3	-3.391	-3.157	-2.57
	A/Perth/16/2009 (H3N2)	HA	-2.474	-2.467	-2.346	-2.552	-2.601	-2.484	-2.03
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	-2.705	-2.606	-2.555	-2.799	-2.848	-2.832	-2.20
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	-2.432	-2.308	-2.301	-2.526	-2.557	-2.592	-1.92
Zika virus	MR766	ENV	-4.31	-4.199	-4.117	-4.438	-4.494	-4.466	-3.62

Supplemental Table 4-4 Average Detrimental Effects of SN and MTN Substitutions

Species	Strain	Gene	Overall	SN	2 MTN	3 MTN
HIV-1	BF520	ENV	-1.546	-1.863	-2.096	-2.255
	BG505	ENV	-2.01	-1.481	-1.591	-1.596
	LAI	ENV	-1.828	-1.681	-1.921	-1.955
Influenza A virus	A/PR8/1934 (H1N1)	NP	-4.065	-3.575	-4.428	-4.422
	A/Aichi/2/1968 (H3N2)	NP	-3.278	-3.095	-3.414	-3.397
	A/WSN/1933 (H1N1)	HA	-3.191	-2.91	-3.378	-3.493
	A/Perth/16/2009 (H3N2)	HA	-2.474	-2.295	-2.585	-2.744
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	-2.705	-2.506	-2.869	-2.68
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	-2.432	-2.25	-2.582	-2.386
Zika virus	MR766	ENV	-4.31	-4.028	-4.527	-4.268

Supplemental Table 4-5 Average Beneficial Effects of Radical/Conservative Substitutions

Species	Strain	Gene	Overall	Con. Charge	Radical Charge	Con. Polarity	Radical Polarity	Con. Polarity & Size	Radic Polar & Siz
HIV-1	BF520	ENV	0.304	0.385	0.409	0.333	0.407	0.429	0.43
	BG505	ENV	0.408	0.302	0.303	0.279	0.305	0.31	0.30
	LAI	ENV	0.904	0.82	0.918	0.832	0.888	0.922	1.02
Influenza A virus	A/PR8/1934 (H1N1)	NP	0.726	0.739	0.713	0.715	0.742	0.73	0.69
	A/Aichi/2/1968 (H3N2)	NP	0.732	0.652	0.742	0.674	0.718	0.752	0.85
	A/WSN/1933 (H1N1)	HA	0.55	0.567	0.549	0.497	0.55	0.568	0.52
	A/Perth/16/2009 (H3N2)	HA	0.525	0.479	0.523	0.514	0.527	0.527	0.58
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	0.467	0.432	0.45	0.434	0.487	0.476	0.51
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	0.47	0.433	0.457	0.434	0.485	0.48	0.51
Zika virus	MR766	ENV	0.829	0.811	0.804	0.742	0.866	0.867	0.8

Supplemental Table 4-6 Average Beneficial Effects of SN and MTN Substitutions

Species	Strain	Gene	Overall	SN	2 MTN	3 MTN
HIV-1	BF520	ENV	0.304	0.425	0.383	0.471
	BG505	ENV	0.408	0.31	0.292	0.33
	LAI	ENV	0.904	0.921	0.883	0.967
Influenza A virus	A/PR8/1934 (H1N1)	NP	0.726	0.752	0.696	0.817
	A/Aichi/2/1968 (H3N2)	NP	0.732	0.716	0.753	0.696
	A/WSN/1933 (H1N1)	HA	0.55	0.565	0.536	0.543
	A/Perth/16/2009 (H3N2)	HA	0.525	0.506	0.543	0.52
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	0.467	0.47	0.454	0.549
	A/Green-winged Teal/Ohio/175/1986 (Avian)	PB2	0.47	0.475	0.461	0.503
Zika virus	MR766	ENV	0.829	0.801	0.841	0.912

Chapter 5 Discussion

Overview

The projects in my thesis have resolved major puzzles on how mutational effects scale across levels of biological organization (Figure 5-1). At the lowest level, the impact of individual mutations can be understood deterministically from biological principles, just as the trajectory of an individual air molecule can be predicted from its collisions with several other molecules. However, at a higher phenotypic level, this determinism breaks down into seeming randomness, or idiosyncrasy. My first chapter quantifies this idiosyncrasy and explains why it occurs. A single site contributes to a phenotype via a potentially astronomical number of interactions. Consequently, a single mutation has unpredictable effects in the context of other mutations, just as the deterministic movement of a single air molecule scales to the random outcome of a die roll in a room full of molecules. At the highest level, due to the same statistical laws that govern a game of dice, we find that randomness in mutational effects leads to universal patterns in their collective effects across fitness levels and in evolutionary trajectories during adaptation and drift.

While idiosyncrasy is a major influence on fitness landscapes, it is not the only one. My second and third chapters revisit the deterministic biological effects of mutations and their impact on fitness and evolutionary patterns (Figure 5-1, right). I find that radical versus conservative

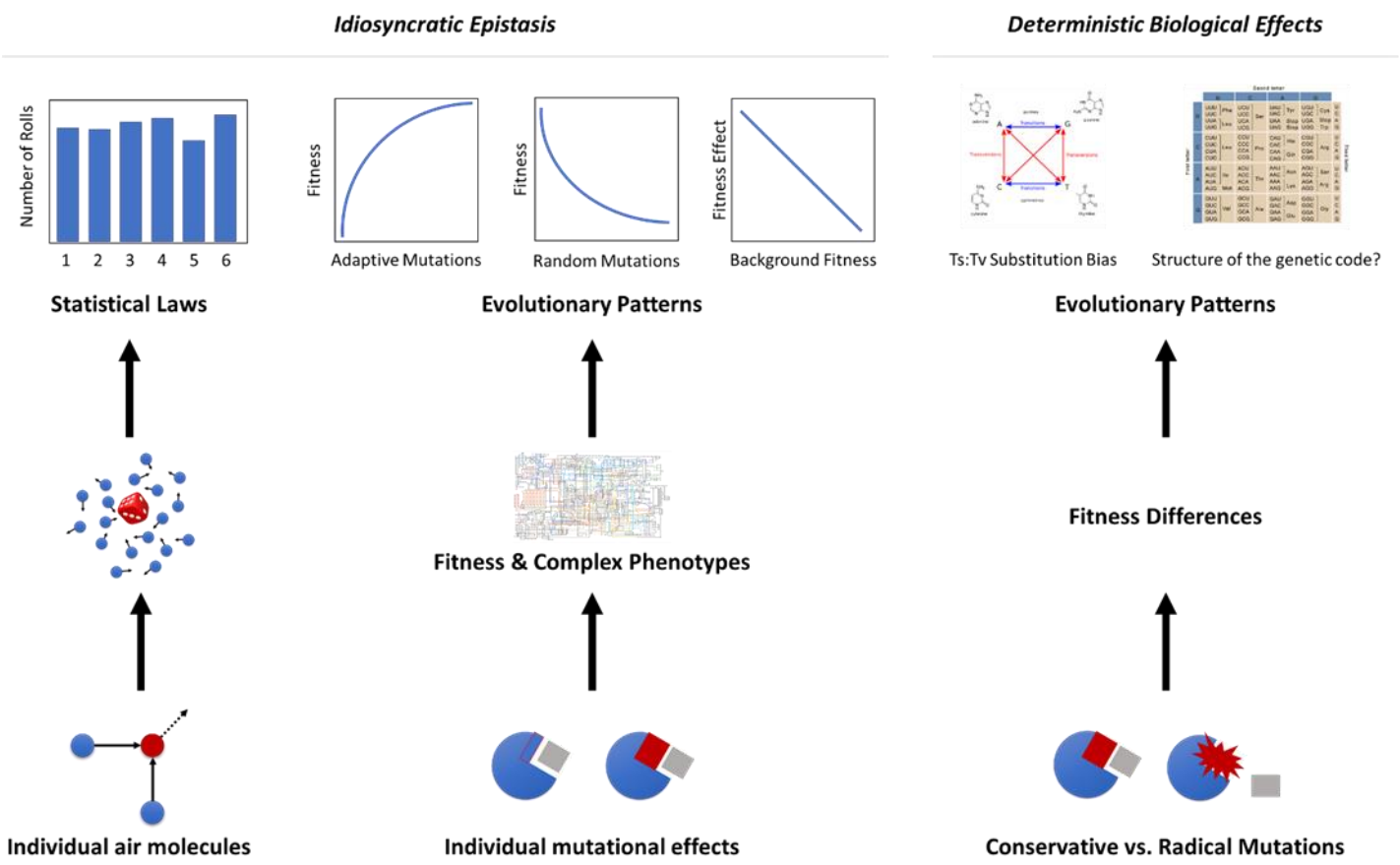


Figure 5-1 The role of the idiosyncratic epistasis theory and conservative versus radical mutations in the scaling of mutational affects across levels of biology.

Left and middle: the idiosyncratic epistasis theory posits that individual mutational effects are like the interactions between individual air molecules. At the lowest level, mutations have predictable consequences, but seeming randomness arises at a higher level due to complex interactions. However, predictability re-emerges as statistical laws at the level of evolution. Right: Conservative versus radical amino acid substitutions represent a deviation from idiosyncrasy, as this distinction at the lowest level leads to predictable fitness differences and biased phylogenetic patterns and evolution.

changes in amino acid properties have consistent differences in fitness effects, helping to explain widespread phylogenetic patterns and supporting the error-minimization hypothesis for the evolution of the genetic code.

How can we use the theory of idiosyncratic epistasis to further understand fitness landscapes? What are the further implications of the differences between radical and conservative amino acid substitutions?

Idiosyncratic epistasis as a null theory

Idiosyncratic epistasis could be used as a null model to uncover biased forms of epistasis and their impact on evolution.

First, evidence for various biologically-based theories of epistasis may need re-examination (Kacser and Burns 1981; Martin et al. 2007; Sanjuán and Nebot 2008; Gros et al. 2009; Wei and Zhang 2019). Most theories posit a bias towards *either* negative *or* positive epistasis and find support from observations involving beneficial or detrimental mutations, respectively. Our theory shows that that the patterns involving beneficial and detrimental mutations must be understood together. They also arise naturally from idiosyncrasy without the need for specific biological mechanisms or selection for robustness as posited by various theories.

However, many of the assumptions of these biologically-based theories still hold – biological systems do show modularity as in the modular life theory (Wei and Zhang 2019), viruses do encode less functional redundancy than eukaryotes as in the multiple hit theory (Sanjuán and Elena 2006), and so on. Thus, they are likely useful for understanding cases in which our null expectations are violated or why idiosyncrasy levels vary between fitness landscapes (Figure 2-1C). For example, idiosyncrasy predicts a negative correlation between mutational effects and background fitness, but we find that the strength of this correlation varies across mutations, and is even positive for some mutations (Figures 2-2B, S2A). This variation is likely not all due to limited empirical information. What are the properties of organisms, genes, sites or individual mutations that determine their level of idiosyncrasy and cause them to deviate from our null expectations? The biological mechanisms other theories posit may hold answers. However,

more work will be needed to understand how such mechanisms combine with idiosyncrasy. Mathematical models of fitness landscapes like the NK-model may be helpful to test hypotheses; for example, how modular genetic architectures may affect idiosyncrasy.

In comparison to an additive null model, the finding of negative epistasis/diminishing returns among beneficial mutations is commonly thought to slow adaptation (Couce and Tenaillon 2015). However, an idiosyncratic landscape may be a more appropriate null model to investigate the role of epistasis during adaptation. Consider a hypothetical organism that has evolved biological mechanisms which in isolation would lead to positive epistasis between beneficial mutations. However, these mechanisms must work against the inherent diminishing returns due to idiosyncrasy, perhaps resulting in overall negative epistasis between beneficial mutations, but less so than our null expectation. Comparing adaptation in this organism to an additive null model would reveal diminished adaptability, when in reality, the organism had evolved to lessen the inherent constraints of idiosyncrasy.

A recent paper on adaptation in a fungus made this same, potentially misguided conclusion (Schoustra et al. 2016). Fisher's geometric model was able to predict the negative epistasis among adaptive mutations while a House-of-Cards landscape (a completely idiosyncratic landscape) predicted stronger negative epistasis than was found. The authors concluded that the pattern of epistasis in the fungus constrained its adaptation, when in fact the bias towards less negative epistasis could have sped up its adaptation relative to our null prediction.

Our theory also makes a null prediction for the relationship between adaptability and robustness (the insensitivity to detrimental mutations). Lower fit organisms will have both higher adaptability and higher robustness due to the negative correlation between mutational fitness effects and background fitness. Some theories that posit the opposite relationship, based on the idea that insensitivity of fitness to detrimental mutations cuts both ways, coinciding with insensitivity to beneficial mutations (Ancel and Fontana 2000; Lenski et al. 2006; Sumedha et al. 2007). Other theories resolve the tension between robustness and evolvability by noting that robustness enhances evolvability by enabling exploration of neutral sequence space and providing a bank of diversity upon environmental change (Wagner 2008). This could combine synergistically with the higher evolvability of more robust, lower fit organisms due to idiosyncrasy. Deviations from our null expectation may reveal interesting forms of epistasis or other biological mechanisms. For example, our finding that radical amino acid substitutions are both more detrimental and more beneficial in the same genotype fits better with the theory that (in)sensitivity to mutation cuts both ways.

It will be essential to create a model of idiosyncratic epistasis that can be fit to empirical data. This would enhance our ability to uncover deviations from null expectations and compare the importance of idiosyncrasy relative to other biological mechanisms. However, the *n*-order model is not easily adapted for this purpose. Because of the vast number of potential terms, the model cannot be simulated beyond 20 or so interacting sites. In some ways there is also too much flexibility in the model; for example, the same level of idiosyncrasy could be achieved by many different schemes for weighting the interaction terms. Our idiosyncrasy index may provide a starting point, but it is unclear how the size of a landscape, missing data, or variation in

idiosyncrasy among sites affects its calculation. Thus, much work is needed to develop the theory of idiosyncrasy into a model capable of direct comparison with empirical data.

Implications of Radical/Conservative Fitness Differences

We have shown that substitutions which radically change amino acid properties are both more detrimental and more highly beneficial than conservative changes, at least in RNA viruses. The difference in detrimental effects likely contributes to the widespread transition-transversion (Ts/Tv) substitution bias. It is interesting to speculate on the impact of radical/conservative differences on other evolutionary processes.

There is widespread evidence of a transitional mutation bias across taxa (Zhang and Gerstein 2003; Jiang and Zhao 2006; Baer et al. 2007; Denver et al. 2009; Pauly et al. 2017). For example, the experimentally determined Ts/Tv mutational bias in influenza is 2-3.6 (Bloom 2014; Pauly et al. 2017). Is it possible that the more radical nature of transversions has led to selection for a reduced transversion mutation rate? The causation could work in reverse, or in both ways – an underlying higher transition rate could have selected for robustness to transitional substitutions via codon usage bias or other mechanisms. Simulations could determine whether the transversion/transition fitness differences are high enough to select for reduced transversion rates (Lynch 2010). This could indicate ongoing selection on the mutation rate which could be verified. Mutational effects in rare organisms in which the transversion rate is higher could also illuminate this question (Keller et al. 2007).

Similarly, the evolution of the genetic code is hypothesized to have involved selection for error-minimization (Woese 1965; Woese et al. 1966; Crick 1968; Haig and Hurst 1991; Freeland and

Hurst 1998; Geyer and Mamlouk 2018; Tripathi and Deem 2018). Our results support this hypothesis, showing that single nucleotide codon mutations indeed cause less damage than the more radical, multi-nucleotide mutations. However, radical mutations are also more beneficial. Did selection for robustness in the genetic code consign all subsequent life forms to lower adaptability? Alternatively, this particular genetic code could have been under selection for evolvability as well as robustness. For example, recent studies have questioned the error-minimization hypothesis because only slightly different genetic codes can minimize radical changes even more than the universal code (Di Giulio and Medugno 2001). However, this study ignored adaptability – perhaps the current genetic code is a more optimal solution to this tradeoff than these alternatives.

The fitness differences between radical/conservative and Ts/Tv substitutions have implications for viral evolution. Radical charge changes showed the least differences from conservative ones in terms of their detrimental effects (Figures 2-4, 4-1, 4-2A-B). This was particularly true for the viral surface proteins HA and ENV as compared to the internal NP proteins (Figures 2-4, 4-1, 4-2A-B). These surface proteins have a history of immune selection and exhibit tolerance to mutations allowing for immune escape (Stephens and Waelbroeck 1999; Plotkin and Dushoff 2003; Thyagarajan and Bloom 2014; Doud and Bloom 2016; Visher et al. 2016). Charge changes were also some of the most beneficial type of radical changes (Figure 4-2); it is interesting to speculate whether this advantage is enhanced under selection for immune escape. DMS data is available to answer this question, which could be incorporated into further work. Furthermore, studies of codon usage have also found a bias towards non-synonymously mutating codons in the HA protein (Plotkin and Dushoff 2003) or that enhance evolvability and robustness (Lauring et

al. 2012), but a potential bias towards codons with a propensity for charge changes was not evaluated. DMS data could be used to further investigate the advantages or disadvantages of the wild-type codon usage compared to simulated alternatives. The role of radical changes in cluster-transition mutations, which are those that drastically change the antigenic properties of influenza and prompt vaccine updates (Koelle et al. 2006), could also be investigated.

Conclusion

In contrast to biased forms of epistasis in other theories, the crucial idea of idiosyncratic epistasis is that the immense number of interactions that determine a phenotype leads to unpredictability of mutational effects. Phrased another way, more idiosyncratic epistasis makes quite different phenotypes accessible by fewer mutational steps. From this perspective, idiosyncrasy may promote the evolvability and great phenotypic diversity of life. Perhaps it contributes to the surprising genetic similarity of quite different organisms (King and Wilson 1975) or speeds up phenotypic evolution in response to environmental changes. Furthermore, as any phenotype becomes more accessible from any genotype, idiosyncrasy also means that very different genotypes may have similar phenotypes. Perhaps this underlies convergence at the phenotypic level. Indeed, ancient proteins with conserved functions have continued to diverge over billions of years seemingly without limit on sequence dissimilarity (Povolotskaya and Kondrashov 2010). In the midst of the idiosyncrasy we find, it is surprising that simple radical versus conservative distinctions in amino acid substitutions have predictable differences. However, if such biological features never mattered to fitness – if idiosyncrasy was complete – the fitness landscape may become too rugged to sustain adaptation for long and any drift would be drastically deleterious. Similar to Kauffman's idea that the most evolvable systems lie at the

boundary “between order and chaos” (Kauffman 1993), might the right balance of idiosyncrasy distinguish evolvable complex systems from non-living ones?

References

Ancel LW, Fontana W. 2000. Plasticity, evolvability, and modularity in RNA. Journal of Experimental Zoology. 288(3):242–283. doi:10.1002/1097-010X(20001015)288:3<242::AID-JEZ5>3.0.CO;2-O.

Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat Rev Genet. 8(8):619–631. doi:10.1038/nrg2158.

Bloom JD. 2014. An Experimentally Determined Evolutionary Model Dramatically Improves Phylogenetic Fit. Mol Biol Evol. 31(8):1956–1978. doi:10.1093/molbev/msu173.

Couce A, Tenaillon OA. 2015. The rule of declining adaptability in microbial evolution experiments. Front Genet. 6. doi:10.3389/fgene.2015.00099. [accessed 2018 Dec 5]. https://www.frontiersin.org/articles/10.3389/fgene.2015.00099/full.

Crick FHC. 1968. The origin of the genetic code. Journal of Molecular Biology. 38(3):367–379. doi:10.1016/0022-2836(68)90392-6.

Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK, Lewis SC, Okamoto K, Thomas WK, Lynch M, et al. 2009. A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. PNAS. 106(38):16310–16314. doi:10.1073/pnas.0904895106.

Di Giulio M, Medugno M. 2001. The Level and Landscape of Optimization in the Origin of the Genetic Code. J Mol Evol. 52(4):372–382. doi:10.1007/s002390010167.

Doud MB, Bloom JD. 2016. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. Viruses. 8(6):155. doi:10.3390/v8060155.

Freeland SJ, Hurst LD. 1998. The Genetic Code Is One in a Million. J Mol Evol. 47(3):238–248. doi:10.1007/PL00006381.

Geyer R, Mamlouk A. 2018. On the efficiency of the genetic code after frameshift mutations. PeerJ. 6:e4825–e4825. doi:10.7717/peerj.4825.

Gros P-A, Nagard HL, Tenaillon O. 2009. The Evolution of Epistasis and Its Links With Genetic Robustness, Complexity and Drift in a Phenotypic Model of Adaptation. Genetics. 182(1):277–293. doi:10.1534/genetics.108.099127.

Haig D, Hurst LD. 1991. A quantitative measure of error minimization in the genetic code. J Mol Evol. 33(5):412–417. doi:10.1007/BF02103132.

Jiang C, Zhao Z. 2006. Mutational spectrum in the recent human genome inferred by single nucleotide polymorphisms. Genomics. 88(5):527–534. doi:10.1016/j.ygeno.2006.06.003.

Kacser H, Burns JA. 1981. The Molecular Basis of Dominance. Genetics. 97(3–4):639–666.

Kauffman SA. 1993. The Origins of Order: Self-organization and Selection in Evolution. Oxford University Press.

Keller I, Bensasson D, Nichols RA. 2007. Transition-Transversion Bias Is Not Universal: A Counter Example from Grasshopper Pseudogenes. PLOS Genet. 3(2):e22. doi:10.1371/journal.pgen.0030022.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science. 188(4184):107–116. doi:10.1126/science.1090005.

Koelle K, Cobey S, Grenfell B, Pascual M. 2006. Epochal Evolution Shapes the Phylodynamics of Interpandemic Influenza A (H3N2) in Humans. Science. 314(5807):1898–1903. doi:10.1126/science.1132745.

Lauring AS, Acevedo A, Cooper SB, Andino R. 2012. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. Cell Host and Microbe. 12(5):623–632. doi:10.1016/j.chom.2012.10.008.

Lenski RE, Barrick JE, Ofria C. 2006. Balancing Robustness and Evolvability. PLOS Biology. 4(12):e428. doi:10.1371/journal.pbio.0040428.

Lynch M. 2010. Evolution of the mutation rate. Trends in Genetics. 26(8):345–352. doi:10.1016/j.tig.2010.05.003.

Martin G, Elena SF, Lenormand T. 2007. Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nat Genet. 39(4):555–560. doi:10.1038/ng1998.

Pauly MD, Procario MC, Lauring AS. 2017. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. eLife Sciences. 6:e26437. doi:10.7554/eLife.26437.

Phillips PC. 2008. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 9(11):855–867. doi:10.1038/nrg2452.

Plotkin JB, Dushoff J. 2003. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. PNAS. 100(12):7152–7157. doi:10.1073/pnas.1132114100.

Povolotskaya IS, Kondrashov FA. 2010. Sequence space and the ongoing expansion of the protein universe. Nature. 465(7300):922. doi:10.1038/nature09105.

Sanjuán R, Elena SF. 2006. Epistasis correlates to genomic complexity. PNAS. 103(39):14402–14405. doi:10.1073/pnas.0604543103.

Sanjuán R, Nebot MR. 2008. A Network Model for the Correlation between Epistasis and Genomic Complexity. PLOS ONE. 3(7):e2663. doi:10.1371/journal.pone.0002663.

Schoustra S, Hwang S, Krug J, Visser JAGM de. 2016. Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. Proc R Soc B. 283(1837):20161376. doi:10.1098/rspb.2016.1376.

Stephens CR, Waelbroeck H. 1999. Codon Bias and Mutability in HIV Sequences. J Mol Evol. 48(4):390–397. doi:10.1007/PL00006483.

Sumedha, Martin OC, Wagner A. 2007. New structural variation in evolutionary searches of RNA neutral networks. Biosystems. 90(2):475–485. doi:10.1016/j.biosystems.2006.11.007.

Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. eLife. 3:e03300. doi:10.7554/eLife.03300.

Tripathi S, Deem MW. 2018. The Standard Genetic Code Facilitates Exploration of the Space of Functional Nucleotide Sequences. J Mol Evol. 86(6):325–339. doi:10.1007/s00239-018-9852-x.

Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS. 2016. The Mutational Robustness of Influenza A Virus. PLOS Pathog. 12(8):e1005856. doi:10.1371/journal.ppat.1005856.

Wagner A. 2008. Robustness and evolvability: a paradox resolved. Proceedings of the Royal Society B: Biological Sciences. 275(1630):91–100. doi:10.1098/rspb.2007.1137.

Wei X, Zhang J. 2019. Patterns and Mechanisms of Diminishing Returns from Beneficial Mutations. Mol Biol Evol. 36(5):1008–1021. doi:10.1093/molbev/msz035.

Woese CR. 1965. Order in the genetic code. PNAS. 54(1):71–75. doi:10.1073/pnas.54.1.71.

Woese CR, Dugre DH, Dugre SA, Kondo M, Saxinger WC. 1966. On the Fundamental Nature and Evolution of the Genetic Code. Cold Spring Harb Symp Quant Biol. 31:723–736. doi:10.1101/SQB.1966.031.01.093.

Zhang Z, Gerstein M. 2003. Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. Nucleic Acids Res. 31(18):5338–5348. doi:10.1093/nar/gkg745.