HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2022 August 31.

Published in final edited form as:

Science. 2022 May 20; 376(6595): 823–830. doi:10.1126/science.abn6895.

Epistatic drift causes gradual decay of predictability in protein evolution

Yeonwoo Park[1], **Brian P. H. Metzger**[2], **Joseph W. Thornton**[1,2,3,*]

[1]Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL, USA.

[2]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA.

[3]Department of Human Genetics, University of Chicago, Chicago, IL, USA.

Abstract

Epistatic interactions can make the outcomes of evolution unpredictable, but no comprehensive data are available on the extent, direction, rate, and consequences of changes in the effects of mutations as protein sequences evolve. Here we characterize the temporal dynamics of epistatic change by using deep mutational scanning to measure the functional effect of every possible amino-acid mutation in a phylogenetic series of reconstructed ancestral and extant proteins, using the steroid receptor DNA-binding domain as a model. Across a 700-million-year historical trajectory, the effects of most mutations became completely or partially decorrelated from their initial effects. Epistatic interactions caused windows of evolutionary accessibility for most mutations to open and close transiently, shaping the historical fate not only of the mutations that fixed during history but also the far greater number that never did. Most mutations' effects evolved under Brownian motion: gradual change without directional bias, at a rate that was largely constant across time but varied dramatically among mutations, indicating a neutral process caused by many weak interactions. Protein sequences therefore drift inexorably into contingency and unpredictability, but that process itself is statistically predictable, given sufficient phylogenetic and experimental data.

One Sentence Summary:

*Corresponding author: joet1@uchicago.edu.

Author contributions:
Conceptualization: YP, BPHM, JWT
Experiments: YP
Data analysis: YP
Development of methodology: YP, JWT, BPHM
Writing – original draft: YP, JWT
Writing – review & editing: YP, JWT, BPHM
Funding acquisition: JWT

Competing interests: The authors declare that they have no competing interests.

Supplementary Materials:
Materials and Methods
Figs. S1 to S10
Tables S1 and S2
References (37–54)
Data S1

Epistatic drift causes the effect and fate of most mutations to gradually become unpredictable during protein evolution.

A mutation's evolutionary fate depends on its phenotypic effects. If the effects are stable over time, knowledge of them in the present can help predict the future course of evolution and explain the causes of evolutionary change in the past. Epistatic interactions, however, may cause a mutation's effects to change over time and its evolutionary accessibility to become contingent on the particular sequence changes that preceded it during history (1, 2).

Despite a recent tide of information about epistatic interactions within proteins, we lack a comprehensive understanding of changes in the effects of mutations (the set of all potential amino-acid changes) caused by interactions with substitutions (the subset of mutations that fix during evolution). What fraction of mutations change in their effects over evolutionary time, and how drastically? Do they change gradually or episodically, and at what rate? What are the consequences for evolutionary outcomes? Deep mutational scanning (DMS) experiments have detected epistasis among mutations within present-day proteins (3–9), but these studies do not address interactions with historical substitutions or reveal changes in mutations' effects over evolutionary time. Some mutations have different effects when introduced into various present-day proteins, implying epistatic interactions with the substitutions that occurred as these proteins diverged from each other (10–14), but without polarizing and calibrating these differences with respect to time, it is not possible to illuminate the rate, direction, or regularity of the process by which mutations' effects changed during evolution. Ancestral protein reconstruction studies have shown that the effects of particular mutations changed during particular phylogenetic intervals (15–22), but these works have examined only the beginning and end of an interval and therefore cannot reveal the temporal dynamics of epistasis.

Here we address this knowledge gap by using DMS to comprehensively assess the effect of introducing every possible amino-acid mutation into a series of reconstructed ancestral and extant proteins along a densely sampled phylogenetic trajectory. We used as a model the DNA-binding domain (DBD) of steroid hormone receptors, a family of essential transcription factors in bilaterian animals that mediate the actions of sex and adrenal steroids by binding to specific DNA sequences and regulating the expression of target genes (23–25). This approach allowed us to measure changes in the functional effect of every possible amino-acid mutation during a series of defined intervals across 700 million years of DBD evolution. To analyze these data, we developed a quantitative framework that treats each mutation's effect as a trait that evolves probabilistically on a phylogeny, which we used to characterize the temporal dynamics, evolutionary consequences, and underlying genetic architecture of epistatic interactions.

Results

Phylogenetic deep mutational scanning

We first inferred the phylogeny of steroid and related receptors (Fig. 1A and fig. S1) and reconstructed the maximum a posteriori protein sequences of 7 ancestral DBDs: the

ancient progenitor protein whose duplication and divergence gave rise to the first steroid receptor (AncNR3), the ancestor of all extant steroid receptors (AncSR, which existed in the ancestor of all bilaterians), and 5 descendants of AncSR along two lineages – one leading to human glucocorticoid receptor (GR) and the other to the steroid receptor of the annelid *Capitella teleta*, which are among the most diverged of all functionally characterized extant DBDs (Fig. 1B and fig. S2). These 9 DBDs are separated by 8 phylogenetic intervals, each involving 3 to 42% sequence divergence. We constructed a yeast strain carrying a GFP reporter driven by a DNA response element for these DBDs and confirmed that all reconstructed ancestral DBDs bind to it, as expected based on prior studies (25). GFP fluorescence in this strain correlates well with binding affinity previously measured using fluorescence anisotropy (fig. S4D).

For each of the 7 ancestral and 2 extant DBDs, we generated a library of variants that contains all 19 possible amino-acid mutations at all 76 sites (fig. S3). We used a bulk assay of fluorescence-activated cell sorting (FACS) coupled with deep sequencing to quantify the GFP fluorescence of each variant with very high repeatability ($r^2 = 0.99$ across 3 replicates; Fig. 1C and figs. S4 and S5). We calculated the effect of a mutation as the difference in the mean $\log_{10}$-GFP fluorescence (ΔF) between variants that differ by a single amino acid; we applied this approach to all mutations from the wild-type amino acid in any of the 9 DBDs to all other 19 amino acids.

Differences in the effect of a mutation between successive nodes on the phylogeny ($\Delta\Delta F$) indicate that the mutation interacts with historical substitutions that occurred during that interval (Fig. 1D). We normalized mutations' effects to remove global background-dependence caused by different wild-type activity levels (fig. S6); after this correction, differences in a mutation's effect among the 9 DBDs are attributable to specific epistatic interactions with intervening substitutions on the phylogeny.

Pervasive random changes in the effects of mutations

To analyze the evolutionary dynamics of epistasis over time, we adapted a classic quantitative framework for modeling trait evolution on phylogenies (26, 27), including the extent, direction, and rate of evolutionary change of the trait, the underlying genetic architecture, and the relative roles of selection and genetic drift. Our approach treats the phenotypic effect of each mutation as a trait that changes probabilistically across phylogenetic intervals, allowing us to ask these questions about epistatic change during historical DBD evolution.

Sixty percent of all mutations display significantly different effects among the 9 backgrounds, and 22% differ in the direction of their effects (FDR ≤ 0.1; Fig. 2A). Most of the mutations that show no evidence of epistasis destroy protein function regardless of genetic background (ΔF always at the lower bound of measurement, −1.3). Only 5% of mutations have a nondestructive effect that did not vary significantly across the phylogeny.

Epistatic changes occurred during all 8 phylogenetic intervals (Fig. 2B and fig. S7A). Even in the shortest interval – during which there were only two sequence substitutions – the effects of more than 200 mutations changed significantly. During the other intervals,

even more mutations changed in effect. On average, each substitution is associated with significant changes in the effects of about 60 mutations (fig. S7B).

These epistatic changes were unbiased over time. Changes in the effects of mutations ($\Delta\Delta F$) are distributed almost symmetrically around 0 (mean = −0.01; Fig. 2C). The fraction of mutations that reduce activity was nearly constant among the 9 intervals, as was the fraction of mutations that destroy activity (Fig. 2D). No individual mutations had effects that changed with a significant bias in either direction over time (fig. S7C). These data indicate that directional selection did not drive long-term epistatic changes in the effects of mutations, and mutational robustness did not change systematically over time. Further, the variance of the distribution of $\Delta\Delta F$ in each interval increased linearly with sequence divergence, rather than plateauing (Fig. 2, E and F), suggesting no role for stabilizing selection to maintain the effects of mutations within defined limits.

The effects of most mutations drifted gradually

To test whether epistatic change was gradual or episodic, we fit probabilistic models of trait evolution to the trajectory of changes in the effect of each mutation. Brownian motion represents a simple model of gradual evolution at a constant rate without directional bias: changes in the trait value among phylogenetic intervals are normally distributed when normalized for the length of the interval, with a mean change of zero and constant variance per unit sequence divergence (which represents the rate of evolution). In the alternative model of episodic evolution, the normalized variance is a free parameter for each interval, which allows the rate to differ among intervals (Fig. 3A). We fit both models to the 8 $\Delta\Delta F$ values of each mutation and used a likelihood-ratio test to compare the fit of the two models.

We found that the Brownian motion model was the best-fit model for 92% of mutations that changed epistatically (Fig. 3, B and C), irrespective of whether mutations' effects changed rapidly or slowly (Fig. 3D). For the 8% of mutations best fit by the episodic model, effects were nearly constant in most intervals with dramatic changes during one or a few intervals. The functional effects of most mutations therefore evolved as a random variable that changes gradually along the phylogeny at a characteristic rate and without bias. We call this process epistatic drift.

Phylogenetic cross-validation confirmed that the effects of most mutations evolved at a steady rate across the phylogeny (Fig. 3E). For each mutation, we predicted the epistatic change expected in each of the 8 intervals given the rate of epistatic change estimated from the 7 other intervals and then compared these predictions to experimental observations, pooling mutations with similar estimated rates (Fig. 3F). Predicted and observed epistatic changes were strongly correlated (Spearman's $\rho \geq 0.94$ for every interval; Fig. 3G), indicating that mutations' relative rates of epistatic change did not strongly vary along the phylogeny. The absolute rate of epistatic change, however, was systematically faster than predicted in some intervals and slower in others: the mean rate of epistatic change for all mutations in each interval ranges from 0.7 to 1.4 of the average across the phylogeny (Fig. 3H). Epistatic change in the effect of each mutation therefore varies stochastically across intervals (consistent with Brownian motion), but this variation is correlated among mutations; as a result, the total amount of epistatic change across all

mutations is systematically greater in some intervals than others. This pattern is likely to arise because the total epistatic change depends on the particular substitutions that fixed during an interval, and some substitutions are more epistatic than others, interacting more strongly or with a larger number of mutations. The mean rate of epistatic change was not systematically different during intervals following gene duplications.

These observations have two major implications for evolution and the genetic architecture of epistatic interactions (Fig. 3I). First, epistatic interactions within the DBD are dense: most mutations' effects changed gradually because of weak interactions with many substitutions, and each substitution typically modified the effects of many mutations (Fig. 2B). If most epistatic changes were triggered by rare, large-effect modifiers, the distribution of $\Delta\Delta F$ would be enriched near zero and at extreme values, a pattern that we observed for only a small fraction of mutations. Most historical contingency is therefore the cumulative result of many small-effect epistatic modifications. Against this background of gradual epistatic drift, a few mutations occasionally undergo dramatic changes in their effects.

Second, some mutations are more epistatically sensitive than others, with effects that diverged more rapidly as substitutions accumulated. Conversely, some substitutions are more epistatic than others, changing the effects of more target mutations or causing changes of greater magnitude. As a result, there are systematic differences among intervals in the average rate of epistatic change across all mutations.

Mutations vary in memory length and the timescale of contingency

Because the effect of each mutation drifts at random at a steady rate, there should be a characteristic time period after which the mutation's effect can no longer be reliably predicted from its known effect at some other time. We call this period the mutation's memory length, the measure of which is the memory half-life – the amount of sequence divergence over which the correlation of a mutation's effect is reduced by half. To estimate the memory half-life, we partitioned mutations into deciles by their rate of epistatic change and calculated for each decile how correlated the effects of mutations are between each pair of DBDs (Fig. 4, A to D). We modeled the correlation coefficient as an exponentially decaying function of sequence divergence. We then used this relationship to estimate the memory half-life of each mutation from its rate of epistatic change.

Reflecting the wide variation in the rate of epistatic change among mutations, memory half-lives range from just 3% sequence divergence to virtually infinite (Fig. 4E). Mutations with the shortest half-lives therefore forget the effects they had in the past after just a few sequence substitutions at other sites: at any moment, their effect and likely fate depend primarily on the substitutions that occurred most recently during their history.

Relative to the timescale of DBD evolution, about one fourth of all mutations have short memories (half-life < 50% sequence divergence); in this group, the effects in present-day human GR are almost completely independent of their initial effects in AncSR ($r^2 = 0.10$, Fig. 4F). 20% of mutations have medium memory (half-life 50 to 200%), with present-day effects that can be partially predicted from their initial effects ($r^2 = 0.68$). The remaining

54% of mutations have long memories (>200% divergence) and interacted negligibly with historical substitutions, retaining their initial effects throughout DBD evolution ($r^2 = 0.98$).

Contingency of historical sequence evolution

We next focused on the subset of mutations that occurred during historical DBD evolution. We first assessed the functional effects of the 79 substitutions that occurred during the phylogenetic intervals that we experimentally characterized (Fig. 5, A and B). When measured in the ancestral background in which they historically occurred, substitutions that reduce activity by $\Delta F < -0.2$ were nearly absent; of the few exceptions, most fixed during intervals immediately after gene duplication (fig. S8A). This represents a 29-fold depletion compared to the set of all mutations, the majority of which have $\Delta F < -0.2$. These results imply that the DBD evolved primarily under purifying selection against mutations that strongly reduce activity, and they establish $\Delta F = -0.2$ as a boundary that roughly defines the evolutionary accessibility of mutations under purifying selection.

Epistasis shaped the fate of most historical substitutions, which occurred during limited windows when they were transiently accessible. Of all substitutions that fixed between AncSR and any extant steroid receptor on our phylogeny, 43% have short or medium memories (Fig. 5C). Among the short-memory substitutions, the majority were inaccessible in AncSR ($\Delta F_{\mathrm{AncSR}} < -0.2$), implying that they became accessible in one or more descendant proteins because of permissive epistatic substitutions, which render otherwise deleterious mutations neutral or advantageous. The remaining short-memory substitutions were accessible in AncSR ($\Delta F_{\mathrm{AncSR}} \geq -0.2$), but almost all of these became subsequently inaccessible because of restrictive substitutions, which render previously neutral or advantageous substitutions deleterious (fig. S8B). By contrast, 95% of long-memory substitutions were accessible in AncSR and remained so across the entire phylogeny (fig. S8B). Medium-memory substitutions displayed an intermediate pattern. The evolutionary fate of long-memory substitutions could therefore have been reliably predicted from their initial effects, but the accessibility of substitutions with short or medium memory depended on other substitutions that occurred during history.

Epistasis also shaped the fate of the many mutations that did not become substitutions. Of all short-memory mutations that were accessible in AncSR, 90% became inaccessible in one or more descendant proteins, indicating that evolutionary paths to them were closed by restrictive substitutions (Fig. 5D). Conversely, 55% of the short-memory mutations that were inaccessible in AncSR subsequently became accessible because of permissive substitutions. Overall, two-thirds of short-memory mutations and one-third of medium-memory mutations changed in accessibility among the 9 DBDs we tested, with each category of mutations being accessible in 2.4 and 4.9 of the 9 DBDs on average (Fig. 5E).

These data indicate that epistatic interactions with the particular set of substitutions that occurred along the phylogeny contingently determined the evolutionary fate not only of the mutations that fixed historically because of permissive substitutions, but also of those that did not have the opportunity to fix because of restrictive substitutions. Studying only the sequence changes that occurred during evolution therefore underestimates the role of historical contingency: doing so cannot detect the many evolutionary roads that were closed

off contingently, but which could have been taken if the trajectory of sequence changes at interacting sites had unfolded differently.

Causes of variation in memory length

Finally, we sought to identify the factors that determine a mutation's memory length. Some variation in memory length is attributable to the sequence site at which a mutation occurs: the median memory half-life of mutations to any of the 19 mutant states at the same site varies among sites from 11% to >200% divergence (Fig. 6A). But this variation is not associated with any obvious structural or functional properties: the median memory half-life of a site is poorly correlated with relative solvent accessibility, rate of substitution, rate of substitution at physically adjacent sites, distance to the DNA-binding residues, and distance to the dimerization interface ($r^2 < 0.1$ for every factor; fig. S9). Further, memory length varies extensively within each site, with 59 of 76 sites in the DBD containing both short- and long-memory mutations. As a consequence, predicting a mutation's memory half-life by the median of all mutations at that site achieves r^2 of only 0.25 (Fig. 6B).

Another possibility is that certain types of mutations (to and from the same pair of states) might be consistently associated with memory length, irrespective of the sites at which they occur. But predicting the memory length of individual mutations from the median memory length of all mutations of the same type at any site achieved r^2 of only 0.13 (Fig. 6C). Explaining memory length variation therefore requires analysis of each particular mutation at each site in the protein.

Estimating a mutation's memory length requires experiments in multiple genetic backgrounds across a phylogenetic trajectory. But how many backgrounds are necessary? When the rate of epistatic change of mutations is estimated from 2 backgrounds randomly chosen from the 9 we assayed, the correlation with the rate measured using all 9 backgrounds is on average $r^2 = 0.40$, and the rate of epistatic change is systematically underestimated (Fig. 6, D and E). The correlation improves as more backgrounds are sampled and reaches 0.8 when estimates are based on 5 backgrounds. A moderate number of experiments is therefore sufficient to provide a rough estimate of a mutation's rate of epistatic change and hence its memory length.

Robustness to uncertainty in ancestral sequence reconstruction

Our conclusions are robust to uncertainties in the reconstruction of ancestral sequences. The ancestral DBDs were generally inferred with high confidence and contained zero to 10 sites at which more than one amino acid state is plausible. For each ancestral DBD, we generated an "Alt-All" reconstruction, which contains the alternative plausible amino acid at all ambiguously reconstructed sites (28); this sequence represents the least likely of all plausible reconstructions and allows a conservative estimate of robustness to sequence uncertainty. We then constructed a complete DMS library of each Alt-All protein and repeated all of our experiments and analyses. Although the effects of some mutations differ between the Alt-All and maximum-posterior-probability reconstructions, all conclusions concerning the temporal dynamics of epistasis were unchanged (fig. S10).

Quantifying evolutionary unpredictability

Prior experimental studies have identified cases in which the functional effects of a few mutations changed dramatically during particular intervals of evolutionary history (12, 15–18, 29, 30). Our observations show that such rare, large-effect epistatic modifications occur against a background of pervasive gradual drift in the effects of the majority of mutations (11, 20, 31–36). Most epistatic changes across DBD history were of small magnitude when they occurred, but across an evolutionary trajectory of moderate length (<50% sequence divergence), they were sufficient to completely or partially decorrelate the effects of the majority of mutations from their initial effects and dramatically alter the set of available opportunities for future sequence change. Because the fold and function of all proteins depend on interactions among many residues, we expect that epistatic drift will be a widespread feature of protein evolution, but the temporal dynamics and distribution of memory lengths may depend on each protein's structural architecture, function, and the selective regime under which it evolved.

Our findings establish strong limits on the ability to predict future evolution and interpret evolutionary history, but they also provide a quantitative framework for understanding those limits. Classical evolutionary theories assume that the constraints imposed by purifying selection do not change as sequences diverge, so the effects and evolutionary fate of mutations can be predicted or retroactively inferred based on their effects measured in the present. Our results show that this assumption of constancy and independence is wrong for about half of DBD mutations, which have short or medium memories. Because epistatic modification occurs at a mostly constant rate for each mutation, however, an estimate of memory length from experimental data across phylogenetic time can quantify the extent to which any mutation's effect can be predicted at any point in time, either future or past. Further, although point projections of the effects of short- and medium-memory mutations across long timescales are unreliable, a probability distribution of those effects can be generated if we know any mutation's memory length and its effect at some other time. Ancestral protein reconstruction can replace predictions with experimental knowledge, but only for proteins in the past.

A probabilistic description of contingency and uncertainty using memory length does not require detailed knowledge of the particular genetic interactions that cause epistatic change. If we had microscopic knowledge of all the interactions that modify each mutation's effect and a dense phylogenetic reconstruction of past trajectories of sequence change, we could reliably predict the effect of every possible mutation in any genetic background. But even complete knowledge like this would not be sufficient to predict future evolutionary trajectories: the accessibility of each future mutation depends on the chain of epistatic substitutions that occur before it, many of which will occur by chance. We can use experimental and phylogenetic data to tame evolutionary uncertainty by recognizing and quantifying it, but the future will always surprise.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank the members of the Thornton laboratory for discussion and comments, R. Ranganathan and R. Kanan for assistance with Illumina sequencing, the University of Chicago Cytometry and Antibody Technology Facility (RRID: SCR_017760) for assistance with FACS, and the University of Chicago Research Computing Center for high-performance computing.

Funding:

National Institutes of Health grant R01GM131128 (JWT)

National Institutes of Health grant R01GM121931 (JWT)

National Institutes of Health grant R01GM139007 (JWT)

Samsung Scholarship (YP)

National Institutes of Health grant F32GM122251 (BPHM).

Data and materials availability:

All data are available in the main text or the supplementary materials. Computer code is available on GitHub (https://github.com/JoeThorntonLab).

References and Notes

1. Starr TN, Thornton JW, Epistasis in protein evolution. Protein science 25, 1204–1218 (2016). [PubMed: 26833806]
2. Blount ZD, Lenski RE, Losos JB, Contingency and determinism in evolution: Replaying life's tape. Science 362, eaam5979 (2018). [PubMed: 30409860]
3. Melamed D, Young DL, Gamble CE, Miller CR, Fields S, Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. RNA 19, 1537–1551 (2013). [PubMed: 24064791]
4. Olson CA, Wu NC, Sun R, A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. Curr Biol 24, 2643–2651 (2014). [PubMed: 25455030]
5. Podgornaia AI, Laub MT, Pervasive degeneracy and epistasis in a protein-protein interface. Science 347, 673–677 (2015). [PubMed: 25657251]
6. Sarkisyan KS et al. , Local fitness landscape of the green fluorescent protein. Nature 533, 397–401 (2016). [PubMed: 27193686]
7. Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R, Adaptation in protein fitness landscapes is facilitated by indirect paths. Elife 5, e16965 (2016). [PubMed: 27391790]
8. Diss G, Lehner B, The genetic landscape of a physical interaction. Elife 7, e32472 (2018). [PubMed: 29638215]
9. Pokusaeva VO et al. , An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. PLoS Genet 15, e1008079 (2019). [PubMed: 30969963]
10. Kondrashov AS, Sunyaev S, Kondrashov FA, Dobzhansky–Muller incompatibilities in protein evolution. Proceedings of the National Academy of Sciences 99, 14878–14883 (2002).
11. Lunzer M, Golding GB, Dean AM, Pervasive cryptic epistasis in molecular evolution. PLoS Genet 6, e1001162 (2010). [PubMed: 20975933]
12. Natarajan C et al. , Epistasis among adaptive mutations in deer mouse hemoglobin. Science 340, 1324–1327 (2013). [PubMed: 23766324]
13. Doud MB, Ashenberg O, Bloom JD, Site-Specific Amino Acid Preferences Are Mostly Conserved in Two Closely Related Protein Homologs. Mol Biol Evol 32, 2944–2960 (2015). [PubMed: 26226986]

14. Haddox HK, Dingens AS, Hilton SK, Overbaugh J, Bloom JD, Mapping mutational effects along the evolutionary landscape of HIV envelope. Elife 7, e34420 (2018). [PubMed: 29590010]
15. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW, Crystal structure of an ancient protein: evolution by conformational epistasis. Science 317, 1544–1548 (2007). [PubMed: 17702911]
16. Bridgham JT, Ortlund EA, Thornton JW, An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature 461, 515–519 (2009). [PubMed: 19779450]
17. Bloom JD, Gong LI, Baltimore D, Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. Science 328, 1272–1275 (2010). [PubMed: 20522774]
18. Gong LI, Suchard MA, Bloom JD, Stability-mediated epistasis constrains the evolution of an influenza protein. Elife 2, e00631 (2013). [PubMed: 23682315]
19. Natarajan C et al. , Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. Science 354, 336–339 (2016). [PubMed: 27846568]
20. Starr TN, Flynn JM, Mishra P, Bolon DNA, Thornton JW, Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. Proceedings of the National Academy of Sciences 115, 4453–4458 (2018).
21. Karageorgi M et al. , Genome editing retraces the evolution of toxin resistance in the monarch butterfly. Nature 574, 409–412 (2019). [PubMed: 31578524]
22. Xie VC, Pu J, Metzger BPH, Thornton JW, Dickinson BC, Contingency and chance erase necessity in the experimental evolution of ancestral proteins. Elife 10, e67336 (2021). [PubMed: 34061027]
23. Whitfield GK, Jurutka PW, Haussler CA, Haussler MR, Steroid hormone receptors: evolution, ligands, and molecular basis of biologic function. Journal of cellular biochemistry 75, 110–122 (1999).
24. Helsen C et al. , Structural basis for nuclear hormone receptor DNA binding. Molecular and cellular endocrinology 348, 411–417 (2012). [PubMed: 21801809]
25. McKeown AN et al. , Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. Cell 159, 58–68 (2014). [PubMed: 25259920]
26. Harvey PH, Pagel MD, The comparative method in evolutionary biology (Oxford university press Oxford, 1991).
27. Harmon L, Phylogenetic comparative methods: learning from trees. (2018).
28. Eick GN, Bridgham JT, Anderson DP, Harms MJ, Thornton JW, Robustness of reconstructed ancestral protein functions to statistical uncertainty. Molecular biology and evolution 34, 247–261 (2017). [PubMed: 27795231]
29. Blount ZD, Borland CZ, Lenski RE, Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. Proceedings of the National Academy of Sciences 105, 7899–7906 (2008).
30. Harms MJ, Thornton JW, Historical contingency and its biophysical basis in glucocorticoid receptor evolution. Nature 512, 203–207 (2014). [PubMed: 24930765]
31. Povolotskaya IS, Kondrashov FA, Sequence space and the ongoing expansion of the protein universe. Nature 465, 922–926 (2010). [PubMed: 20485343]
32. Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA, Epistasis as the primary factor in molecular evolution. Nature 490, 535–538 (2012). [PubMed: 23064225]
33. Pollock DD, Thiltgen G, Goldstein RA, Amino acid coevolution induces an evolutionary Stokes shift. Proceedings of the National Academy of Sciences 109, E1352–E1359 (2012).
34. Shah P, McCandlish DM, Plotkin JB, Contingency and entrenchment in protein evolution under purifying selection. Proceedings of the National Academy of Sciences 112, E3226–E3235 (2015).
35. Sailer ZR, Harms MJ, Molecular ensembles make evolution unpredictable. Proceedings of the National Academy of Sciences 114, 11938–11943 (2017).
36. Morrison AJ, Wonderlick DR, Harms MJ, Ensemble epistasis: thermodynamic origins of nonadditivity between mutations. Genetics 219, iyab105 (2021). [PubMed: 34849909]
37. Bridgham JT et al. , Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. PLoS biology 8, e1000497 (2010). [PubMed: 20957188]

38. Hochberg GKA et al. , A hydrophobic ratchet entrenches molecular complexes. Nature 588, 503–508 (2020). [PubMed: 33299178]
39. Zerbino DR, Birney E, Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research 18, 821–829 (2008). [PubMed: 18349386]
40. Schulz MH, Zerbino DR, Vingron M, Birney E, Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28, 1086–1092 (2012). [PubMed: 22368243]
41. Haas BJ et al. , De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols 8, 1494–1512 (2013). [PubMed: 23845962]
42. Edgar RC, MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32, 1792–1797 (2004). [PubMed: 15034147]
43. Stamatakis A, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014). [PubMed: 24451623]
44. Weigert A et al. , Illuminating the base of the annelid tree using transcriptomics. Molecular biology and evolution 31, 1391–1401 (2014). [PubMed: 24567512]
45. Yang Z, PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution 24, 1586–1591 (2007). [PubMed: 17483113]
46. Zilliacus J, Wright AP, Norinder U, Gustafsson J-A, Carlstedt-Duke J, Determinants for DNA-binding site recognition by the glucocorticoid receptor. Journal of Biological Chemistry 267, 24941–24947 (1992). [PubMed: 1459998]
47. Starr TN, Picton LK, Thornton JW, Alternative evolutionary histories in the sequence space of an ancient protein. Nature 549, 409–413 (2017). [PubMed: 28902834]
48. Maclean CJ et al. , Deciphering the genic basis of yeast fitness variation by simultaneous forward and reverse genetics. Molecular biology and evolution 34, 2486–2502 (2017). [PubMed: 28472365]
49. Anderson DW, McKeown AN, Thornton JW, Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. Elife 4, e07864 (2015). [PubMed: 26076233]
50. Fowler DM, Stephany JJ, Fields S, Measuring the activity of protein variants on a large scale using deep mutational scanning. Nature protocols 9, 2267–2284 (2014). [PubMed: 25167058]
51. Wang W, Yan J, Shape-Restricted Regression Splines with R Package splines2. Journal of Data Science 19, (2021).
52. Otwinowski J, McCandlish DM, Plotkin JB, Inferring the shape of global epistasis. Proceedings of the National Academy of Sciences 115, E7550–E7558 (2018).
53. Kabsch W, Sander C, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers: Original Research on Biomolecules 22, 2577–2637 (1983).
54. Guindon S et al. , New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology 59, 307–321 (2010). [PubMed: 20525638]

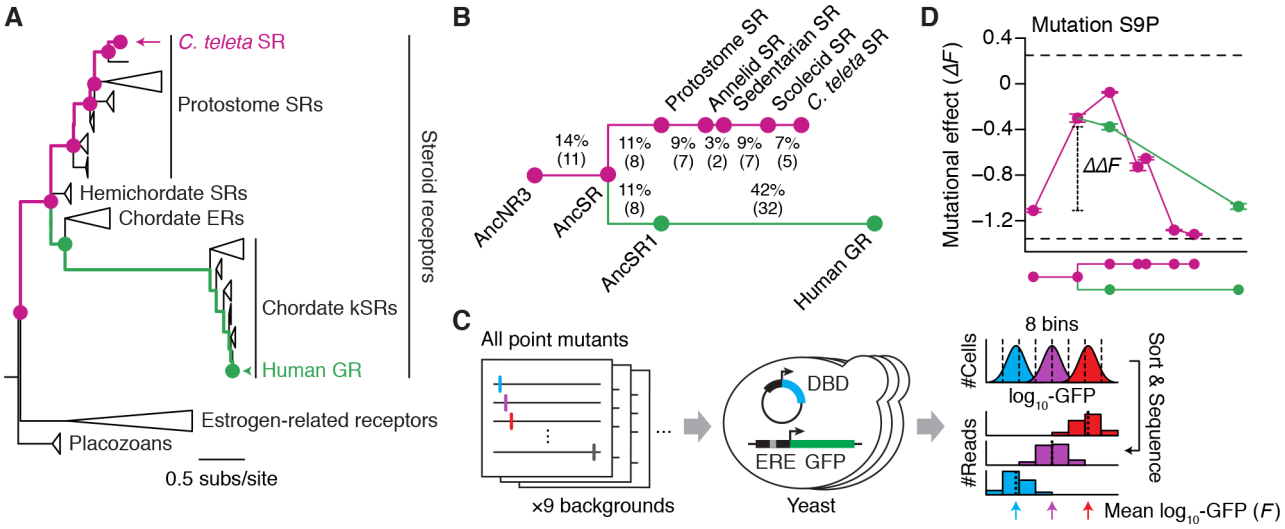

Fig. 1. Phylogenetic deep mutational scanning.

(**A**) Phylogeny of the DNA-binding domain (DBD) of steroid and related receptors. Circles, DBDs characterized here by deep mutational scanning. SRs, steroid receptors; ERs, estrogen receptors; kSRs, ketosteroid receptors—including glucocorticoid receptor (GR). Complete phylogeny in fig. S1. (**B**) Phylogenetic relations among the 9 characterized DBDs. Colors distinguish trajectories to *C. teleta* SR and human GR. Sequence divergence (percent) and number of sequence differences (parentheses) in each interval are shown. (**C**) Sort-seq assay for DBD activity. For each DBD, a library containing all possible single-amino acid mutations was generated using microarray-based synthesis and cassette assembly (fig. S3) and cloned into yeast carrying a GFP reporter; ERE, estrogen response element. Activity of each mutant was measured by sorting the library of cells into fluorescence bins, inferring the distribution of each mutant among bins by sequencing, and calculating the mean $\log_{10}$-GFP fluorescence (F). Hypothetical distributions for 3 variants with high, medium, and low F are shown. (**D**) Tracing epistatic change in mutational effect across the phylogeny using example mutation S9P. The effect on each DBD's activity (points) was quantified as the change in mean $\log_{10}$-GFP fluorescence (ΔF). Horizontal axis, each DBD in order on the phylogeny, positioned by sequence divergence and colored by trajectory. $\Delta\Delta F$, change in the mutation's effect between a pair of DBDs, caused by epistatic interactions with intervening substitutions. Error bars, SEM ($n = 3$). Dashed lines, upper and lower measurement bounds.

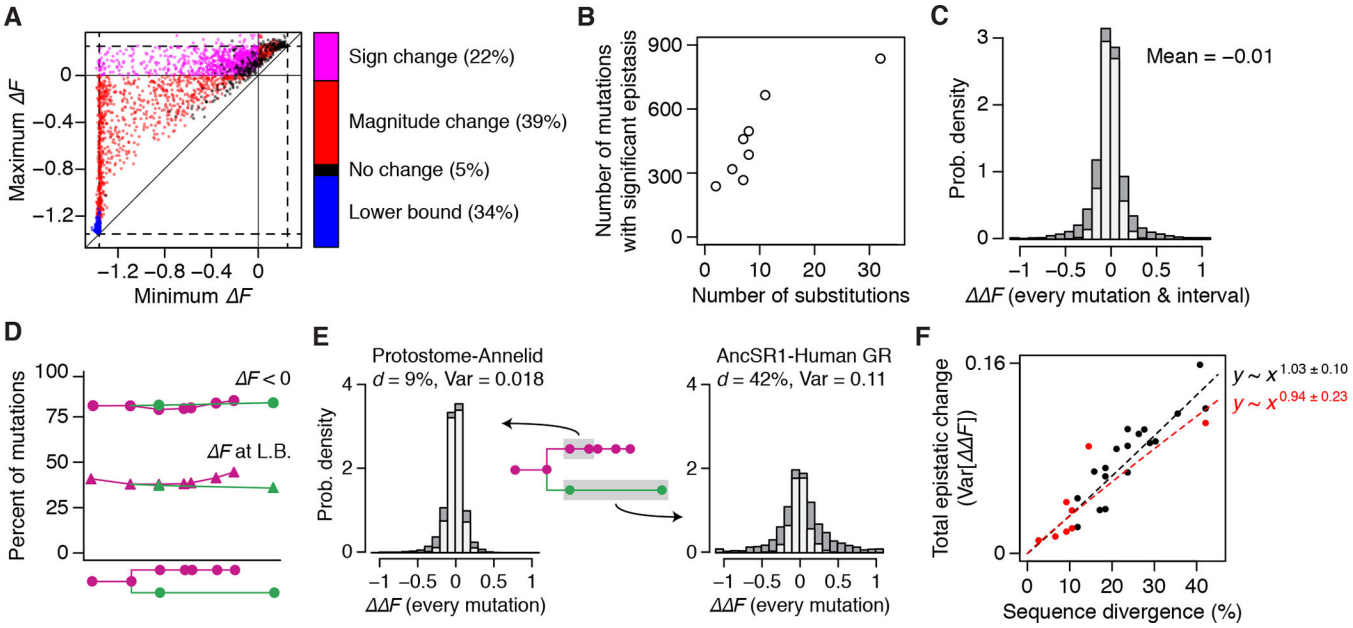

Fig. 2. Pervasive random changes in the effects of mutations.

(**A**) Maximum and minimum effect of each mutation (points) across the 9 DBDs, colored according to the stacked column at right, which shows the proportion of mutations in four categories: pink, significant effect of DBD background on Δ*F* and the sign of Δ*F* different between the maximum and minimum; red, significant effect of background but no sign difference; black, no significant effect of background and Δ*F* within measurement limits; blue, Δ*F* at the lower bound of measurement in all 9 DBDs. Significance was evaluated by Welch's ANOVA, Benjamini-Hochberg FDR ≤ 0.1. (**B**) Number of mutations in each phylogenetic interval that changed significantly in Δ*F* (*t*-test between parent and child node, FDR ≤ 0.1), plotted versus the number of amino acids that diverged in the interval. (**C**) Distribution of epistatic change in the effect of every mutation during every phylogenetic interval (ΔΔ*F*). Dark grey, ΔΔ*F* significantly different from 0. Mutations always at the lower bound of measurement were excluded. (**D**) Fraction of mutations in each DBD with Δ*F* < 0 (circles) or Δ*F* at the lower bound of measurement (triangles). (**E**) Distribution of ΔΔ*F* of all mutations for the protostome-annelid interval or the AncSR1-human GR interval. The variance of the distribution (Var) quantifies the total epistatic change in the effects of all mutations during an interval. *d*, sequence divergence. (**F**) Total epistatic change as a function of sequence divergence across the phylogeny. Red dots, each of the 8 independent phylogenetic intervals between characterized DBDs; black, all composite intervals. Dashed lines, best-fit power function for all (black) or the 8 independent intervals (red).

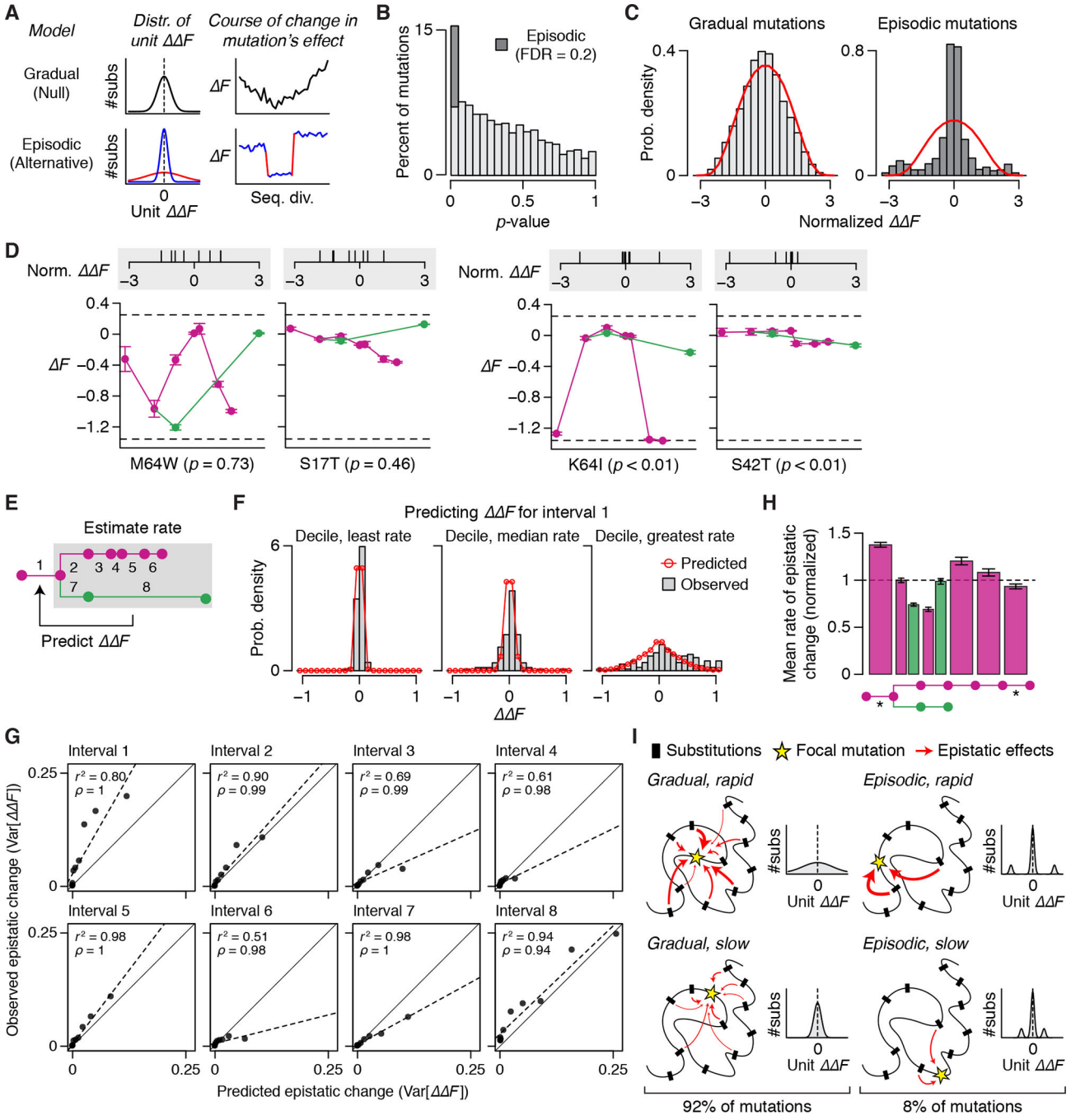

Fig. 3. Effects of most mutations changed gradually at characteristic rates.

(**A**) Models of the tempo of epistatic change. Null model, the amount of change in a mutation's effect per substitution in an interval (unit $\Delta\Delta F$) is randomly drawn from a normal distribution centered at 0; the variance is the same among intervals, so the mutation's effect changes gradually at a constant expected rate as substitutions accrue. Alternative model, the variance may differ among phylogenetic intervals (blue vs. red), leading to episodic changes in a mutation's effect. (**B**) Distribution of the *p*-value of the likelihood-ratio test (LRT) comparing gradual and episodic models for each mutation. Darker grey, mutations for

which the gradual model is rejected (FDR ≤ 0.2). Mutations always at the lower bound of measurement were excluded from this analysis. (**C**) Distribution of the normalized amount of epistatic change in each interval, for all mutations better fit by the gradual model (*left*) or the episodic model (*right*). Normalized $\Delta\Delta F$, $\Delta\Delta F$ of a mutation in an interval divided by $\sigma d^{1/2}$, where σ is that mutation's average rate of epistatic change and d is the length of the interval. Gray columns, observed data; red line, distribution expected under the null model. (**D**) Trajectory of changes in the effect of two example mutations that are better fit by the gradual model (*left*) or episodic model (*right*); in each category, one evolves rapidly and the other slowly. Each mutation's p-value in the LRT is shown; gray box, normalized changes in the mutation's effect across each of the 8 intervals. (**E**) Phylogenetic cross-validation. In the example shown, $\Delta\Delta F$ in interval 1 is predicted from the average rate of epistatic change measured across intervals 2–8 (grey box). (**F**) Distribution of observed $\Delta\Delta F$ during interval 1 (gray columns) and predicted by cross validation (red line). Mutations were grouped into deciles by their rate of epistatic change across intervals 2–8; predictions are shown for deciles with the slowest, median, or fastest rates. (**G**) Mutations' relative rates of epistatic change are consistent across phylogenetic intervals. Points, deciles of mutations grouped by the predicted rate of epistatic change; observed epistatic change in an plotted against that predicted by cross-validation. r, Pearson's correlation coefficient; ρ, Spearman's rank correlation; dashed line, linear regression. (**H**) Among-interval differences in average rate of epistatic change. Each column shows the mean rate of epistatic change of all mutations in one phylogenetic interval, normalized so that the mean across all intervals equals 1. Error bars, estimated standard deviation obtained by bootstrap-resampling of mutations. Asterisks, intervals immediately following gene duplication. **(I)** Inferring the architecture of epistatic interactions between substitutions (black boxes) and a focal mutation (star) from phylogenetic DMS. *Left,* gradual changes in the mutation's effect during evolution arise if many substitutions act as epistatic modifiers (arrows, with thickness showing the strength of interaction), yielding a normal distribution of $\Delta\Delta F$ per substitution. *Right,* episodic changes arise from interactions with only a few substitutions, yielding a distribution heavy at zero and the tails. In either case, strong vs. weak interactions cause rapid *(top)* vs. slow (*bottom*) epistatic change. The fraction of all mutations in each category in our experiments is shown.

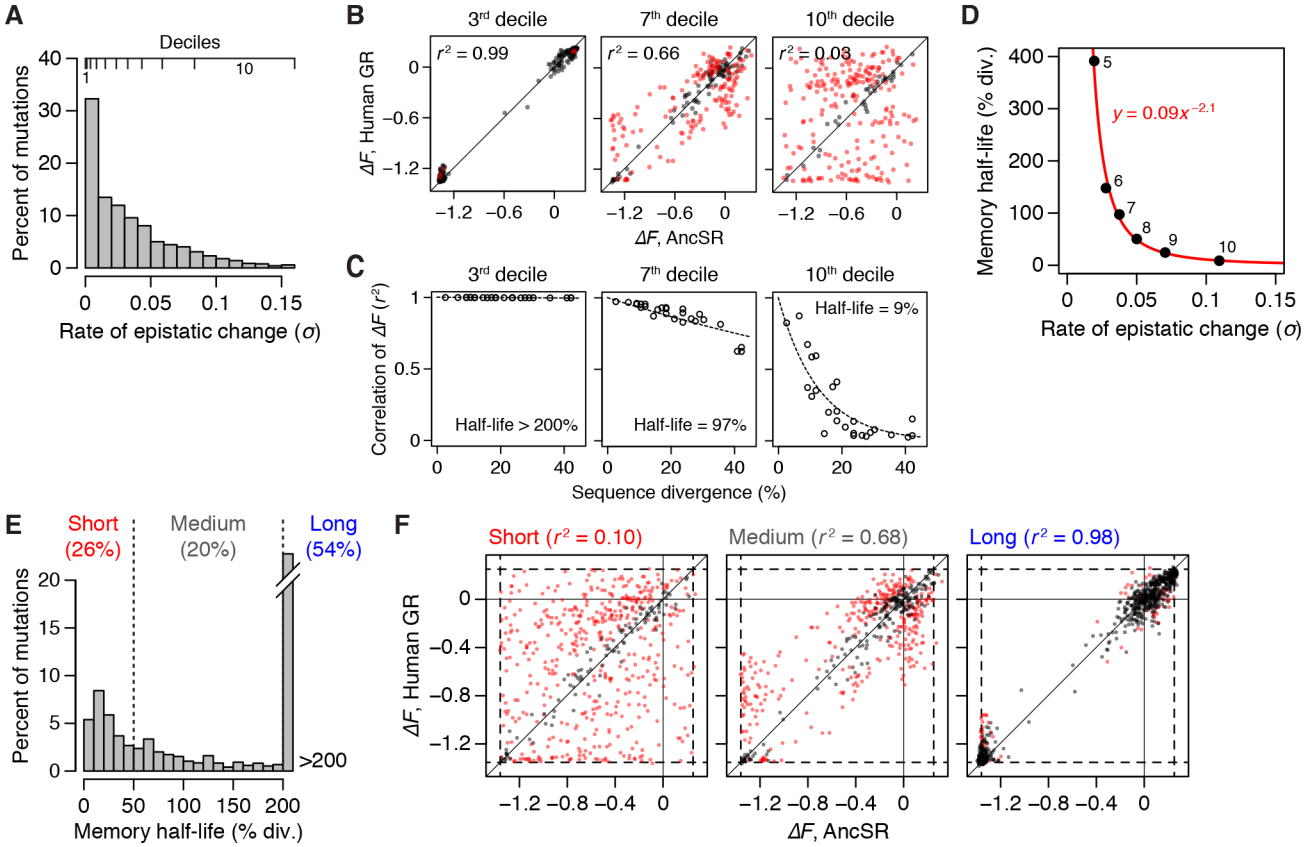

Fig. 4. Memory length of mutations and the timescale of historical contingency.

(**A-D**) Measuring the memory length of mutations. (**A**) Mutations were grouped into deciles by their rate of epistatic change (σ, expected standard deviation of $\Delta\Delta F$ per 1% sequence divergence). (**B**) The effects of mutations in each decile were compared between every pair of DBDs; shown are comparisons between AncSR and human GR (42% divergence). (**C**) The squared Pearson correlation coefficient (r^2) for each DBD pair was plotted against the sequence divergence of that pair. Dotted line, best-fit exponential decay curve; memory half-life, sequence divergence at which $r^2 = 0.5$. (**D**) Relationship between the rate of epistatic change and memory half-life inferred by fitting a power function (red) to the mean rate of epistatic change and memory half-life of the deciles. This relationship was used to calculate the memory half-life of each mutation from its rate of epistatic change. **E**) Distribution of memory half-life among mutations. Mutations were classified into short-, medium-, and long-memory categories using cutoffs of 50% and 200% divergence. (**F**) Comparing the effects of mutations between AncSR and human GR (42% divergence) for each memory category. Red dots, mutations with significant difference in ΔF (t-test, FDR ≤ 0.1); black, no significant difference.

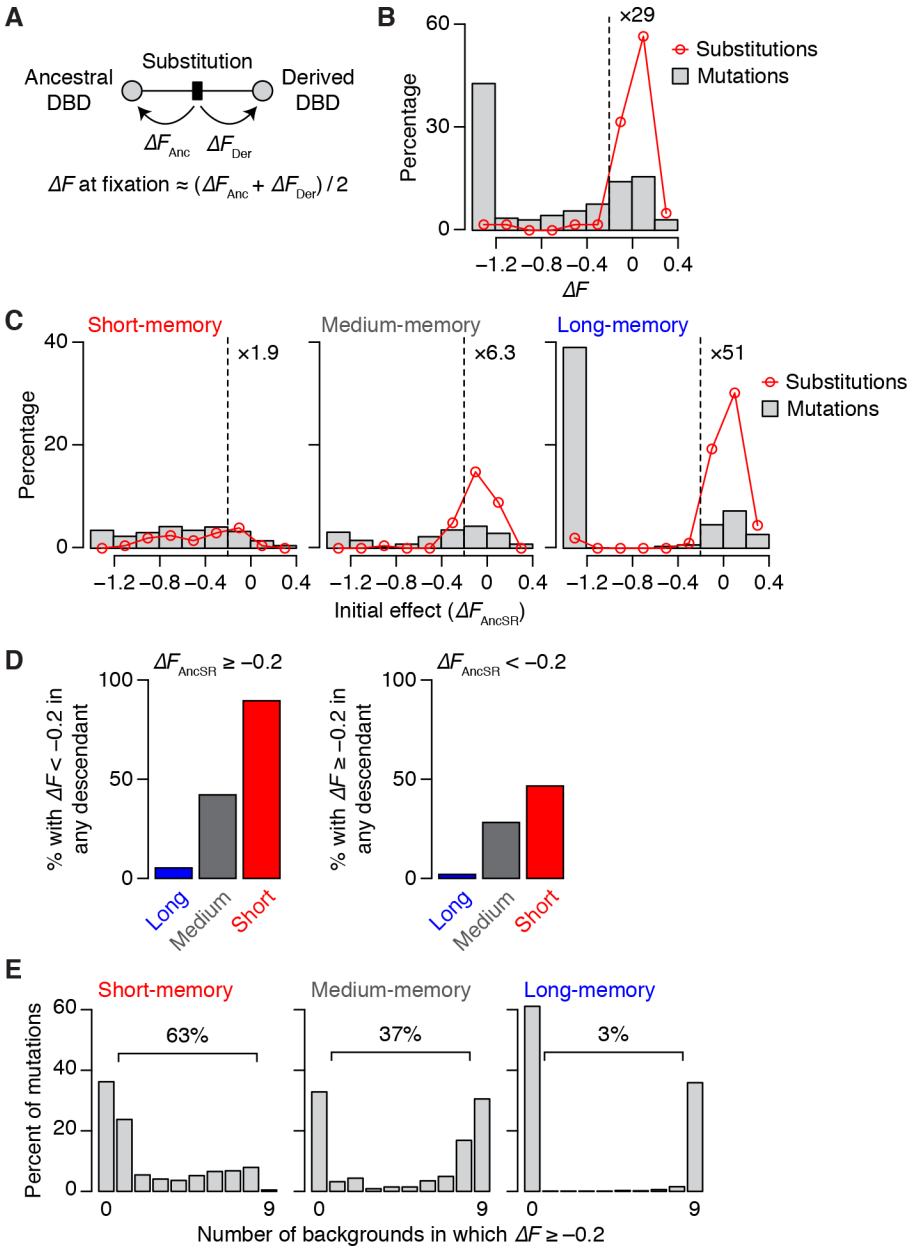

Fig. 5. Impact on sequence evolution of memory length and initial functional effect.

(**A**) The effect of a substitution at the time it fixed during history was calculated as the mean of Δ*F*s measured by DMS in the nearest ancestral and descendant nodes. (**B**) Comparing the effects of the 79 substitutions that occurred along the phylogenetic trajectories we characterized to the effects of all possible mutations. Substitutions are 29-fold enriched for $\Delta F \geq -0.2$ compared to mutations, providing an estimate of the threshold of accessibility during DBD evolution. (**C**) Distribution of the initial effect (Δ*F* on AncSR) of 275 substitutions that fixed between AncSR and any extant DBD in our phylogeny. Distributions are shown by memory half-life category. Enrichment of substitutions with $\Delta F \geq -0.2$ relative to mutations is shown. (**D**) *Left,* proportion of initially accessible mutations ($\Delta F_{\text{AncSR}} \geq -0.2$) that become inaccessible in at least one descendant DBD. *Right,* proportion of

initially inaccessible mutations that become accessible in at least one descendant DBD. (**E**) Distribution of the number of characterized DBDs in which each mutation is accessible ($\Delta F \geq -0.2$), classified by memory-length category. The percentage of mutations that were accessible in some but not all DBDs is shown.

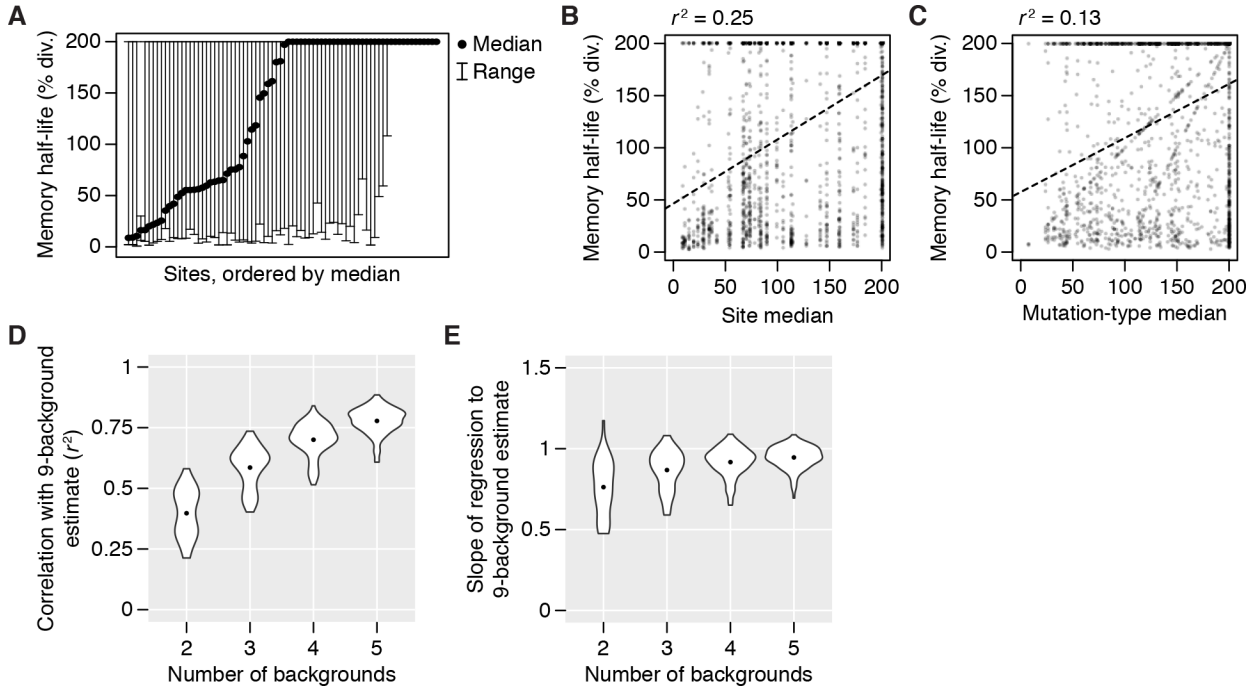

Fig. 6. Variation of memory half-life of mutations among and within sites.

(**A**) Distribution of memory half-life among sites. Each line shows the range of memory half-life of all mutations at one site in the DBD sequence. (**B-C**) Predicting the memory half-life of a mutation (points) by the median memory half-life of all possible mutations at the same site (**B**) or by the median of mutations of the same type (between the same wild type and mutant amino acid) at all sites (**C**). Dashed line, linear regression. (**D-E**) Effect of number of DBDs characterized by DMS on estimates of rate of epistatic change. The rate of epistatic change of every mutation was estimated using a subset of the 9 DMS experiments; the relationship between the estimated rate from each subset to that estimated from all 9 experiments was analyzed by linear regression. The graphs show the distribution of correlation coefficient **(D)** and best-fit regression slope (**E**) across every possible subset of a given size.