Chance, contingency, and necessity in the experimental evolution of ancestral proteins

Victoria Cochran Xie[1,4], Jinyue Pu[1,4,*], Brian P.H. Metzger[2,4], Joseph W. Thornton[2,3,*], Bryan C. Dickinson[1,*]

[1]Department of Chemistry, University of Chicago
[2]Department of Ecology and Evolution, University of Chicago
[3]Department of Human Genetics, University of Chicago
[4]These authors contributed equally: Victoria Cochran Xie, Jinyue Pu, Brian P.H. Metzger

*email: pujy@uchicago.edu; joet1@uchicago.edu, dickinson@uchicago.edu

Keywords: PACE, continuous directed evolution, protein-protein interactions, genetic variance, epistasis, BCL-2, ancestral protein reconstruction

ABSTRACT

The extent to which evolutionary outcomes reflect the unpredictable influences of chance and contingency is a central but unanswered question in evolutionary biology[1–4]. A precise characterization requires evolutionary trajectories to be repeated multiple times under identical environmental conditions from multiple starting points across history, a scenario that rarely, if ever, occurs in nature. Here we combine continuous experimental evolution[5] with ancestral protein reconstruction[6] and manipulative genetic experiments to identify the causes and consequences of chance and contingency in the genetic outcomes of molecular evolution. By repeatedly evolving ancestral proteins in the B-cell lymphoma-2 (BCL-2) family of apoptosis regulators[7] to acquire the same protein-protein interaction specificities that evolved during history, we found that contingency and chance interact to make sequence evolution increasingly unpredictable over phylogenetic timescales. Although replicates from the same starting genotype sometimes share mutations – indicating partial predictability – there are multiple alternative sets of changes that can alter specificity, and chance decides which of these paths is taken. Contingency has a stronger effect: when trajectories are initiated from different starting points, outcomes are even more divergent, because substitutions that occurred during phylogenetic history repeatedly changed the potential of other mutations to confer new binding specificities. The impact of contingency increased steadily with phylogenetic distance and magnified the effects of chance, resulting in a >3-fold increase in genetic variance among evolutionary trajectories initiated from different starting points across the timescale of metazoan evolution. Our findings show how a particular cascade of chance evolutionary steps throughout history makes the outcomes of molecular evolution increasingly idiosyncratic and unpredictable, even under strong selection.

INTRODUCTION

Whether the world is a necessary outcome of fate-like processes has long fascinated philosophers, historians, and scientists[1,8,9]. In biology, if a single optimal state is accessible from all starting points, then natural selection is expected to predictably drive the realization of that state, deterministically and independently of initial conditions. Chance and contingency reduce the necessity of evolutionary outcomes in distinct ways and arise from different aspects of the genotype-phenotype relationship. If multiple adaptive outcomes are accessible from a given genetic starting point, then which one is realized depends on chance; a process free of chance is deterministic and predictably leads to the same outcome again and again[10,11]. If the accessibility of outcomes differs among starting genotypes, then the realized outcome of evolution is contingent on the specific starting point; even in the absence of chance, the course of evolution can be predicted only given full knowledge of the starting genotype and the particular outcomes to which it can lead[12] (Fig. 1a, Fig. S1a).

To systematically characterize how chance and contingency interacted to affect historical protein evolution, we would have to travel back in time, re-launch evolution repeatedly from each of various starting points during history, and compare the outcomes across replicates from the same starting point (for chance) and across starting points (for

contingency). This design cannot be executed directly, but we can come close by reconstructing ancestral proteins as they existed in the deep past and using them as starting points for replicated evolutionary trajectories in the laboratory under selection to evolve the same molecular functions they acquired during history. Numerous studies have provided insights into the effects of chance and contingency on sequence evolution, but their effects have never been systematically evaluated within a single system using historical starting points[13,14]. Experimental studies of the repeatability of sequence evolution have typically replicated trajectories from just one or a few starting points[15–20]– illuminating chance but limiting their assessment of contingency – and they have not used ancestral proteins ordered across time to make their findings historically relevant[21–25]. Mechanistic studies have established that particular historical mutations have different effects when introduced into different ancestral backgrounds, suggesting contingency, but history happened only once, so these studies do not elucidate the role of chance[26–29]. Cases of natural phenotypic convergence, usually among closely related populations or species, sometimes involve the same gene, and occasionally the same mutations, suggesting a degree of necessity, but these studies do not involve replicate lineages from multiple starting points, so chance and contingency cannot be disentangled[13,14].
We performed multi-replicate experimental evolution using reconstructed ancestral proteins in the B-cell lymphoma-2 (BCL-2) family of proteins, which promote or inhibit apoptosis in eukaryotes[30,31]. The functions of BCL-2 family proteins are regulated by specific protein-protein interactions (PPIs) with coregulators[32,33], the structural and biochemical basis of which is well understood[7,34]. BCL-2 family proteins are distributed throughout Metazoa[35,36], with the anti-apoptotic proteins divided into two major classes: the Myeloid Cell Leukemia Sequence 1 (MCL-1) group, which bind both the BID and NOXA BH3-only proteins, and the BCL-2 group, which bind BID but not NOXA (Fig. 1b)[37]. The two classes are structurally similar and use the same cleft to interact with the different BH3-only proteins (Fig. 1c, Fig. S1b), and while studies have identified determinants of BH3-only protein binding[38], little is known about how the BCL-2 and MCL-1 proteins evolved their distinct binding profiles during history.

RESULTS

Historical evolution of BCL-2 family proteins

We first characterized the historical evolution of PPI specificity in the BCL-2 family of proteins using ancestral protein reconstruction. Using maximum likelihood phylogenetics, we recovered the expected sister relationship between the metazoan BCL-2-class and the MCL-1-class sequences (Fig. S2a). We inferred the maximum a posteriori sequence of multiple ancestral proteins along the lineages extending from the human BCL-2 (hsBCL-2) and human MCL-1 (hsMCL-1) proteins back to their most recent common ancestor (AncMB1), a gene duplication that occurred before the last common ancestor of all animals (Fig. S2b). We then used our previously-developed proximity-dependent split RNA polymerase (RNAP) luciferase assay to measure BID and NOXA binding (Fig. 1d, e)[39] and found that AncMB1 bound both BID and NOXA, just as hsMCL-1 does (Fig. 1f, Fig. S3a). NOXA binding specificity in the BCL-2 class therefore evolved when this ancestral binding to NOXA was lost, an event that occurred between AncB2 (the BCL-2 class progenitor in the ancestral eumetazoan) and AncB4 (in the

ancestral deuterostome). This history is consistent with the distribution of protein binding among extant family members (Fig. 1f, Fig. S3b). It is also robust to uncertainty in the ancestral sequence reconstruction, because experiments on "AltAll" proteins at each ancestral node, which contains all plausible alternative amino acid states (PP>0.2) in a single "worst-case" alternative reconstruction, also showed that BID specificity arose between AncB2 and AncB4 (Fig. S3c, d).

Repeated experimental evolution of protein binding specificity

We next developed an experimental evolution system to repeatedly recapitulate or reverse historic changes in PPI specificity among BCL-2 family proteins using phage-assisted continuous evolution (PACE)[5] (Fig. S4a, b). To select for the evolution of specificity, we used two orthogonal proximity-dependent split RNAPs, allowing us to simultaneously select for binding of a target protein and against binding of a counterselection protein (Fig. 2a). For positive selection, BCL-2 family proteins were fused with the N-terminal portion of a split, proximity-dependent T7 RNAP and placed in a bacteriophage[39], and the target binding peptide fragment was fused to the RNAP C-terminal portion and placed into an *E.coli* host; binding by the BCL-2 family protein reconstitutes the RNAP and drives expression of the viral gIII protein, producing infectious phage. To select against an undesired interaction, a second C-terminal RNAP that specifically binds a different promoter was fused to the counter selection peptide and expressed in *E. coli*; binding this peptide reconstitutes the RNAP and activates expression of a dominant negative gIII gene, impairing the phage's capacity to infect[40]. We optimized this system using activity-dependent plaque assays and phage growth assays to recapitulate the natural binding profiles of hsBCL-2 and hsMCL-1 (Fig. S4c, d).

Characterizing chance, contingency, and necessity

To characterize the extent of chance, contingency, and necessity on the outcomes of evolution, we used our new directed evolution system to drive extant and reconstructed ancestral proteins to repeatedly recapitulate or reverse the historical loss of NOXA binding. Three proteins with the ancestral phenotype—hsMCL-1, AncM6, and AncB1—were evolved to acquire the derived BCL-2 phenotype, losing NOXA binding but retaining BID binding (Fig. S5a). Conversely, hsBCL-2, AncB5, and AncB4 were evolved to revert to the ancestral phenotype, gaining NOXA binding (Fig. S5b). For each starting genotype, we performed three to four replicate experimental evolution trajectories (Supplementary Table 1). From the evolved populations, we confirmed that the selected PPI specificity had been acquired using activity-dependent plaque assays as well as *in vivo* and *in vitro* binding assays of randomly isolated variants (Fig. 2b, S5c,d, S6, S7, Supplementary Table 1). All trajectories produced the target phenotype and high-throughput sequencing of the phage populations revealed the amino acid allele frequencies at each site in each replicate experiment (Fig. 2c, d, Fig. S5e, f, Supplementary Table 13).

We found very limited necessity in the sequence outcomes. Across all evolutionary trajectories in both selection regimes, 100 mutant amino acid states (at 75 different sites) evolved to frequency >5% in at least one replicate (Fig. 2e, Fig. S8a). Of these, 27 appeared in more than one replicate, which is significantly greater than expected by chance alone ($P<10^{-5}$ by permutation test, given the same number of mutations and replicates). Most of these repeated

mutations were observed in replicate trajectories from the same starting genotype, indicating partial determinism in evolutionary outcomes from a given starting genotype. Only 4 mutations were observed in more than one replicate from distinct starting genotypes, however, indicating considerable contingency, and only a single mutation was observed in every replicate from more than one starting point, which was not sufficient to confer the desired phenotype (Fig. S7g). When mapped onto the protein structure, all repeatedly-mutated sites either directly contact the bound peptide or are on secondary structural elements that do so (Fig. S7f, g), suggesting that determinism, to the extent that it exists, reflects a limited number of structural mechanisms by which PPIs can be altered.

To quantify the effects of chance and contingency on the genetic outcomes of evolution, we compared the genetic variance within and between replicates started from the same and different genotypes. We found that the variance among replicates that evolved from the same initial genotype was on average 1.3-fold greater than the within-replicate variance, indicating that chance causes evolution to produce substantially different genetic outcomes (Fig. 3a). Contingency had an even larger effect, increasing the variance among trajectories that evolved from different starting genotypes by a further 1.8-fold beyond that among replicates from the same starting point. Together, chance and contingency increased genetic variance between replicates from different starting points by a factor of 2.4-fold (averaged across all starting points) compared to the within-replicate genetic variance.

To determine the evolutionary tempo and mode by which chance and contingency exert their effects, we evaluated how the effects of chance and contingency changed as the phylogenetic distance between starting genotypes increased. Across all pairs of starting genotypes, the combined effects of chance and contingency increased significantly with phylogenetic distance (slope = 0.208, $P = 1\text{x}10^{-4}$) (Fig. 3b, Fig. S8b-e); this relationship remained statistically significant after accounting for phylogenetic non-independence by comparing only successive ancestors along phylogenetic lineages ($P = 0.03$). Across the metazoan evolutionary timescale of our experiments, contingency and chance together increased genetic variance more than three-fold. This relationship was driven almost entirely by an increase in the effects of contingency with phylogenetic distance (slope = 0.128, $P = 5\text{x}10^{-4}$), with the effects of chance displaying a very weak increase (slope=0.019, p=0.002). The combined effect of chance and contingency increased faster with phylogenetic distance than contingency alone did. Contingency and chance therefore magnify each other's effects, increasing the unpredictability, and reducing necessity, in evolutionary outcomes as sequences diverge through history.

Sources of contingency

Contingency arises when sequence outcomes differ between genetic starting points; this conditionality arises when epistasis exists – that is, when the effects of mutations depend on the sequence background into which they are introduced. Epistasis could produce contingency if mutations that confer selected phenotypes in some backgrounds fail to confer that phenotype in others or are not tolerated at all; alternatively, epistasis might merely change the probability that selection favors one mutation over another by altering the relative magnitude of their effect on the phenotype. To experimentally distinguish between these possibilities, we transferred sets of mutations that arose repeatedly during experimental evolution into other starting genotypes and then measured their effects on BID and NOXA binding (Fig. 3c, d, Fig.

S9). Eleven of the 12 "swaps" failed to confer the PPI specificity on other genetic backgrounds that they did on their evolved genetic background. Of the six swaps of mutation sets that arose when MCL-1-like proteins were evolved to lose NOXA binding and preserve BID binding, five compromised BID binding and three had no effect on NOXA binding when introduced into other backgrounds. Of the six swaps of mutation sets that caused BCL-2-like proteins to gain NOXA binding, four failed to gain any detectable NOXA binding and one compromised BID binding in other backgrounds. The only case in which mutations that conferred the target phenotype during experimental evolution had the same effect in another background was the swap of mutations evolved on AncB5 into AncB4, and these two genotypes are more similar to each other than any other pair. Thus, contingency arose because of strong epistatic interactions between the mutations that confer new specificities in our experiments and the historical substitutions that occurred during the intervals between ancestral proteins; these substitutions transiently opened and blocked routes to adaptive phenotypes by making specificity-changing mutations subsequently—and also previously—deleterious or invisible to selection.

To identify when these epistatic effects arose, we mapped these incompatibilities onto the phylogeny. We identified historical intervals in which PACE-derived mutations altered binding on an ancestral (permissive and potentiating mutations) or descendent (restrictive and depotentiating) background for either BID binding (permissive or restrictive) or NOXA binding (potentiating or depotentiating) (Fig. 3e). We found that all four types of epistatic effects occurred, with multiple types of contingency-inducing substitutions present on most branches. The only exception—the branch from AncB4 to AncB5, on which only depotentiating substitutions occurred—is the branch immediately after NOXA function changed during history, the shortest of the branches examined, and the one with the smallest effect of contingency on genetic variance (Fig. 3b). Even across this branch, the PACE mutations that restore the ancestral PPI specificity in AncB4 can no longer do so in AncB5. These results indicate that the paths through sequence space that allow new NOXA functions to evolve repeatedly changed during the BCL-2 family's history, even during intervals when the proteins' PPI binding profiles did not change.

Consistent with this conclusion, the mutations that altered NOXA binding during PACE almost never recapitulated or reversed substitutions that occurred during the historical interval in which NOXA binding changed (Fig. S8f, S10, S11). The sole exceptions, f160L and y259F, arose when AncB4 was evolved to regain the ancestral NOXA binding and are reversals to ancestral amino acids observed in AncB2, directly prior to the historic loss of NOXA binding. No PACE trajectories for the loss of NOXA binding recapitulated substitutions that occurred during this interval. These observations suggest that the substitutions that changed PPI specificity during historical evolution had the capacity to confer this function—and, if reverted, to restore the ancestral function—only during a very limited temporal window.

Sources of chance and determinism

For chance to strongly influence the outcomes of adaptive evolution, multiple paths to a selected phenotype must be accessible with reasonably similar probabilities of being taken; such a situation could arise if several different mutations (or sets of mutations) can confer a new function, or if there are mutations that have no effect on function that can accompany function-changing mutations by chance. To distinguish between these possibilities, we

measured the functional effects of sets of mutations that arose in different replicates when hsMCL-1 was evolved to lose NOXA binding (Fig. 3f, Fig. S12a, S13). One mutation (v189G) was found at high frequency in all 4 replicates, but it was always accompanied by other mutations, which varied among trajectories. We found that the apparently-deterministic mutation v189G was a major contributor to the loss of NOXA binding, but it had this effect only in the presence of the other mutations, which did not decrease NOXA binding on their own. v189G therefore required permissive mutations, and there are multiple sets of mutations that can exert that effect, so which ones occur in any replicate is a matter of chance. All permissive mutations were located near the NOXA binding cleft (Fig. 3g). Other starting genotypes showed a similar pattern of multiple key mutations capable of conferring the selected function (Fig. S12b,c, S14).

To gain further insight into the causes of chance and determinism, we performed PACE experiments in which we evolved hsBCL-2 to retain its BID binding, without selection on NOXA binding, and then screened for variants that fortuitously gained NOXA binding by activity-dependent plaque assay (Fig. S17a, Fig. S15a-d). This strategy allowed us to distinguish between the number of mutations that can confer this phenotype and the influence of selection in favoring a subset of mutations that most rapidly increase NOXA binding. All four replicate populations produced clones that gained NOXA binding at a frequency of about ~0.1% to 1% – lower than when NOXA binding was selected for, but five orders of magnitude higher than when NOXA binding was selected against (Fig. 4a, Fig. S16). From each replicate, we sequenced three NOXA-binding clones and found that all but one of them contained mutation r165L (Fig. 4b), a mutation that also occurred at high frequency in hsBCL-2 trajectories when NOXA binding was selected for. We introduced r165L into hsBCL-2 and found that it conferred significant NOXA binding with little effect on BID binding (Fig. S17b, Fig. S15e). Though the high frequency mutant profiles looked different among the experiments for NOXA, for BID, and against NOXA (Fig. S17c), additional mutations in individual clones that fortuitously acquired NOXA binding when evolved to maintain BID binding were also found in some hsBCL-2 replicates evolved to gain NOXA binding. A similar pattern was observed when AncB4 and AncB5 clones fortuitously or selectively evolved NOXA binding (Fig. S17d-g, Fig. S15f, g). The genetic basis by which NOXA binding was gained was therefore highly similar among replicates from the same starting genotype, whether or not NOXA binding was selected for. This indicates that the determinism observed in these experiments arises because there are few genotypes that can increase NOXA binding while retaining BID binding from each starting point, rather than because there are many such genotypes but a few are favored over others under strong selection.

Chance and contingency alter accessibility of new functions

Although we found a strong influence of chance and contingency in the sequence outcomes of evolution across ancestral starting points, they had little influence on the acquisition of the historical functions per se, because all replicates from all starting points acquired the selected BCL-2- or MCL-1-like specificity. However, changes in function that did not evolve during history might be subject to more chance or contingency. Using PACE and several proteins that can bind both BID and NOXA as starting points, we selected for variants that could bind NOXA but not BID, a phenotype not known in nature. hsMCL-1 readily evolved the selected phenotype, but two experimentally-evolved variants of hsBCL-2 that had acquired NOXA binding went extinct under the same selection conditions (Fig. 4c, Fig. S18, Fig. S19). The

inability of the derived hsBCL-2 genotypes to acquire NOXA specificity was not attributable to a general lack of evolvability, because these same genotypes successfully evolved in a separate PACE experiment to lose their NOXA binding but retain BID binding (Fig. 4c, Fig. S20). These results illustrate that contingency can influence the ability to evolve new functions, even between closely related backgrounds.

CONCLUSIONS

Without chance, contingency in history would be inconsequential, because all phylogenetic lineages launched from a common ancestor would always lead to the same intermediate steps and thus the same ultimate outcomes. Without contingency, chance events would be inconsequential, because the mutations that happen to occur in any time interval would have no effect on the next set of available steps or ultimate outcomes; every path that was ever open would remain forever so. Our work shows that, across the timescale of BCL-2 sequence evolution, the interaction of chance and contingency eliminates virtually all traces of necessity and repeatability in sequence evolution under strong selection, except for sites that appear constrained.

To systematically assess chance and contingency in BCL-2 evolution, we developed an efficient new method to simultaneously select for particular interactions and against others, which allowed us to drive the evolution of new PPI specificities in multiple replicates without severe bottlenecks in just days. By applying this technology to reconstructed ancestral proteins, our experiments directly illuminate how the sequence divergence that occurred during BCL-2 historical evolution generates chance and contingency in the experimental evolution of the same functional specificities that evolved during the family's history. Our design does not directly reveal chance and contingency during the historical evolution of these proteins, because the selection pressures, environmental conditions, and population parameters that pertained in the deep past are unknown. It is likely, however, that chance and contingency were at least as significant during history as in our experiments for several reasons. First, our experiments likely favor determinism in the evolutionary process due to the very large population sizes, extremely strong selection pressures, and high mutation rates, all of which were directed at a single gene; if, as seems likely, BCL-2 historical evolution involved smaller populations, weaker selection, lower mutation rates, and a larger genetic "target size" for adaptation, then chance would have played an even larger role during history than in our experiments. Second, chance and contingency both arise from the relationship between genotype and function—chance from the number of accessible genotypes that confer a selected-for function, and contingency from epistasis-induced differences in the accessibility of paths among starting genotypes. The fold and structural basis for binding of BH3 containing proteins is conserved across extant BCL-2 family members, so it is unlikely that these factors, and the ways that they lead to chance and contingency, were dramatically different between the evolutionary processes underlying changes in BCL-2 specificity during history and in our experiments.

Epistasis is a common feature of protein structure and function, so we expect that the accumulating effect of contingency that we observed across phylogenetic time among BCL-2 family members will be a general feature of protein evolution, although its rate and extent is likely to vary among folds and functions[41–44]. The influence of chance is likely to depend on the

particular function that is evolving: more determinism is expected for functions with very narrow sequence-structure-function-constraints (e.g.[13,15,25,45–47]), than those for which sequence requirements are less strict (e.g.[19,23,48,49]). In the extreme, when diffuse selection pressures are imposed on whole organisms, virtually no repeatability has been observed among replicates, because loci across the entire genome are potential sources of adaptive mutations[21,22].

Our results have implications for protein biochemistry, engineering, and evolution. First, we found no evidence that ancestral proteins were more "evolvable" than extant proteins: the selected-for phenotypes readily evolved from both extant and ancestral proteins; further, the effect of chance in these processes was virtually constant across about a billion years of evolution, indicating that the number of accessible mutations that can confer a selected-for function was no greater in the past than it is in the present. Second, the strong effect of contingency suggests that efforts to produce proteins with new functions by design or directed evolution will be most effective if they use multiple different protein sequences as starting points, ideally separated by long intervals of sequence evolution. Third, our finding that affinity for new partners can evolve fortuitously during selection to maintain existing binding indicates that new interactions can sometimes be acquired neutrally and then, if they become functionally significant, be amplified or preserved by selection; conversely, maintaining specificity during natural and directed evolution requires selection against off-target interactions[50].

Finally, our observations suggest that the sequence-structure-function associations apparent in sequence alignments are, to a significant degree, the result of shared but contingent constraints that were produced by chance events during history[27–29,48]. Present-day proteins are physical anecdotes of a particular history: they reflect the interaction of accumulated chance events during descent from common ancestors with necessity imposed by physics, chemistry and natural selection. Apparent "design principles" in extant or evolved proteins express not how things must be—or even how they would be best—but rather the contingent legacy of the constraints and opportunities that those molecules just happen to have inherited.

References

1. Gould, S. J. *Wonderful life: the Burgess Shale and the nature of history*. (Norton, 1990).
2. Travisano, M. *et al.* Experimental Tests of the Roles of Adaptation , Chance , and History in Evolution. *Science* **267**, 87–90 (1995).
3. Jablonski, D. Approaches to Macroevolution: 1. General Concepts and Origin of Variation. *Evol. Biol.* **44**, 427–450 (2017).
4. Beatty, J. Chance Variation and Evolutionary Contingency: Darwin, Simpson, The Simpsons, and Gould. *Oxford Handb. Philos. Biol.* 1–22 (2009).
5. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
6. Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375 (2004).
7. Kale, J., Osterlund, E. J. & Andrews, D. W. BCL-2 family proteins: Changing partners in the dance towards death. *Cell Death Differ.* **25**, 65–80 (2018).
8. Aristotle. *Categories. On Interpretation. Prior Analytics.* (Harvard University Press, 1938).
9. Ramsey, G. & Pence, C. H. *Chance in Evolution*. (The University of Chicago Press, 2016).
10. Monod, J. *Chance & Necessity. Translation of le hasard et la necessite.* (First Vintage Books Edition, 1972).
11. Lobkovsky, A. E. & Koonin, E. V. Replaying the tape of life: Quantification of the predictability of evolution. *Front. Genet.* **3**, 1–8 (2012).
12. Blount, Z. D., Lenski, R. E. & Losos, J. B. Contingency and determinism in evolution: Replaying life's tape. *Science* **362**, eaam5979 (2018).
13. Storz, J. F. Causes of molecular convergence and parallelism in protein evolution. *Nat. Rev. Genet.* **17**, 239–250 (2016).
14. Orgogozo, V. Replaying the tape of life in the twenty-first century. *Interface Focus* **5**, 20150057 (2015).
15. Meyer, J. R. *et al.* Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda. *Science* **335**, 428–432 (2012).
16. Couñago, R., Chen, S. & Shamoo, Y. In Vivo Molecular Evolution Reveals Biophysical Origins of Organismal Fitness. *Mol. Cell* **22**, 441–449 (2006).
17. Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M. & Liu, D. R. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proc. Natl. Acad. Sci.* **110**, 9007–9012 (2013).
18. Bollback, J. P. & Huelsenbeck, J. P. Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. *Genetics* **181**, 225–234 (2009).
19. Zheng, J., Payne, J. L. & Wagner, A. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science* **365**, 347–353 (2019).
20. Kacar, B., Ge, X., Sanyal, S. & Gaucher, E. A. Experimental Evolution of Escherichia coli Harboring an Ancient Translation Protein. *J. Mol. Evol.* **84**, 69–84 (2017).
21. 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).
22. Wünsche, A. *et al.* Diminishing-returns epistasis decreases adaptability along an evolutionary trajectory. *Nat. Ecol. Evol.* **1**, 1–6 (2017).
23. Blount, Z. D., Barrick, J. E., Davidson, C. J. & Lenski, R. E. Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* **489**, 513–518 (2012).
24. Baier, F. *et al.* Cryptic genetic variation shapes the adaptive evolutionary potential of enzymes. *Elife* **8**, 1–20 (2019).
25. Salverda, M. L. M. *et al.* Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.* **7**, e1001321 (2011).
26. Ortlund, E. A., Bridgham, J. T., Redinbo, M. R. & Thornton, J. W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**, 1544–1548 (2007).
27. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife* **2013**, 1–19 (2013).
28. Harms, M. J. & Thornton, J. W. Historical contingency and its biophysical basis in glucocorticoid receptor

evolution. *Nature* **512**, 203–207 (2014).
29. Starr, T. N., Flynn, J. M., Mishra, P., Bolon, D. N. A. & Thornton, J. W. Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. *Proc. Natl. Acad. Sci.* **115**, 4453–4458 (2018).
30. Chipuk, J. E., Moldoveanu, T., Llambi, F., Parsons, M. J. & Green, D. R. The BCL-2 Family Reunion. *Mol. Cell* **37**, 299–310 (2010).
31. Danial, N. N. & Korsmeyer, S. J. Cell Death: Critical Control Points. *Cell* **116**, 205–219 (2004).
32. Chen, L. *et al.* Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. *Mol. Cell* **17**, 393–403 (2005).
33. Lomonosova, E. & Chinnadurai, G. BH3-only proteins in apoptosis and beyond: An overview. *Oncogene* **27**, S2–S19 (2008).
34. Petros, A. M., Olejniczak, E. T. & Fesik, S. W. Structural biology of the Bcl-2 family of proteins. *Biochim. Biophys. Acta - Mol. Cell Res.* **1644**, 83–94 (2004).
35. Banjara, S., Suraweera, C. D., Hinds, M. G. & Kvansakul, M. The Bcl-2 family: Ancient origins, conserved structures, and divergent mechanisms. *Biomolecules* **10**, 1–21 (2020).
36. Lanave, C., Santamaria, M. & Saccone, C. Comparative genomics: The evolutionary history of the Bcl-2 family. *Gene* **333**, 71–79 (2004).
37. Certo, M. *et al.* Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. *Cancer Cell* **9**, 351–365 (2006).
38. Dutta, S. *et al.* Determinants of BH3 Binding Specificity for Mcl-1 versus Bcl-xL. *J. Mol. Biol.* **398**, 747–762 (2010).
39. Pu, J., Zinkus-Boltz, J., Dickinson, B. C., Boltz, J. Z.- & Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nat. Chem. Biol.* **13**, 1–27 (2017).
40. Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat. Chem. Biol.* **10**, 216–222 (2014).
41. Chandler, C. H., Chari, S. & Dworkin, I. Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends Genet.* **29**, 358–366 (2013).
42. Storz, J. F. Compensatory mutations and epistasis for protein function. *Curr. Opin. Struct. Biol.* **50**, 18–25 (2018).
43. Harms, M. J. & Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **14**, 559–571 (2013).
44. Shah, P., Plotkin, J. B., McCandlish, D. M. & Plotkin, J. B. Contingency and entrenchment in protein evolution under purifying selection. *Proc. Natl. Acad. Sci.* **112**, E3226–E3235 (2015).
45. Hawkins, N. J., Bass, C., Dixon, A. & Neve, P. The evolutionary origins of pesticide resistance. *Biol. Rev.* **94**, 135–155 (2019).
46. Karageorgi, M. *et al.* Genome editing retraces the evolution of toxin resistance in the monarch butterfly. *Nature* **574**, 409–412 (2019).
47. Menéndez-Arias, L. Molecular basis of human immunodeficiency virus drug resistance: An update. *Antiviral Res.* **85**, 210–231 (2010).
48. Starr, T. N., Picton, L. K. & Thornton, J. W. Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* **549**, 409–413 (2017).
49. Yokoyama, S., Tada, T., Zhang, H. & Britt, L. Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 13480–13485 (2008).
50. Levin, K. B. *et al.* Following evolutionary paths to protein-protein interactions with high affinity and selectivity. *Nat. Struct. Mol. Biol.* **16**, 1049–55 (2009).

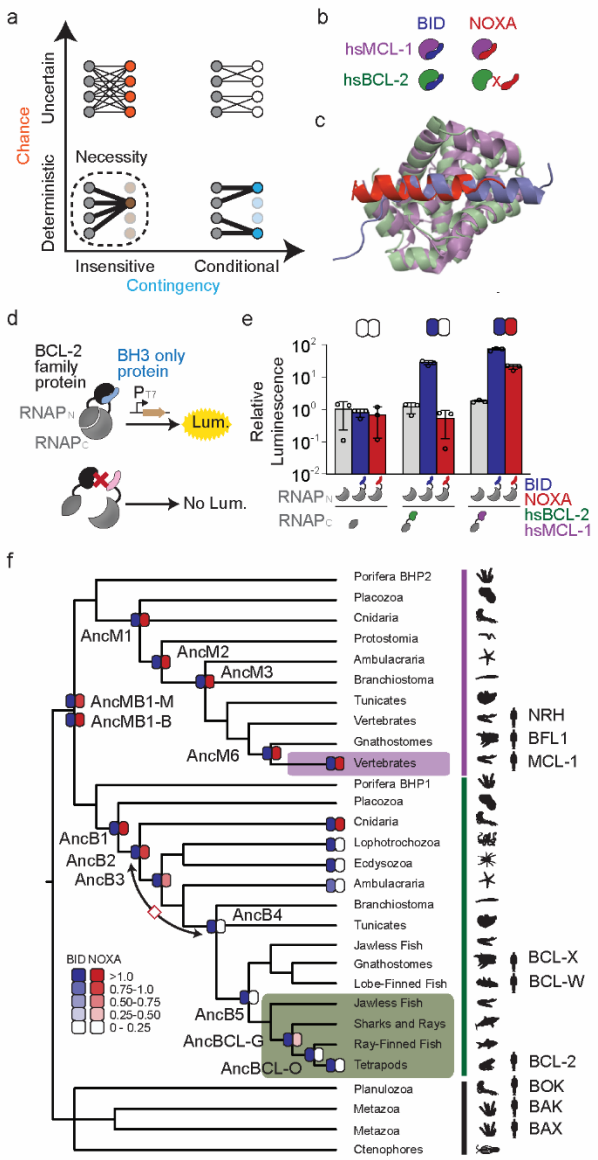

Fig. 1: BID specificity is a novel derived trait in the BCL-2 lineage. a, Chance, contingency, and necessity in evolutionary outcomes. The outcomes of evolution from a set of starting genotypes (gray circles) can be influenced by both chance events (y-axis) and contingency (x-axis). Black lines connect starting genotypes to evolutionary outcomes, with thickness proportional to the probability that a path is taken. Faded circles are inaccessible from the starting genotypes. Evolutionary outcomes can be influenced by chance (orange), contingency (teal), both chance and contingency (white), or neither (brown). This later case corresponds to evolutionary necessity (dashed lines). **b,** Protein binding specificities of extant BCL-2 family members. Human MCL-1 (purple) binds BID (blue) and NOXA (red), while human BCL-2 (green) binds BID but not NOXA. **c,** Crystal structures of MCL-1 (purple) bound to NOXA (red, PDB: 2nla) and BCL-xL (a vertebrate paralog of BCL-2, green) bound to BID (blue, PDB: 4qve). BID and NOXA bind at the same interface. **d,** Schematic of the luciferase reporter assay to assess PPIs. If a BCL-2 family protein binds a given peptide from a BH3-only protein, the split T7 RNAP biosensor assembles and drives luciferase expression. **e,** *In vivo* luciferase assay of hsBCL-2 (green), hsMCL-1 (purple), and no BCL-2 family protein interacting with BID (blue), NOXA (red), or no BH3-only protein (gray). All activity is scaled relative to the control of no BCL-2 family protein with no BH3-only protein. Bars are the mean of three

biological replicates (circles); error bars, SD. The shaded boxes above the bars indicate how this data is represented in heatmap form, where BID signal is normalized with the BID with hsBCL-2 output and NOXA signal is normalized with the NOXA with hsMCL-1 output. **f,** Cladogram of BCL-2 family proteins. Silhouettes are of a representative species in each clade. Heatmaps indicate BID (blue) and NOXA (red) binding measured using the *in vivo* luciferase reporter assay as in (**d**) and (**e**). Each shaded box represents the normalized mean of three biological replicates. The double-headed arrow and white diamond with red outline indicates the interval in which NOXA binding was historically lost, yielding BID specificity in the BCL-2 clade. Purple bar, MCL-1 like clade. Green bar, BCL-2 like clade. Black bar, pro-apoptotic family members. Purple shaded box, vertebrate MCL-1. Green shaded box, vertebrate BCL-2.

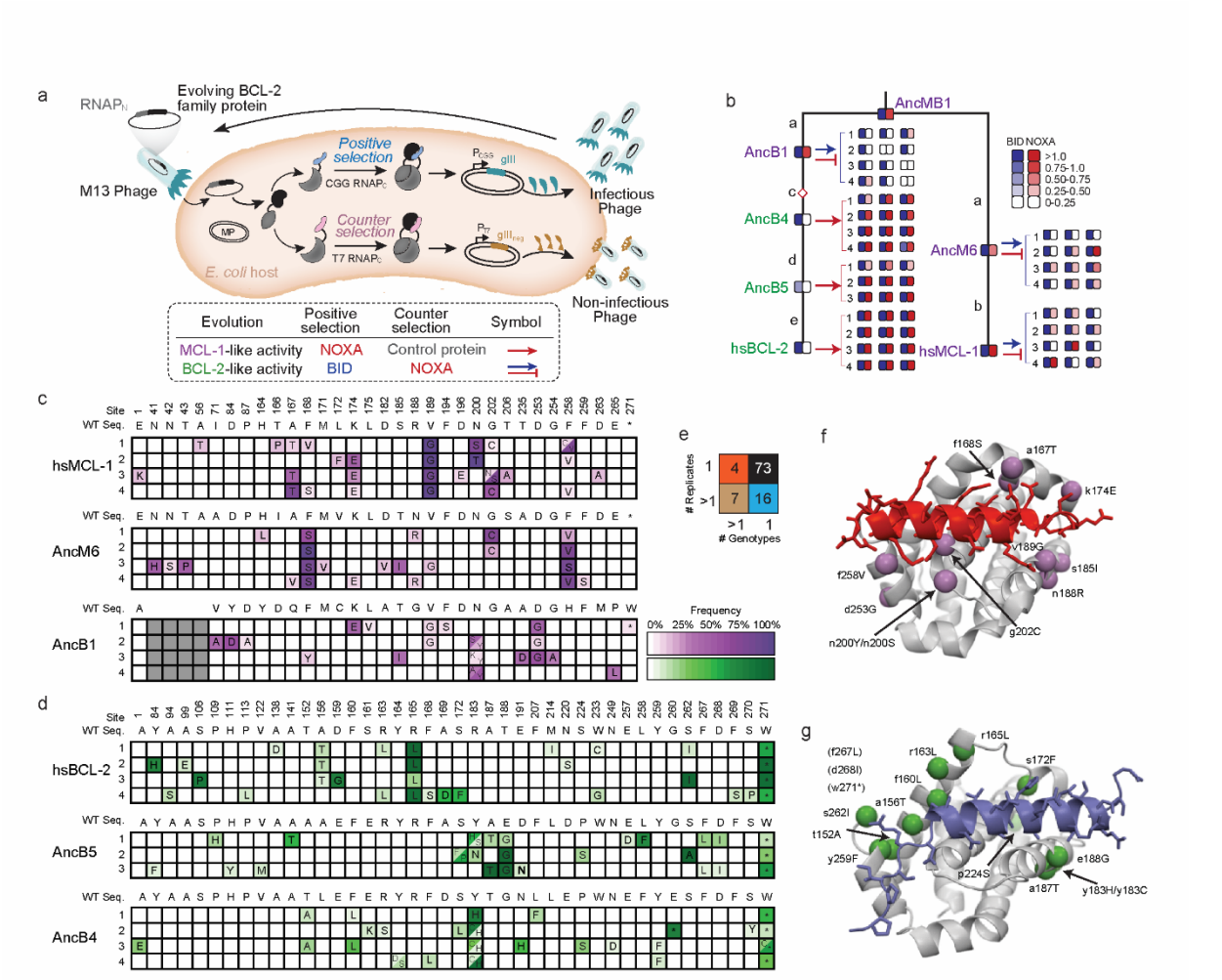

Fig. 2: Chance and contingency shape evolutionary outcomes. a, PACE system for evolving specificity in PPIs. Gene III (gIII, teal) is required for the production of infectious M13 phage. To evolve binding of BCL-2 family proteins (black) to a desired BH3-only protein (light blue), phage lacking gIII are engineered to contain a fusion of the N-terminal portion of an evolved proximity-dependent split T7 RNA polymerase ($RNAP_N$, gray) to the BCL-2 family protein. Host *E. coli* cells carry a positive selection plasmid containing $T7_{CGG}$-promoter (P_{CGG}) driven gIII and the C-terminal portion of $T7_{CGG}$ RNA polymerase (CGG $RNAP_C$) fused to a peptide from the target BH3-only protein. Upon host infection, binding of the BCL-2 protein to the target peptide reconstitutes intact $RNAP_{CGG}$ and drives expression of gIII, generating infectious phage. To evolve BCL-2 family proteins against binding of a peptide from an undesired BH3-only protein (pink), host *E. coli* also carry a counter selection plasmid containing T7 promoter (P_{T7}) driven dominant-negative gIII ($gIII_{neg}$, gold) and the C-terminal portion of T7 RNA polymerase (T7 RNAP) fused to the undesired petide. Binding of the BCL-2 protein to the undesired peptide reconstitutes intact T7 RNAP and drives expression of $gIII_{neg}$, generating non-infectious phage. An arabinose-inducible mutagenesis plasmid (MP) increases the mutation rate. **b,** Phenotypic outcomes of PACE experiments. Evolutionary relationships among the extant and ancestral genotypes used as PACE starting points are shown on a reduced phylogeny. White diamond with red outline indicates the historical loss of NOXA binding. Letters indicate specific phylogenetic intervals between starting genotypes. Each starting genotype was evolved in multiple independent replicate trajectories (numbers) under the selection conditions shown as arrows

and blunt bars: arrow, selection for binding to BID (blue) or NOXA (red); blunt bar, selection against binding a peptide. Starting genotypes in green were evolved to gain NOXA binding, while starting genotypes in purple were evolved to lose NOXA binding. Heatmaps show activity of BID (blue) and NOXA (red) in the luciferase assay for starting genotypes and three variants randomly isolated from each replicate population. Each shaded box represents the normalized mean of three biological replicates. **c, d,** Allele frequency of non-wildtype states from PACE experiments. Replicate populations started from hsMCL-1, AncM6, or AncB1 and evolved to lose NOXA binding. Gray boxes are sites that do not exist in a particular sequence. (**c**). Replicate populations started from hsBCL-2, AncB5, or AncB4 and evolved to gain NOXA binding (**d**). Non-wildtype amino acid states that reached > 5% in frequency are shown, with frequency proportional to color saturation. Site numbers and wildtype (WT) amino acid states are listed above each sequence. Each row represents an independent replicate. Split cells show populations with multiple non-WT states > 5%. **e,** Categories of the 100 non-WT states observed. White, number of mutant states observed in only one replicate. Teal, number of mutant states observed in multiple replicates from the same starting genotype. Orange, number of mutant states observed in a single replicate from multiple different starting genotypes. Brown, number of mutant states observed in multiple replicates from the same starting genotype and from multiple different starting genotypes. **f,** Location of repeated mutations when hsMCL-1, AncM6, and AncB1 were evolved to lose NOXA binding (purple spheres). Structure is of MCL-1 (light gray) bound to NOXA (red, PDB: 2nla). **g,** Location of repeated mutations when evolving hsBCL-2, AncB5, and AncB4 to gain NOXA binding (green spheres). Structure is of hsBCL-xL (light gray) bound to BID (blue, PDB: 4qve).

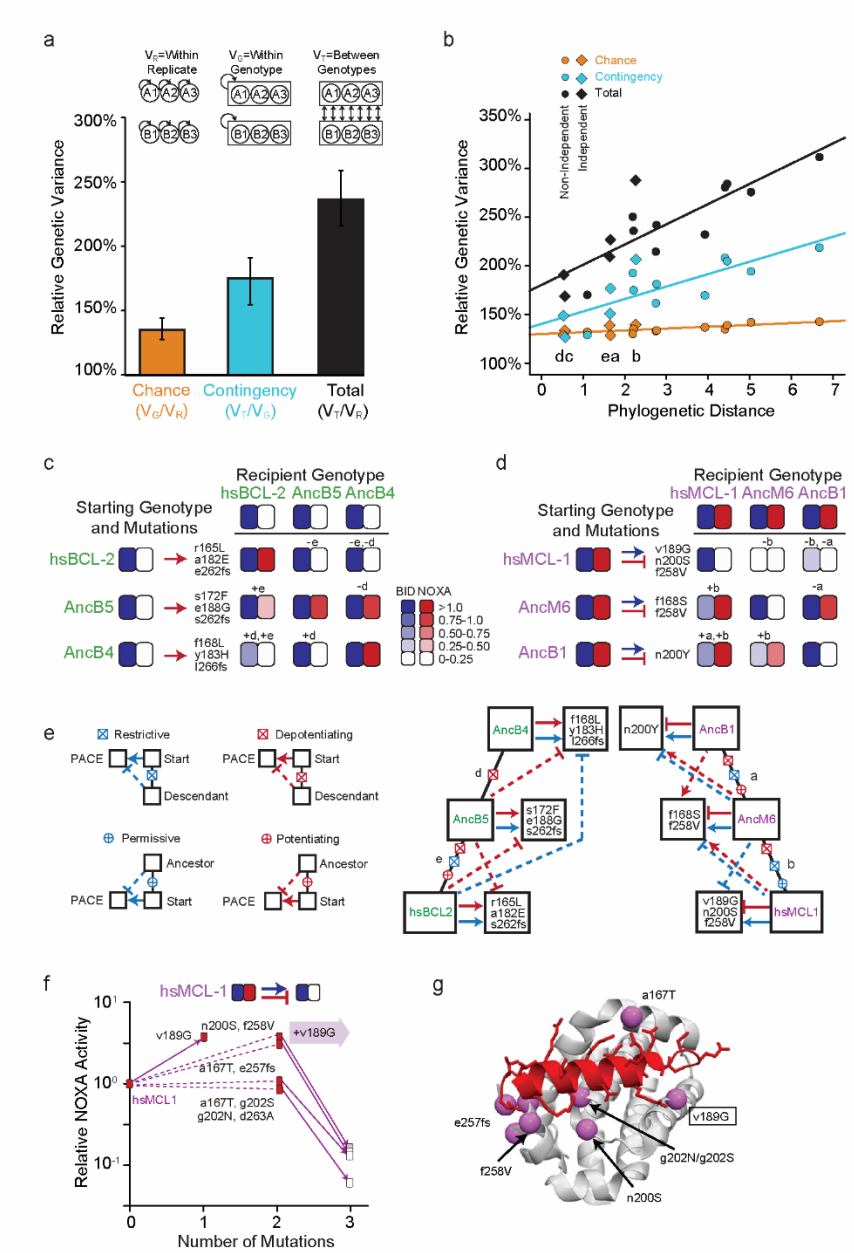

Fig. 3: Genetic sources of chance and contingency. a, Estimated variation in evolutionary outcomes due to chance (orange), contingency (teal), and both (black). Error bars, 95% confidence intervals on the mean as estimated by bootstrapping PACE replicates. Inset: schematic for estimating the roles of chance and contingency on evolutionary outcomes. The effects of chance were estimated as the average genetic variance among replicates from the same starting genotype (V_G) compared to the within replicate genetic variance (V_R). The effects of contingency were estimated as the average genetic variance among replicates from different starting genotypes (V_T) compared to the average genetic variance among replicates from the same starting genotype (V_G). The combined effects of chance and contingency were estimated as the average genetic variance among replicates from different starting genotypes (V_T) compared to the within replicate genetic variance (V_R). **b,** Change in chance and contingency over time. Relationship between phylogenetic distance between pairs of starting genotypes for experimental evolution (ancestral or extant proteins, as the total branch lengths separating them)

and the effects of chance (orange), contingency (teal), or both (black) on the outcomes of evolution between them. Solid lines are best fits from linear models. Diamonds, points that are phylogenetically independent (i.e. branches connecting direct ancestors and descendants). Circles, points that are non-independent. Letters indicate the specific branch from Fig. 2b. **c, d,** Effects of mutations in different genotypic backgrounds. BID and NOXA activity for variants isolated when evolving hsBCL-2, AncB5, and AncB4 to gain NOXA binding (**c**) and of hsMCL-1, AncM6, and AncB1 to lose NOXA binding (**d**). Each starting genotype was evolved under the selection conditions shown as arrows (positive selection) and blunt bars (counter selection) for and against BID (blue) and NOXA (red) binding. Listed variants are the set of mutations isolated from a single PACE variant and introduced into recipient genotypes (top). Heatmaps show binding to BID (blue) and NOXA (red) in the luciferase assay for starting genotypes (on the left) and recipient genotypes. Each shaded box represents the normalized mean of three biological replicates. Letters above the shaded boxes indicate the specific phylogenetic branch from Fig. 2b that contains historical changes between compared sequences. Mutations placed on descendent sequence (plus). Mutations placed on ancestral sequence (minus). **e,** Interactions between historical substitutions and PACE mutations. Green, starting genotypes evolved to gain NOXA binding. Purple, starting genotypes evolved to lose NOXA binding. Arrow, gain or maintenance of binding. Bar, loss of binding. Solid lines, selection conditions. Dashed lines, outcomes that differ from the selection conditions. Red, NOXA. Blue, BID. Left inset: schematic defining restrictive, permissive, depotentiating, and potentiating substitutions. If PACE mutations decrease BID activity when introduced into a protein's descendants, then restrictive substitutions must have occurred historically between them. Conversely, if PACE mutations compromise or abolish BID activity when introduced into a protein's ancestor, then historical permissive substitutions occurred between them. If PACE mutations confer the selected change in NOXA activity in one protein but fail to have that effect in its ancestors, then potentiating substitutions occurred on the phylogeny. Finally, if PACE mutations fail to have the effect on NOXA activity in a protein's descendants, then depotentiating substitutions occurred during history. **f,** Effects of mutations that occurred in PACE trajectories of hsMCL-1 evolved to lose NOXA binding. Values are normalized means of three biological replicates. Solid lines show the effect of the v189G mutation. Dotted lines show the effect of double mutations lacking v189G. The four trajectories of dotted lines followed by solid lines illustrate four unique individual variants. **g,** Mutations found in isolated hsMCL-1 variants evolved to lose NOXA binding (purple spheres). Structure is of MCL-1 (light gray) bound to NOXA (red, PDB: 2nla). The v189G mutation (box) was found in all four isolated variants.

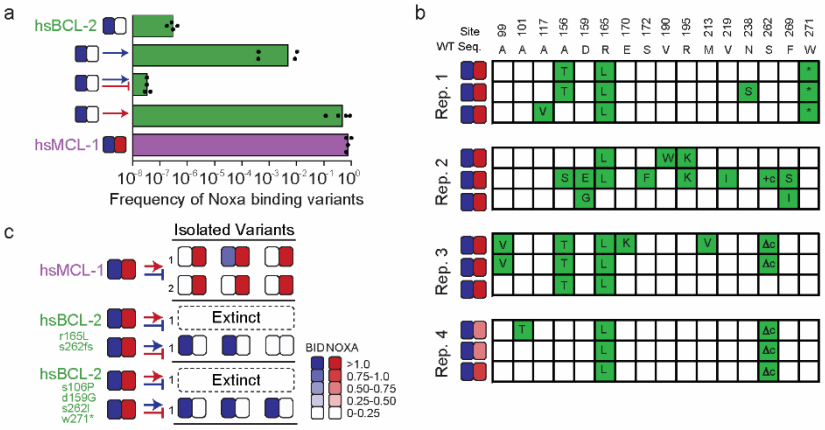

Fig. 4: Evolutionary consequences of chance and contingency. a, Frequency of NOXA-binding phage. Frequency was calculated as the ratio of plaque forming units (PFU) per milliliter on *E. coli* cells that require NOXA binding to form plaques to the PFU on cells that require BID binding to form plaques. Wildtype hsBCL-2 (green) and hsMCL-1 (purple) are shown as controls. Arrow, positive selection for function. Bar, counter selection against function. Blue, BID. Red, NOXA. **b,** Phenotypes and genotypes of individual BCL-2 variants evolved in the absence of selection on NOXA and then screened for a gain in NOXA activity. Sites and WT amino state are indicated at top. For each variant, non-WT states acquired are shown in green. Heatmaps on the left show binding to BID (blue) and NOXA (red) in the luciferase assay for each variant. Each shaded box represents the normalized mean of three biological replicates. **c,** Phenotypic outcomes when selecting for novel functions in PACE. Heatmaps show binding to BID (blue) and NOXA (red) in the luciferase assay for each starting genotype (on the left) and for three individual variants picked from that experiment. Each shaded box represents the normalized mean of three biological replicates. Colored arrows and bars indicate PACE selection pressures. Arrow, positive selection for function. Bar, counter selection against function. Blue, BID. Red, NOXA.

Methods

Cloning and general methods

Plasmids were constructed as previously described[51]. Briefly, Q5® High-Fidelity DNA Polymerase (NEB) was used to amplify fragments that were then ligated via Gibson Assembly. Primers were obtained from IDT and all plasmids were sequenced at the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility. Vectors used are listed in Supplementary Tables 2-9, and gene sequences are listed in Supplementary Tables 10-11. Fully annotated vector maps are accessible via Benchling. Key vectors are deposited at Addgene and all vectors are available upon request. The following working concentrations of antibiotics were used: 50 µg/mL carbenicillin, 50 µg/mL spectinomycin, 40 µg/mL kanamycin, and 33 µg/mL chloramphenicol. Protein structures and alignments were generated using the program PyMOL[52].

Phylogenetics

The amino acid sequences of the human BCL-2, BCLW, BCL-xL, MCL-1, NRH, BFL1, BAK, BAX, and BOK paralogs were used as starting points for identifying BCL-2 family members in other species. For each paralog, tblastn and protein BLAST on NCBI BLAST were used to identify orthologous sequences between January and March of 2018[53]. Sequences for each paralog were aligned using MAFFT (G-INS-I) with the --allowshift option and --unalignlevel set at 0.1. For each paralog, phylogenetic structure was determined using fasttree 2.1.11 within Geneious 10.1.3. Missing clades based on known species relationships were then identified and additional tblastn searches were used within Afrotheria (taxid:311790), Marsupials (taxid:9263), Monotremes (taxid:9255), Squamata (taxid:8509), Archosauria (taxid:8492), Testudinata (taxid:8459), Amphibia (taxid:8292), Chondrichthyes (taxid:7777), Actinopterygii (taxid:7898), Dipnomorpha (taxid:7878), Actinistia (taxid:118072), Agnatha (taxid:1476529), Cephalochordata (taxid:7735), and Tunicata (taxid:7712) as needed. Additional sequences were added by downloading genome and transcriptome data for tuatara[54], sharks and rays[55], gar[56], ray-finned fish[57], lamprey[58], hagfish[59], *Ciona savignyi*[56], tunicates[60], echinoderms[61], porifera [62](https://doi.org/10.7939/R3WH2DV20, https://doi.org/10.7939/R3S000), and ctenophores[63]. In each case, local BLAST databases were made in Geneious and searched using tblastn. Finally, we used BCL-2DB to add any missing groups and downloaded sequences as needed[64].

After collection of sequences, each paralog was realigned using MAFFT (G-INS-I) with the --allowshift option and --unalignlevel set at 0.1. Based on known species relationships, lineage specific insertions were removed and gaps manually edited. Only a single sequence was kept among pairs of sequences differing by a single amino acid and sequences with more than 25% of missing sites were removed. For difficult to align sequences, sequences were modeled on the structures of human BCL-2 family members using SWISS-Model to help identify likely locations of gaps[65]. Finally, paralogs were profile aligned to each other and paralog specific insertions were identified.

In total, 151 amino acid sites from 745 taxa were used to infer the phylogenetic relationships among BCL-2 paralogs. PROT Test 3.4.2 was used to identify the best fit model among JTT, LG, and WAG, with combinations of observed amino acid frequencies (+F), gamma distributed rate categories (+G), and an invariant category (+I). From this, JTT + G + F had the highest likelihood and lowest AIC. RAXML-ng 0.6.0 was then used to identify the maximum likelihood tree using JTT+G12+F0 (12 gamma rate categories with maximum likelihood estimated amino acid frequencies)[66]. Finally, we enforced monophyly within each paralog for the following groups: lobe-finned fish (9), ray-finned fish (9), jawless fish (5), cartilaginous fish (8), tunicates (4), branchiostoma (4), chordates (5), ambulacraria (5, hemichordata + echinodermata), deuterostomia (5), protostomia (5), cnidaria (5), and porifera (4) (values in parenthesis are number of paralogs in each group) and used RAXML-ng with JTT + G12 + F0 to identify the best tree given these constraints (Supplementary Data Phylogenetic.Data.zip).

Overall, we recovered three clades; a pro-apoptotic clade, a clade containing the BCL-2, BCLW, and BCLX paralogs, and a clade containing the MCL-1, BFL1, and NRH paralogs. We used the pro-apoptotic clade as the outgroup to the two anti-apoptotic clades. Within the BCL-2 clade, the majority of vertebrates contained all three copies. However, the exact relationship among the paralogs was unclear; only two copies were identified within jawless fish and their phylogenetic placement had weak support. Non-vertebrate clades tended to have good support and only a single copy. However, support for these groups following established species relationships was often limited. The MCL-1 clade contained the fastest evolving paralogs of the BCL-2 family. As with the BCL-2-like clade, only two copies were found within the jawless fish and the exact sister relationships among paralogs was unclear. Non-vertebrates contained only a single copy, but as with the BCL-2-like clade, support for relationships following established species relationships was often weak.

The BCL-2-like and MCL-1-like paralogs formed a clade with the BHP1 and BHP2 sequences from porifera. The sister relationships among these four clades were unresolved. In addition, we recovered a sister relationship between the BAK and BAX paralogs. While both paralogs contained copies from porifera, these clades evolved quickly and had relatively low support, and they may be artifactual. We identified only a single clade of ctenophores. Finally, the placement of BOK was unresolved; BOK may be sister to the BAK/BAX clade or an outgroup to all clades and the most ancient copy of the BCL-2 family.

Ancestral Reconstruction

Posterior probabilities of each amino acid at each site were inferred using Lazarus[67] to run codeml within PAML[68]. We used the same model and alignment as used to infer the phylogeny. We used the branch lengths and topology of the constrained maximum likelihood phylogeny found by raxml-ng.

We first reconstructed the last common ancestors (LCA) of all BCL-2 and MCL-1 like sequences, AncMB1-M, using the maximum likelihood state for each alignable site and the human MCL-1 sequence to fill in the N and C-terms and the loop between the first and second alpha helices. We then reconstructed a series of ancestors from AncMB1 to modern human MCL-1. These included AncM1 (last common ancestor of MCL-1 related sequences), AncM2 (LCA of MCL-1 related Deuterostomes and Protostomes), AncM3 (LCA of MCL-1 related Deuterostomes), AncM4 (LCA of MCL-1 related Urochordates and Chordates), AncM5 (LCA of MCL-1, BFL1, and NRH like copies in vertebrates), AncM6 (LCA of MCL-1 and BFL1 like copies), AncMCL-1 (LCA of MCL-1 like copies), AncMCL-1-G (LCA of MCL-1 like Gnathostomes), AncMCL-1-O (LCA of MCL-1 like Osteichthyes), and AncMCL-1-T (LCA of MCL-1 like Tetrapods), AncMCL-1-A (LCA of MCL-1 like Amniotes), and AncMCL-1-M (LCA of MCL-1 like Mammals). In each case, the sequence of each ancestor used the maximum likelihood state at each site, with gaps inserted based on parsimony. We used the modern sequences of human MCL-1 to fill in portions of the sequence that showed poor alignment and could not be reconstructed. These included both the N and C-terms, as well as the loop between the first and second alpha-helices. Average posterior probabilities for ancestors in the MCL-1 clade ranged from 0.73 (AncM6) to 0.98 (AncMCL-1-M) with an average of 0.83 (sd 0.08) (Supplementary Table 12).

For BCL-2, we also reconstructed AncMB1, this time using human BCL-2 sequence to fill in the N and C-terms and the loop between the first and second alpha helices (AncMB1-B). We then reconstructed sequences from AncMB1 to modern human BCL-2. These included AncB1 (last common ancestor of BCL-2 related sequences), AncB2 (LCA of BCL-2 related Bilaterian and Cnidaria), AncB3 (LCA of BCL-2 related Deuterostomes and Protostomes), AncB4 (LCA of BCL-2 Deuterostomes), AncB5 (LCA of BCL-2, BCLW, and BCLX like copies in vertebrates), AncBCL-2 (LCA of BCL-2 like copies), AncBCL-2-G (LCA of BCL-2 like Gnathostomes), AncBCL-2-O (LCA of BCL-2 like Osteichthyes), and AncBCL-2-T (LCA of BCL-2 like Tetrapods) using human BCL-2 sequences for the N and C-terms and the loop between the first and second alpha helices. Average posterior probabilities for ancestors in the BCL-2 clade ranged from 0.87 (AncB1) to 0.95 (AncBCL-2-T) with an average of 0.9 (sd 0.04).

Luciferase assay

Cloned expression vectors contained the following: 1) a previously-evolved, isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible N-terminal half of T7 RNAP[69] fused to a BCL-2 family protein; 2) the C-terminal half of T7 RNAP fused to a peptide from a BH3-only protein; and 3) T7 promoter-driven luciferase reporter. Chemically-competent S1030 *E. coli* cells[40] were prepared by culturing to an OD_{600} of 0.3, washing twice with a calcium chloride/HEPES solution (60 mM $CaCl_2$, 10 mM HEPES pH 7.0, 15% glycerol), and then resuspending in the same solution. Vectors were transformed into chemically competent S1030 cells via heat shock at 42 ºC for 45 sec, followed by recovery in 3x volume of 2xYT media, and then plated on agar with the appropriate antibiotics (carbenicillin, spectinomycin, and chloramphenicol) to incubate overnight at 37 ºC. Individual colonies (3-4 biological replicates per condition) were picked and cultured in 1 mL of LB media containing the appropriate antibiotics overnight at 37 ºC in a shaker. The next morning, 50 µL of each culture was diluted into 450 µL of fresh LB media containing the appropriate antibiotics, as well as 1 µM of IPTG. The cells were incubated in a shaker at 37 ºC, and OD_{600} and luminescence measurements were recorded between 2.5 and 4.5 hours after the start of the incubation. Measurements were taken on a Synergy Neo2 Microplate Reader (BioTek) by transferring 150 µL of the daytime cultures into Corning® black, clear-bottom 96-well plates. Data were analyzed in Microsoft Excel and plotted in GraphPad Prism[39].

Test of robustness of ancestral inference

To determine the robustness of our conclusions on the phenotype of ancestral sequences, we synthesized and cloned alternative reconstructions for key ancestors. In each case, sequences contained the most likely alternative state with posterior probability > 0.2 for all such sites where such a state existed. Alternative reconstructions contained an average of 24 alternative states and represent a conservative test of function (min: 4, max: 44, Supplementary Table 12). In our luciferase assay, all but two alternative reconstructions retained similar BID and NOXA binding as the maximum likelihood ancestral sequences. The first alternative reconstruction that differed from the ML reconstruction was AltAncB3, which bound both BID and NOXA, while the ML for AncB3 bound BID, but NOXA only weakly. As a result, the exact branch upon which NOXA binding was lost historically cannot be fully resolved.

The second alternative reconstruction that differed from the ML reconstruction was AltAncMB1-B, which had weaker NOXA binding than the ML reconstruction. To further test the robustness of AncMB1-B to alternative reconstructions, we synthesized and tested additional reconstructions that included only alternative amino acids with posterior probabilities greater than 0.4 (3), 0.35 (7), 0.3 (13), and 0.25 (18), and compared these to AncMB1-B and the 0.2 AltAncMB1-B (21) (values in parentheses are number of states that differ from the ML state). We found that the 0.4, 0.35, and 0.3 alternative reconstructions bound both BID and NOXA, while the 0.25 and 0.2 alternative reconstructions had diminished NOXA binding.

Finally, we synthesized and tested modern sequences from key groups to determine the robustness of our inference on the timing of NOXA binding loss. These included BCL-2 related sequences from groups that diverged prior to the predicted loss of NOXA binding (*Trichoplax adhaerens* and *Hydra magnapapillata*), sequences from groups that diverged around the time of predicted NOXA binding loss (*Octopus bimaculoides* and *Stegodyphus mimosarum*), or sequences from groups predicted to have diverged after NOXA binding lost (*Saccoglossus kowalevskii* and *Branchiostoma belcheri*). While the *T. adhaerens* and *B. belcheri* sequences were completely non-functional in our luciferase assays, binding neither BID nor NOXA, *H. magnapapillata* bound both BID and NOXA and the remaining sequences bound only BID, suggesting a loss of NOXA binding prior to the divergence of protostomes and deuterostomes in the BCL-2 related clade, consistent with the conclusion drawn using reconstructed proteins.

Protein expression

hsBCL-2, hsMCL-1, and evolved variants were constructed as N-terminal 6xHis-GST tagged proteins. The recombinant proteins were expressed in BL21 *E. coli* (NEB) and purified following standard Ni-NTA resin purification protocols (ThermoFisher Scientific)[70]. Briefly, BL21 *E. coli* containing an N-terminal 6xHis-GST tagged BCL-2 family protein were cultured in 5 mL LB with carbenicillin overnight. The following day, the culture was added to 0.5 L of LB with carbenicillin, incubated at 37 ºC until it reached an OD_{600} of 0.6, induced with IPTG (final concentration: 200 µM), and cultured overnight at 16 ºC. The cell pellet was harvested by centrifugation followed by resuspension in 30 mL of lysis buffer (50 mM Tris 1 M NaCl, 20% glycerol, 10 mM TCEP, pH 7.5) supplemented by protease inhibitors (200 nM Aprotinin , 10 µM Bestatin , 20 µM E-64, 100 µM Leupeptin, 1 mM AEBSF, 20 µM Pepstatin A). Cells were lysed via sonication and were then centrifuged at 12,000 g for 40 min at 4 ºC. Solubilized proteins, located in the supernatant, were incubated with His60 Ni Superflow Resin (Takara), and the protein was eluted using a gradient of imidazole in lysis buffer (50-250 mM). Protein was concentrated, purified via a desalting column with storage buffer (50mM Tris-HCl (pH 7.5), 300 mM NaCl, 10% glycerol, 1 mM DTT), then further concentrated. The concentration of the purified BCL-2 family proteins was determined by BCA assay (ThermoFisher Scientific), and they were flash-frozen in liquid nitrogen and stored at -80ºC.

Fluorescent polarization binding assays

Fluorescent polarization (FP) was used to measure the affinity of BCL-2 family proteins with peptide fragments of the BH3-only proteins in accordance with previously-described methods[71]. hsBCL-2, hsMCL-1, and evolved variants were purified as described above. The fluorescent NOXA and BID peptides (95+% purity) were synthesized by GenScript and were N-terminally labeled with 5-FAM-Ahx and C-terminally modified by amidation. These peptides were dissolved and stored in DMSO. Corning® black, clear-bottom 384-well plates were used to measure FP, and three replicates were prepared for each data point. Each well contained the following 100 µL reaction: 20 nM BH3-only protein, 0.05 nM to 3 µM of BCL-2 family protein (1/3 serial dilutions), 20 mM Tris (pH 7.5), 100 mM NaCl, 1 mM EDTA, and 0.05% pluronic F-68. FP values (in milli-polarization units; mFP) of each sample were read by a Synergy Neo2 Microplate Reader (BioTek) with the FP 108 filter (485/530) at room temperature 5-15 minutes after mixing all the components. Data were analyzed in GraphPad Prism 8, using the following customized fitting equation, to calculate K_d [72] :

$$y = B + C\ (D + K_d + x - \sqrt{(D + K_d + x)^2 - 4Dx})$$

where *y* is normalized measured FP, *x* is the concentration of BCL-2 protein, *D* is the concentration of the BH3-only protein, *B* and *C* are parameters related to the FP value of free and bound BH3-only protein, and K_d is the dissociation constant.

Phage-assisted continuous evolution

Phage-assisted continuous evolution (PACE) was used to evolve hsBCL-2, hsMCL-1, and ancestral proteins in accord with previously reported technical methods[5,39,40,73] using a new vector system. Briefly, combinations of accessory plasmids and the MP6 mutagenesis plasmid[74] were transformed into S1030 *E. coli*., plated on agar containing the appropriate antibiotics (carbenicillin, kanamycin, and chloramphenicol) and 10 mM glucose, and incubated overnight at 37 ºC. Colonies were grown overnight in 5 mL of LB containing the appropriate antibiotics and 20 mM glucose. Davis Rich media[40] was prepared in 5-10 L carboys and autoclaved, and the PACE flasks and corresponding pump tubing were autoclaved as well. The following day, PACE was set up in a 37 ºC environmental chamber (Forma 3960 environmental chamber, ThermoFisher Scientific). For each replicate, an overnight culture was added to ~150 mL of Davis Rich carboy media in chemostats and grown for 2-3 h until reaching an OD_{600} of approximately 0.6. Lagoons containing 20 µL of phage from saturated phage stocks (10^8-10^9 phage) were

then connected to the chemostat. Magnetic stir bars were used to agitate chemostats and lagoons. The chemostat cultures were flowed into the lagoons at a rate of approximately 20 mL/h. Waste output flow rates were adjusted to maintain a constant volume of 20 mL in the lagoons, 150 mL in the chemostat, and an OD_{600} close to 0.6 in the chemostat. A 10% w/v arabinose solution was pumped into the lagoons at a rate of 1 mL/h. If the experiment included a mixing step (two separate chemostats flowed together into one lagoon for a mixed selection pressure), a chemostat was prepared the next day (as described above) and connected to the lagoons. During this step, lagoon volumes were increased to 40 mL, and the arabinose inflow rate was increased to 2 mL/h. After disconnecting the first chemostat the next day, the lagoon volumes and arabinose inflow were both lowered to 20 mL and 1 mL/h, respectively. During the experiment, samples were collected from the lagoons every 24 h, centrifuged to collect the phage-containing supernatant, as well as the cell pellet for DNA extraction. PACE experiments are listed in Supplementary Table 1.

Plaque assays

Plaque assays were performed on 1059 *E. coli* cells[40,75], which supply gene III (gIII) to phage in an activity-independent manner, to measure phage titers. Additionally, activity-dependent plaque assays were done on S1030 *E. coli* containing the desired accessory plasmids to determine the number of phage encoding a BCL-2 family protein with a given peptide binding profile. All cells were grown to an OD_{600} of approximately 0.6. Four serial dilutions were done in Eppendorf tubes by serially pipetting 1 µL of phage into 50 µL of cells to yield the following dilutions: 1/50, 1/2500, 1/125000, and 1/6250000. 650 µL of top agar (0.7% agar with LB media) was added to each tube, which was then immediately spread onto a quad plate containing bottom agar (1.5% agar with LB media). Plates were incubated overnight at 37 ºC. Plaques were counted the following day, and plaque forming units (PFU) per mL was calculated using the following equation:

$$PFU = 1000 * A * 50^{4-B}$$

Where *A* is the number of plaques in a given quadrant, and *B* is the quadrant number where the phage were counted, in which 1 is the least dilute quadrant and 4 is the most dilute quadrant.

Phage growth assays

Phage growth assays were performed by adding the following to a culture tube and shaking at 37 ºC for 6 h: 1 mL of LB with the appropriate antibiotics (carbenicillin and kanamycin), 10 µL of saturated S1030 *E. coli* containing the accessory plasmids of interest, and ~1000 phage. Phage were then isolated by centrifugation, and PFU was determined by plaque assays using 1059 *E. coli* and the plaque assay protocol described above.

High-throughput sequencing library construction

PACE samples were collected from each lagoon every 24 hours. The lagoon samples were centrifuged at 13,000 rpm for 3 min on a bench top centrifuge to separate supernatant and cell pellet. The phage-containing supernatants were stored at 4 ºC prior to the creation of sequencing libraries. To prepare Illumina sequencing libraries, each phage sample was cultured overnight with 1059 *E. coli* cells, followed by phage DNA purification (Qiagen plasmid purification reagents buffer P1 (catalog number 19051), P2 (catalog number 19052), N3 (catalog number 19064), PE (catalog number 19065), and spin column for DNA (EconoSpin™, catalog number 1920-250). The resulting DNA concentration was ~50 ng/µL. Freshly generated DNA samples were then used as template for PCR amplification. For each library sample, we amplified three overlapping fragments of the BCL-2 family protein, which are 218~241bp in length (Fig. S21). Each primer also included 6-9 "N"s to introduce length variation (Supplementary Table 13). In total, 12 PCR reactions were used for each library. Phusion DNA polymerases and buffers (ThermoFisher Scientific, catalog number F518L) were used in the first PCR

round to amplify all three fragments for all library sequencing. The 25 µL reaction contained: 0.5 µL of 50 mM $MgCl_2$, 0.75 µL of 10 mM dNTP, 0.75 µL Phusion DNA polymerase, 20 ng library DNA, 0.5 µL of 10µM primer (each). The PCR reactions were run on a C1000 Touch Thermal Cycler (ThermoFisher Scientific), with the following parameters: 98°C for 1 min, followed by 16 cycles of 98°C for 12 sec, 58°C for 15 sec and 72°C for 45 sec, and finally 72°C for 5 min. PCR reactions were purified using the ZYMO DNA clean and concentrator kit (catalog number D4013) and 96 well filter plate (EconoSpin, catalog number 2020-001). The DNA products were dissolved in 30 µL ddH_2O. All 12 reactions for each library were combined and 1 µL was used as the template for a second PCR round. PCR reaction components and thermocycler parameters were the same as above, except that the annealing temperature was 56 °C, and only 15 rounds of amplification were conducted. The primer and sample combinations are listed in Supplementary Table 13**.** PCR reactions were then purified following the same procedure as previous step. Equal volumes of all 72 library samples were combined and concentration was measured using a Qubit 4 Fluorometer. The total DNA sample was 2.68 ng/µL (equivalent to 10 nM, according to the average length of PCR fragments). DNA samples were diluted to 4 nM from step 4 following the Illumina MiSeq System Denature and Dilute Libraries Guide and then diluted to 12 pM for high-throughput sequencing. The final sample contained 100 µL of 20 pM PhiX spike-in plus 500 µL of the 12 pM library sample. Sequencing was performed on the Illumina MiSeq System using MiSeq Reagent Kit v3 (600-cycle) with paired-end reads according to the manufacturer's instructions.

Processing of Illumina data

Illumina sequencing yielded 22 million reads, 13 million of which could be matched to a specific sample (Supplementary Table 13). To process this data, we first used Trim Galore with default settings to trim reads based on quality. (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Then, we used BBMerge[76], a script in BBTools (https://jgi.doe.gov/data-and-tools/bbtools/), to merge paired-end reads. Next, we used Clumpify to remove repeated barcode sequences. We then used Seal to identify and bin reads by sample and fragment (Figure S21). Finally, we used BBDuk to remove any primer or adapter sequence present. Scripts and reference sequences are available on Github (https://github.com/JoeThorntonLab/BCL2.ChanceAndContingency)

Illumina Sequencing Analysis

Reads were first binned by experiment and then aligned to the appropriate WT sequence using Geneious (low Sensitivity, 5 iterations, gaps allowed). Sequences were then processed in R to remove sequences containing "N"s or that were not full length. Insertions found in less than 1% of the population and sites that extended outside of the coding region were also removed from all sequences. Remaining gaps were then standardized among replicates and within an experiment. Finally, allele frequencies were calculated for each site and amino acid, as well as remaining insertions and deletions.

Estimating the effects of chance and contingency on the outcomes of evolution

Estimating the effects of chance

Allele frequency differences among replicates started from the same genotype can only be caused by chance events. Thus, to determine the effects of chance (C_1) on the outcomes of evolution, we compared allele frequencies from replicate PACE experiments started from the same genotype. We compared allele frequencies of individual replicates to the average allele frequencies among replicates started from the same genotype by estimating the probability that two randomly chosen alleles would be different, i.e. the genetic variance, for each replicate individually and the pooled sample of all replicates from a given starting genotype:

$$C_1 \equiv \frac{1 - Q_g}{1 - Q_r}$$

where Q_r is the probability that two randomly chosen alleles in the same replicate are identical in state and Q_g is the probability that two randomly chosen alleles from the pooled replicates started from the same genotype are identical in state. C_1 is related to Wright's F-statistics[77] as:

$$F_{is} \equiv F_{rg} \equiv \frac{Q_r - Q_g}{1 - Q_g}$$
$$1 - F_{rg} = \frac{1 - Q_g - Q_r + Q_g}{1 - Q_g} = \frac{1 - Q_r}{1 - Q_g}$$
$$\frac{1}{1 - F_{rg}} = \frac{1 - Q_g}{1 - Q_r} = C_1$$

We used count data from Illumina sequencing to estimate allele frequencies and followed the approach of Hivert *et al.* 2018, which developed a methods of moments estimator, $\hat{F}_{st}^{pool}$, that is appropriate for pooled data and accounts for both the sampling of individuals within a population and the sampling of reads during sequencing. We treated each amino acid site independently and defined the following:

$R_{1:rgs} \equiv$ # of reads in replicate r of starting genotype g at site s
$\hat{\pi}_{a:rgs} \equiv$ observed allele frequency of allele a in replicate r of starting genotype g at site s
$\hat{\pi}_{a:gs} \equiv$ observed allele frequency of allele a in pooled replicates of starting genotype g at site s

Using these values, we used the estimator of $\hat{F}_{st}^{pool}$ defined in Hivert et al. 2018 to estimate F_{rg} for a single site:

$$\hat{F}_{st}^{pool} \equiv \hat{F}_{rg}^{s} \equiv \frac{MSG_{gs} - MSR_{gs}}{MSG_{gs} + (\mathcal{R}_{gs} - 1)MSR_{gs}}$$

Where:
$MSG_{gs} \equiv \frac{1}{\mathbb{R}_{1:gs} - \mathbb{R}_{2:gs}} \sum_{a=1}^{21} \sum_{r=1}^{n_{r:g}} R_{1:rgs}\left(\hat{\pi}_{a:rgs} - \hat{\pi}_{a:gs}\right)^2$ is the mean sum-of-squares for pooled replicates,
$MSR_{gs} \equiv \frac{1}{R_{1:gs} - \mathbb{R}_{1:gs}} \sum_{a=1}^{21} \sum_{r=1}^{n_{r:g}} R_{1:rgs}\, \hat{\pi}_{a:rgs}(1 - \hat{\pi}_{a:rgs})$ is the mean sum-of-squares within replicates,
$\mathcal{R}_{gs} \equiv \frac{R_{1:gs} - \frac{R_{2:gs}}{R_{1:gs}}}{\mathbb{R}_{1:gs} - \mathbb{R}_{2:gs}}$, is the effective number of individuals after accounting for sampling, with
$R_{1:gs} \equiv \sum_{r=1}^{n_{r:g}} R_{1:rgs}$,
$R_{2:gs} \equiv \sum_{r=1}^{n_{r:g}} R_{1:rgs}^2$,
$\mathbb{R}_{1:gs} \equiv \sum_{r=1}^{n_{r:g}} \frac{R_{1:rgs} + n_{i:rg} - 1}{n_{i:rg}}$, and
$\mathbb{R}_{2:gs} \equiv \frac{1}{R_{1:gs}} \sum_{r=1}^{n_{r:g}} \frac{R_{1:rgs}(R_{1:rgs} + n_{i:rg} - 1)}{n_{i:rg}}$.

Here, $n_{r:g}$ is the number of replicates started from genotype g and $n_{i:rg}$ is the number of individual phage in replicate r of starting genotype g in the sample used to make the sequencing library.

From the relationship between C_1 and F_{rg} we approximated the site-specific effects of chance for a particular starting genotype as:

$$\hat{C}_{1:g}^{s} \equiv \frac{1}{1-\hat{F}_{rg}^{s}} = \frac{1}{1-\frac{MSG_{gs}-MSR_{gs}}{MSG_{gs}+(\mathcal{R}_{gs}-1)MSR_{gs}}} = \frac{1}{\frac{MSG_{gs}+(\mathcal{R}_{gs}-1)MSR_{gs}-MSG_{gs}+MSR_{gs}}{MSG_{gs}+(\mathcal{R}_{gs}-1)MSR_{gs}}}$$
$$= \frac{MSG_{gs}+(\mathcal{R}_{gs}-1)MSR_{gs}}{\mathcal{R}_{gs}MSR_{gs}} = 1+\frac{MSG_{gs}-MSR_{gs}}{\mathcal{R}_{gs}MSR_{gs}}$$

When there were replicates from more than one starting genotype, we calculated MSG_{gs}, MSR_{gs}, and $\mathcal{R}_{gs}$ separately for each starting genotype and averaged these values together, using weights proportional to the number of replicates for that genotype. Thus:

$$\hat{C}_{1}^{s} = 1+\frac{\sum_{g=1}^{n_g} n_{r:g} \sum_{g=1}^{n_g} n_{r:g}(MSG_{gs}-MSR_{gs})}{\sum_{g=1}^{n_g} n_{r:g}\mathcal{R}_{gs} \sum_{g=1}^{n_g} n_{r:g}MSR_{gs}}$$

Where n_g is the number of distinct starting genotypes.

We then took the average numerator and average denominator as suggested by Hivert et al. 2018 and Weir and Cockerham 1984 for estimating F_{st}:

$$\hat{C}_{1} = 1+\frac{\sum_{s=1}^{n_s}\sum_{g=1}^{n_g} n_{r:g} \sum_{g=1}^{n_g} n_{r:gs}(MSG_{gs}-MSR_{gs})}{\sum_{s=1}^{n_s}\sum_{g=1}^{n_g} n_{r:g}\mathcal{R}_{gs} \sum_{g=1}^{n_g} n_{r:g}MSR_{gs}}$$

Where n_s is the number of sites.

Estimating the effects of contingency

To determine the effects of contingency (C_2) on the outcomes of evolution, we compared the average allele frequency of replicate PACE experiments between different starting genotypes. For each starting genotype, we pooled all replicates started from that genotype and treated it as a single sample. We compared allele frequencies among genotypes by estimating the probability that two randomly chosen alleles in a sample would be different if they were both drawn from the same starting genotype or drawn from different starting genotypes:

$$C_2 \equiv \frac{1-Q_{g1\neq g2}}{1-Q_{g1=g2}}$$

where $Q_{g1=g2}$ is the probability that two randomly chosen alleles from the same starting genotype are identical in state and $Q_{g1\neq g2}$ is the probability that two randomly chosen alleles from different starting genotypes are identical in state. To calculate $1-Q_{g1\neq g2}$ we note that the probability of two randomly drawn alleles being different when chosen irrespective of starting genotype is simply the average of the probability of two randomly drawn alleles being different when they are drawn from the same and different starting genotypes, i.e.:

$$(1-Q_t) = \frac{1}{2}\left(1-Q_{g1=g2}\right) + \frac{1}{2}\left(1-Q_{g1\neq g2}\right)$$

where Q_t is the probability that two randomly chosen alleles irrespective of starting genotype are identical in state. From this we have:

$$2(1-Q_t) = \left(1-Q_{g1=g2}\right) + \left(1-Q_{g1\neq g2}\right)$$
$$\left(1-Q_{g1\neq g2}\right) = 2(1-Q_t) - \left(1-Q_{g1=g2}\right)$$
$$\left(1-Q_{g1\neq g2}\right) = 1-2Q_t + Q_{g1=g2}$$

Using this and the fact that $Q_{g1=g2}$ is equivalent to Q_g used above to calculate the effects of chance, we have:

$$C_2 = \frac{1-2Q_t+Q_{g1=g2}}{1-Q_{g1=g2}} = \frac{1-Q_t}{1-Q_{g1=g2}} + \frac{Q_{g1=g2}-Q_t}{1-Q_{g1=g2}} = \frac{1-Q_t}{1-Q_g} + \frac{Q_g-Q_t}{1-Q_g}$$

Again, this statistic is related to Wright's F-statistics as:

$$F_{st} \equiv F_{gt} \equiv \frac{Q_g-Q_t}{1-Q_t}$$
$$1-F_{gt} = \frac{1-Q_t-Q_g+Q_t}{1-Q_t} = \frac{1-Q_g}{1-Q_t}$$

and $1+F_{gt} = \frac{1-Q_t}{1-Q_t} + \frac{Q_g-Q_t}{1-Q_t}$

$$\frac{1+F_{gt}}{1-F_{gt}} = \frac{\frac{1-Q_t}{1-Q_t} + \frac{Q_g-Q_t}{1-Q_t}}{\frac{1-Q_g}{1-Q_t}} = \frac{1-Q_t}{1-Q_g} + \frac{Q_g-Q_t}{1-Q_g}$$

As with the effects of chance, we used the method of moments estimator defined in Hivert et al. 2018 to estimate the effects of contingency:

$$\hat{F}_{gt} \equiv \frac{MST_s - MSG_s}{MST_s + (\mathcal{R}_s - 1)MSG_s}$$

Where:

$MST_s \equiv \frac{1}{\mathbb{R}_{1:s}-\mathbb{R}_{2:s}} \sum_{a=1}^{21} \sum_{g=1}^{n_g} R_{1:gs}\left(\hat{\pi}_{a:gs} - \hat{\pi}_{a:s}\right)^2$ is the mean sum-of-squares for the entire pooled sample

$MSG_s \equiv \frac{1}{R_{1:s}-\mathbb{R}_{1:s}} \sum_{a=1}^{21} \sum_{g=1}^{n_g} R_{1:gs}\, \hat{\pi}_{a:gs}(1-\hat{\pi}_{a:gs})$ is the mean sum-of-squares for a starting genotype,

$\mathcal{R}_{ts} \equiv \frac{R_{1:s} - \frac{R_{2:s}}{R_{1:s}}}{\mathbb{R}_{1:s}-\mathbb{R}_{2:s}}$, is the effective number of individuals after accounting for sampling, with

$\hat{\pi}_{a:s} \equiv$ the observed allele frequency of allele a among all genotypes at site s,

$R_{1:s} \equiv \sum_{g=1}^{n_g} R_{1:gs}$,

$R_{2:s} \equiv \sum_{g=1}^{n_g} R_{1:gs}^2$,

$\mathbb{R}_{1:s} \equiv \sum_{g=1}^{n_g} \frac{R_{1:gs}+n_{i:g}-1}{n_{i:g}}$, and
$\mathbb{R}_{2:s} \equiv \frac{1}{R_{1:s}} \sum_{g=1}^{n_g} \frac{R_{1:gs}(R_{1:gs}+n_{i:g}-1)}{n_{i:g}}$.

With $n_{i:g}$ being the number of individual phage used to make the libraries for starting genotype g.

From the relationship between F_{gt} and C_2, we approximated the effects of contingency as:

$$\hat{C}_2^s = \frac{1+\hat{F}_{gt}^s}{1-\hat{F}_{gt}^s} = \frac{1+\frac{MSP_{ts}-MSI_{ts}}{MSP_{ts}+(\mathcal{R}_{ts}-1)MSI_{ts}}}{1-\frac{MSP_{ts}-MSI_{ts}}{MSP_{ts}+(\mathcal{R}_{ts}-1)MSI_{ts}}}$$

$$= \frac{\frac{MSP_{ts}+(\mathcal{R}_{ts}-1)MSI_{ts}+MSP_{ts}-MSI_{ts}}{MSP_{ts}+(\mathcal{R}_{ts}-1)MSI_{ts}}}{\frac{MSP_{ts}+(\mathcal{R}_{ts}-1)MSI_{ts}-MSP_{ts}+MSI_{ts}}{MSP_{ts}+(\mathcal{R}_{ts}-1)MSI_{ts}}}$$

$$= \frac{MSP_{ts}+(\mathcal{R}_{ts}-1)MSI_{ts}+MSP_{ts}-MSI_{ts}}{MSP_{ts}+(\mathcal{R}_{ts}-1)MSI_{t}s-MSP_{ts}+MSI_{ts}}$$

$$= \frac{\mathcal{R}_{ts}MSI_{ts}+2MSP_{ts}-2MSI_{ts}}{\mathcal{R}_{ts}MSI_{ts}}$$

$$= 1+2\frac{MSP_{ts}-MSI_{ts}}{\mathcal{R}_{ts}MSI_{ts}}$$

Again, we treated all sites as independent and summed the numerator and denominators to estimate the effects of contingency:

$$\hat{C}_2 = 1+2\frac{\sum_{s=1}^{n_s}(MSP_{ts}-MSI_{ts})}{\sum_{s=1}^{n_s}\mathcal{R}_{ts}MSI_{ts}}$$

Combined effect of chance and contingency

To determine the combined effects of chance and contingency (C_3) on the outcomes of evolution, we compared allele frequencies from individual replicates to the average allele frequency among replicates from different starting genotypes. In each case, we pooled replicates started from a genotype and treated it as a single sample and compared it to the individual replicates started from different genotypes. We compared allele frequencies by estimating the probability that two randomly chosen alleles would be different if they were both drawn from the same replicate or if they were drawn from a different starting genotype:

$$C_3 \equiv \frac{1-Q_{g1\neq g2}}{1-Q_r} = \frac{1-Q_g}{1-Q_r} * \frac{1-Q_{g1\neq g2}}{1-Q_g} = C_1 * C_2$$

We thus used:

$$\hat{C}_3 = \hat{C}_1 * \hat{C}_2$$

as our estimate of the combined effects of chance and contingency.

Citations

50. Levin, K. B. *et al.* Following evolutionary paths to protein-protein interactions with high affinity and selectivity. *Nat. Struct. Mol. Biol.* **16**, 1049–55 (2009).
51. Pu, J., Chronis, I., Ahn, D. & Dickinson, B. C. A panel of protease-responsive RNA polymerases respond to biochemical signals by production of defined RNA outputs in live cells. *J. Am. Chem. Soc.* **137**, 15996–15999 (2015).
52. Schrödinger, L. *The {PyMOL} Molecular Graphics System, Version~2.0.7*. (2018).
53. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
54. Miller, H. C., Biggs, P. J., Voelckel, C. & Nelson, N. J. De novo sequence assembly and characterisation of a partial transcriptome for an evolutionarily distinct reptile, the tuatara (Sphenodon punctatus). *BMC Genomics* **13**, (2012).
55. Wyffels, J. *et al.* SkateBase, an elasmobranch genome project and collection of molecular resources for chondrichthyan fishes. *F1000Research* **3**, (2014).
56. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
57. Hughes, L. C. *et al.* Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc. Natl. Acad. Sci.* 201719358 (2018). doi:10.1073/pnas.1719358115
58. Smith, J. J. *et al.* The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat. Genet.* **50**, 270–277 (2018).
59. Takechi, M. *et al.* Overview of the transcriptome profiles identified in hagfish, shark, and bichir: Current issues arising from some nonmodel vertebrate taxa. *J. Exp. Zool. Part B Mol. Dev. Evol.* **316 B**, 526–546 (2011).
60. Delsuc, F. *et al.* A phylogenomic framework and timescale for comparative studies of tunicates. *BMC Biol.* **16**, 1–14 (2018).
61. Reich, A., Dunn, C., Akasaka, K. & Wessel, G. Phylogenomic analyses of echinodermata support the sister groups of Asterozoa and Echinozoa. *PLoS One* **10**, 1–11 (2015).
62. Riesgo, A., Farrar, N., Windsor, P. J., Giribet, G. & Leys, S. P. The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Mol. Biol. Evol.* **31**, 1102–1120 (2014).
63. Moroz, L. L. *et al.* The ctenophore genome and the evolutionary origins of neural systems. *Nature* **510**, 109–114 (2014).
64. De Laval, V. R., Deléage, G., Aouacheria, A. & Combet, C. BCL2DB: Database of BCL-2 family members and BH3-only proteins. *Database* **2014**, 1–7 (2014).
65. Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
66. Kozlov, A. M. & Stamatakis, A. Using RAxML-NG in Practice. *Preprints* 1–25 (2019). doi:10.20944/preprints201905.0056.v1
67. Finnigan, G. C., Hanson-Smith, V., Stevens, T. H. & Thornton, J. W. Evolution of increased complexity in a molecular machine. *Nature* **481**, 394–398 (2012).
68. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
69. Zinkus-Boltz, J., Devalk, C. & Dickinson, B. C. A Phage-Assisted Continuous Selection Approach for Deep Mutational Scanning of Protein-Protein Interactions. *ACS Chem. Biol.* **14**, 2757–2767 (2019).
70. Zhou, H. *et al.* Evolution of a reverse transcriptase to map N 1-methyladenosine in human messenger RNA. *Nat. Methods* **16**, 1281–1288 (2019).

71. Zhang, H., Nimmer, P., Rosenberg, S. H., Ng, S. C. & Joseph, M. Development of a high-throughput fluorescence polarization assay for Bcl-xL. *Anal. Biochem.* **307**, 70–75 (2002).
72. Zhou, H. *et al.* Characterizing Watson-Crick versus Hoogsteen Base Pairing in a DNA-Protein Complex Using Nuclear Magnetic Resonance and Site-Specifically 13C- and 15N-Labeled DNA. *Biochemistry* **58**, 1963–1974 (2019).
73. Pu, J., Disare, M. & Dickinson, B. C. Evolution of C-Terminal Modification Tolerance in Full-Length and Split T7 RNA Polymerase Biosensors. *ChemBioChem* **20**, 1547–1553 (2019).
74. Badran, A. H. & Liu, D. R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015).
75. Hubbard, B. P. *et al.* Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat. Methods* **12**, 939–942 (2015).
76. Bushnell, B., Rood, J. & Singer, E. BBMerge – Accurate paired shotgun read merging via overlap. *PLoS One* **12**, 1–15 (2017).
77. Wright, S. THE GENETICAL STRUCTURE OF POPULATIONS. *Ann. Eugen.* 323–354 (1951).
78. Hivert, V., Leblois, R., Petit, E. J., Gautier, M. & Vitalis, R. Measuring Genetic Differentiation from Pool-seq Data. *Genetics* **210**, genetics.300900.2018 (2018).
79. Weir, B. S. & Cockerham, C. C. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358–1370 (1984).

Acknowledgements
We thank members of the Dickinson and Thornton groups for helpful comments on the manuscript, Somayeh Ahmadiantehrani for editing, and R. Ranganathan for the use of the Illumina MiSeq instrument. This work was supported by CAREER Award #1749364 from the National Science Foundation (NSF) to B.C.D and NIH R01GM131128 and R01GM121931 to J.W.T. V.C.X was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1746045. B.P.H.M. was supported by a National Institutes of Health NRSA award (F32GM122251).

Contributions
All authors contributed to conception of the project. BCD, VCX, and JP designed the PACE dual-selection system, which VCX and JP engineered, optimized, and implemented. VCX and JP performed PACE, biochemical assays, and sequencing experiments, with input from BPHM. BPHM and JWT developed and designed the evolutionary and genetic analyses. BPHM led and performed the phylogenetic, genetic, and evolutionary analyses, with input from VCX and JP. All authors contributed to writing the manuscript.

Ethics Declaration
J.P. and B.C.D. have a patent on the proximity-dependent split RNAP technology used in this work. The content is solely the responsibility of the authors and the funders had no input on the study design, analysis, or conclusions.

Data availability
Raw high-throughput sequencing data are available on SRA (accession number PRJNA647218). Processed HTS data are available on Dryad (https://datadryad.org/stash/share/Ty32n-2c8nUZuiFegxDdy7wNeGDEsqlwtL8tOKz1RWU).

Code availability
Code used for analysis and phylogenetic data are available at
https://github.com/JoeThorntonLab/BCL2.ChanceAndContingency