Synthetic biology as an empirical tool for evolutionary theory

Authors:

Xueying Li[1], Paco Majic[2] & Cauã Antunes Westmann[3,4*]

*Corresponding author email: caua.antuneswestmann@epfl.ch

Affiliations:

[1] Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, China

[2] European Molecular Biology Laboratory, 69117 Heidelberg, Germany

[3] Institute of Bioengineering, School of Engineering, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland

[4] Swiss Institute of Bioinformatics, Quartier Sorge-Batiment Genopode, 1015, Lausanne, Switzerland

Abstract

Evolutionary biology has traditionally inferred process from patterns in extant organisms and the fossil record, leaving many foundational questions constrained by their historical nature. Over the past two decades, synthetic and high-throughput approaches — including deep mutational scanning, genome editing, ancestral sequence reconstruction, engineered mutators, and random-sequence assays — have made it possible to test these questions directly by constructing, perturbing, and replaying evolutionary systems. Here, we review how these approaches reshape several foundational questions: the distribution of mutational effects, the structure and navigability of fitness landscapes, the evolution of evolvability, developmental constraint, historical contingency, and the engineering of evolutionary systems. Across these domains, synthetic experiments are exposing unexpected mechanistic detail that refines and extends classical theory — revealing, for example, how strongly mutational effects depend on environmental and genetic context, how ruggedness can coexist with broad accessibility on fitness landscapes, how genotype–phenotype maps are intrinsically biased and heterogeneous, and how random sequences carry latent functional potential that may serve as raw material for later innovation. As these technologies continue to expand the empirical reach of evolutionary biology, theory in turn sharpens the questions they are best suited to address — an iterative dialogue between experiment and theory that brings us closer to understanding how life evolves.

Significance Statement:

Evolutionary biology has long faced a fundamental limitation: evolution unfolds over deep time, and we can study only the organisms and fossils that happen to have survived — leaving many central ideas about how evolution works difficult to test directly. New laboratory tools now allow

scientists to build, modify, and replay evolutionary systems from scratch, providing direct experimental tests of these ideas. This review shows that such experiments are refining — and sometimes overturning — long-held views, for example by revealing that mutational effects depend strongly on context, that many evolutionary trajectories are shaped by historical contingencies, and that even random DNA can carry biological function. Together, these advances shift the field from observing evolution to experimentally probing its underlying molecular mechanisms.

Keywords: Evolutionary Theory; Molecular Evolution; Synthetic Biology; Adaptive Landscapes; Historical Contingencies; Ancestral Sequence Reconstruction

1. Introduction

Evolutionary biology has long been characterized by a productive interplay between theory and empirical observation, with each continually informing and refining the other. Early empirical challenges to transformist ideas, such as comparisons of ancient and extant organisms, foreshadowed the more systematic synthesis achieved by Charles Darwin in *On the Origin of Species*[1], where conceptual reasoning was tightly interwoven with diverse lines of evidence. Subsequent discoveries, notably the fossil of *Archaeopteryx*, reinforced evolutionary predictions by revealing transitional forms. Over time, major theoretical advances have provided unifying frameworks: the foundations of population genetics laid by Ronald A. Fisher[2], J. B. S. Haldane[3,4], and Sewall Wright[5,6] formalized the roles of mutation, selection, drift, and migration, while later developments—including the neutral theory advanced by Motoo Kimura[7–9] and Tomoko Ohta[10], as well as models of adaptive dynamics and evolutionary game theory—extended evolutionary analysis into new domains. In parallel, empirical support has expanded across genetics, paleontology, comparative biology, biogeography, and controlled experiments, from early work in *Drosophila* to long-term microbial evolution studies. Together, these intertwined empirical and theoretical traditions have transformed evolutionary biology into a quantitatively grounded discipline capable of explaining and predicting patterns of biological diversity.

Yet the sources of empirical information have been limited by the inescapable fact that evolution is ultimately a historical process. Thus, many of the questions in evolutionary theory have been constrained by inferring what processes have underlain the observations of "design features of 'finished products'"[11]. The study of evolutionary processes has been greatly improved by the capacity to run evolutionary simulations using computational programmes[11–14]. Though useful to investigate dynamical processes and to act as proof-of-principle of potential evolutionary avenues, such simulations were still restricted to stark simplifications of biological parameters. Over the last few decades, however, new molecular and synthetic approaches have provided complementary experimental tools that allow us to revisit foundational ideas directly in the laboratory. Techniques such as high-throughput DNA sequencing, genome editing, and synthetic gene construction now enable researchers to directly test classical evolutionary hypotheses, often affirming their relevance while also challenging long-held assumptions[15–17].

These approaches, which we will broadly refer to as "synthetic approaches", as opposed to the "traditional" empirical evidence based on sampling natural variation, have become a strong experimental tool that has allowed us to empirically fill the observational gaps produced by extinction, and to explore the space of possibilities beyond the realm biological evolution has explored over the eons that precede us. In an analogous way to comparative morphology, sequence comparisons between different species allow identification of possible changes that relate to phenotypic changes that lineages followed. But the power of synthetic biology is that, by reconstructing the mutational steps that a lineage took to evolve a novel trait in the context of other organisms, it is possible to assay the extent at which a genetic change may have caused phenotypic traits. For example, synthetic biology has been useful to examine the genetic causes of the limblessness of snakes[18], the eyespots of butterfly wings[19], the chiropatagium of bats[20], the wing-like fins of flying fish[21], or the exaggeration of sexually selected traits like the boxing first pair of legs of *Drosophila prolongata*[22].

More broadly, synthetic biology is propelling evolutionary research into new empirical territory, not only by dissecting the mechanistic basis of morphological and physiological change but also by enabling direct tests of the rules governing evolutionary processes. In this way, it has become a powerful tool for advancing evolutionary theory. In this review, we revisit classical questions in the field and examine how cutting-edge experimental approaches are providing fresh insights into them. Focusing on six major questions in evolutionary theory, we highlight representative examples in which synthetic methods bridge theoretical concepts and empirical data and emphasize how these emerging tools are reshaping our understanding of evolutionary dynamics, concluding with a forward-looking perspective on open questions and future directions.

2. Foundational questions revisited through synthetic approaches

2.1. The interplay between mutations and selection

Genetic mutations play a central role in evolution by providing the ultimate source of heritable variation upon which natural selection and other evolutionary forces act. Despite this fundamental role, the evolutionary importance of mutations themselves—their rates, spectra, and biases—has been a recurrent point of debate. During the early twentieth century, mutationist perspectives gained prominence during the so-called "eclipse of Darwinism"[23], emphasizing the creative role of mutations in shaping evolutionary change. These views were later largely relegated by a selectionist framework, grounded in the assumptions that individual mutational effects are typically small relative to the strength of selection, and that populations harbor sufficient standing genetic variation to respond rapidly to environmental change. The elucidation of the structure of DNA and the genetic code transformed our understanding of the nature of mutations and their evolutionary consequences, leading to major conceptual advances such as the neutral theory of molecular evolution, renewed interest in gene duplication as a driver of novelty, and the rise of comparative and population genomics. These molecular advances have promoted the idea of "mutation-driven evolution" [24].

The ability to study mutations in a synthetic and highly controlled manner has further highlighted the evolutionary relevance of the mutational process itself. High-throughput generation of large numbers of sequence variants, their efficient introduction into biological systems, and precise manipulation of the molecular machinery that governs mutation rates and spectra now allow direct experimental interrogation of how mutational processes shape evolutionary trajectories. In combination with experimental evolution approaches, artificial genetic constructs and engineered mutational regimes make it possible to explore evolutionary dynamics under conditions that both recapitulate and extend beyond those encountered in natural systems (**Figure 1**).

The distribution of fitness effects of mutations is dynamic

The relative importance of selection and mutational input on the potential of affecting the evolutionary process depends primarily on (i) the rate at which mutations occur, and (ii) the phenotypic and fitness effects of individual mutations. Our capacity to estimate both of these parameters has substantially improved over the last decades. In the case of the effects of mutations (for an overview of advances in mutation rate estimations see Padilla-Iglesias and Majic 2026[25]), synthetic biology has been fundamental. Early efforts to estimate the distribution of fitness effects of mutations (DFE) relied on indirect approaches, most prominently mutation–accumulation experiments in *Drosophila*[26,27], bacteria[28], and yeast[29]. By relaxing natural selection and allowing mutations to accumulate, these studies quantified the resulting declines in fitness and inferred that most new mutations are deleterious, with a wide distribution of effects. Parallel insights came from comparative genomics, where patterns of synonymous and nonsynonymous substitutions were interpreted through the lens of the neutral and nearly neutral theories[7,10,30,31]. Together, these classical approaches established the prevailing view that the DFE is skewed toward neutrality and deleteriousness, with beneficial mutations being rare—even though the precise shape of the distribution remained difficult to resolve.

Synthetic biology techniques have added much-needed nuance to classical interpretations of mutational effects. In particular, the advent of deep mutational scanning has enabled direct, large-scale measurement of the fitness consequences of hundreds of thousands of specific gene variants in parallel. These studies challenge the simplistic view that most mutations are strongly deleterious, instead revealing different properties of DFEs such as: (i) most single mutations are neutral or slightly deleterious, with only a small fraction being beneficial[32,33]; (ii) the shape of the DFE depends strongly on environmental context [34,35]; (iii) saturating proteins with mutations has revealed pervasive epistasis, whereby the effect of a mutation depends on the presence of others [36,37]; (iv) many DFEs are well described by truncated exponential distributions, such that small shifts in the mean markedly affect the probability of rare beneficial variants, highlighting the importance of the distribution's "tail" for adaptation[30,38]; and (v) the mean effect of new mutations often declines with increasing background fitness, suggesting that physiological and environmental constraints modulate the DFE [39–41].

Together, these results illustrate that DFEs are not static properties of mutations alone, but dynamic features shaped by molecular context, environment, and evolutionary history.

The fitness effect of synonymous mutations

The classical neutral theory proposed by Motoo Kimura treated most mutations at the molecular level as selectively neutral and regarded synonymous substitutions in particular as a paradigmatic example of neutral variation, given that they do not alter amino acid sequences[7,8]. Synonymous changes, however, are now recognized as potentially having substantial functional and fitness consequences[33,42–47], by altering transcription, mRNA splicing, stability, and secondary structure, as well as translation efficiency and co-translational protein folding[42]. However, recent evidence based on high-throughput experiments has been controversial: in yeast, many synonymous variants were found strongly non-neutral[48,49], although technical concerns have been raised with the finding[50]. Similar results have been observed in bacteria, where synonymous edits in *E. coli* can impair growth in a condition-dependent manner[45]. Beyond deleterious outcomes, synonymous mutations have also been implicated in adaptation, for example in experimentally evolved *Pseudomonas fluorescens* populations[44]. However, high-throughput genome-editing experiments in human cells supported the traditional view that synonymous mutations rarely have fitness effects[51]. Therefore, a conclusive understanding about the neutrality of synonymous mutations awaits more empirical data.

Mutational biases may shape evolutionary trajectories

Classical population genetics often assumed mutation to be isotropic, with equal probability of change in all directions of sequence space, but this view has been overturned by accumulating evidence that mutation processes are inherently biased (e.g., transition–transversion asymmetries and pervasive GC→AT biases). While it was traditionally argued that natural selection would dominate over such biases except under neutrality, recent theoretical and experimental work shows that mutation bias can influence the trajectory of adaptation by shaping the pool of available variants. Some of this evidence comes from experiments focused on synthetic biology approaches. A clear experimental demonstration comes from work by Deepa Agashe and colleagues in *Escherichia coli*, who engineered strains with altered mutation spectra by modifying DNA replication machinery[52]. By effectively rotating the direction of mutational bias, they showed that the supply and distribution of beneficial mutations depend strongly on mutation spectrum. Strains biased toward mutation types rare in the wild type accessed a broader set of beneficial mutations, whereas those exaggerating the native bias rapidly depleted accessible beneficial variation, exhibiting up to an order-of-magnitude reduction in beneficial mutations in the distribution of fitness effects[52]. These results provide direct empirical support for the prediction that shifts in mutation bias can redirect adaptive evolution by altering which mutational paths are available, allowing populations with novel biases to access adaptive solutions that might otherwise remain unexplored.

Gene duplication as a source of innovation

Gene duplication has long been considered a central engine of evolutionary innovation. Susumu Ohno proposed that duplication frees one gene copy from selective constraint, allowing it to accumulate mutations that may eventually confer a new function (“neofunctionalization”), while the other copy preserves the ancestral role[53]. Synthetic approaches have begun to test this scenario directly. The Schaerli group[54] constructed *E. coli* strains carrying either one or two copies of a fluorescent protein gene and evolved them under selection for altered color output. Two-copy populations indeed accumulated more mutations and tolerated greater genetic diversity, consistent with relaxed constraint, yet they did not consistently evolve the target trait faster than single-copy strains. Often, one duplicate decayed into a pseudogene before it could be repurposed, illustrating that duplication enables but does not guarantee innovation. These findings nuance Ohno’s model by showing that the fate of duplicates depends on additional factors—selection regime, regulatory context, and the chance order of mutations.

A complementary framework, the escape from adaptive conflict (EAC) model, proposes that an ancestral gene performing multiple functions can be constrained by trade-offs, and that duplication allows each copy to specialize and resolve this conflict[55,56]. Support for EAC comes largely from ancestral reconstruction studies (see Section 2.5), which show how combining forward evolution and historical approaches refines classical views of innovation. Classical models further assume that loss-of-function mutations in one copy are neutral. However, a recent study has shown that when duplicated genes interact, a defective copy can impair its partner, meaning purifying selection can maintain both copies longer than expected. For example, a deep mutational scan of a tetrameric protein found that ~6% of harmful mutations disrupt the second copy, helping preserve redundancy over evolutionary time [57].

2.2. The navigability of fitness landscapes

Theoretical efforts to understand adaptation have long relied on spatial metaphors that map evolutionary change onto abstract representations of fitness. Sewall Wright's fitness landscape provided an intuitive framework in which populations move across adaptive peaks and valleys under the combined influences of mutation, selection, drift, and recombination[5,6]. Complementing this view, Fisher's geometric model of adaptation formalized how mutations of varying effect sizes influence fitness in high-dimensional phenotypic space, predicting that small-effect mutations are more likely to be beneficial as populations approach an adaptive optimum[2]. Subsequent theoretical work explored how the topology of fitness landscapes influences evolutionary dynamics. Models of rugged fitness landscapes, including those developed by Kauffman, predicted that increasing epistasis and landscape ruggedness can constrain adaptive walks, trap populations on local optima, and generate complex evolutionary trajectories[58–61]. The fitness landscape metaphor treats evolution as populations moving on a high-dimensional map, where the location corresponds to genotype and the elevation corresponds to fitness[2,5,6]. A central question is ruggedness—the prevalence of local peaks (genotypes whose single-mutational neighbors are all lower in fitness) and valleys (steps that reduce fitness)—and whether ruggedness traps populations on suboptimal solutions[58–61]. Classic theory debated how populations could traverse valleys.

Advances in synthetic biology have enabled direct empirical mapping of fitness landscape topology at unprecedented scale. For example, an exhaustive study of the antibiotic resistance enzyme dihydrofolate reductase in *Escherichia coli* quantified the fitness of over 260,000 genotypes using genome editing and competitive assays, revealing a highly rugged landscape with numerous local peaks[62] (**Figure 2A**). Despite this extensive epistasis, the global fitness optimum remained broadly accessible, with many mutational trajectories proceeding uphill without traversing deep fitness valleys[62]. Similar "rugged yet navigable" architectures have been observed in regulatory DNA, including transcription factor binding sites mapped via massively parallel reporter assays (**Figure 2B**), where compensatory changes in core motifs, flanking regions, or spacing create ridge-like paths that maintain function[63–65] (**Figure 2C**). Other high-throughput studies across fluorescent proteins[32], enzymes[33,66], regulatory elements[67,68], and viral proteins[69,70] likewise report pervasive epistasis alongside neutral networks[8,31,71] and multiple accessible optima[72,73]. Together, these findings suggest that while biological fitness landscapes are complex, their structure often permits evolutionary exploration through "multidimensional bypasses"[74], and connected ridges and plateaus, mitigating the constraining effects of local peaks. Rather than trapping populations, ruggedness may instead reduce predictability, as stochastic mutational paths can lead different populations to distinct adaptive peaks.

High-throughput fitness assays in microbes have also enabled systematic characterization of genotype-by-environment (G×E) interactions, revealing dynamic "fitness seascapes"[75–79] (**Figure 2C**) in which peaks and ridges shift with external conditions such as drug concentration or temperature. This is exemplified by high-throughput mutational studies of the molecular chaperone Hsp90 across diverse stressors (e.g., heat, salinity, oxidative stress)[34,80,81], as well as by analyses of promoter and transcription factor binding site variation[82]. These studies show that both ruggedness and accessibility are environment-dependent: a given genotypic neighborhood may be highly constrained in one condition yet readily navigable in another. Accordingly, predictive power improves when models are formulated in joint GxE space, incorporating environment-dependent global epistasis. The environmental dependence of the distribution of fitness effects is thus mirrored at the landscape level, with context reshaping not only mutational effects but also the topography traversed during adaptation.

2.3. The evolution of evolvability

Dawkins asked whether evolvability can itself evolve[12]. Essentially, organismal features can be selected such that they pre-condition the range of adaptive solutions that become available. Dawkins referred to this as the "evolution of evolvability"—certain features act as "evolutionary watersheds" that can occasionally "open the floodgates to future evolution".

Cryptic variation

The concept of cryptic genetic variation as an evolutionary resource gained molecular footing when Rutherford and Lindquist showed that the chaperone Hsp90 acts as a phenotypic capacitor in Drosophila: under stress, previously buffered genetic variation is released, producing diverse morphological phenotypes[83]. Similar capacitor behaviour has since been documented for chromatin regulators, prions, and DNA methylation[84]. But can cryptic variation actually accelerate adaptation? Recent experiments have directly tested the role of cryptic variation in evolution. In one study, researchers accumulated mutations in a gene encoding a fluorescent protein in *E. coli* under conditions where those mutations had no effect on the protein's function (thus building up cryptic variation in the population). They then challenged these lineages to evolve a new function—shifting the protein's fluorescence from yellow to green under strong selection[85]. The results were striking: populations that harbored cryptic genetic variation evolved novel adaptive phenotypes faster and reached higher fitness peaks than populations without that cryptic pre-variation [85]. The cryptically mutagenized populations explored more diverse genotypic solutions and ended up with distinct mutations, whereas the non-cryptic (unmutated) populations tended to converge on the same limited set of changes. In essence, cryptic variation "opened up" access to diverse adaptive peaks, allowing different replicates to discover otherwise inaccessible high-fitness genotypes. This empirically demonstrates a long-suspected principle: genetic variation that is initially neutral can set the stage for later evolutionary leaps when the environment shifts. It also showed that cryptic variation increases the contingency and unpredictability of evolution—different populations, even in identical conditions, took divergent paths and outcomes because they had different hidden mutations available to exploit [85]. This finding has deep theoretical implications. It affirms that the past evolutionary history (even neutral drift) can profoundly influence future adaptation (a form of evolutionary "memory"). It also provides a mechanistic basis for evolvability: robustness (tolerance of mutations) and adaptability are two sides of the same coin, where robustness allows accumulation of cryptic changes that later give adaptability when conditions change.

Evolvability as an object of selection

Evolvability itself can also be a direct target of selection. In a recent long-term experiment with the bacterium *Pseudomonas fluorescens*, Barnett et al.[86] (2025) forced populations to repeatedly switch back and forth between two ecological states, to select for populations with greater capacity to generate changes. Over time, one lineage evolved a special kind of "hypermutable" DNA sequence in a regulatory gene: a short repeat that made it very easy for mutations to occur. This repeat acted like a genetic switch, producing rapid and reversible changes that allowed the bacteria to adapt again and again. This experiment showed, for the first time in real time, that natural selection can favor genetic features that make a lineage more adaptable, proving that evolvability is not just a by-product of evolution but a trait that can itself evolve. In a related vein, experiments on protein evolvability have revealed how selection can shape future potential. One study[74] evolved yellow fluorescent protein under strong selection, then shifted to select a new color. The variants that emerged under strong selection were also more robust to mutations (better foldability), which allowed them to evolve the new function more quickly. Similarly, Bloom et al.[87] showed that proteins evolving in large populations tend to become more robust and stable, improving evolvability. These findings support the idea that selective regimes—including alternation of functional pressures or variation in selection strength—can shape not only present adaptation but also the potential for future adaptation.

2.4. Developmental constraints

The question of evolvability is further related to the topology of genotype–phenotype maps. For much of the 20th century, evolutionary theory treated variation as essentially random with respect to need—the Modern Synthesis emphasized natural selection as the creative force that shapes organisms, given raw material from mutations. However, not all phenotypic variation is equally likely: the way organisms are built (their developmental biology) can bias the variants that appear[88,89]. Developmental processes can favor some forms over others, not only limiting what is possible, but also facilitating the capacity to adapt and diversify[90].

Synthetic biology tools have been useful to explore this question empirically at the molecular level. A long-standing argument holds that regulatory changes drive morphological evolution because they are more modular and less pleiotropic than coding mutations[91–93], but recent evidence suggests that *cis*-regulatory mutations are often more pleiotropic than previously appreciated. Deep mutational scanning is beginning to resolve this debate: DMS of the yeast *TDH3* promoter shows that *cis*-regulatory mutations alter expression level and noise but cause milder fitness effects and lower pleiotropy than *trans*-regulatory mutations[94,95]. By contrast, DMS in developmental contexts—made tractable by high-throughput imaging—reveals that single point mutations in *Drosophila* enhancers can have large effects on the level and pattern of target gene expression[96,97], challenging the assumption of low pleiotropy and highlighting the dense regulatory information packed into developmental enhancers.

To explore developmental constraint experimentally, deep mutational scans in developmental contexts become a useful tool. In terms of regulatory mutations—which are prime drivers of developmental evolution—it is possible to explore the effects of mutations using reporter constructs that become active in developmental contexts. Some examples using these systems have shown how developmental constraint manifests itself in highly pleiotropic enhancer mutations and the difficulty in generating new patterns with a few mutations[96–98]. Random sequences may have high regulatory potential in *Drosophila* embryos, but their regulatory activity may be highly contingent on pre-established chromatin accessibility[98]. More empirical evidence is still needed for evaluating the ruggedness of the fitness landscape of developmental networks, but current evidence hints that it could be highly constrained. It would also be of interest to compare developmental networks to gene regulatory networks underlying other traits, e.g., metabolism or behavior, in order to understand whether high constraint is a unique property of development or a general feature of complex networks.

2.5. The historical contingency of evolutionary change

Despite their enduring influence, the power of selection as the driving force of trajectories over fitness landscapes has been debated. Gould and Lewontin famously questioned the adaptationist tendency to interpret biological traits as optimal solutions to selective problems, emphasizing instead the roles of historical contingency, developmental constraint, and correlated trait evolution[99] —all of which can affect the ruggedness, dimensionality, and accessibility of fitness landscapes and can, therefore, profoundly shape evolutionary outcomes, independent of selection strength alone. In the context of this criticism, Gould famously framed this issue through the metaphor of "replaying the tape of life," asking whether evolution would retrace similar paths if history were reset, or whether chance events would lead to fundamentally different outcomes[100]. This question lies at the heart of debates about evolutionary determinism and about the extent to which selection, given similar starting conditions, leads to convergent solutions versus historically contingent trajectories[101].

A common challenge in evolutionary biology is that we can observe only the present, not the past—yet evolution is inherently a historical process. An interesting approach to address these questions has come from our capacity to recreate the paths evolution took, but to also synthesize the paths it could have taken. Integrating this with experimental evolution and developmental and structural approaches offers a path toward understanding how contingency, constraint, and bias jointly determine the replayability—and predictability—of evolutionary change.

Ancestral sequence reconstruction (ASR) is a powerful molecular approach that seeks to infer ancient biomolecules (DNA, RNA, proteins) from extant sequences, synthesize them in the lab, and experimentally characterize their functions[102,103] (**Figure 2D**). In essence, ASR lets us "replay" evolution in retrospect: we can resurrect long-extinct proteins or gene sequences and examine how their properties differ from modern ones, thereby testing hypotheses about what changed and why. Among other interesting conclusions, findings from ASR have particularly informed about the role of contingency in molecular evolution.

Reconstructing a molecule's evolutionary path allows direct observation of evolutionary constraint. By recreating ancient enzymes, scientists can introduce historical mutations one by one to see how function changes, revealing epistatic interactions that constrained certain evolutionary paths and channeled others. A famous example is the evolution of cortisol specificity in glucocorticoid receptors: ASR revealed that a particular mutation required to confer high specificity was only tolerated after another "permissive" mutation occurred, explaining why evolution took the path it did (historical contingency at the molecular level)[104–107]. These kinds of studies marry well with theoretical models of sequence space and adaptive walks, providing concrete data on how many mutations, and in what order, are needed for new functions (**Figure 2D**).

Ancestral reconstruction has also illuminated how gene duplication generates new function through the resolution of adaptive conflict (see Section 2.1). Reconstruction of an ancestral steroid hormone receptor before a duplication event showed that it bound multiple hormones with intermediate specificity, consistent with a single gene juggling multiple roles[104]. After duplication, one copy acquired degenerative mutations that abolished hormone-activated signaling but preserved DNA binding, converting it into a competitive repressor that fine-tunes the activity of its sister copy[104]. This division of labor resolved an apparent regulatory conflict and demonstrated that ostensibly deleterious loss-of-function mutations can drive innovation when they occur in a duplicated background[104]. Such studies provide direct experimental support for the escape from adaptive conflict model and show how synthetic resurrection complements forward-evolution experiments in dissecting the mechanistic origins of new functions.

Ancestral reconstruction has also allowed the mapping of ancient genotype–phenotype relationships. Herrera-Álvarez et al.[108] generated comprehensive libraries of ancestral transcription factors and their DNA response elements, assaying all possible combinations at historically variable sites. They found that the resulting genotype–phenotype maps were strongly biased: some phenotypes were far more likely to appear than others (anisotropy), and the set of phenotypes accessible from one genotype differed greatly from those around another (heterogeneity). These intrinsic biases steered ancient proteins toward the lineage-specific DNA-binding specificities that later evolved, showing that the structure of genotype–phenotype maps themselves can act as causal determinants of evolutionary outcomes. This work extends the fitness landscape framework by demonstrating that ruggedness and accessibility are not only properties of extant molecules, but also of the ancestral maps that shaped long-term evolutionary history.

Beyond informing molecular evolution, ASR can record ancient environmental conditions. Reconstructed enzymes from cyanobacteria and land plants show thermostabilities consistent with Earth's surface cooling from ~75 °C three billion years ago to ~35 °C by the Devonian, providing biophysical evidence that complements geological paleotemperature estimates[109]. Overall, ASR gives evolutionary biologists a remarkable time machine: instead of merely inferring the past, we can experiment on it. The insights from these resurrected molecules often uphold classical narratives (e.g., confirming a predicted ancestral function) but also add layers of nuance (e.g., revealing previously unknown intermediate states or latent functions). By directly observing "missing links" at the molecular level, we enrich evolutionary theory with a more detailed understanding of how new biological functions and complexities have arisen through deep time.

3. The evolving toolkit for engineering evolutionary systems

Beyond high-throughput measurement of natural sequence variation, synthetic biology offers a fundamentally different empirical mode: building biological systems *de novo* and observing how evolution acts on them. Three classes of approach have proven especially productive.

Synthetic gene circuits and minimal genomes provide controlled platforms for evolutionary experimentation[110,111]. For example, the study of synthetic genetic circuits in *E. coli* has shown that certain gene regulatory network topologies are more evolvable towards specific functions than others[110,111] (**Figure 2E**). By introducing engineered regulatory circuits or metabolic pathways into living cells, researchers can ask whether selection tunes their components toward predicted optima, thereby probing principles of adaptation in tractable systems[112–114]. The assembly of a minimal bacterial genome of approximately 473 genes (JCVI-syn3.0) provides a stripped-down chassis in which to ask whether selection regenerates discarded complexity[115–118]. Adaptive evolution on cells with such reduced genome showed that very few generations are needed for fitness to be improved (**Figure 2F)**. On the eukaryotic side, the synthetic yeast genome project *Saccharomyces cerevisiae* 2.0[119–121] incorporates scrambleable genome segments designed to allow controlled rearrangement under selection[109]. Such systems test classical questions about the evolution of robustness, modularity, and genome architecture in regimes that natural genomes do not readily afford.

Engineered mutators decouple variation from natural mutation processes[122]. Programmable in vivo platforms such as EvolvR[122,123] can target hypermutation to specific loci using CRISPR-guided DNA polymerases, while systems such as TF-HighEvo[124] in *Drosophila* fuse a transcription factor to an activation-induced deaminase to concentrate germline mutations around its binding sites, raising local mutation rates roughly tenfold and generating heritable phenotypic diversity. These tools allow direct experimental control over the supply, location, and spectrum of variation, providing tests of how mutation rate and bias shape adaptation in regimes that natural mutagenesis cannot easily produce.

Random sequences interrogate the null expectation for biological function. What fraction of sequence space is functional, and how readily can selection convert noise into signal (**Figure 2G**)? Random RNAs and peptides expressed in *E. coli* often show measurable bioactivity, modulating growth in unexpected ways[125]. Random sequences also display surprisingly high structural propensity: ~40% of random 105-residue proteins adopt compact, protease-resistant conformations whether built from the full 20-amino-acid alphabet or from a reduced 10-residue prebiotic set [126], and de novo designs using only the ten "early" amino acids can reconstruct ancient protein folds with thermostability and mutational robustness comparable to their full-alphabet counterparts [127]; the canonical alphabet itself appears to have been shaped by these foldability constraints, since prebiotically abundant non-canonical residues yield polypeptides that fold less reliably than the early canonical set [128].

Random DNA sequences have also been used to study the potential of de novo evolution of gene regulatory elements. In bacteria, roughly 10% of random short DNA sequences placed upstream of a reporter in *E. coli* function as promoters without any evolution, and 60% become active after a single mutation[129]. Biophysical models calibrated on random mutant libraries reach similar conclusions, predicting that ~20% of arbitrary sequences drive measurable expression and that ~82% of inactive sequences lie one mutation away from a functional σ^{70} promoter[130]. This propensity for de novo regulatory emergence is even greater in random DNA than in extant genomic sequence: a recent massively parallel reporter assay found that promoters arise roughly threefold more readily from random DNA than from native *E. coli* sequences, because the genome contains fewer proto-binding sites for transcriptional activators than random DNA does [131]. Screens of approximately 100 million random promoters in yeast similarly show that random sequences frequently contain transcription factor binding motifs and span much of the natural expression range[132,133]. In Drosophila embryos, random sequences readily drive reporter expression at mid- and late-stage development but rarely in early embryos, where pre-established chromatin accessibility constrains regulatory activity[98]. Together, these results suggest that, in many contexts, evolution must act as much to suppress unwanted activity as to create new function—and that natural genomes may even bear the signature of selection trimming away regulatory potential that random DNA possesses by default—reframing a longstanding question about the origins of regulatory novelty.

4. Conclusion

Bringing together these threads, it is evident that we are in a golden age of empirical evolutionary biology that is deeply informed by and informing theory. The dialogue between past and future—between classical ideas and modern approaches—is enriching both. We find that many foundational concepts were essentially on the right track, but reality is often more quantitatively nuanced and interconnected than early models assumed. What do these advances provide for the future of evolutionary theory? The progression from measuring sequence–function relationships (**Figure 2A, B**) to characterizing fitness seascapes (**Figure 2C**) and reconstructing evolutionary history (**Figure 2D**), to testing architectural constraints and de novo function with synthetic systems (**Figure 2E–G**), to predicting evolutionary outcomes (**Figure 2H**) mirrors the broader intellectual trajectory of the field: from cataloguing evolutionary possibilities to forecasting which of those possibilities will be realized.

As empirical fitness landscapes and genotype–phenotype datasets expand, machine learning approaches are beginning to capture latent biophysical constraints and generalize across sequence space[133,134], and deep neural models trained on exhaustive landscape data can infer fitness for unmeasured genotypes and guide experimental design[135]**.** Yet these advances also caution against overconfidence: replay experiments and large-scale maps consistently show that the order of mutations, clonal competition, and environmental fluctuations make evolution partly unpredictable[136]. A more realistic goal is probabilistic forecasting — estimating which outcomes are most likely and how probable different evolutionary paths are, rather than predicting single trajectories.

The horizon of tools keeps expanding. Long-read sequencing and ancient DNA may allow us to directly observe evolution over longer timescales (thousands of years) in natural populations, bridging the gap between lab time and geological time. CRISPR-based lineage tracking in multicellular organisms could open experimental evolution to plants and animals, not just microbes—one could envision tracking somatic evolution or virus evolution within a host in unprecedented detail[137]. These will pose new theoretical questions about heterogeneous environments and multilevel selection (e.g., cell-level vs. organism-level fitness). Synthetic biology also offers the possibility of constructing a minimal life form and seeing what directions evolution takes it in various worlds, akin to simulating alternative histories. Furthermore, as empirical fitness landscapes and genotype–phenotype datasets expand, the prospect of predictive evolution—anticipating likely adaptations in pathogens, tumors, or crops—moves from speculation toward practice.

Acknowledgements

This review originated from the symposium "Synthetic approaches to evolutionary theory: a dialogue between the past and the future" held at SMBE 2025. We thank the Society for Molecular

Biology and Evolution for the opportunity to organize this session and the participants for stimulating discussions that shaped the ideas presented here.

Funding

X.C.L. is supported by The Fundamental Research Funds for the Central Universities and Young Scientists Fund of the National Natural Science Foundation of China (#32400495). P.M. is supported by a fellowship from the EMBL Interdisciplinary Postdoctoral Programme (EIPOD) under Marie Skłodowska-Curie Actions COFUND (grant agreement number 101217201). C.A.W. is supported by the EPFL School of Life Sciences ELISIR scholarship.

Author contributions

X.C.L., P.M., and C.A.W. conceived the review. All authors contributed to writing and revising the manuscript. C.A.W. designed the figures. All authors read and approved the final version.

Competing interests

The authors declare no competing interests.

Data Availability Statement

No data was generated for this manuscript. All references and materials are available from the corresponding publishers.

References:

1. Darwin, C. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*. (John Murray, London, 1859).

2. Fisher, R. A. *The Genetical Theory of Natural Selection*. (Clarendon Press, Oxford, 1930). doi:10.5962/bhl.title.27468.

3. Haldane, J. B. S. A Mathematical Theory of Natural Selection. Part VIII. Metastable Populations. *Mathematical Proceedings of the Cambridge Philosophical Society* **27**, 137–142 (1931).

4. Haldane, J. B. S. *The Causes of Evolution*. 222 (Princeton University Press, 1990).

5. Wright, S. Evolution in Mendelian Populations. *Genetics* **16**, 97 (1931).

6. Wright, S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc of the 6th International Congress of Genetics* Preprint at (1932).

7. Kimura, M. Evolutionary Rate at the Molecular Level. *Nature 1968 217:5129* **217**, 624–626 (1968).

8. Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge University Press, 1983). doi:10.1017/CBO9780511623486.

9. Kimura, M. On the Probability of Fixation of Mutant Genes in a Population. *Genetics* **47**, 713–719 (1962).

10. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).

11. Dennett, D. C. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. (Simon & Schuster, New York, 1995).

12. Dawkins, R. The Evolution of Evolvability. in *Artificial Life* 201–220 (Routledge, 2019). doi:10.4324/9780429032769-10.

13. Wagner, G. P. & Altenberg, L. PERSPECTIVE: COMPLEX ADAPTATIONS AND THE EVOLUTION OF EVOLVABILITY. *Evolution (N. Y).* **50**, 967–976 (1996).

14. Hansen, T. F. & Wagner, G. P. The Evolution of Evolvability. in *Evolvability* 121–146 (The MIT Press, 2023). doi:10.7551/mitpress/14126.003.0009.

15. Ornelas, M. Y., Cournoyer, J. E., Bram, S. & Mehta, A. P. Evolution and synthetic biology. *Curr. Opin. Microbiol.* **76**, 102394 (2023).

16. Yi, X. & Dean, A. M. Adaptive Landscapes in the Age of Synthetic Biology. *Mol. Biol. Evol.* **36**, 890–907 (2019).

17. Crocker, J. & Ilsley, G. R. Using synthetic biology to study gene regulatory evolution. *Curr. Opin. Genet. Dev.* **47**, 91–101 (2017).

18. Kvon, E. Z. *et al.* Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167**, 633-642.e11 (2016).

19. Murugesan, S. N. *et al.* Butterfly eyespots evolved via cooption of an ancestral gene-regulatory network that also patterns antennae, legs, and wings. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2108661119 (2022).

20. Schindler, M. *et al.* Comparative single-cell analyses reveal evolutionary repurposing of a conserved gene programme in bat wing development. *Nature Ecology & Evolution 2025 9:9* **9**, 1626–1642 (2025).

21. Daane, J. M. *et al.* Modulation of bioelectric cues in the evolution of flying fishes. *Current Biology* **31**, 5052-5061.e8 (2021).

22. Audet, T. *et al.* Genetic architecture of the developing forelegs of Drosophila prolongata; an exaggerated weapon and ornament. *bioRxiv* 2025.10.09.681446 (2025) doi:10.1101/2025.10.09.681446.

23. Huxley, J. *Evolution, the Modern Synthesis*. (George Allen & Unwin., London, 1942).

24. Nei, Masatoshi. *Mutation-Driven Evolution*. (Oxford University Press, 2014).

25. Padilla-Iglesias, C. & Majic, P. What is the human germline mutation rate? methodological innovations, challenges, and evolutionary implications. https://doi.org/10.32942/X2T083 (2026) doi:10.32942/X2T083.

26. Mukai, T. THE GENETIC STRUCTURE OF NATURAL POPULATIONS OF DROSOPHILA MELANOGASTER. I. SPONTANEOUS MUTATION RATE OF POLYGENES CONTROLLING VIABILITY. *Genetics* **50**, 1–19 (1964).

27. Ohnishi, O. Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in Drosophila melanogaster. III. Heterozygous effect of polygenic mutations. *Genetics* **87**, 547–556 (1977).

28. Kibota, T. T. & Lynch, M. Estimate of the genomic mutation rate deleterious to overall fitness in E. coli. *Nature* **381**, 694–696 (1996).

29. Lynch, M., Butcher, D., Bürger, R. & Gabriel, W. The mutational meltdown in asexual populations. *Journal of Heredity* **84**, 339–344 (1993).

30. Eyre-Walker, A. & Keightley, P. D. High genomic deleterious mutation rates in hominids. *Nature* **397**, 344–347 (1999).

31. Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275–276 (1977).

32. Sarkisyan, K. S. *et al.* Local fitness landscape of the green fluorescent protein. *Nature 2015 533:7603* **533**, 397–401 (2016).

33. Firnberg, E., Labonte, J. W., Gray, J. J. & Ostermeier, M. A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape. *Mol. Biol. Evol.* **31**, 1581–1592 (2014).

34. Hietpas, R. T., Jensen, J. D. & Bolon, D. N. A. Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7896–7901 (2011).

35. Hietpas, R. T., Bank, C., Jensen, J. D. & Bolon, D. N. A. Shifting fitness landscapes in response to altered environments. *Evolution (N. Y).* **67**, 3512–3522 (2013).

36. Bank, C. Epistasis and Adaptation on Fitness Landscapes. *https://doi.org/10.1146/annurev-ecolsys-102320-112153* **53**, 457–479 (2022).

37. Di Bari, L., Bisardi, M., Cotogno, S., Weigt, M. & Zamponi, F. Emergent time scales of epistasis in protein evolution. *Proceedings of the National Academy of Sciences* **121**, e2406807121 (2024).

38. Kassen, R. & Bataillon, T. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat. Genet.* **38**, 484–488 (2006).

39. Couce, A. *et al.* Changing fitness effects of mutations through long-Term bacterial evolution. *Science (1979).* **383**, (2024).

40. Ardell, S. M., Martsul, A., Johnson, M. S. & Kryazhimskiy, S. Environment-independent distribution of mutational effects emerges from microscopic epistasis. *Science* **386**, 87 (2024).

41. Aguilar-Rodríguez, J. *et al.* Massively parallel interrogation of the fitness of natural variants in ancient signaling pathways reveals pervasive local adaptation. *bioRxiv* 2024.10.30.621178 (2025) doi:10.1101/2024.10.30.621178.

42. Zhang, J. & Qian, W. Functional synonymous mutations and their evolutionary consequences. *Nat. Rev. Genet.* 1–16 (2025) doi:10.1038/S41576-025-00850-1;SUBJMETA.

43. Bao, K., Melde, R. H. & Sharp, N. P. Are mutations usually deleterious? A perspective on the fitness effects of mutation accumulation. *Evol. Ecol.* **36**, 753–766 (2022).

44. Bailey, S. F., Morales, L. A. A. & Kassen, R. Effects of Synonymous Mutations beyond Codon Bias: The Evidence for Adaptive Synonymous Substitutions from Microbial Evolution Experiments. *Genome Biol. Evol.* **13**, (2021).

45. Lebeuf-Taylor, E., McCloskey, N., Bailey, S. F., Hinz, A. & Kassen, R. The distribution of fitness effects among synonymous mutations in a gene under directional selection. *Elife* **8**, (2019).

46. Zwart, M. P. *et al.* Unraveling the causes of adaptive benefits of synonymous mutations in TEM-1 β-lactamase. *Heredity (Edinb).* **121**, 406–421 (2018).

47. Liu, Y., Yang, Q. & Zhao, F. Synonymous but Not Silent: The Codon Usage Code for Gene Expression and Protein Folding. *Annu. Rev. Biochem.* **90**, 375–401 (2021).

48. Shen, X., Song, S., Li, C. & Zhang, J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature* **606**, 725–731 (2022).

49. Shen, X., Song, S., Li, C. & Zhang, J. Further Evidence for Strong Nonneutrality of Yeast Synonymous Mutations. *Mol. Biol. Evol.* **41**, (2024).

50. Kruglyak, L. *et al.* Insufficient evidence for non-neutrality of synonymous mutations. *Nature 2023 616:7957* **616**, E8–E9 (2023).

51. Rao, Y., Sokirniy, I., O'Brien, E. & Pritchard, J. High-throughput Experimental Validation of Synonymous Mutation Neutrality in the Human Genome. *bioRxiv* 2025.08.30.673250 (2025) doi:10.1101/2025.08.30.673250.

52. Sane, M., Parveen, S. & Agashe, D. Mutation bias alters the distribution of fitness effects of mutations. *PLoS Biol.* **23**, e3003282 (2025).

53. OHNO, S., WOLF, U. & ATKIN, N. B. EVOLUTION FROM FISH TO MAMMALS BY GENE DUPLICATION. *Hereditas* **59**, 169–187 (1968).

54. Mihajlovic, L. *et al.* A direct experimental test of Ohno's hypothesis. *Elife* **13**, (2024).

55. Des Marais, D. L. & Rausher, M. D. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**, 762–765 (2008).

56. Sikosek, T., Chan, H. S. & Bornberg-Bauer, E. Escape from adaptive conflict follows from weak functional trade-offs and mutational robustness. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14888–14893 (2012).

57. Cisneros, A. F. *et al.* Paralog interference contributes to the preservation of genetic redundancy. *Current Biology* **36**, 1373-1384.e10 (2026).

58. Kauffman, S. A. & Weinberger, E. D. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.* **141**, 211–245 (1989).

59. Obolski, U., Ram, Y. & Hadany, L. Key issues review: evolution on rugged adaptive landscapes. *Reports on Progress in Physics* **81**, 012602 (2018).

60. Kauffman, S. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **128**, 11–45 (1987).

61. Krug, J. & Oros, D. Evolutionary accessibility of random and structured fitness landscapes. *Journal of Statistical Mechanics: Theory and Experiment* **2024**, 034003 (2024).

62. Papkou, A., Garcia-Pastor, L., Escudero, J. A. & Wagner, A. A rugged yet easily navigable fitness landscape. *Science (1979).* **382**, eadh3860 (2023).

63. Antunes Westmann, C., Goldbach, L. & Wagner, A. The adaptive landscapes of three global Escherichia coli transcriptional regulators. *eLife* vol. 14 Preprint at https://doi.org/10.7554/eLife.103774.1 (2025).

64. Westmann, C. A., Goldbach, L. & Wagner, A. The highly rugged yet navigable regulatory landscape of the bacterial transcription factor TetR. *Nature Communications* **15**, (2024).

65. Westmann, C. A., Goldbach, L. & Wagner, A. Entangled adaptive landscapes facilitate the evolution of gene regulation by exaptation. *bioRxiv* 2024.11.10.620926 (2024) doi:10.1101/2024.11.10.620926.

66. Weeks, R. & Ostermeier, M. Fitness and Functional Landscapes of the E. coli RNase III Gene rnc. *Mol. Biol. Evol.* **40**, (2023).

67. Kuo, S. T. *et al.* Global fitness landscapes of the Shine-Dalgarno sequence. *Genome Res.* **30**, 711–723 (2020).

68. Aguilar-Rodríguez, J., Payne, J. L. & Wagner, A. A thousand empirical adaptive landscapes and their navigability. *Nat. Ecol. Evol.* **1**, 0045 (2017).

69. Schulz, S., Tan, T. J. C., Wu, N. C. & Wang, S. Epistatic hotspots organize antibody fitness landscape and boost evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **122**, e2413884122 (2025).

70. Adams, R. M., Kinney, J. B., Walczak, A. M. & Mora, T. Epistasis in a fitness landscape defined by antibody-antigen binding free energy. *Cell Syst.* **8**, 86 (2019).

71. Reidys, C. M. & Stadler, P. F. Neutrality in fitness landscapes. *Appl. Math. Comput.* **117**, 321–350 (2001).

72. Pahujani, S. & Krug, J. Complexity and accessibility of random landscapes. https://arxiv.org/pdf/2502.05896 (2025).

73. Srivastava, M., Louis, A. A. & Martin, N. S. Predicting the topography of fitness landscapes from the structure of genotype-phenotype maps. *bioRxiv* 2025.02.14.638275 (2025) doi:10.1101/2025.02.14.638275.

74. Wu, N. C., Dai, L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *Elife* **5**, (2016).

75. Trubenová, B., Krejca, M. S., Lehre, P. K. & Kötzing, T. Surfing on the seascape: Adaptation in a changing environment. *Evolution* **73**, 1356 (2019).

76. Mustonen, V. & Lässig, M. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends in Genetics* **25**, 111–119 (2009).

77. Li, C. & Zhang, J. Multi-environment fitness landscapes of a tRNA gene. *Nature Ecology & Evolution 2018 2:6* **2**, 1025–1032 (2018).

78. Ogbunugafor, C. B., Wylie, C. S., Diakite, I., Weinreich, D. M. & Hartl, D. L. Adaptive Landscape by Environment Interactions Dictate Evolutionary Dynamics in Models of Drug Resistance. *PLoS Comput. Biol.* **12**, e1004710 (2016).

79. Aubé, S. & Landry, C. R. How genotype-by-environment interactions on fitness emerge. *Nature Ecology & Evolution 2024 8:12* **8**, 2157–2158 (2024).

80. Flynn, J. M. *et al.* Comprehensive fitness maps of Hsp90 show widespread environmental dependence. *Elife* **9**, (2020).

81. Bank, C., Matuszewski, S., Hietpas, R. T. & Jensen, J. D. On the (un)predictability of a large intragenic fitness landscape. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14085–14090 (2016).

82. Chattopadhyay, G., Papkou, A. & Wagner, A. The fitness landscape of the E.coli lac operator is highly rugged in two different environments. *bioRxiv* 2025.07.23.666252 (2025) doi:10.1101/2025.07.23.666252.

83. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).

84. Gandara, L. *et al.* Developmental phenomics suggests that H3K4 monomethylation confers multi-level phenotypic robustness. *Cell Rep.* **41**, (2022).

85. Zheng, J., Payne, J. L. & Wagner, A. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science* **365**, 347–353 (2019).

86. Barnett, M., Meister, L. & Rainey, P. B. Experimental evolution of evolvability. *Science (1979).* **387**, (2025).

87. Bloom, J. D. *et al.* Evolution favors protein mutational robustness in sufficiently large populations. *BMC Biol.* **5**, 1–21 (2007).

88. Smith, J. M. *et al.* DEVELOPMENTAL CONSTRAINTS AND EVOLUTION: Perspective from the Mountain Lake Conference on Development and Evolution. *Quarterly Review of Biology* **60**, (1985).

89. Alberch, P. From genes to phenotype: dynamical systems and evolvability. *Genetica* **84**, 5–11 (1991).

90. Uller, T., Moczek, A. P., Watson, R. A., Brakefield, P. M. & Laland, K. N. Developmental bias and evolution: A regulatory network perspective. *Genetics* **209**, 949–966 (2018).

91. Carroll, S. B. Endless forms: The evolution of gene regulation and morphological diversity. *Cell* **101**, 577–580 (2000).

92. Prud'homme, B. *et al.* Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* **440**, 1050–1053 (2006).

93. Prud'homme, B., Gompel, N. & Carroll, S. B. Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8605–8612 (2007).

94. Wittkopp, P. J. Contributions of mutation and selection to regulatory variation: lessons from the Saccharomyces cerevisiae TDH3 gene. *Philosophical Transactions of the Royal Society B: Biological Sciences* **378**, 20220057 (2023).

95. Van de Zande, P., Siddiq, M. A., Hodgins-Davis, A., Kim, L. & Wittkopp, P. J. Active compensation for changes in TDH3 expression mediated by direct regulators of TDH3 in Saccharomyces cerevisiae. *PLoS Genet.* **19**, e1011078 (2023).

96. Fuqua, T. *et al.* Dense and pleiotropic regulatory information in a developmental enhancer. *Nature* **587**, 235–239 (2020).

97. Le Poul, Y. *et al.* Regulatory encoding of quantitative variation in spatial activity of a Drosophila enhancer. *Sci. Adv.* **6**, (2020).

98. Galupa, R. *et al.* Enhancer architecture and chromatin accessibility constrain phenotypic space during Drosophila development. *Dev. Cell* **58**, 51-62.e4 (2023).

99. Gould, S. J. & Lewontin, R. C. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. B. Biological Sciences* **205**, 581–598 (1979).

100. Gould, S. J. Wonderful Life; The Burgess Shale and the Nature of History. *J. Hist. Biol.* **24**, 163–170 (1992).

101. Jacob, F. Evolution and Tinkering. *Science (1979).* **196**, 1161–1166 (1977).

102. Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nature Reviews Genetics 2004 5:5* **5**, 366–375 (2004).

103. Hochberg, G. K. A. & Thornton, J. W. Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annual Review of Biophysics* vol. 46 247–269 Preprint at https://doi.org/10.1146/annurev-biophys-070816-033631 (2017).

104. Bridgham, J. T., Carroll, S. M. & Thornton, J. W. Evolution of hormone-receptor complexity by molecular exploitation. *Science (1979).* **312**, 97–101 (2006).

105. Eick, G. N., Colucci, J. K., Harms, M. J., Ortlund, E. A. & Thornton, J. W. Evolution of Minimal Specificity and Promiscuity in Steroid Hormone Receptors. *PLoS Genet.* **8**, e1003072 (2012).

106. Eick, G. N. & Thornton, J. W. Evolution of steroid receptors from an estrogen-sensitive ancestral receptor. *Mol. Cell. Endocrinol.* **334**, 31–38 (2011).

107. Thornton, J. W., Need, E. & Crews, D. Resurrecting the ancestral steroid receptor: Ancient origin of estrogen signaling. *Science (1979).* **301**, 1714–1717 (2003).

108. Herrera-Álvarez, S., Patton, J. E. J. & Thornton, J. W. The structure of an ancient genotype–phenotype map shaped the functional evolution of a protein family. *Nat. Ecol. Evol.* **9**, 1656–1669 (2025).

109. Garcia, A. K., Schopf, J. W., Yokobori, S. I., Akanuma, S. & Yamagishi, A. Reconstructed ancestral enzymes suggest long-Term cooling of Earth's photic zone since the Archean. *Proc. Natl. Acad. Sci. U. S. A.* **114**, (2017).

110. Santos-Moreno, J., Tasiudi, E., Kusumawardhani, H., Stelling, J. & Schaerli, Y. Robustness and innovation in synthetic genotype networks. *Nature Communications 2023 14:1* **14**, 1–17 (2023).

111. Schaerli, Y. *et al.* Synthetic circuits reveal how mechanisms of gene regulatory networks constrain evolution. *Mol. Syst. Biol.* **14**, e8102 (2018).

112. Nørholm, M. H. H. Meta synthetic biology: controlling the evolution of engineered living systems. *Microb. Biotechnol.* 1751-7915.13350 (2018) doi:10.1111/1751-7915.13350.

113. Renda, B. A., Hammerling, M. J. & Barrick, J. E. Engineering reduced evolutionary potential for synthetic biology. *Mol. Biosyst.* **10**, 1668–1678 (2014).

114. Castle, S. D., Grierson, C. S. & Gorochowski, T. E. Towards an engineering theory of evolution. *Nature Communications 2021 12:1* **12**, 1–12 (2021).

115. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science (1979).* **351**, (2016).

116. Gibson, D. G. Programming biological operating systems: genome design, assembly and activation. *Nat. Methods* **11**, 521–6 (2014).

117. Moger-Reischer, R. Z. *et al.* Evolution of a minimal cell. *Nature* **620**, 122–127 (2023).

118. Sandberg, T. E. *et al.* Adaptive evolution of a minimal organism with a synthetic genome. *iScience* **26**, 107500 (2023).

119. James, J. S., Dai, J., Chew, W. L. & Cai, Y. The design and engineering of synthetic genomes. *Nat. Rev. Genet.* **26**, 298–319 (2025).

120. Richardson, S. M. *et al.* Design of a synthetic yeast genome. *Science (1979).* **355**, 1040–1044 (2017).

121. Dai, J., Boeke, J. D., Luo, Z., Jiang, S. & Cai, Y. Sc3.0: Revamping and minimizing the yeast genome. *Genome Biol.* **21**, 1–4 (2020).

122. Molina, R. S. *et al.* In vivo hypermutation and continuous evolution. *Nature Reviews Methods Primers 2022 2:1* **2**, 36- (2022).

123. Halperin, S. O. *et al.* CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature 2018 560:7717* **560**, 248–252 (2018).

124. Li, X. C. *et al.* TF-High-Evolutionary: In Vivo Mutagenesis of Gene Regulatory Networks for the Study of the Genetics and Evolution of the Drosophila Regulatory Genome. *Mol. Biol. Evol.* **41**, (2024).

125. Neme, R., Amador, C., Yildirim, B., McConnell, E. & Tautz, D. Random sequences are an abundant source of bioactive RNAs or peptides. *Nat. Ecol. Evol.* **1**, 1–7 (2017).

126. Tretyachenko, V. *et al.* Modern and prebiotic amino acids support distinct structural profiles in proteins. *Open Biol.* **12**, 220040 (2022).

127. Giacobelli, V. G. *et al.* Ancient amino acid sets enable stable protein folds. *bioRxiv* 2025.10.29.685319 (2025) doi:10.1101/2025.10.29.685319.

128. Makarov, M. *et al.* Early Selection of the Amino Acid Alphabet Was Adaptively Shaped by Biophysical Constraints of Foldability. *J. Am. Chem. Soc.* **145**, 5320–5329 (2023).

129. Yona, A. H., Alm, E. J. & Gore, J. Random sequences rapidly evolve into de novo promoters. *Nat. Commun.* **9**, 111880 (2018).

130. Lagator, M. *et al.* Predicting bacterial promoter function and evolution from random sequences. *Elife* **11**, (2022).

131. Fuqua, T. & Wagner, A. The latent cis-regulatory potential of mobile DNA in Escherichia coli. *Nature Communications 2025 16:1* **16**, 4740- (2025).

132. de Boer, C. G. *et al.* Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nature Biotechnology 2019 38:1* **38**, 56–65 (2019).

133. Vaishnav, E. D. *et al.* The evolution, evolvability and engineering of gene regulatory DNA. *Nature 2022 603:7901* **603**, 455–463 (2022).

134. Sandhu, M. *et al.* Computational and Experimental Exploration of Protein Fitness Landscapes: Navigating Smooth and Rugged Terrains. *Biochemistry* https://doi.org/10.1021/ACS.BIOCHEM.4C00673 (2025) doi:10.1021/ACS.BIOCHEM.4C00673.

135. Wagner, A. Genotype sampling for deep-learning assisted experimental mapping of a combinatorially complete fitness landscape. *Bioinformatics* **40**, (2024).

136. Wortel, M. T. *et al.* Towards evolutionary predictions: Current promises and challenges. *Evol. Appl.* **16**, 3–21 (2023).

137. Askary, A. *et al.* The lives of cells, recorded. *Nature Reviews Genetics 2024 26:3* **26**, 203–222 (2024).

Figures:

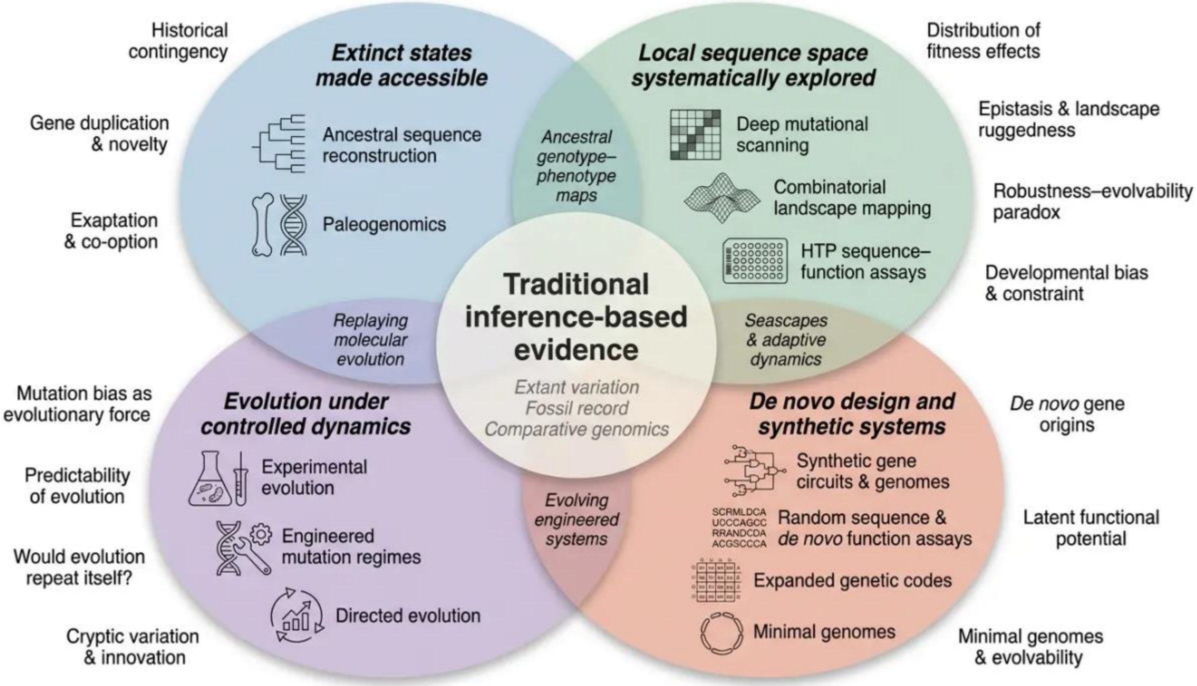

Figure 1. Synthetic and experimental approaches expand evolutionary access beyond natural observation. Evolutionary biology has traditionally relied on inference from extant variation, the fossil record, and comparative genomics (centre). Four complementary classes of approaches now extend this empirical reach. *Extinct states made accessible* (upper left): ancestral sequence reconstruction and paleogenomics resurrect ancient molecules and genotypes, enabling direct functional tests of evolutionary history. *Local sequence space systematically explored* (upper right): deep mutational scanning, combinatorial libraries, and high-throughput sequence–function assays map genotype–phenotype and fitness landscapes around natural sequences with unprecedented resolution. *Evolution under controlled dynamics* (lower left): experimental evolution, engineered mutation regimes, and directed evolution allow populations to be replayed, redirected, or exposed to defined selective pressures, testing how mutation bias, contingency, and repeatability shape adaptive trajectories. *De novo design and synthetic systems* (lower right): synthetic gene circuits, minimal genomes, expanded genetic codes, and random-sequence assays probe functional regions of sequence space that natural evolution may never have sampled, addressing what evolution could do rather than only what it did. Overlap zones highlight questions that are most effectively addressed by combining approaches: ancestral genotype–phenotype maps emerge from coupling sequence reconstruction with systematic landscape mapping; replaying molecular evolution integrates resurrected ancestors with forward experimental evolution; seascapes and adaptive dynamics link empirical landscape characterisation with controlled evolutionary experiments; de novo regulatory landscapes connect high-throughput assays with synthetic and random sequence libraries; and evolving engineered systems combines experimental evolution with synthetic circuits and minimal genomes. Classical evolutionary questions are shown at the periphery. Together, these approaches shift the field from observing evolutionary outcomes to experimentally testing evolutionary possibilities.

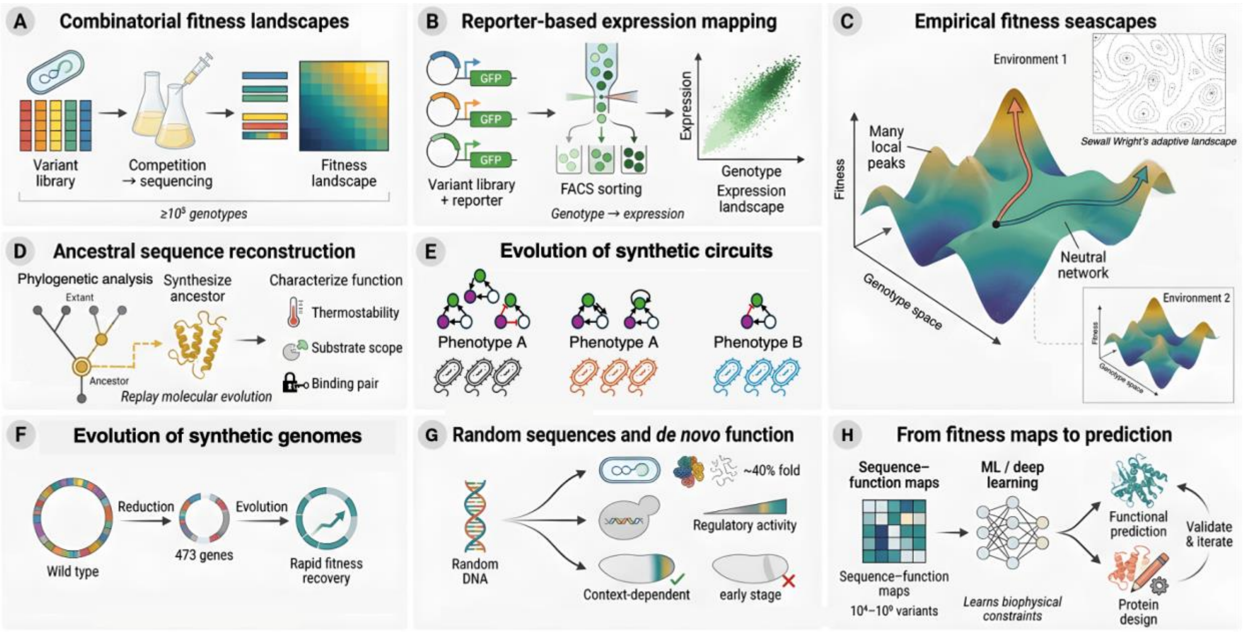

Figure 2. Experimental and synthetic approaches for studying evolution. The figure is organized in three rows that follow a logical progression: the top row shows how high-throughput genotype–fitness maps are constructed and what the resulting landscapes look like, the middle row illustrates how evolutionary history is reconstructed and how synthetic systems test architectural constraints on evolvability, and the bottom row shows how synthetic genomes evolve, how random sequences reveal latent functional potential, and how empirical data feed predictive models. **A. Combinatorial fitness landscapes.** Large combinatorial variant libraries (≥10^5 genotypes) are introduced into cells by genome editing and subjected to growth competition under defined selective conditions. Deep sequencing yields empirical fitness landscapes that quantify *in vivo* fitness for each genotype. **B. Reporter-based expression mapping.** Libraries of regulatory sequence variants are cloned upstream of a fluorescent reporter (e.g., GFP), expressed in cells, and sorted by fluorescence intensity using FACS. Sequencing of each sorted bin links sequence variation to quantitative regulatory output, producing high-resolution genotype–expression landscapes. **C. Empirical fitness seascapes.** In contrast to Wright's original (faint grey silhouette), empirical landscapes reveal richly structured topographies with many local peaks, ridges, and neutral networks. Two trajectories from a shared starting genotype illustrate that most peaks are reachable without crossing deep valleys: a direct adaptive path (coral arrow) climbs steeply to one peak, whereas neutral exploration (teal arrow) traverses a plateau before ascending a different peak. Crucially, landscape topography is environment-dependent: the same genotype space exhibits a different fitness surface under alternative conditions (inset, Environment 2), producing fitness "seascapes" in which peaks, valleys, and accessible paths shift with external context. **D. Ancestral sequence reconstruction.** Phylogenetic analysis of extant sequences is used to infer ancestral states, which are then synthesized in the laboratory and experimentally characterized for properties such as thermostability, substrate scope, and binding specificity — effectively replaying molecular evolution. **E Evolution of synthetic circuits.** Synthetic gene regulatory networks with distinct topologies are introduced into cells and evolved under selection. Different circuits producing the same phenotype (Phenotype A) can employ different underlying architectures, while alternative wiring can produce a different phenotype (Phenotype B), demonstrating that network topology constrains which evolutionary outcomes are accessible. **F. Evolution of synthetic genomes.** A wild-type bacterial genome is radically reduced to a minimal set of ~473 essential genes, then subjected to experimental evolution. Rapid fitness recovery demonstrates that even highly simplified genomes retain substantial evolvability and can reinvent functions that were stripped during reduction. **G. Random sequences and *de novo* function.** Random DNA sequences, never exposed to natural selection, are tested for biological activity across three systems. In *E. coli*, ~40% of random polypeptides adopt compact, folded conformations. In yeast, random sequences frequently contain transcription factor binding sites that drive measurable regulatory activity. In *Drosophila* embryos, random sequences can drive reporter expression at late developmental stages but not at early stages, revealing that functional potential is widespread but context-dependent. **H. From fitness maps to prediction.** High-throughput sequence–function maps encompassing 10^4–10^9 variants feed machine-learning and deep-learning models that learn latent biophysical constraints. These models enable functional prediction and protein design, with iterative experimental validation closing the loop between prediction and observation.