



In vivo hypermutation and continuous evolution

Rosana S. Molina^{1,25}, Gordon Rix^{2,25}, Amanuella A. Mengiste³, Beatriz Álvarez⁴, Daeje Seo⁵, Haiqi Chen^{6,7}, Juan E. Hurtado⁸, Qiong Zhang⁸, Jorge Donato García-García⁹, Zachary J. Heins^{10,11}, Patrick J. Almhjell¹², Frances H. Arnold^{12,13}, Ahmad S. Khalil^{10,11,14}, Andrew D. Hanson¹⁵, John E. Dueber^{8,16,17}, David V. Schaffer^{8,16,18,19,20}, Fei Chen^{21,22}, Seokhee Kim¹⁵, Luis Ángel Fernández⁴, Matthew D. Shoulders³ and Chang C. Liu^{1,2,23,24} ✉

Abstract | Directed evolution has revolutionized biomolecular engineering by applying cycles of mutation, amplification and selection to genes of interest (GOIs). However, classical directed evolution methods that rely on manually staged evolutionary cycles constrain the scale and depth of the evolutionary search that is possible. We describe genetic systems that achieve cycles of rapid mutation, amplification and selection fully inside living cells, enabling the continuous evolution of GOIs as cells grow. These systems advance the scale, evolutionary search depth, ease and overall power of directed evolution and access important new areas of protein evolution and engineering.

Directed evolution

A method that employs the evolutionary process of mutation, amplification and screening or selection to improve a protein or other biomolecule towards a desired function on laboratory timescales.

Hypermutation

A marked increase in the mutation rate of a DNA sequence.

Throughout the history of life, evolution has relied on the basic processes of random mutation and natural selection to yield a diverse array of biomolecules with remarkable functions. The field of directed evolution has long sought to leverage the power of evolution to engineer novel biomolecular functions^{1,2}. However, the mutation rate of DNA replication in a typical bacterial, yeast or human cell is 10^{-10} – 10^{-9} substitutions per base³, or a mutation within a gene of average length (~1 kb) occurring approximately once in every 1 million to 10 million cell divisions. At such low rates of mutation, it is difficult to sample even simple single mutations that improve a gene of interest (GOI) — and the RNA or protein it encodes — towards a desired function.

Directed evolution has traditionally turned to diversity generation in vitro, where high rates of mutation can be imposed on a GOI using error-prone PCR or randomized oligonucleotide pools². The resulting libraries of GOI variants are then transformed into cells where they are expressed as RNAs and proteins and subjected to selection or screening. Enriched GOI variants serve as templates for the next round of in vitro diversification, transformation and selection or screening, advancing the evolutionary cycle (FIG. 1a). Although directed evolution has revolutionized biomolecular engineering — particularly fluorescent protein, enzyme and antibody engineering^{2,4} — its classical reliance on manually staged evolutionary steps limits the accessible depth and scale of evolutionary search. By requiring in vitro GOI diversification, classical directed evolution forfeits

the autonomous and decentralized operation of natural evolution and restricts directed evolution campaigns to a few evolutionary cycles at the scale of a few independently evolving populations. To rapidly evolve GOIs but remain in vivo requires the targeting of hypermutation to specific genetic material inside the cell while leaving the large host genome alone (BOX 1).

In vivo continuous evolution does this by relying on the construction of targeted hypermutation systems that selectively and durably mutate GOIs inside cells. With such systems, full evolutionary cycles consisting of rapid diversification, selection and amplification can run perpetually and automatically as cells replicate (FIG. 1a). New types of biomolecular evolution experiment characterized by extensive search depth and scale are accessible through continuous evolution (FIG. 1b). For example, continuous evolution can traverse long mutational pathways along rugged fitness landscapes to reach ambitious biomolecular function targets (depth), and researchers can evolve many GOIs in parallel or one GOI in many replicates, making it possible to access larger sets of target functions, probe the rules of evolution and map sequence–function relationships with greater statistical power. These exciting opportunities provide motivation and aspirations for the broad application of the in vivo continuous evolution methods described.

The main types of in vivo continuous evolution system are viral systems and cellular systems, differentiated by the unit of selection. In viral systems (BOX 2), most prominently phage-assisted continuous evolution

✉e-mail: ccl@uci.edu

<https://doi.org/10.1038/s43586-022-00119-5>

Author addresses

- ¹Department of Biomedical Engineering, University of California, Irvine, CA, USA.
²Department of Molecular Biology and Biochemistry, University of California, Irvine, CA, USA.
³Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA.
⁴Department of Microbial Biotechnology, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CNB-CSIC), Madrid, Spain.
⁵Department of Chemistry, Seoul National University, Seoul, South Korea.
⁶Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA.
⁷Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX, USA.
⁸Department of Bioengineering, University of California, Berkeley, Berkeley, CA, USA.
⁹Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Jalisco, Mexico.
¹⁰Biological Design Center, Boston University, Boston, MA, USA.
¹¹Department of Biomedical Engineering, Boston University, Boston, MA, USA.
¹²Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA.
¹³Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA.
¹⁴Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA.
¹⁵Horticultural Sciences Department, University of Florida, Gainesville, FL, USA.
¹⁶Innovative Genomics Institute, University of California Berkeley and San Francisco, Berkeley, CA, USA.
¹⁷Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
¹⁸Department of Chemical and Biomolecular Engineering, University of California Berkeley, Berkeley, CA, USA.
¹⁹Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA.
²⁰Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA.
²¹Broad Institute of MIT and Harvard, Cambridge, MA, USA.
²²Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA.
²³Department of Chemistry, University of California, Irvine, CA, USA.
²⁴Center for Synthetic Biology, University of California, Irvine, CA, USA.
²⁵These authors contributed equally: Rosana S. Molina, Gordon Rix.

(PACE)⁵, the unit of selection is the virus, so evolvable GOI functions are those that can be linked to viral fitness. In cellular systems, the unit of selection is the cell, so evolvable GOI functions are those that can be linked to cellular fitness. Although there is overlap in the types of function that can be linked to these two units of selection and although viral fitness depends on cellular fitness, cellular systems carry unique advantages. They support a more direct link between GOI functions and cell metabolism or physiology, useful in the evolution of metabolic enzymes and pathways, tolerance and antibiotic resistance. They are more appropriate for evolving GOI functions that will, ultimately, be applied in a cellular context, such as strain or therapeutic cell engineering. They enable screens or selections based on the physical properties of cells, such as cell sorting. They may permit evolution in complex settings disrupted or made irrelevant by viral propagation, such as the evolution of GOI function in microbial communities or multicellular tissues and animals. Additionally, cellular systems include practical advantages such as their ease of cultivation, translating to increased accessibility and scalability of evolution experiments. For these reasons, and because viral systems have been extensively

reviewed elsewhere^{6–8}, this Primer focuses on cellular systems.

Since the early 2000s, researchers have worked on developing cellular systems for in vivo continuous evolution^{7–19}. In these systems, the GOI is encoded in the genome of the cell, on plasmids or within other types of DNA elements^{20,21}. Molecular machinery is engineered to target the GOI for hypermutation while sparing other DNA elements such as the host cell's genome. We focus on three recently developed systems: OrthoRep, MutaT7 and EvolvR (FIG. 1c). In the OrthoRep system, a special error-prone DNA polymerase (DNAP) replicates a linear plasmid encoding the GOI. OrthoRep achieves targeting through orthogonal replication: the error-prone DNAP does not replicate the genome and host DNAPs do not replicate the linear plasmid^{13,18}. In MutaT7 systems, a nucleobase deaminase is fused to T7 RNA polymerase (T7RNAP). T7RNAP is specifically recruited to a T7 promoter placed next to the GOI and, as T7RNAP transcribes, the deaminase alters the GOI. The term MutaT7, coined in the first publication of such a hypermutation strategy¹⁶, serves as an umbrella term for all systems applying a similar approach^{22–26}. In EvolvR, an error-prone DNAP is fused to nickase Cas9 (nCas9)²⁷. At a target site dictated by a guide RNA (gRNA), nCas9 makes a single-stranded DNA break from which the error-prone DNAP extends with low fidelity and limited processivity^{17,28}. With these in vivo hypermutation systems in place, if the activity of the GOI is linked to increased cell fitness, simply culturing cells under selection drives the evolution of improved GOI variants.

This Primer is for scientists looking to do their own in vivo continuous evolution experiments where the cell is the unit of selection. We describe experimental considerations, expected results and successful applications.

Experimentation

There are five basic steps to completing an in vivo continuous evolution experiment: choosing the starting GOI sequence(s), host cell and hypermutation system (FIG. 2a); designing the selection or screen for the desired GOI activity (FIG. 2b); setting up the hypermutation system in the chosen host cell (FIG. 2c); planning and executing the evolution campaign (FIG. 2d); and collecting and analysing results (FIG. 2e), described in the next section. The first three steps should be approached concurrently to exploit interdependencies.

Choosing a starting point for evolution

As is true for any directed evolution experiment, choosing the GOI sequences from which to start evolution is a critical step. Important considerations include the activities of the starting sequences as well as whether strategies to maximize the scale of experimentation should be leveraged initially or after some pilot experimentation.

In classical directed evolution, a typical precondition is that the starting GOI sequence — or at least one member of a library of variants built from the starting sequence — has detectable activity for the function being evolved. Whereas this is also ideal for in vivo continuous evolution, this condition may be relaxed because

Clonal interference

When one clone with a (new) beneficial mutation fails to fix because another lineage with a (new) beneficial mutation arises in the same population, common in asexual populations when mutation rates are high.

Integration cassettes

Pieces of DNA designed to integrate into a specific location within another piece of DNA such as a genome or a plasmid.

the population sizes and diversity that can be accumulated through in vivo targeted hypermutation can be much higher than the diversity that can be transformed into cells for selection in classical directed evolution. Nevertheless, characterization of activity in the starting GOI sequence is recommended to begin any evolution experiment. Specifically, the starting GOI sequence should be measured to see whether its detectability increases the growth rate (or selective advantage) of the cell under selection for the desired function; if not, proceed with moderated expectations.

With in vivo continuous evolution, one can leverage experimental scalability to evolve from multiple starting GOIs in separate experiments or one starting GOI in independent replicates. Collections of different GOI starting points (such as orthologues of an enzyme found in nature or computationally designed libraries) create distinct opportunities to evolve the desired function^{29,30}. Their separation into independent evolution experiments ensures that a given GOI starting point with weak initial activity is not immediately outcompeted by another GOI starting point that has higher initial activity^{31–33}. Likewise, separating a single evolution experiment from one starting GOI into multiple smaller replicates of the experiment can limit the influence of clonal interference³⁴. In both cases, separation into independent evolution experiments should favour the exploration of a greater number of evolutionary paths, increasing the chance of finding the most exceptional functional outcomes^{30,35}.

Setting up the hypermutation system

The cellular hypermutation systems OrthoRep, MutaT7 and EvolvR each have special properties and specific set-up requirements. When choosing the best hypermutation system for a particular experiment, it is important to first consider the host cell appropriate for expressing the GOI and the GOI function being evolved. OrthoRep currently only functions in yeast cells whereas MutaT7 and EvolvR systems function in *Escherichia coli*, yeast and higher eukaryotes. Other important aspects requiring consideration include durability and ease of implementation. OrthoRep is unique in that hypermutation of the GOI is enforced and the host genome does not experience any elevation in mutation rate. These properties make it possible to durably mutate GOIs over prolonged continuous evolution experiments, as discussed in detail in previous literature³⁶. EvolvR and MutaT7 systems are distinguished by their ease of implementation as they rely on standard parts such as nCas9 and expression elements from the T7RNAP ecosystem. For additional considerations on choosing a particular system, see the Limitations and optimizations section.

Special aspects of OrthoRep. OrthoRep is derived from a natural plasmid system found in the yeast *Kluyveromyces lactis* and ported to *Saccharomyces cerevisiae*¹³. The natural system comprises two linear plasmids in the cytoplasm, p1 (8.9 kb) and p2 (13.4 kb)³⁷. Each plasmid is replicated by its own dedicated DNAP through a unique protein-primed mechanism in which the DNAP recognizes terminal proteins covalently linked to the 5' ends of the plasmids to begin replication. The wild-type p1

plasmid encodes the DNAP that exclusively replicates p1, in addition to a toxin and its antitoxin. The p2 plasmid encodes the DNAP that exclusively replicates p2 (REF.³⁸), in addition to associated replication components and transcription machinery for cytoplasmic expression from p1 and p2.

In OrthoRep, the p1 DNAP has been engineered to be highly error prone so that GOIs encoded on p1 experience an elevated mutation rate¹⁸. Owing to the orthogonal replication mechanism, hypermutation is exclusive to p1 and does not affect the genome. The mutation rate of the most error-prone orthogonal DNAP engineered to date is 10⁻⁵ substitutions per base, or 100,000-fold above the genomic mutation rate. There are two error-prone orthogonal DNAPs in regular use, available as pAR-Ec633 and pAR-Ec611 on Addgene, and referred to as 633 and 611, respectively. The 633 DNAP contains the mutations L477V, L640Y, I777K and W814N; and the 611 DNAP contains the mutations I777K and L900S. The 611 DNAP sustains a higher p1 copy number and a lower mutation rate than 633. As a higher copy number leads to higher expression of the p1-encoded GOI, it is sometimes advisable to use the 611 DNAP.

OrthoRep uses an orthogonal transcription system. The p2 plasmid, which can be considered an accessory plasmid for OrthoRep, encodes an RNAP that recognizes special promoters driving GOIs on p1. Various promoters have been engineered to drive the expression of p1-encoded GOIs at strengths matching moderately expressed host genes³⁹.

OrthoRep-specific set-up. OrthoRep requires the GOI to be integrated into the p1 plasmid. This starts with a *S. cerevisiae* strain that already harbours p1 and p2, such as strain F102-2 (REFS^{37,40}). Integration cassettes can be designed to replace the DNAP and toxin-antitoxin genes present on wild-type p1 with any GOI alongside an antibiotic or auxotrophy selection marker. By transforming cells with linearized versions of such cassettes and selecting for the integration product, one obtains strains that contain a recombinant p1 with the GOI encoded. Then, an error-prone orthogonal DNAP encoded on a nuclear plasmid (such as pAR-Ec633 or pAR-Ec611) can be transformed into cells and the GOI will undergo autonomous hypermutation. The error-prone orthogonal DNAP can also be transformed concurrently with the p1 integration cassette.

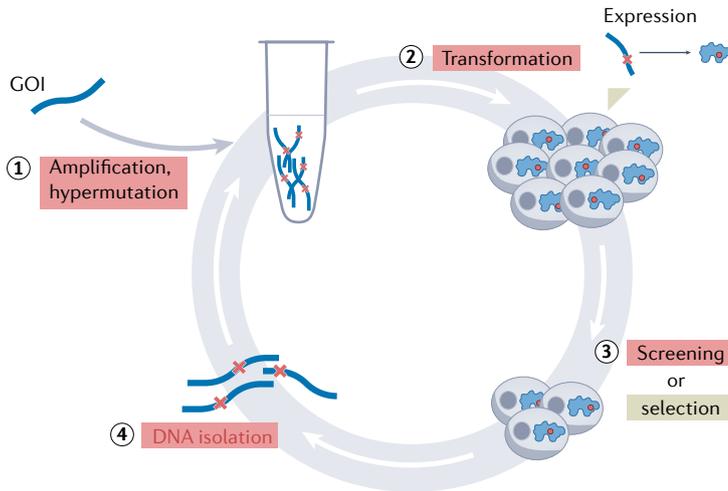
A nuance in this procedure is that p1 is a multicopy plasmid. Therefore, when the GOI is integrated into p1, resulting cells can carry both wild-type and recombinant p1. Once the error-prone DNAP is added to the cell, the wild-type p1 may no longer be required and can be lost over time. This process can be accelerated by using CRISPR-Cas9 to degrade wild-type p1 (REF.⁴¹) or designing a recombinant p1 that is smaller than wild-type p1 so it has a replicative advantage. It is often the case, however, that wild-type p1 will remain, because it allows for higher p1 copy numbers and higher expression of p1-encoded genes under selection.

OrthoRep is compatible across all *S. cerevisiae* strains tested⁴¹, and there are various strains available upon request that contain a landing pad p1 to receive GOIs.

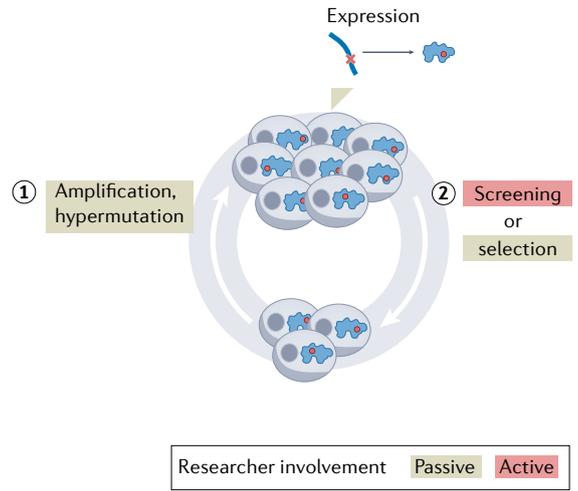
It is also straightforward to transfer recombinant p1 and p2 plasmids from one strain to another by protoplast fusion⁴¹. This is recommended if a pre-existing, extensively engineered host strain is needed for selection of the desired GOI function.

Special aspects of MutaT7. MutaT7 systems have been applied in various model organisms, including *E. coli* (the original MutaT7 system¹⁶, eMutaT7 (REF.²⁴) and T7-DIVA (REF.²³)), yeast (TRIDENT²⁵), plants²⁶ and mammalian cells (TRACE²²). The main component of these systems

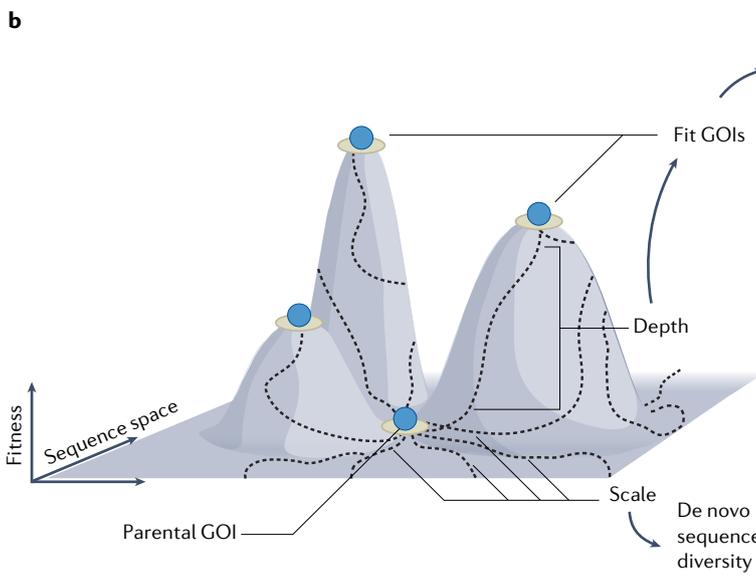
a Classical directed evolution



In vivo continuous evolution



Researcher involvement Passive Active



Efficient enzymes

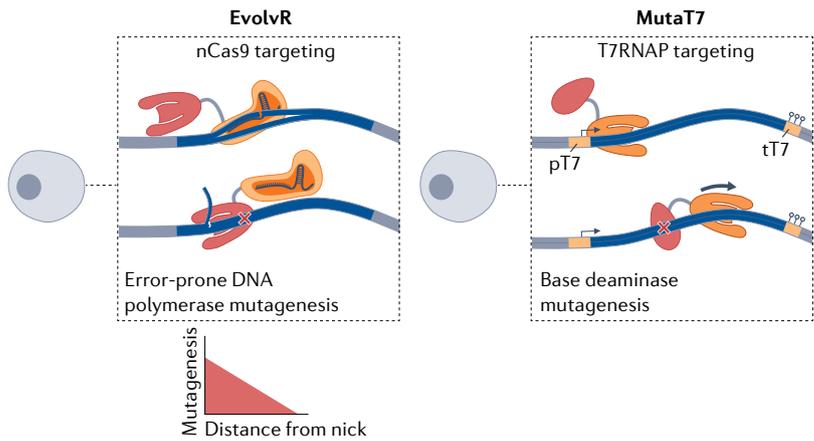
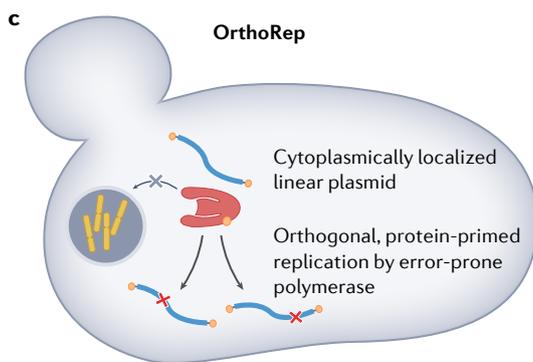
High dynamic range biosensors

High-affinity binding proteins

Statistical learning

Evolutionary dynamics

Inferred sequence–structure–function relationships



Processivity

The ability of an enzyme to catalyse multiple consecutive reactions without releasing its substrate.

is a protein fusion comprising an amino-terminal nucleobase deaminase enzyme and T7RNAP. T7RNAP, derived from the T7 bacteriophage^{42,43}, is highly specific for the T7 promoter, a 23-bp sequence not native to genomes of standard research organisms, and can transcribe almost any DNA downstream of its cognate promoter with high processivity^{44,45}. Unlike OrthoRep and EvolvR, which rely on mutagenesis by error-prone DNAPs (FIG. 3a,b), MutaT7 systems rely on the recruitment of deaminase–T7RNAP fusions to loci adjacent to T7 promoters (FIG. 3a,c). Once the T7RNAP domain of a fusion protein recognizes and binds to the T7 promoter, it unwinds a small portion of double-stranded DNA and initiates transcription. As transcript elongation proceeds, it is the non-template DNA strand that predominantly exists as single-stranded DNA within the transcription R-loop⁴⁶ and becomes exposed to the deaminase domain of the fusion protein, resulting in hypermutation (FIG. 3c). The template strand usually hydrogen bonds with the nascent RNA, and is therefore deaminated somewhat less frequently. The end of the target region of mutagenesis is delineated by a T7 terminator array¹⁶ or catalytically dead Cas9 (dCas9)²³ directed with a CRISPR RNA (crRNA) array to block transcriptional elongation.

Nucleobase deaminases used in MutaT7 systems are either cytidine or adenosine deaminases that accept single-stranded DNA substrates^{47–52}. As their names indicate, nucleobase deaminases catalyse the hydrolysis of exocyclic amines on deoxycytidine (dC) or deoxyadenosine (dA) to generate deoxyuridine (dU) or deoxyinosine (dI), respectively (FIG. 3a). The resultant dU or dI bases

invert the hydrogen bonding properties of the original nucleotides, leading to temporary mismatches at deaminated positions. Unless these deaminated bases are eliminated by DNA repair systems such as uracil-DNA N-glycosylase for dU⁵³ or endonuclease V for dI⁵⁴, these mismatches are resolved as permanent mutations when cellular DNA replication machinery reads dU and dI as deoxythymidine (dT) and deoxyguanosine (dG), respectively^{55,56}. As a result, the deaminase–T7RNAP fusion proteins randomly generate all four possible base pair transition mutations (C•G→T•A, G•C→A•T, T•A→C•G, A•T→G•C) by deaminating the non-template and, somewhat less frequently, template strands. This strand bias can be mitigated by placing T7 promoters on either side of the target region facing inwards and installing terminator arrays just beyond the reciprocal T7 promoters¹⁶.

To date, three cytidine deaminases have been used in MutaT7 systems: rat apolipoprotein B mRNA editing catalytic polypeptide 1 (rAPOBEC1)^{57–59}; activation-induced deaminase (AID), required for antibody maturation in the adaptive immune system^{60,61}; and *Petromyzon marinus* cytidine deaminase 1 (*pmCDA1*), an AID homologue from sea lamprey⁶². The adenosine deaminases used so far for in vivo MutaT7-based hypermutation, as first demonstrated in the T7-DIVA platform²³, are derived from *E. coli* TadA, a tRNA-specific adenosine deaminase that has been evolved to accept single-stranded DNA as a substrate^{63,64}. Although there are now even more active TadA variants (collectively known as the TadA8s)⁶⁵, these adenosine deaminases have yet to be implemented in the context of MutaT7.

The choice of deaminase will largely depend on the desired mutagenesis profile and the host organism. The original MutaT7 (MutaT7^{C→T}) employed rAPOBEC1 (REF.¹⁶), and it was later demonstrated that the *pmCDA1*–T7RNAP fusion was 7-fold to 20-fold more mutagenic in *E. coli*²⁴. A similar relatively higher mutation rate of *pmCDA1*–T7RNAP was concurrently observed with the T7-DIVA platform, which showed that the mutagenic activity of different fusions follows the hierarchy of AID < rAPOBEC1 < *pmCDA1* in *E. coli*²². Demonstrating host dependence of base deaminases showed that AID*Δ (a hyperactive mutant of AID)–T7RNAP fusions were more active than rAPOBEC1–T7RNAP fusions in HEK293T cells with TRACE²². For the yeast MutaT7 system TRIDENT, Cravens et al. employed *pmCDA1*–T7RNAP and also optimized a TadA variant for yeast, yeTadA1.0 (REF.²⁵). In this publication, the group also showed that recruiting DNA repair factors involved in somatic hypermutation to deaminase–T7RNAP fusions can enhance mutagenic diversity by an apparent increase in editing of the template strand.

MutaT7-specific set-up. To carry out a MutaT7 experiment, one encodes the GOI on a plasmid or in the genomes of host cells with a T7 promoter as the recognition element to recruit MutaT7 machinery. In *E. coli*, and in mammalian cells if an internal ribosome entry site is inserted before the GOI, the GOI can be translated directly from the T7 RNA transcript⁶⁶. The T7 promoter can also be placed adjacent to the GOI in the antisense direction if GOI expression should not be

- ◀ **Fig. 1 | Continuous evolution with cellular systems.** a | Comparison of classical directed evolution and in vivo continuous evolution. Classical directed evolution relies upon discrete steps performed in vitro (for example, hypermutation through error-prone PCR) and in vivo. Transformation, DNA isolation and PCR-based amplification and hypermutation are required to complete classical directed evolution cycles. By contrast, in vivo continuous evolution cycles can be carried out autonomously as cells grow. Required steps are each categorized as either requiring limited or no researcher involvement (passive) or requiring extensive researcher involvement (active). Although amplification and hypermutation require cell culturing and, therefore, some researcher intervention, this dilution task is trivial enough to be considered passive. b | Depth and scale in continuous evolution enable broad applications. Mutational depth can yield gene of interest (GOI) variants that are highly fit to perform a desired activity, such as enzymatic activity, binding and biosensing. Scale at which experiments can be carried out provides access to diverse evolutionary pathways, and resulting broad sequence diversity in de novo generated sequences may be used for machine learning on structure and function and to better understand evolutionary trajectories. c | Depictions of OrthoRep, EvolvR and MutaT7. Each system achieves targeted hypermutation. OrthoRep and EvolvR both use error-prone DNA replication-based mutagenesis (pink), with the distinction that in EvolvR error-prone replication is not essential for propagation of the GOI as error-prone replication occurs in addition to replication by host DNA polymerases. MutaT7 utilizes base modification by a nucleotide base deaminase for mutagenesis. All three systems utilize distinct targeting methods (orange). In OrthoRep, the terminal protein fused to error-prone polymerase and linear p1 plasmid provides means of protein-primed replication, a unique mechanism of replication initiation that contributes to orthogonality and makes replication specific to p1. In EvolvR, Cas9 portion of fusion protein localizes error-prone polymerase to the guide RNA (gRNA)-targeted site. In MutaT7, T7 RNA polymerase (T7RNAP) localizes fused deaminase to region between T7 promoter and T7 terminator. OrthoRep has only been applied in yeast, whereas EvolvR and MutaT7 have been used in multiple model organisms. OrthoRep accessory plasmid p2 not depicted for simplicity. Note that termination of transcription and hypermutation by MutaT7 systems may be accomplished by terminators (depicted) or catalytically dead Cas9 (dCas9) (not depicted). nCas9, nickase Cas9; pT7, T7 RNA polymerase promoter; tT7, T7 RNA polymerase terminator array.

Box 1 | Genomic error thresholds

Evolution works best when the mutation rate is high enough to explore sequence space quickly. However, if it is too high, deleterious or lethal mutations are virtually guaranteed in every replication event. This creates the notion of error thresholds. Error thresholds are limits on how high mutation rates can be without escaping the hold of selection^{35,141} or substantially decreasing fitness by ensuring the sampling of lethal or deleterious mutations along with beneficial mutations. Error thresholds scale (roughly) inversely with genome size, as large genomes encode more essential information than small genomes. From the perspective of a single gene of interest (GOI), high mutation rates are therefore allowed and, in fact, needed to explore the GOI's sequence space on laboratory timescales, but these same mutation rates may not be tolerable for the cellular genomes that are typically larger by orders of magnitude. To bypass genomic error thresholds, classical directed evolution approaches settle on diversification in vitro, whereas in vivo continuous evolution strategies specifically target hypermutation to the GOI.

driven from a T7 promoter. To define the end point of hypermutation, a T7 terminator array is inserted downstream of the T7 promoter, or a triple crRNA array/dCas9 targeted to the desired end point in the GOI can be inserted to limit mutation to a section of the GOI. Once these cloning or genome engineering operations are complete, the mutagenesis machinery (such as the deaminase–T7RNAP fusion protein) is introduced. The mutagenesis machinery can be expressed genomically^{16,25} or from plasmids^{22–25} and can also be placed under inducible promoters to achieve varying levels of maximum expression and mutagenesis activity at controlled times^{16,22,23}.

Importantly, when using a cytidine deaminase, the activity of the DNA repair enzyme uracil *N*-glycosylase (UNG) should be neutralized. This enzyme eliminates uracil from DNA to initiate base excision repair, thus suppressing cytidine deaminase-induced mutations. Deletion of the host *ung* gene can prevent this activity, as demonstrated in *E. coli*^{16,23}. Alternatively, the uracil–DNA glycosylase inhibitor (UGI) from bacteriophage PBS2 can be expressed in the host^{67–71}, as was done in eMutaT7 and TRACE.

Special aspects of EvolvR. The EvolvR system comprises a Cas9 nickase (nCas9)²⁷ fused to a low-fidelity DNAP¹⁷. EvolvR diversifies GOIs by recruiting error-prone DNAP activity to single-stranded breaks generated by nCas9 at locations dictated by gRNAs. After nCas9 nicks and dissociates from its target sequence, the fused nick–translating error-prone DNAP initiates DNA extension from the 3' end of the nick, displacing the incumbent strand while unidirectionally generating substitution errors according to the polymerase's error rate. Unlike OrthoRep and MutaT7, EvolvR can target any locus with an adjacent protospacer adjacent motif (PAM) site without the need for prior engineering of the target sequence. As long as the target site remains sufficiently intact for recognition by the gRNA–Cas9 complex, hypermutation will continuously occur.

As EvolvR relies on nCas9 kinetics and DNA polymerization for generating mutations, its exact substitution rate and window length are modular and are determined by the properties of nCas9 and the error-prone DNAP. In its initial design, EvolvR was composed of a nCas9 (*Streptococcus pyogenes* Cas9 containing a D10A mutation) fused to the N terminus of a low-fidelity variant of *E. coli* DNAP I (PolI) harbouring the mutations D424A, I709N and A759R (PolI3M)^{10,17}. A mutated version of EvolvR's nCas9 domain (enCas9) was also made to increase nCas9 dissociation from DNA after nicking, thereby increasing EvolvR's activity by allowing the DNAP to extend more efficiently from the nick. Different variants of EvolvR have been created by changing the fused DNAP to meet different needs. To increase the targeted hypermutation rate, a more error-prone PolI containing mutations F742Y and P796H in addition to those in PolI3M was developed (PolI5M)¹⁷. To increase the length of the editing window, several variants of the more processive bacteriophage Phi29 DNA polymerase (Phi29) were tested. Although using Phi29 increased the targeted window length, it also reduced mutation rates. To increase the length of the editing window while maintaining a high mutation rate, the EvolvR variant nCas9–PolI3M–TBD (thioredoxin-binding domain of bacteriophage T7 DNA polymerase) was constructed, increasing the hypermutation window to reach at least 56 bp downstream of the nick¹⁷; the TBD domain was previously shown to increase the processivity of PolI when inserted into the thumb domain of PolI in the presence of thioredoxin from *E. coli*⁷². The EvolvR variants nCas9 (D10A)–PolI3M or nCas9 (D10A)–PolI5M have been used in most of the experiments carried out with EvolvR so far. EvolvR was initially developed in *E. coli*, and more recently extended to *S. cerevisiae*²⁸.

EvolvR-specific set-up. To use EvolvR, the first step is to design gRNAs to recruit EvolvR to target GOIs for hypermutation. A unique advantage of EvolvR is that one can target endogenous loci in addition to GOIs introduced exogenously in plasmids or integrated into the host genome. Mutations introduced by EvolvR occur at the highest frequency between the nCas9 (D10A)-generated nick and 20–40 bp 3' of the nick, so the desired hypermutation region should be placed within ~20 bp of the gRNA spacer region. If the region of interest is longer than ~40 nucleotides, the region of interest can be tiled with additional gRNAs. In this case, we recommend targeting the same strand, as the expression of two gRNAs that nick separate strands at nearby genomic locations generates double-strand breaks, which are lethal in *E. coli* and may abolish targeting in other organisms. Nicking the same strand at adjacent locations avoids these double-strand break problems.

To express the components of EvolvR, distinct expression cassettes for nCas9–DNAP and gRNAs are included on a plasmid and transformed into the organism of interest. When porting EvolvR into a different strain or organism, we recommend testing different expression strengths of nCas9–DNAP in order to maximize mutation rates on the target GOI while minimizing off-target elevation of mutation rates outside the GOI.

Cheater mutations

Mutations that allow a cell to satisfy selection without actually improving the desired function of the biomolecule under evolution.

For example, in *S. cerevisiae*, a panel of promoters driving EvolvR expression was tested, including pREV1, pRET2, pRPL18B, pTEF1 and pTDH3 in order of increasing promoter strength²⁸. Among them, the highest hypermutation rate at the target GOI was already reached at pTEF1 expression levels whereas the stronger promoter, pTDH3, increased off-target mutation rates at genomic loci outside the GOI without further increasing mutation rates at the GOI. Therefore, in this case, pTEF1-controlled expression of nCas9–DNAP should be favoured over pTDH3-controlled expression.

Selection design

To take full advantage of in vivo continuous evolution, we recommend setting up selections that link the desired GOI function to cell fitness and/or survival. This is straightforward in cases where the GOI function being evolved is already essential to the cell (such as the production of essential amino acids, tolerance to new environmental conditions such as temperature or the presence of toxins, drug resistance, metabolism from new carbon sources, production of cofactors and so on) but less straightforward in cases where the desired GOI function is arbitrary with respect to the natural essential biology of the cell. In the latter case, an engineered genetic or biomolecular circuit is required to link the desired GOI function to the expression of a selectable marker or the activation of an essential protein's function. The advantage of survival-based selections is that when they are coupled to in vivo hypermutation systems, evolution experiments simply involve the serial culturing of cells under selection. Another viable approach is to link GOI function to an optical output for high-throughput screening via fluorescence-activated cell sorting (FACS). The use of high-throughput screening breaks the cycle of continuous evolution into discontinuous steps, but, even so, in vivo hypermutation allows staged cycles of

diversification, selection and amplification to occur in a highly streamlined fashion.

A detailed discussion of selection design^{73,74} is beyond the scope of this Primer, but here we describe some basic principles. Evolution generally works best when selection pressure for the desired GOI activity can be increased over time. Therefore, selection strength should ideally be titratable, for example by altering the concentration of a chemical in the growth medium. The selection should also exhibit a high dynamic range so that higher activity is distinguishable from lower activity across the relevant range. The upper end of the selection, not just the fitness landscape on which the GOI evolves, will limit the possible results of the experiment. Mock selection experiments, in which GOI variants of known fitness are pooled and selection is applied without mutagenesis, can be used to confirm that the selection is capable of enriching fitter GOI variants and serve as a benchmark to evaluate the dynamic range of selection.

Sometimes, it can be helpful to select for an intermediate function that can act as a stepping stone to the final desired function. For example, in the continuous evolution of T7RNAP to recognize new promoter sequences, hybrid promoters containing only some parts of the target promoter sequence were used as stepping stones^{5,75}.

The durability of selection is an especially important consideration for in vivo continuous evolution. As cells are the unit of selection and the typical continuous evolution experiment involves passaging cells over many generations under selection, opportunities for the emergence and fixation of cheater mutations that compromise the link between GOI function and cellular fitness compound. The danger of potential cheater mutations, which by definition occur outside the GOI being evolved, is generally mitigated by the fact that the GOI is hypermutating, giving it most of the opportunity to satisfy the selected function. Still, before embarking on a full evolution experiment, it can be helpful to start with some small pilot experiments to determine an appropriate selection schedule that does not yield frequent cheaters.

Negative selections, in which undesired individuals are actively suppressed in the population, should also be considered when an undesired GOI activity can be selected for by the primary, positive selection. Although negative selections have not yet been demonstrated in cellular continuous evolution systems, they have been employed in both traditional directed evolution and in the viral continuous evolution system, PACE, to engineer specificity in tRNAs⁷⁶, RNA polymerases⁷⁷ and proteases⁷⁸. They will likely be similarly useful in future continuous evolution campaigns with cellular systems.

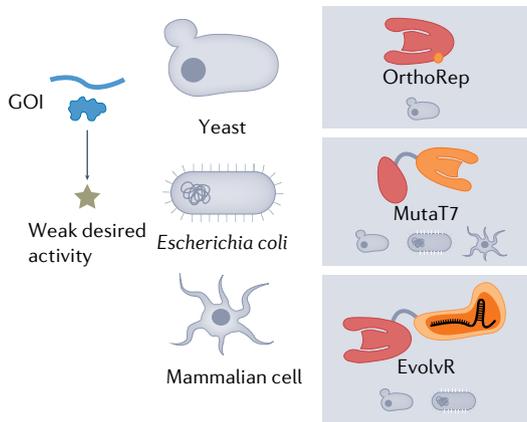
Evolution campaign

During a typical evolution campaign, cells are cultured under increasing selection pressure and the fixation of new mutations in the GOI is observed as cellular fitness improves. Once fitness stops improving, the GOI has reached a local fitness peak and the experiment is stopped in the simple case. There are several considerations at play and experimental variations that deserve attention when executing an evolution campaign.

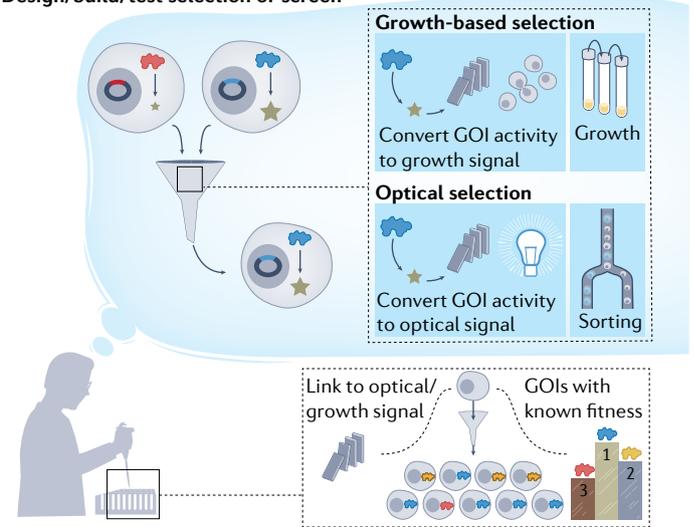
Box 2 | Virus-based in vivo continuous evolution

With viral in vivo continuous evolution systems, the desired gene of interest (GOI) activity is linked to the fitness of viruses as they propagate through cells. The most successfully applied viral system is phage-assisted continuous evolution (PACE), in which the GOI is encoded on a bacteriophage genome engineered to lack essential components for the production of infective virus⁵⁶. The engineered phages are used to infect inducibly mutagenic *Escherichia coli* cells. The *E. coli* cells also contain a genetic selection circuit designed to supply the missing phage components only if the GOI achieves a desired biomolecular function, thus linking the GOI's function to the amount of infective phage produced. In a PACE experiment, phages are serially or continuously diluted in fresh supplies of the mutagenic *E. coli*, resulting in the continuous evolution of the GOI function. Here, *E. coli* cells serve as reagents that do not have to replicate and are, instead, used, discarded and resupplied. Therefore, genome-wide high mutation rates that guarantee lethal mutations in the host genome are allowable so long as those high mutation rates do not destroy the small phage genome containing the GOI. As the cell is not the replicating unit of selection, acting instead as a reagent whose mutations do not persist through evolutionary cycles, phages are the only entities that accumulate mutations, leading to a form of targeted hypermutation. From its initial publication in 2011, researchers have used PACE to evolve a wide range of targets including RNA polymerases⁷⁷, biosensors¹⁴², proteases^{78,143}, insecticides¹⁴⁴, orthogonal tRNA/aminoacyl-tRNA synthetase pairs⁸⁴, transcription factor motifs¹⁴⁵ and genome editors^{65,146,147}, demonstrating PACE's enormous power. More recently, the PACE architecture has been successfully extended to mammalian cells through the development of adenovirus¹⁴⁸ and sindbis virus¹⁴⁹ versions that allow for the continuous evolution of mammalian biomolecules, expanding the scope of addressable problems.

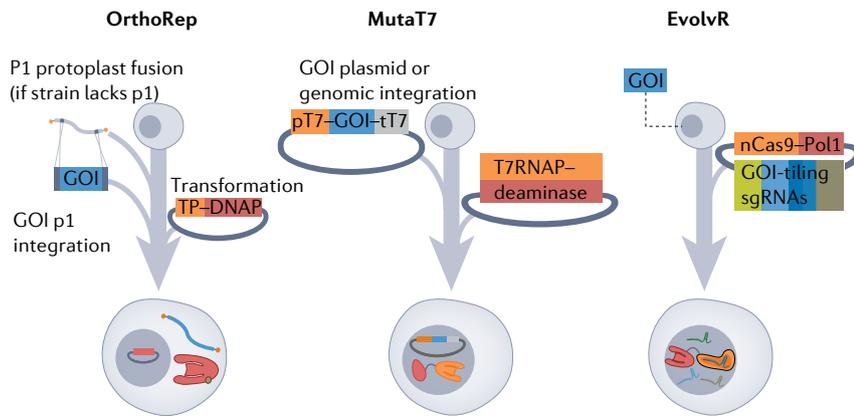
a Choose GOI, host cell and hypermutation system



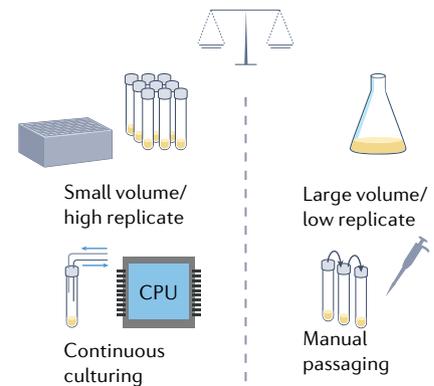
b Design/build/test selection or screen



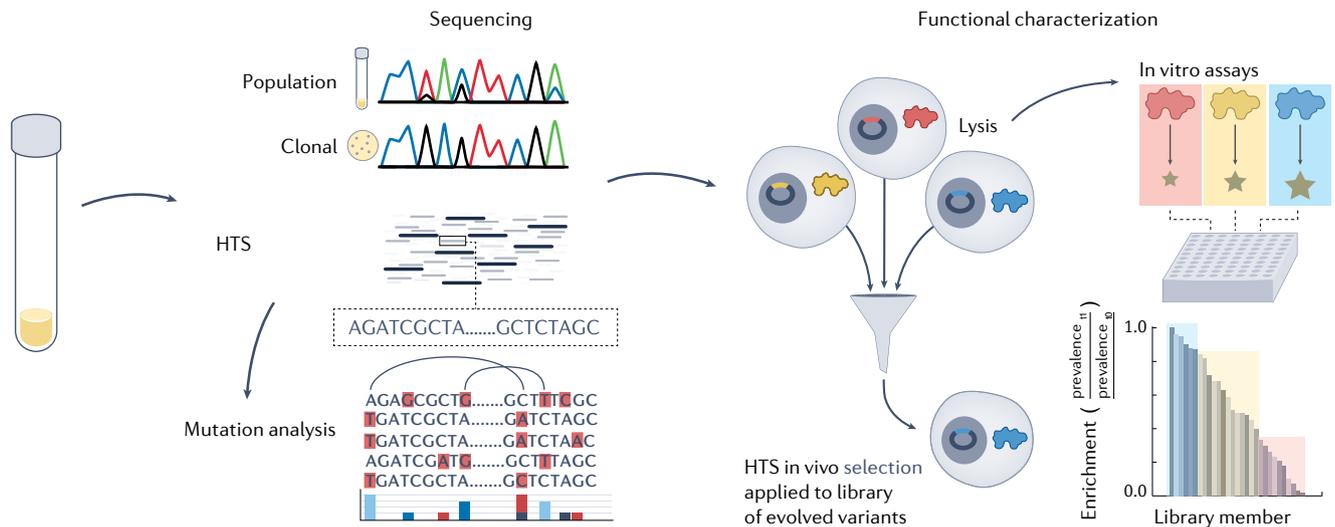
c Generate continuous evolution cell line



d Apply continuous evolution strategies



e Characterize evolutionary outcomes



Within a cell culture, the mutational diversity of the GOI is determined by the size of the culture, the time throughout which mutations have been accumulating and the mutation spectrum and rate of the hypermutation system. The specifics depend on the chosen hypermutation system (Supplementary Table 1), but,

generally, the larger the culture size and the longer mutations are allowed to accumulate, the higher the coverage of sequence space at any given point during evolution⁷⁹. Another aspect to consider is the number of experimental replicates. Evolving several spatially separated populations at the same time can lead to several

◀ Fig. 2 | **Generalized steps for carrying out a continuous evolution experiment.**

a | Gene of interest (GOI) chosen for its predicted potential to evolve a desired activity. Often, ability to carry out the desired activity at a low level is a promising starting point, although activities that are similar to desired activity may be suitable as well. Host cell and hypermutation system also chosen at this stage. **b** | Researcher must design, build and test a selection or screen that is capable of enriching cells encoding GOIs with the desired function. This requires some cell-based mechanism of converting activity of GOI into a growth or optical signal, symbolized as falling dominos. If available, GOI variants with known propensity for desired activity, or fitness, may be used to test whether screen or selection is effective. **c** | GOI and necessary components of hypermutation system encoded in host cell. **d** | Continuous evolution strategies applied to evolve the desired activities. Different strategies offer different advantages and disadvantages and should be considered against each other. Labour, resources, feasible population size, durability and other factors can vary depending on the chosen strategy. **e** | Following an evolution experiment that appears successful, GOIs contained in evolved populations should be isolated and characterized via sequencing and functional assays. Chosen approach will depend on specific demands of the experiment and application. In vitro assays are useful for direct or indirect measurement of GOI function and properties but can only be applied at low throughput. Enrichment studies utilize in vivo selection for GOI function applied to a large library of GOI variants, followed by high-throughput sequencing (HTS) of the library before and after selection to determine level of enrichment of each GOI variant, which serves as a proxy for fitness. DNAP, DNA polymerase; nCas9, nickase Cas9; pT7, T7 RNA polymerase promoter; sgRNA, single guide RNA; T7RNAP, T7 RNA polymerase; tT7, T7 RNA polymerase terminator array.

useful solutions, because each replicate population may be dominated by different evolutionary trajectories^{34,80}. Additionally, this can prevent rare cheaters from overtaking the experiment, as cheater mutations, similar to all mutations, arise stochastically and may not occur or fix in every replicate. There is a practical trade-off between the culture size and the number of replicates an experimenter can manage, however, so these should be balanced based on the culture size needed to achieve reasonable diversity and the expected benefits of many replicates. Finally, continuous evolution campaigns can be run with complex selection histories such as alternating phases of selection and neutral drift, or even alternating selection environments, both of which can act to maximize the crossing of fitness valleys in the search for superior optima^{30,81}.

During a typical evolution campaign, the only hands-on demands of the researcher are to tune the selection parameters, usually by adjusting the composition of the media, and to keep cells propagating over time, usually by serial passaging into fresh media. The only hardware associated with this stereotypical process is standard equipment and materials, the very same non-specialized test tubes, consumables and devices needed for the routine task of inoculating media with microbial or mammalian cell stocks and growing them to saturation or confluence. Indeed, an advantage of cellular systems for in vivo continuous evolution is that typical evolution campaigns involve conventional laboratory hardware, as all of the complex machinery for mutation and selection is autonomously running inside the self-replicating cell. Optionally, one can invest in specialized equipment to automate serial passaging to achieve a fully hands-free evolution campaign. Specifically, a continuous culture bioreactor — a vessel that maintains a culture with an equal inflow and outflow of media — to passage evolving cells under selection can be used. The possibilities of automated culturing have been expanded with eVOLVER (not to

be confused with the hypermutation system EvolvR), an open-source platform to continuously culture tens to hundreds of separate populations under independently controlled growth and selection conditions^{82,83}. With eVOLVER, the researcher can programme a closed feedback loop to adjust the selection pressure on each evolving population based on its measured growth rate or other parameters. In this way, each population is challenged or allowed to drift based on the fitness it has achieved. In addition to enabling automation, this can outperform a predetermined selection schedule that, in some cases, could lead to extinction or a suboptimal fitness plateau during GOI evolution^{83,84}.

Results

Data collection and analysis of results can be divided into sequencing and functional validation steps. Although these are both required in all cases, which of these is prioritized depends on the goal of the project. If the primary goal is to obtain an applicable biomolecular function, low-throughput sequencing methods, thorough characterization of evolved GOI fitness and functional studies on evolved biomolecules are usually sufficient. If the goal is to understand the space of accessible evolutionary outcomes or gain statistical sequence–function relationships for a GOI, high-throughput sequencing (HTS) and high-throughput functional enrichment assays that rank a large number of evolved variants in order of fitness are used. Here, there is also a unique synergy possible with computational approaches, especially machine learning, where large data sets comprising diverse evolutionary outcomes train probabilistic models^{85–89}. If the goal is to study the principles and rules of evolution itself, it may be necessary to carry out HTS and high-throughput functional enrichment assays across multiple time points of an evolution experiment. We discuss these basic types of analysis and results for in vivo continuous evolution experiments below.

Sequencing

Sanger sequencing of heterogeneous gene variant mixtures can usually detect mutations whose frequencies exceed 10%. Therefore, Sanger sequencing of bulk GOI DNA extracted from evolving populations at different time points is an easy way to track the most common mutational pathways being traversed during an evolution campaign. Sanger traces are context-dependent, so to obtain accurate estimates of the frequency of particular mutations in a population, one must compare Sanger sequencing traces of an evolved population with traces of the wild-type sequence with a computational tool such as QSVAnalyzer⁹⁰. This provides estimates of population-level mutation frequency similar in accuracy to HTS, but with much lower labour requirements and cost when dealing with the throughput of only a few samples. However, owing to the aggregated population-level nature of Sanger sequencing data, using it to identify which mutations appear together in the same sequence is challenging and only possible in limited contexts, such as when the most common genotypes in a population can be tracked over time with

frequent Sanger sequencing time points⁹¹. However, such methods are ineffective for identifying rare mutational pathways and extracting information on covariation and

epistatic relationships^{84,86} embedded in the rich sequence data sets that in vivo continuous evolution experiments typically generate.

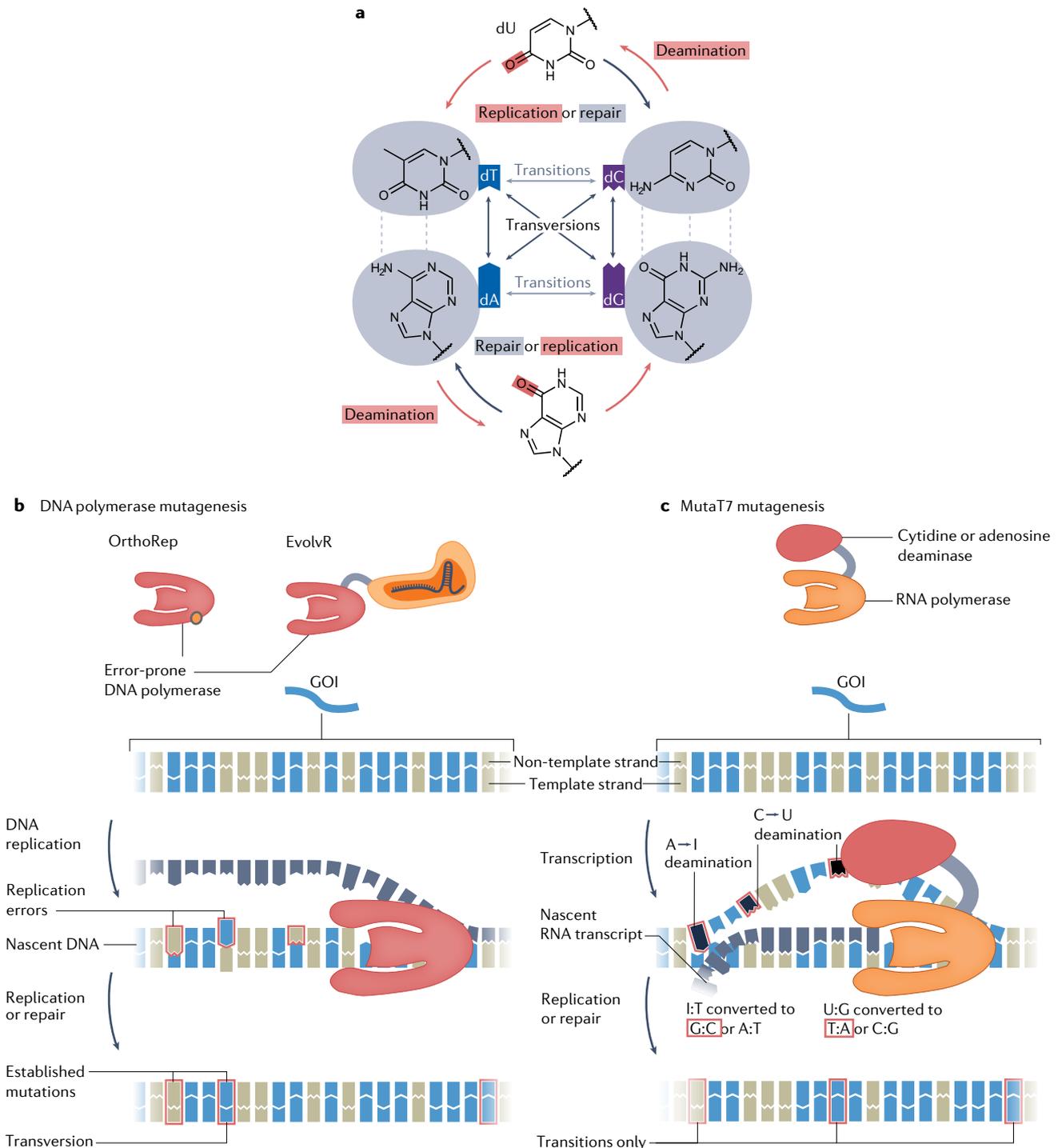


Fig. 3 | **Mutagenesis mechanisms for in vivo hypermutation systems.**

a | Types of nucleotide substitution. Transition describes a mutation that interconverts between nucleotide bases with the same aromatic ring structure (purine to purine or pyrimidine to pyrimidine), whereas transversion describes a mutation that interconverts between nucleotide bases with distinct aromatic ring structures (pyrimidine to purine or vice versa). Enzymatic deamination of deoxycytidine (dC) or deoxyadenosine (dA) forms deoxyuridine (dU) or deoxyinosine (dI), which can serve as the equivalent of deoxythymidine (dT) or deoxyguanosine (dG) in

non-Watson–Crick base-pairing, yielding either a dC→dT or a dA→dG transition mutation, respectively, after DNA replication of the deaminated base. **b,c** | Following mutation by DNA replication error (part **b**) or deamination (part **c**), the nucleotide at the corresponding position on the opposite DNA strand is not yet altered. This mismatch will result in a mutation if it is used as a template in a subsequent DNA replication event. Alternatively, mismatch repair machinery of the cell may correctly repair this mismatch, preventing establishment of the mutation. GOI, gene of interest.

Unlike Sanger sequencing, HTS can detect low-frequency mutations and linkage among mutations on a large scale. DNA for HTS is generated by PCR amplification of targets from distinct samples using primer-appended barcodes to demarcate different evolutionary cultures or time points. These can then be combined prior to sequencing preparation and, subsequently, demultiplexed during analysis steps, saving in preparation and sequencing costs. Additionally, clonal populations used to seed replicate evolution experiments may be uniquely barcoded at the start of evolution to enable multiplexing prior to DNA isolation and PCR.

There are several HTS platforms that are viable options for continuous evolution sequencing projects^{92–94}. The choice of which to use depends on both the length of the target gene and the project goals. Short-read HTS platforms are sufficient if the GOI under evolution is less than ~450 bp in length or if long-range mutation correlations are not of interest, in which case subsections of the GOI can be sequenced independently.

Total sequencing yields and sequencing error rates vary across sequencing platforms and methods of library construction. Illumina's short-read sequencing platform MiSeq produces sequencing yields up to 15 Gb, read lengths of ~500 bp and a raw read accuracy of 99.5%, or roughly 5 sequencing errors per 1 kb (REFS^{94,95}). Long-read sequencing platforms provided by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) offer sequencing yields of 50 Gb and contiguous read lengths of more than 10 kb, but with accuracy ranging from 90 to 98%^{93,96,97}. The relatively low accuracy of raw sequencing reads, which typically guarantees several errors per sequence, can still be valuable for certain applications. For instance, in engineering-focused applications, high-throughput mutant data coupled with Sanger sequencing of selected clones can reveal the most common and most consequential mutations. Methods to process sequencing data to obtain higher accuracy at the cost of read depth can also be considered.

Higher accuracy than that of raw reads is achievable by combining multiple independent reads of the same original sequence to form a consensus sequence. This is typically accomplished either through circular consensus or through unique molecular identifier (UMI) consensus sequencing. For circular consensus, a template is circularized prior to amplification, resulting in concatemeric reads containing multiple copies of the original linear template to form a consensus sequence. For UMI consensus, UMIs composed of random DNA barcodes are appended to the template prior to amplification and sequencing, and consensus sequences are derived from reads with the same or similar UMIs.

These error correction methods allow long-read sequencing platforms to compete in accuracy with short-read sequencing platforms. PacBio is currently the more accessible long-read sequencing platform due to its standardized error correction procedures⁹⁸, but recent examples of error correction methods for ONT sequencing data demonstrate the potential utility of this platform^{97,99,100}, particularly considering the relatively low cost of sequencing device ownership, which can facilitate rapid data generation.

HTS data sets can be analysed using freely available tools for the several necessary steps such as demultiplexing (for example, *Axe*¹⁰¹), alignment (for example, *Minimap2* (REF.¹⁰²)) and variant calling (for example, *VarScan 2* (REF.¹⁰³)). When multiplexing samples, care should be taken to prevent or measure the degree of template switching, which can result in erroneous demultiplexing assignments. Data processing pipelines such as *Breseq*¹⁰⁴ unify some of these steps and improve reproducibility.

Validating evolved activities

A successful continuous evolution experiment will produce a population of GOI variants that satisfy the selection and/or screening conditions. However, cheater mutations that circumvent the selection and/or screen are possible, and deleterious on-target mutations generated towards the end of the experiment may not have had sufficient time to be purged from the population. The expected phenotypes for individual GOI variants must be confirmed in host cells that have not been subject to selection. This can be accomplished through PCR amplification of GOI variants, cloning into a plasmid backbone in library format and, then, transforming this library into host cells to obtain individual clones for analysis. In such uniform fresh strain backgrounds, phenotypic differences will be representative of GOI function. Variants can then be evaluated fairly, ideally using multiple distinct assays.

At a minimum, evolved variants should be compared with unevolved variants using the same selection or screen for the evolution experiment. This comparison can also be performed in high throughput by using functional enrichment assays where barcoded evolved variants are pooled and subject to growth under selection. Enrichment scores for each variant can then be calculated by measuring barcode frequencies before and after selection via HTS¹⁰⁵. In such experiments, it is often necessary to, first, use long-read sequencing to match barcodes to specific GOI sequences and, then, use short-read sequencing — where greater read numbers are available — to track the enrichment of barcodes. The resulting data provide a measurement of relative fitness for many GOI variants, which can be compared with parental GOI variants present in the library.

If engineering is the primary goal of the study, characterization beyond fitness-based assays should be performed. Individual GOI variants with high fitness in clonal populations can be isolated and the biomolecules encoded by those variants purified for in vitro biochemical studies or biological assays.

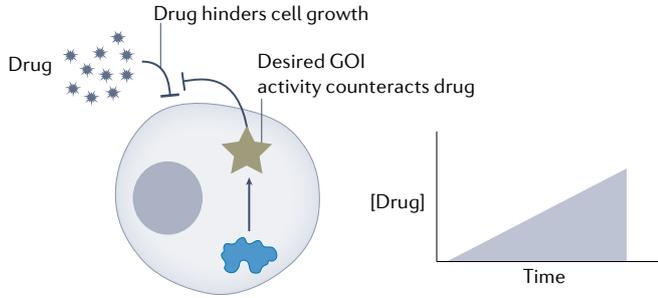
Applications

Successful applications of cellular systems for in vivo continuous evolution thus far fall under three categories: studying pathways to drug resistance, enzyme engineering and FACS-based evolution. There is some overlap among these categories, but they have been chosen for organizational clarity. Potential applications are much broader and we discuss them further in the Outlook section.

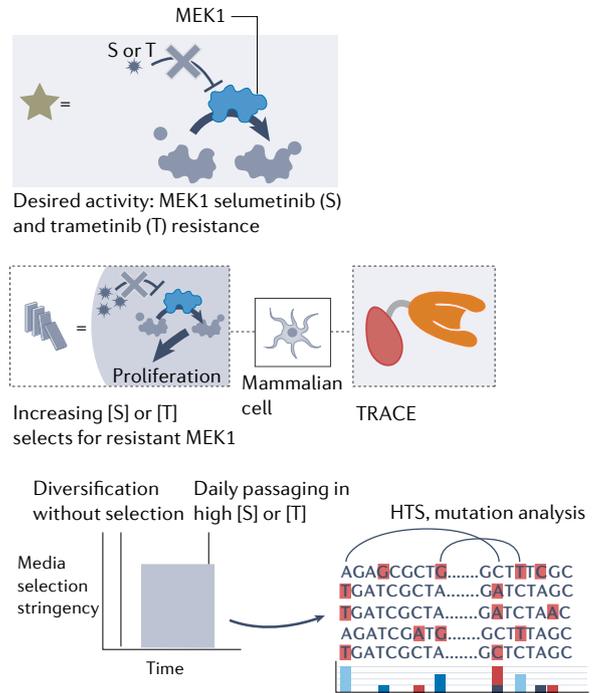
Unique molecular identifier (UMI). A random barcode added to sequencing libraries to differentiate individual molecules from each other before amplification.

Consensus sequencing. An approach used in high-throughput sequencing (HTS) that corrects errors by sequencing a particular sequence multiple times and taking the consensus.

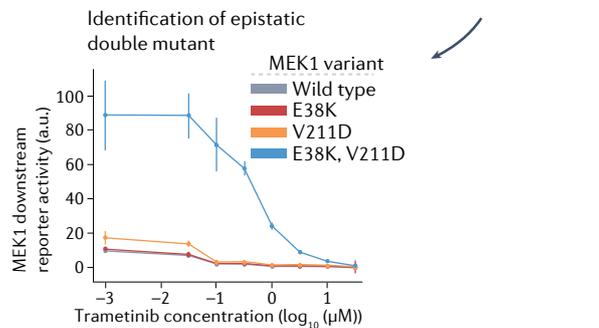
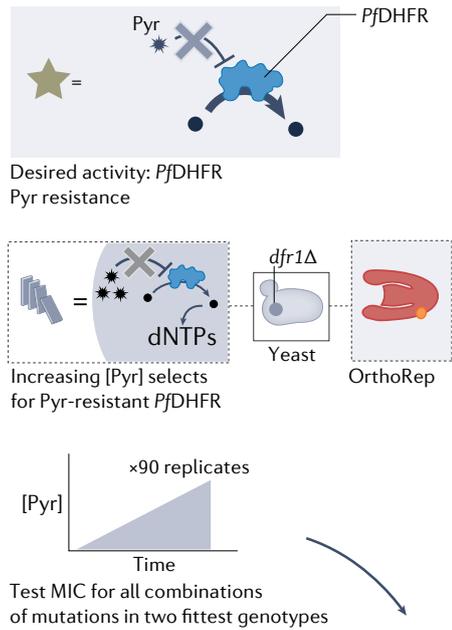
a Drug resistance



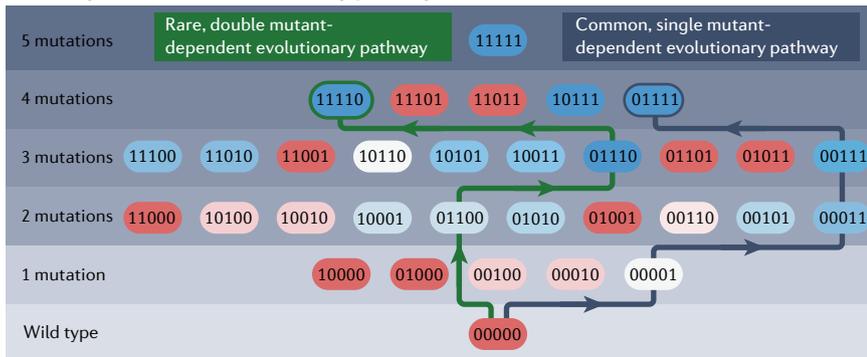
b



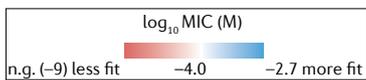
c



Discovery of alternative evolutionary pathways



PfDHFR genotype: 10000 = C50R 01000 = D54N 00100 = Y57H 00010 = C59R 00001 = S108N



Studying pathways to drug resistance

Uncovering drug-resistance mutations in clinically relevant targets is an important application area of directed and experimental evolution. Such efforts predict how a drug may lose effectiveness over time, inform strategies that can limit the development of drug resistance and

reveal basic principles of evolution and evolutionary dynamics. Cellular systems for in vivo continuous evolution are particularly effective in this application space because they can drive the evolution of resistant drug target variants on short laboratory timescales in multiple replicates. This allows researchers to sample the scope of

◀ Fig. 4 | **Application of continuous evolution for studying drug resistance.** **a** | Drug-resistance continuous evolution experiments. Cycle with only one step shown but is meant to represent continuous growth following desired selection schedules, for example serial passaging into media with increasing drug concentration. **b** | Continuous evolution of mitogen activated protein kinase kinase 1 (MEK1) with the TRACE MutaT7 system. Native activity of MEK1 is essential for proliferation in the mammalian cell line used (HEK293T) and is therefore drug selectable. MEK1 targeted for mutagenesis and grown for a short initial period without selection for drug resistance to generate mutational diversity. Selection then applied using a static high concentration of either of the two drugs being studied. **c** | Continuous evolution of dihydrofolate reductase (DHFR) to study drug resistance using OrthoRep. DHFR activity is essential for nucleotide biosynthesis, and therefore selectable in cells deleted for the native enzyme. Ninety replicate cultures were passaged in parallel with gradually increasing drug concentration, and Sanger sequencing of populations revealed two genotypes that commonly become fixed in populations. All combinations of mutations in these two genotypes were individually cloned and tested in isolation to determine fitness and understand accessibility of different mutational pathways leading to different outcomes. GOL, gene of interest; HTS, high-throughput sequencing; MIC, minimum inhibitory concentration; *PfDHFR*, *Plasmodium falciparum* dihydrofolate reductase; Pyr, pyrimethamine. Part **b** adapted from REF.²², Springer Nature Limited. Part **c** adapted with permission from REF.¹⁸, Elsevier.

possible mutational pathways leading to drug resistance, and to test the comparative effects of interventions on the emergence of resistance with statistical power, all in controlled experiments (FIG. 4a).

EvolvR has been used to target the endogenous *E. coli* *rspE* gene for hypermutation to identify novel spectinomycin-resistance mutations in the ribosomal unit S5 (REF.¹⁷). A single overnight growth step to diversify the gene followed by selection on agar plates supplemented with varying concentrations of spectinomycin led to the identification of several resistant variants, including new mutations not previously known to confer resistance. These mutations led to the hypothesis that moving Lys26 relative to the spectinomycin-binding pocket is a mechanism for resistance, which prompted the identification of additional resistance mutations that use this mechanism. Similarly, MutaT7^{C→T} was used in *E. coli* to target an episomal copy of *rpsL*, which encodes the ribosomal unit S12, resulting in the evolution of streptomycin-resistant S12 variants after 24 h of growth/mutagenesis¹⁶. Trimethoprim-resistant variants of the *E. coli* dihydrofolate reductase (DHFR) were also evolved by targeting MutaT7^{C→T} to DHFR in a bioreactor. These studies demonstrate the generality of applying continuous evolution to reveal how mutations in drug targets lead to resistance.

With MutaT7 systems in *E. coli*, both T7-DIVA and eMutaT7 were used to evolve TEM-1 β -lactamase for the ability to degrade third-generation cephalosporin antibiotics^{23,24}. With T7-DIVA, two iterative cycles of mutagenesis followed by one selection step produced double mutants with a >1,000-fold increased minimum inhibitory concentration (MIC) of ceftazidime²³. With eMutaT7, serial passaging of batch cultures into increasing antibiotic concentrations was performed, and clones were isolated after 24–32 h with 9–16 mutations and ~10,000-fold increases in MICs to cefotaxime and ceftazidime²⁴. The MutaT7 system TRACE was also used in mammalian cells to identify two functionally correlated mutations in mitogen activated protein kinase kinase 1 (MEK1) that promote resistance to selumetinib and trametinib — two pharmacologically relevant MEK1 inhibitors²². MEK1 was integrated under a T7 promoter

into the genome, diversified through the action of TRACE, and subjected to selection for drug-resistant cells (FIG. 4b).

By taking advantage of the accessible depth and scale of in vivo continuous evolution, multiple mutational pathways across complex evolutionary landscapes can be explored. In a demonstration of this ability, OrthoRep was used to study how DHFR from the malaria-causing parasite *Plasmodium falciparum* (*PfDHFR*) acquires resistance to pyrimethamine in 90 small-volume (0.5 ml) replicates¹⁸ (FIG. 4c). An engineered yeast strain solely dependent on *PfDHFR* encoded on the hypermutating p1 plasmid was used. After 13 passages into increasing concentrations of pyrimethamine, 78 replicates adapted to the highest soluble concentration (3 mM) and yielded new highly resistant variants with 3–6 mutations. Sanger sequencing of each replicate population across time points showed that multiple mutational pathways in *PfDHFR* led to resistance. Intricate interplay among adaptive mutational pathways was elucidated and traced to the existence of greedy mutations, sign epistasis and clonal interference. From these data, population structures and strategies that favour certain pathways over others were predicted and confirmed through additional replicate evolution experiments.

The MutaT7 system TRIDENT was also used to evolve pyrimethamine-resistant *PfDHFR* variants in yeast²⁵. The study with TRIDENT observed the dominance of a single mutation (D54N) that conferred resistance to 3 mM pyrimethamine across 180 replicate cultures. This is in contrast to the OrthoRep experiment, where three to six mutations were necessary to achieve full resistance to pyrimethamine, with S108N, C59R, Y57H and D54N being most dominant¹⁸. A possible explanation for the difference in outcomes is that evolution experiments with OrthoRep and TRIDENT started from different strengths of *PfDHFR* expression. Notably, pyrimethamine-resistant *PfDHFR*s observed in the field are commonly multi-mutant variants containing S108N and C59R (REFS^{106,107}).

Enzyme engineering

By coupling the activity of an enzyme to cell growth, one can apply in vivo continuous evolution to engineer enzymes towards improved and new functions (FIG. 5a). In one example, eMutaT7 was used to evolve the bacterial heat-shock protease DegP to discover mutations that increase its proteolytic activity and understand the fitness consequences of hyperactive DegP variants²⁴. A hypoactive mutant of DegP containing the known activity-reducing mutation A184S was subjected to continuous evolution by increasing temperature over time. Elevated temperatures cause the build up of unfolded or misfolded proteins that harm the cell. As DegP degrades these unfolded or misfolded proteins, high temperature selects for restored activity from the hypoactive DegP A184S mutant. This experiment resulted in the fixation of mutations that compensate for A184S, mutations that by themselves act to yield the desired hyperactive DegPs. In a second example, EvolvR was used to improve the catalytic efficiency of ornithine cyclodeaminase (OCD) for L-proline synthesis from L-ornithine¹⁰⁸ (FIG. 5b).

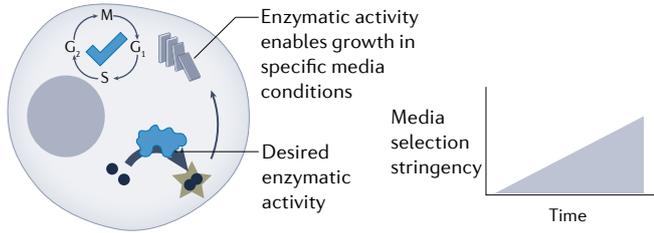
Greedy mutations

Single mutations representing the locally optimal choice for improving the function of a gene during a given stage of evolution.

Sign epistasis

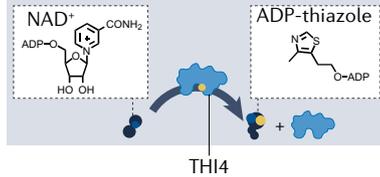
When one mutation that has a particular effect on the desired biomolecular function causes the opposite effect when it is in the presence of another mutation.

a Enzyme engineering

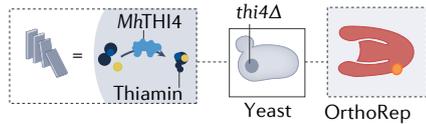
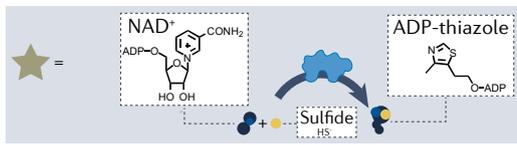


c

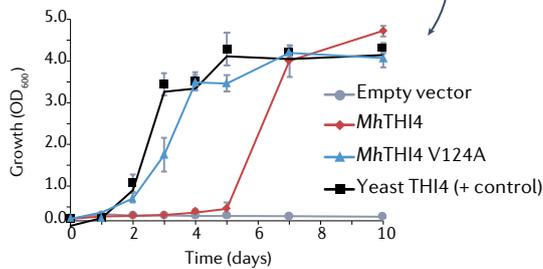
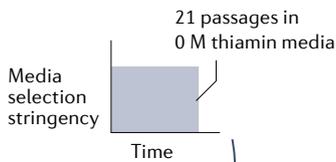
Most THI4 orthologues are only capable of a single catalytic cycle



Desired activity: active multi-turnover THI4

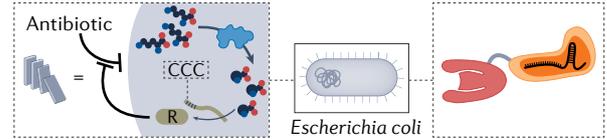
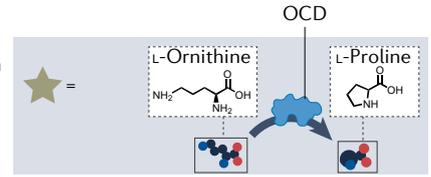


Grow yeast with only a multi-turnover THI4 orthologue without thiamin



b

Desired activity: proline production from ornithine



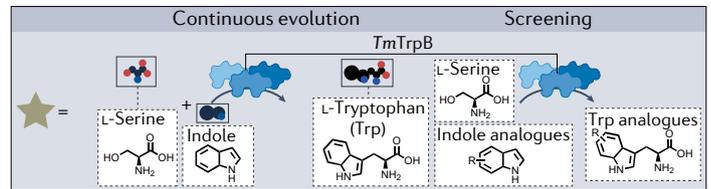
Translation of rare codons in antibiotic resistance gene (R) requires high [proline]

Screen mutants by growth in selective media

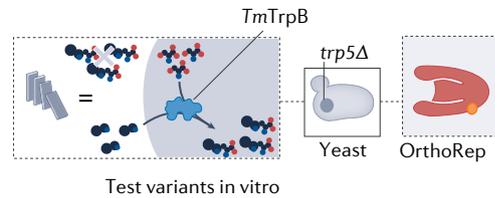
Combine three most promising mutations, test kinetics



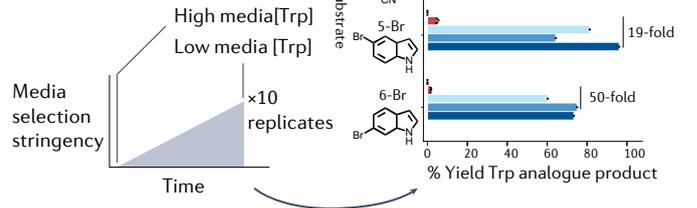
d



Desired activity: synthesis of diverse Trp analogues



Supplementing indole in absence of Trp selects for TrpB activity



A growth-based screen was created in which proline codons in an antibiotic resistance marker were replaced with rare codons, leading to a growth defect that can be rescued by increased L-proline production. After diversifying OCD with EvolvR, variants conferring faster

growth were screened for and three mutations in OCD were found that, when combined, improved enzyme activity by 2.4-fold. In a third example, OrthoRep was used to evolve the thiamin biosynthesis enzyme THI4 from the anaerobic bacterium *Mucinivorans hirudinis*

◀ Fig. 5 | **Application of continuous evolution for enzyme engineering.** **a** | Enzyme engineering through continuous evolution. Cycle with only one step shown, as experiments are typically carried out using growth-based selections. This means all necessary steps for evolution are accomplished by the cell concurrently when provided with a consistent supply of media with increasing selection stringency. Media selection stringency modification is typically accomplished through reducing the concentration of a required nutrient whose production is made to be dependent upon gene of interest (GOI) activity using some natural or engineered linkage between the desired enzymatic activity and growth in specific media conditions (shown as dominos). **b** | Continuous evolution of ornithine cyclodeaminase (OCD) for improved proline production. Following diversification with EvolvR, clones underwent a single round of screening on basis of colony size (dot plot), after which mutations in the three fastest growing clones were manually combined and kinetic parameters of L-ornithine were assessed. Bar plot shows catalytic efficiency for resulting triple mutant in comparison with wild-type enzyme. Michaelis–Menten kinetics determined from measurements taken in at least triplicate for L-ornithine concentrations ranging from 0 to 60 mM. Error bars represent 95% confidence interval. **c** | Adaptation of multi-turnover thiazole synthase from the prokaryote *Mucinivorans hirudinis* (*MhThi4*) for activity in plant-like cellular environment of yeast. After evolution of *MhThi4* for activity in yeast, an evolved variant with V124A mutation was compared with parental multi-turnover. Growth curves show improved complementation by evolved *MhThi4* V124A relative to wild-type *MhThi4* of a *Thi4*-deficient yeast strain. Empty vector and yeast *Thi4* controls are included for comparison. Data are means and standard errors of two to six independent replicates. **d** | Continuous evolution of tryptophan synthase from the thermophile *Thermotoga maritima* (*TmTrpB*) to generate enzymes with diverse substrate promiscuities. OrthoRep first used to evolve *TmTrpB* to function in yeast in ten replicates. Resulting diverse *TmTrpB* variants then expressed in *Escherichia coli* and tested for in vitro activity with a panel of several indole analogue substrates, assessed by HPLC percentage yield of the corresponding Trp analogue. Yields from three of the best performing OrthoRep-evolved TrpB variants shown in comparison with parental *TmTrpB* variant and empty plasmid vector negative control. Dots represent percentage yield for single biological replicate. Bars either means of four independent replicates (empty vector and parental TrpB) or percentage yield for single replicate (evolved TrpBs). Part c adapted from REF.¹¹⁰, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>). Part d adapted from REF.¹¹³, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>).

(*MhThi4*) to function efficiently in aerobic conditions similar to those in plant cells, as a step towards its use to replace the highly inefficient native plant *Thi4* and increase plant productivity^{109,110} (FIG. 5c). Many eukaryotic *Thi4* orthologues, including those of plants and yeast, use an active-site cysteine residue as the sulfur donor for the reaction and can, thus, catalyse only one reaction, making these enzymes energetically costly and a target for replacement with a longer-lived version¹¹¹. *MhThi4* instead uses free sulfide as the sulfur donor and mediates multiple reaction cycles¹¹². However, this orthologue is not fit to function in plants, as it is oxygen-sensitive. To adapt it to function in plant-like conditions, *MhThi4* was encoded into the OrthoRep system in a yeast strain with the native *Thi4* deleted. After 21 passages of 9 starting populations, multiple single and double mutations were obtained that improved growth in the absence of thiamin.

Continuous enzyme evolution can also be automated. OrthoRep combined with the continuous culturing platform eVOLVER was used to adapt an enzyme to a new environment^{82,83}. The set-up, termed automated continuous evolution (ACE), was used to evolve the thermophilic *Thermotoga maritima* HisA enzyme (*TmHisA*) for mesophilic activity in yeast, with implications for industrial biotechnology applications that commonly require changing the temperature optimum of enzymes. The evolution of HisA highlights ACE's potential to realize speed (ACE arrived at HisA

solutions hundreds of hours faster than manual batch culture-mediated selection), scale (ACE autonomously managed replicate cultures at volumes >25 ml with frequent, minimal dilutions, minimizing population bottlenecks, and independently modulated the histidine concentration of each culture based on feedback from real-time growth rates, maintaining optimal selection across the replicates) and depth (evolution occurred over 600 h of continuous selection through long mutational pathways ranging from 5 to 18 mutations, suggesting that ACE can traverse relatively complex fitness landscapes that necessitate a large number of small effect mutations to reach desired activity).

In a final example of enzyme engineering, the scale of experimentation possible with continuous in vivo hypermutation was leveraged to evolve a diverse set of TrpB variants that was then mined for substrate promiscuity that lead to the production of valuable chemicals¹¹³ (FIG. 5d). TrpB and its allosteric partner TrpA make up tryptophan synthase, which mediates the final steps of L-tryptophan (Trp) production¹¹⁴. After receiving indole from TrpA, TrpB synthesizes Trp by coupling the indole to L-serine. TrpB enzymes can also accept indole analogues and readily convert them into Trp analogues, which are useful as biological probes and as scaffolds in the synthesis of pharmaceuticals. Previously, several directed evolution campaigns have been carried out to evolve TrpB to function in the absence of TrpA and expand its substrate scope^{115–118}. Rix et al. reasoned that in vivo continuous evolution could be used to improve and scale this process¹¹³. Using OrthoRep, a thermophilic TrpB enzyme was continuously evolved in yeast to complement the biosynthesis of Trp from exogenously supplied indole in several replicates, resulting in highly active TrpB variants containing up to 16 mutations. A panel of more than 60 TrpB variants from 10 independently evolved populations displayed a diverse range of promiscuous activities, with up to 50-fold improvements in activity at mesophile temperatures, despite selecting only for cognate Trp synthesis activity. Not only are these TrpB variants commercially useful but using this new method for synthetic generation of enzyme orthologues should be general to the expansion of activities and substrate promiscuity profiles of other biosynthetic enzymes.

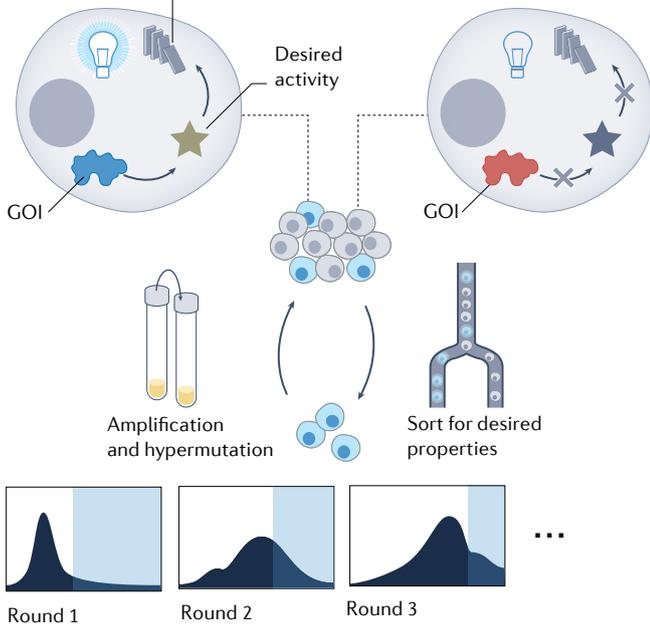
FACS-based evolution

In vivo continuous evolution systems can also streamline the engineering of biomolecules with FACS when the desired function is tied to a fluorescent output (FIG. 6a). The simplest application is to evolve a biomolecule that is itself fluorescent, such as a fluorescent protein. In two such examples, the MutaT7 strategy was applied in yeast to evolve a red-shifted variant of mCherry²⁵ (FIG. 6b) and in mammalian cells to shift the emission spectra of blue fluorescent protein to that of green fluorescent protein (GFP)²².

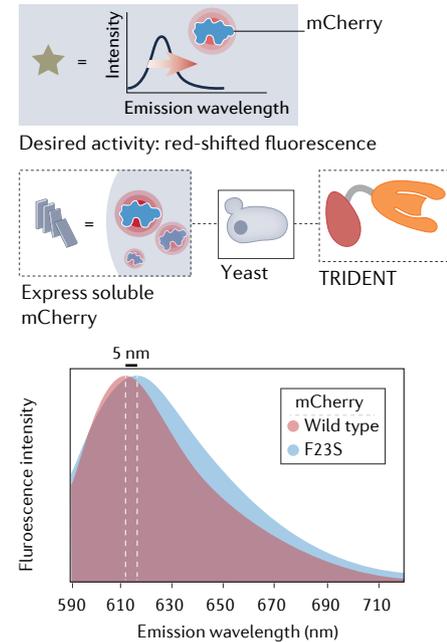
Another way to combine in vivo continuous evolution and FACS is to evolve a GOI whose desired function leads to the expression of a fluorescent reporter gene. For example, OrthoRep was used to evolve the allosteric transcription factor BenM to sense the presence of its

a FACS-based

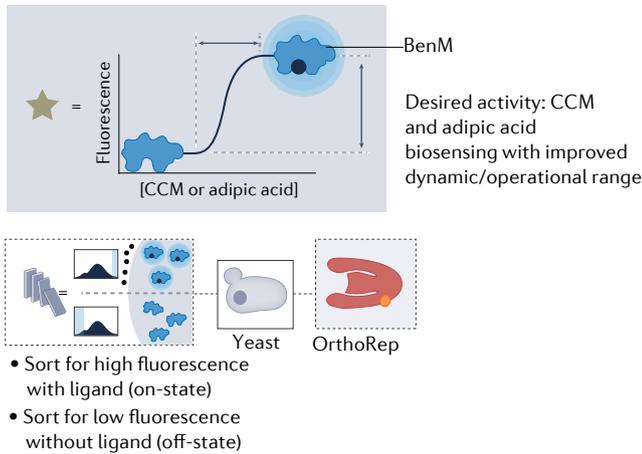
Transduction into optical signal via genetic circuit and/or cell treatment



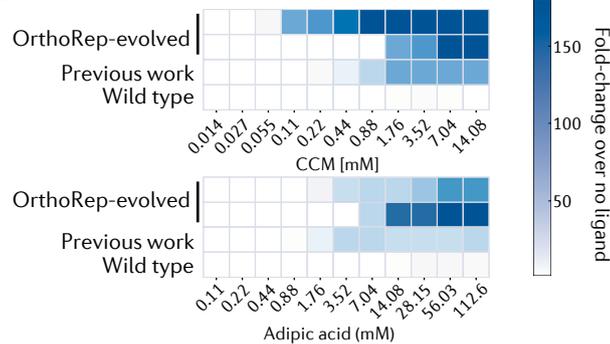
b



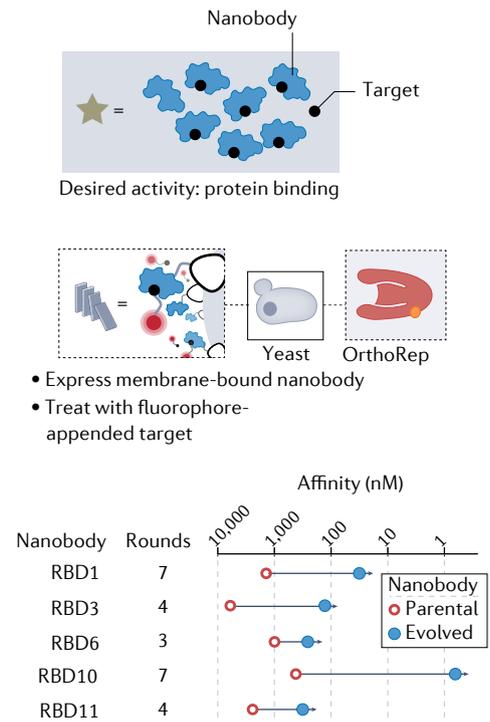
c



BenM variant



d



cognate ligand, muconic acid, as well as a non-cognate ligand, adipic acid¹¹⁹ (FIG. 6c). With BenM encoded on p1 for targeted hypermutation, 11 evolutionary cycles of yeast culturing were carried out, and positive and negative rounds of FACS were used to enrich cells that expressed GFP only in the presence of either muconic

acid or adipic acid. The evolved biosensors displayed broad operational ranges of sensitivity to biologically relevant concentrations of muconic acid and adipic acid, as well as high dynamic ranges up to 180-fold. High-performance biosensors can, in turn, be used as the read-out to evolve synthetic metabolic pathways

◀ Fig. 6 | **Continuous evolution applications employing FACS-based screening.**

a | Continuous evolution experiments that employ fluorescence-activated cell sorting (FACS)-based high-throughput screening to evolve genes of interest (GOIs) with a desired activity. Unlike selections, screening and amplification must be carried out in two discrete steps. Dominos represent an effective screen that must link the desired activity of the GOI (star) to an optical signal (lightbulb). With such a screen, GOI variants capable of carrying out the activity (blue GOI) exhibit an increased optical signal, whereas GOI variants less capable of carrying out the activity (red GOI) exhibit a lower optical signal. Through multiple consecutive rounds of sorting with gates set to enrich for the desired property (blue area in plots), a desired activity can be evolved. **b** | Evolution of red-shifted mCherry using TRIDENT MutaT7 system. Emission spectra of parental wild-type mCherry and evolved F23S mCherry variant. **c** | Evolution of *cis,cis*-muconic acid (CCM) biosensor BenM to exhibit higher dynamic range (top left, horizontal arrows) and broader operational range (top left, vertical arrows) using OrthoRep system. Following multiple rounds of selection for high fluorescence in presence of ligand and low fluorescence without ligand (bottom left), OrthoRep-evolved variants tested for biosensing of CCM and adipic acid by fluorescence. Heat maps depicting fold change in fluorescence with and without ligand for OrthoRep-evolved BenM variants in comparison with BenM variants that have been engineered in previous work and wild-type BenM for both CCM and adipic acid. **d** | SARS-CoV-2 receptor-binding domain (RBD) nanobody evolution using OrthoRep for diversification and yeast display for FACS-based screening. Prior to FACS, cells treated with RBD, which itself binds to a fluorophore. Thus, cells encoding surface-bound nanobodies that tightly bind RBD can be enriched for via FACS. Following several rounds of FACS applied to different RBD nanobody variants, clonal OrthoRep-evolved RBD variants isolated and tested for affinity to RBD by surface plasmon resonance (SPR). Affinity plots compare RBD affinities for evolved nanobodies with their parental nanobodies. Part **b** adapted from REF.²⁵, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>). Part **c** adapted with permission from REF.¹¹⁹, ACS. Part **d** adapted from REF.¹²¹, Springer Nature Limited.

that more efficiently produce the sensed molecule. In a related example, an enzyme from a muconic acid production pathway in yeast was encoded on OrthoRep, and BenM was used as the biosensor to guide selection for higher muconic acid production¹²⁰.

OrthoRep has also been used to drive the rapid evolution of antibodies in a system termed AHEAD (Autonomous Hypermutation yEAST surfACE Display) (FIG. 6d). Here, antibody scaffolds are encoded for yeast surface display from the orthogonal p1 plasmid¹²¹. Culturing of the yeast cells results in the self-diversification of the displayed antibodies such that straightforward cycles of yeast growth, induction of surface display and FACS for cells that bind to a labelled antigen generate high-affinity antibody variants over time. In the midst of the COVID-19 pandemic, AHEAD was used to evolve nanobodies with sub-nanomolar affinity and pseudovirus neutralization potency to the receptor-binding domain (RBD) of SARS-CoV-2. Starting from a naive synthetic nanobody library, eight parental clones were selected with weak binding to the RBD, the sequences were transplanted onto p1 and eight separate evolution experiments were carried out involving cycles of yeast culturing and FACS to affinity mature the parental clones into high-affinity RBD binders. The resulting nanobodies reached sub-nanomolar binding affinities and neutralization potencies by evolving several hundred-fold improvements in some cases. The streamlined nature of AHEAD experiments allowed the eight evolution experiments to be run in parallel, which prevented clonal interference among lineages derived from distinct parents and promoted functional diversity, such as the location of bound RBD, in the set of final binding proteins.

Reproducibility and data deposition

Evolution campaign reporting

To ensure reproducibility, researchers undertaking *in vivo* continuous evolution experiments must report important details of their experimental design as well as how evolved sequences are characterized and annotated. *In vivo* continuous evolution systems are still under active development, so in addition to reporting the specific system that is used, its precise architecture (exact variants of mutagenic polymerases/enzymes, sequences of gRNAs, genetic modifications to publicly available strains) should also be reported. If feasible, researchers should include the exact sequences of plasmids and modified genomic loci present in strains used for evolution.

The selection used for evolution experiments should be well tested and documented. The exact sequences of GOI starting variants should be reported. How selection is applied during evolution, including the number and volume of cultures, how dilutions are carried out, the volume of culture transferred in each passage, the increments used in modifying selection stringency and the criteria used to determine when to increase selection stringency should all be documented. Additionally, controls used to confirm that evolution is working as expected, such as GOI variants with a single inactivating mutation, should be described.

Evolution outcomes reporting

All GOI variants that are characterized individually should be fully sequenced, even if mutagenesis was targeted only to a particular region of the GOI, and full sequences should be included in publication. Lists of amino acid mutations are, of course, necessary, but for convenience of other researchers and to capture synonymous mutations that may have functional significance, complete sequences should be included as well.

Given the wealth of sequence diversity that continuous evolution can generate to the benefit of future researchers, we encourage HTS data to be collected, properly annotated and publicly deposited. Both raw sequencing data and preprocessed data (for example, data that have been demultiplexed or error-corrected) should be deposited in a public database such as the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA), the NCBI BioProject and/or the European Nucleotide Archive (ENA). Ideally, any analysis performed on HTS data should be easily reproduced, for instance using a version-controlled pipeline that is available for download with clear installation instructions. At a minimum, the analysis steps performed should be carefully described, including all non-default options used for command line tools. Any custom scripts that are critical to the conclusions of a study should be publicly accessible, accompanied by a description of the necessary dependencies.

Limitations and optimizations

Host

When deciding which system to use for an *in vivo* continuous evolution experiment, one clear consideration is the host. OrthoRep has only been demonstrated in yeast;

MutaT7 systems have been established in *E. coli*^{16,23,24}, yeast²⁵, plants²⁶ and mammalian cells²²; and EvolvR has been successfully tested in *E. coli*¹⁷ and yeast²⁸. Host choice is typically determined by the biomolecules being evolved — whether they function natively in the host or require host-specific post-translational modifications, for example — as well as the ease of setting up a reliable genetic or cell-based selection in the various hosts being considered. Other considerations include the generation times, population sizes and scale of experimentation possible with different hosts.

OrthoRep, MutaT7 and EvolvR systems are currently being developed and optimized for compatibility with a broader host range. For OrthoRep, it is unknown how difficult it will be to transfer the underlying orthogonal replication machinery into hosts beyond yeast. It may also be possible to establish OrthoRep in bacteria or mammalian cells by using the DNA replication systems of existing bacterial or mammalian viruses that may be (or engineered to be) orthogonal to host DNA replication¹²². For MutaT7 and EvolvR systems that already operate in bacteria, yeast and mammalian cells, areas of optimization include addressing host-specific differences in the mismatch repair systems responding to hypermutation, toxicity or burden of the mutagenesis machinery¹⁰⁹ (such as deaminase–T7RNAP fusion or nCas9–DNAP fusion) and minimizing cargo size, as in the case of EvolvR, for delivery and stable expression of mutagenesis machinery in mammalian cells.

Hypermutation characteristics

The hypermutation rate of in vivo continuous evolution systems determines how long it takes a cell to sample new GOI sequences at any given time during an evolution experiment. The hypermutation profile determines what types of mutation are sampled. Although it is possible to reach a hypermutation rate that will effectively render any GOI inactive in just one cycle of replication, current in vivo continuous evolution systems are far from this lethal mutagenesis rate. Thus, increasing the mutation rates and expanding the mutational spectrum of OrthoRep, MutaT7 and EvolvR are active areas of research. As it stands, one should typically prefer the highest mutation rate and broadest mutational profile when selecting systems. As these characteristics have not always been measured in the same way, it is not straightforward to directly compare them across different systems, but we make an attempt in Supplementary Table 1.

Another consideration for hypermutation is the level of off-target mutagenesis. Off-target mutagenesis increases the chance of genomic adaptation, mutations in the genetic selection system used to guide evolution, mutations that modulate the hypermutation system itself and mutations that are deleterious to cellular fitness. An advantage of OrthoRep is that there is no measurable mutation rate elevation in the host genome when the GOI is being continuously hypermutated¹⁸. This derives from the mechanistic and spatial separation of DNA replication between the orthogonal p1 plasmid and the genome. The error-prone polymerase of EvolvR and the deaminase of MutaT7 cause low but measurable

off-target mutagenesis, currently a few hundred-fold lower than on-target mutagenesis.

Finally, an important feature of in vivo continuous evolution is that evolution should be able to occur for extended periods of time during which continuous operation of mutation, amplification and selection cycles result in the exploration of long mutational paths over many generations. For this to occur, hypermutation must be durable. The durability of hypermutation in OrthoRep is high. Evolution experiments with OrthoRep have been carried out for hundreds of generations with continued evolution^{18,83}. Durability of mutagenesis for MutaT7 and EvolvR systems has not been tested thoroughly but is likely lower than for OrthoRep. This is because the elements recruiting hypermutation machinery, such as the T7 promoter or gRNA target site, can themselves become mutated while still allowing the GOI to be replicated and expressed by host machinery. This may allow the system to reduce its own hypermutation rate over time. The measurable off-target mutation rate of MutaT7 and EvolvR also elevates the chance of mutations in the mutagenesis machinery itself, potentially causing changes in the hypermutation rate over time, especially if there is toxicity associated with MutaT7 and EvolvR parts that create selective pressures for their functional degradation. Indeed, reducing the burden, toxicity and off-target mutagenesis for MutaT7 and EvolvR are areas of ongoing optimization.

Target size and context

The amount of genetic cargo that can be placed on OrthoRep is up to at least 20 kb (REF.⁴¹), although smaller cargo sizes up to ~7 kb are most tractable. MutaT7 systems may tolerate up to at least 25 kb based on the processivity of T7RNAP⁴⁵, although only smaller cargo sizes up to ~2 kb have been tested. The amount of DNA undergoing hypermutation for each targeting gRNA used in EvolvR is less than a few hundred base pairs depending on the EvolvR DNAP used, but one can employ a collection of gRNAs to target multiple loci to expand the effective size of the genetic cargo undergoing mutagenesis. Still, the size of what can be targeted for hypermutation imposes limits for the various in vivo continuous evolution systems.

Another consideration is the context of the target GOI under evolution, by which we mean where the target GOI is encoded. EvolvR has the unique benefit that any locus targetable with gRNAs can be the subject of hypermutation. Therefore, genomic loci in their native context can be continuously evolved, preserving native regulation of expression and reducing engineering requirements. MutaT7 can also target host genomic loci, but the loci must first be engineered to contain a T7 promoter. Whereas installing T7 promoters can be relatively trivial if target regions are on plasmids, doing so can be challenging or infeasible if the desired target regions are genomic, although continued innovation in the genome editing field is making genome engineering routine^{123–125}. OrthoRep is restricted to GOIs encoded on the orthogonal p1 plasmid and cannot target genomic loci for hypermutation. Additionally, the cytoplasmic localization of p1 may complicate the evolution of

RNAs that function in the nucleus. Some GOIs cannot be evolved with *in vivo* continuous evolution in general, namely those that are toxic to the host or a GOI whose function cannot be selected or screened for directly in or on cells.

Finally, targeting only a region of a GOI for evolution (one specific domain in a protein, for example) may be possible with MutaT7 by using dCas9 to terminate MutaT7's action in the middle of a GOI while still allowing the entire GOI to be expressed, with the caveat that this technique does not enable exclusion of both termini of a GOI from hypermutation²³. EvolvR can also achieve partial mutagenesis of a GOI by using gRNAs corresponding to a small region in a GOI along with nCas9–DNAP fusions where the error-prone DNAP has low processivity. OrthoRep cannot selectively target one part of a GOI for hypermutation because the entire GOI needs to be encoded on the orthogonal p1 plasmid for expression as a single protein product. However, it may be possible to split a GOI into domains that are post-translationally joined, for example by using a split intein¹²⁶, in which case one domain can be encoded on p1 for hypermutation and the other domain can be encoded on a host plasmid or in the host genome where it is not hypermutated.

Ease of implementation

An advantage of MutaT7 and EvolvR is their reliance on common synthetic biology parts (T7RNAP, associated promoters and CRISPR/Cas9 components) and procedures (conventional genetics and cloning methods). By contrast, OrthoRep experiments require custom promoters for GOIs and custom genetics for integrating genes onto the cytoplasmic orthogonal p1 plasmid, requiring more specialized knowledge. The ease of setting up MutaT7 and EvolvR to hypermutate the GOI should be balanced against the architectural advantages of OrthoRep in supporting continuous GOI evolution experiments durably over extended periods of time.

Outlook

Directed evolution of GOIs has traditionally used iterative cycles of *in vitro* diversification (such as error-prone PCR) followed by transformation of cells with the resulting GOI mutant libraries for expression and screening or selection. Continuous hypermutation systems bring GOI diversification *in vivo*, allowing GOIs to evolve autonomously as cells propagate under selection. This dramatically transforms the depth and scale of GOI evolution, accessing new avenues for biomolecular engineering and evolution. Although by no means complete, the Applications section of this Primer features recent work that exploits the key features of depth and scale in GOI evolution available to *in vivo* continuous evolution. This includes the traversal of long multi-mutation pathways in the optimization of enzyme function¹⁸ and the replicate evolution of enzymes and antibodies to augment the scale at which we gain new GOI functions and sample diverse regions of sequence space^{113,121}. The continued evolution of proteins with new functions at depth and scale will naturally blossom, defining one part of the future of *in vivo* continuous evolution.

We provide three less obvious but equally tantalizing future directions here.

Expansion into multicellular organisms

With the unit of selection having gone from the replicating RNA molecule (as in the very first continuous RNA evolution experiments^{127,128}) to a virus (as in the case of PACE⁵) and now the cell, the scope of functions that we can evolve a GOI to accomplish has dramatically broadened. In essence, the field of continuous evolution has followed an arc where GOI hypermutation has been made possible in increasingly complex units of selection, accessing broader spectrums of function that a GOI can be pressured to evolve. The logical next step in this arc is to bring continuous GOI evolution to multicellular organisms. If *in vivo* continuous evolution systems can be installed within the cells of a complex animal, we can evolve biomolecules that change the physiology of animals. Short of using an animal as the unit of selection, we can at least carry out cell-based selections in the context of animals where the biomolecular function serves a therapeutic goal occurring in the relevant environment. An example of this would be to continuously evolve receptors encoded in therapeutic T cells within mouse models of cancer. Naturally, ethical concerns must be carefully assessed before initiating any experiments involving evolution with or in multicellular animals, ranging from less ethically challenging organisms such as flies and worms to more ethically fraught mammalian models.

Deep learning and continuous evolution

Concurrent with the development of *in vivo* continuous evolution has been a revolution in the power of artificial intelligence, especially deep learning, to navigate the nearly infinite combinatorial space underlying biomolecular engineering^{87,88,129}. Deep learning is only successful when data from which to learn are abundant. By running deep evolution experiments at a scale of thousands of replicates, as is possible with *in vivo* continuous evolution, we may be able to generate big biomolecular evolution data sets in a systematic manner where the entire evolutionary record is also available. This would allow artificial intelligence to produce probabilistic sequence-to-function models that can predict and generate new sequences with desired functions and functional improvements. In comparison with techniques such as low-throughput classical directed evolution or deep mutational scanning approaches that systematically evaluate the consequences of only one or two-mutation variants of a parent sequence, continuous evolution experiments would sample the contours of fitness landscapes through long mutational trajectories at an unprecedented scale. Such data sets can train generative deep learning models whose outcomes can even be reloaded into continuous evolution systems for further evolution and divergence, creating a virtuous cycle. As natural data sets are incomplete and simulation of RNA and protein function is often inaccurate, we predict that continuous evolution experiments may become the other side of the deep learning coin in the realm of biomolecular engineering.

Going from zero to one and from one to many

The emergence of desired activity where none existed before is a major challenge in the biomolecular engineering field. Strategies that mine diverse gene collections for desired biomolecular functions to bootstrap directed evolution campaigns have acted as an effective solution, but, ultimately, the goal is to gain functional sequences from scratch. We call this the zero-to-one goal: going from zero sequences that have any desired activity to one. An approach to the zero-to-one goal that has proven successful in the RNA enzyme and aptamer evolution fields is to start from staggeringly large random sequence libraries (at 10^{13} variants)^{128,130}. For proteins, such library sizes are traditionally inadmissible because the transformation efficiency of cells can only reach 10^7 – 10^9 . With in vivo continuous evolution, diversity is generated directly inside cells, making it possible to bypass transformation efficiency limitations. With sufficiently high and durable mutation rates on a GOI and the large population sizes accessible in a bioreactor, protein libraries that are 10^{13} in size could conceivably be generated. When such diversity is reached, selection can be imposed to initiate further evolution of low activity sequences in a continuous format.

Another approach to the zero-to-one goal is computational design. The de novo design of desired RNA and protein structures, and to some extent desired

functions, has witnessed major advances over the past 20 years^{87,129,131,132}. However, once a de novo design is generated, its activity almost certainly requires improvement, which directed evolution campaigns can address. Perhaps more important from a computational design perspective is the value of diverging a de novo design into a much larger set of highly dissimilar variants to be able to map the fitness landscape governing the de novo design. The depth and scale of continuous evolution campaigns may be uniquely capable of achieving this. Indeed, this one-to-many goal — going from one sequence with a desired activity to many — is within the unique purview of in vivo continuous evolution and applies to all cases where we find only one example of a sequence with a desired activity. Besides de novo designed biomolecules and sequences isolated from large random libraries, orphan proteins or ribozymes that could represent a lost epoch of life may be turned into rich families of variants through the power of in vivo continuous evolution. Such efforts may give us a comparative understanding of why well-populated natural RNA and protein families have been so successful, and also offer us entry into knowledge-based strategies for engineering de novo^{133–136}, orphan¹³⁷ and ancient biomolecules^{138–140}. Indeed, in vivo continuous evolution presents many exciting opportunities ahead.

Published online: 19 May 2022

- Arnold, F. H. Design by directed evolution. *Acc. Chem. Res.* **31**, 125–131 (1998).
- Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).
- Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* **148**, 1667–1686 (1998).
- Chen, K. & Arnold, F. H. Engineering new catalytic activities in enzymes. *Nat. Catal.* **3**, 203–213 (2020).
- Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
- Miller, S. M., Wang, T. & Liu, D. R. Phage-assisted continuous and non-continuous evolution. *Nat. Protoc.* **15**, 4101–4127 (2020).
- Morrison, M. S., Podracky, C. J. & Liu, D. R. The developing toolkit of continuous directed evolution. *Nat. Chem. Biol.* **16**, 610–619 (2020).
- Hendel, S. J. & Shoulders, M. D. Directed evolution in mammalian cells. *Nat. Methods* **18**, 346–357 (2021).
- Fabret, C. et al. Efficient gene targeted random mutagenesis in genetically stable *Escherichia coli* strains. *Nucleic Acids Res.* **28**, 95 (2000).
- Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proc. Natl Acad. Sci. USA* **100**, 9727–9732 (2003).
- Finney-Manchester, S. P. & Maheshri, N. Harnessing mutagenic homologous recombination for targeted mutagenesis in vivo by TaGTEAM. *Nucleic Acids Res.* **41**, 1–10 (2013).
- Crook, N. et al. In vivo continuous evolution of genes and pathways in yeast. *Nat. Commun.* **7**, 13051 (2016).
- Ravikumar, A., Arrieta, A. & Liu, C. C. An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* **10**, 175–177 (2014).
This work establishes an orthogonal DNA replication system in yeast that enables the elevation of mutation rates on an orthogonal plasmid replicated by a cognate orthogonal error-prone DNAP.
- Hess, G. T. et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).
This work is an early demonstration that attachment of mutagenic machinery, in this case a cytidine deaminase, to dCas9 is an effective strategy for targeting hypermutation to desired loci in mammalian cells.
- Ma, Y. et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).
- Moore, C. L., Papa, L. J. & Shoulders, M. D. A processive protein chimera introduces mutations across defined DNA regions in vivo. *J. Am. Chem. Soc.* **140**, 11560–11564 (2018).
This work establishes the MutaT7 strategy involving the fusion of a DNA-damaging cytidine deaminase to a processive RNA polymerase to achieve in vivo targeted hypermutation of multi-kilobyte DNA sequences, thereby enabling continuous evolution of GOIs inside *E. coli*.
- Halperin, S. O. et al. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560**, 248–252 (2018).
This work demonstrates that fusion of an error-prone DNAP to nCas9 achieves hypermutation at desired gRNA-targeted loci in *E. coli* to support continuous in vivo diversification and evolution of GOIs.
- Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A. & Liu, C. C. Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell* **175**, 1946–1957.e13 (2018).
This work establishes a highly error-prone orthogonal DNA replication system that durably hypermutates an orthogonal plasmid at mutation rates 100,000-fold higher than the genome in yeast, thereby supporting the continuous evolution of GOIs for extended periods of time and at scale, as demonstrated through the replicate evolution of drug resistance by a malarial drug target.
- Yi, X., Khey, J., Kazlauskas, R. J. & Travisano, M. Plasmid hypermutation using a targeted artificial DNA replisome. *Sci. Adv.* **7**, eabg871 (2021).
- Yi, X., Kazlauskas, R. & Travisano, M. Evolutionary innovation using EDGE, a system for localized elevated mutagenesis. *PLoS ONE* **15**, 1–18 (2020).
- Jensen, E. D. et al. A synthetic RNA-mediated evolution system in yeast. *Nucleic Acids Res.* **49**, 1–12 (2021).
- Chen, H. et al. Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. *Nat. Biotechnol.* **38**, 165–168 (2020).
This work presents an extension of the MutaT7 strategy to mammalian cells, enabling the continuous targeted hypermutation and evolution of GOIs inside human cells.
- Álvarez, B., Mencia, M., de Lorenzo, V. & Fernández, L. Á. In vivo diversification of target genomic sites using processive base deaminase fusions blocked by dCas9. *Nat. Commun.* **11**, 6436 (2020).
This work expands the MutaT7 technology through the fusion of new base deaminases to T7RNAP (thereby achieving targeted hypermutation with expanded mutational parameters in *E. coli*) and the addition of dCas9 to terminate polymerization by T7RNAP (thereby providing more control over the window of hypermutation).
- Park, H. & Kim, S. Gene-specific mutagenesis enables rapid continuous evolution of enzymes in vivo. *Nucleic Acids Res.* **49**, e32–e32 (2021).
This work presents an expansion of MutaT7 technology to achieve exceptionally high rates of hypermutation in *E. coli*.
- Cravens, A., Jamil, O. K., Kong, D., Sockolovsky, J. T. & Smolke, C. D. Polymerase-guided base editing enables in vivo mutagenesis and rapid protein engineering. *Nat. Commun.* **12**, 1579 (2021).
This work extends the MutaT7 strategy to yeast, enabling the continuous targeted hypermutation and in vivo continuous evolution of GOIs in *S. cerevisiae*.
- Butt, H., Ramirez, J. L. M. & Mahfouz, M. Synthetic evolution of herbicide resistance using a T7 RNAP-based random DNA base editor. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.11.30.470689> (2021).
- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Tou, C. J., Schaffer, D. V. & Dueber, J. E. Targeted diversification in the *S. cerevisiae* genome with CRISPR-guided DNA polymerase I. *ACS Synth. Biol.* **9**, 1911–1916 (2020).
- Khanal, A., McLoughlin, S. Y., Kershner, J. P. & Copley, S. D. Differential effects of a mutation on the normal and promiscuous activities of orthologs: implications for natural and directed evolution. *Mol. Biol. Evol.* **32**, 100–108 (2015).
- Zheng, J., Payne, J. L. & Wagner, A. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science* **365**, 347–353 (2019).

31. Gupta, R. D. & Tawfik, D. S. Directed enzyme evolution via small and effective neutral drift libraries. *Nat. Methods*. **5**, 939–942 (2008).
32. Salverda, M. L. M. et al. Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.* **7**, e1001321 (2011).
33. Baier, F. et al. Cryptic genetic variation shapes the adaptive evolutionary potential of enzymes. *eLife* **8**, 1–20 (2019).
34. Lang, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*. **500**, 571–574 (2013).
35. Eigen, M., McCaskill, J. & Schuster, P. Molecular quasi-species. *J. Phys. Chem.* **92**, 6881–6891 (1988).
36. Rix, G. & Liu, C. C. Systems for in vivo hypermutation: a quest for scale and depth in directed evolution. *Curr. Opin. Chem. Biol.* **64**, 20–26 (2021).
- This work outlines the value of in vivo continuous evolution systems in accessing new categories of directed evolution experiments characterized by depth and scale.**
37. Gunge, N. & Sakaguchi, K. Intergeneric transfer of deoxyribonucleic acid killer plasmids, pGKI1 and pGKI2, from *Kluyveromyces lactis* into *Saccharomyces cerevisiae* by cell fusion. *J. Bacteriol.* **147**, 155–160 (1981).
38. Arzumanyan, G. A., Gabriel, K. N., Ravikumar, A., Javanpour, A. A. & Liu, C. C. Mutually orthogonal DNA replication systems in vivo. *ACS Synth. Biol.* **7**, 1722–1729 (2018).
39. Zhong, Z., Ravikumar, A. & Liu, C. C. Tunable expression systems for orthogonal DNA replication. *ACS Synth. Biol.* **7**, 2930–2934 (2018).
40. Kämper, J., Esser, K., Gunge, N. & Meinhardt, F. Heterologous gene expression on the linear DNA killer plasmid from *Kluyveromyces lactis*. *Curr. Genet.* **19**, 109–118 (1991).
41. Javanpour, A. A. & Liu, C. C. Genetic compatibility and extensibility of orthogonal replication. *ACS Synth. Biol.* **8**, 1249–1256 (2019).
42. Chamberlin, M., McGrath, J. & Waskell, L. New RNA polymerase from *Escherichia coli* infected with bacteriophage T7. *Nature*. **228**, 227–231 (1970).
43. Tabor, S. & Richardson, C. C. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl Acad. Sci. USA* **82**, 1074–1078 (1985).
44. McAllister, W. T., Morris, C., Rosenberg, A. H. & Studier, F. W. Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection. *J. Mol. Biol.* **153**, 527–544 (1981).
45. Thiel, V., Herold, J., Schelle, B. & Siddell, S. G. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J. Gen. Virol.* **82**, 1273–1281 (2001).
46. Steitz, T. A. The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Curr. Opin. Struct. Biol.* **19**, 685–690 (2009).
47. Conticello, S. G. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **9**, 229 (2008).
48. Gerber, A. P. & Keller, W. An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **286**, 1146–1149 (1999).
49. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell.* **10**, 1247–1253 (2002).
50. Navaratnam, N. & Sarwar, R. An overview of cytidine deaminases. *Int. J. Hematol.* **83**, 195–200 (2006).
51. Cacciamani, T. et al. Purification of human cytidine deaminase: molecular and enzymatic characterization and inhibition by synthetic pyrimidine analogs. *Arch. Biochem. Biophys.* **290**, 285–292 (1991).
52. Chung, S. J., Fromme, J. C. & Verdine, G. L. Structure of human cytidine deaminase bound to a potent inhibitor. *J. Med. Chem.* **48**, 658–660 (2005).
53. Lada, A. G. et al. Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. *Biochem* **76**, 131–146 (2011).
54. Vik, E. S. et al. Endonuclease V cleaves at inosines in RNA. *Nat. Commun.* **4**, 2271 (2013).
55. Krokan, H. E., Drablos, F. & Slupphaug, G. Uracil in DNA — occurrence, consequences and repair. *Oncogene* **21**, 8935–8948 (2002).
56. Alseth, I., Dalhus, B. & Bjørås, M. Inosine in DNA and RNA. *Curr. Opin. Genet. Dev.* **26**, 116–123 (2014).
57. Hirano, K. I., Min, J., Funahashi, T. & Davidson, N. O. Cloning and characterization of the rat *apobec-1* gene: a comparative analysis of gene structure and promoter usage in rat and mouse. *J. Lipid Res.* **38**, 1103–1119 (1997).
58. MacGinnitie, A. J., Anant, S. & Davidson, N. O. Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity. *J. Biol. Chem.* **270**, 14768–14775 (1995).
59. Scott, J., Navaratnam, N., Bhattacharya, S. & Morrison, J. R. The apolipoprotein B messenger RNA editing enzyme. *Curr. Opin. Lipidol.* **5**, 87–93 (1994).
60. Arakawa, H., Hauschild, J. & Buerstedde, J. M. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* **295**, 1301–1306 (2002).
61. Muramatsu, M. et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*. **102**, 553–563 (2000).
62. Rogozin, I. B. et al. Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID–APOBEC family cytosine deaminase. *Nat. Immunol.* **8**, 647–656 (2007).
63. Kim, J. et al. Structural and kinetic characterization of *Escherichia coli* TadA, the wobble-specific tRNA deaminase. *Biochemistry*. **45**, 6407–6416 (2006).
64. Gaudelli, N. M. et al. Programmable base editing of T to G in genomic DNA without DNA cleavage. *Nature*. **551**, 464–471 (2017).
65. Richter, M. F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
66. Martínez-Salas, E. Internal ribosome entry site biology and its use in expression vectors. *Curr. Opin. Biotechnol.* **10**, 458–464 (1999).
67. Wang, Z. & Mosbaugh, D. W. Uracil-DNA glycosylase inhibitor of bacteriophage PBS2: cloning and effects of expression of the inhibitor gene in *Escherichia coli*. *J. Bacteriol.* **170**, 1082–1091 (1988).
68. Wang, Z. & Mosbaugh, D. W. Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. *J. Biol. Chem.* **264**, 1163–1171 (1989).
69. Karran, P., Cone, R. & Friedberg, E. C. Specificity of the bacteriophage PBS2 induced inhibitor of uracil-DNA glycosylase. *Biochemistry*. **20**, 6092–6096 (1981).
70. Bennett, S. E. & Mosbaugh, D. W. Characterization of the *Escherichia coli* uracil-DNA glycosylase-inhibitor protein complex. *J. Biol. Chem.* **267**, 22512–22521 (1992).
71. Bennett, S. E., Schimerlik, M. I. & Mosbaugh, D. W. Kinetics of the uracil-DNA glycosylase/inhibitor protein association. Ung interaction with Ugi, nucleic acids, and uracil compounds. *J. Biol. Chem.* **268**, 26879–26885 (1993).
72. Wang, Y. et al. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.* **32**, 1197–1207 (2004).
73. Tizei, P. A. G., Csibra, E., Torres, L. & Pinheiro, V. B. Selection platforms for directed evolution in synthetic biology. *Biochem. Soc. Trans.* **44**, 1165–1175 (2016).
74. Wang, Y. et al. Directed evolution: methodologies and applications. *Chem. Rev.* **121**, 12384–12444 (2021).
75. 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. USA* **110**, 9007–9012 (2013).
76. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science*. **292**, 498–500 (2001).
77. 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).
78. Blum, T. R. et al. Phage-assisted evolution of botulinum neurotoxin proteases with reprogrammed specificity. *Science*. **371**, 803–810 (2021).
79. Szendro, I. G., Franke, J., De Visser, J. A. G. M. & Krug, J. Predictability of evolution depends nonmonotonically on population size. *Proc. Natl Acad. Sci. USA* **110**, 571–576 (2013).
80. Salverda, M. L. M., Koomen, J., Koopmanschap, B., Zwart, M. P. & de Visser, J. A. G. M. Adaptive benefits from small mutation supplies in an antibiotic resistance enzyme. *Proc. Natl Acad. Sci. USA* **114**, 12773–12778 (2017).
81. Steinberg, B. & Ostermeier, M. Environmental changes bridge evolutionary valleys. *Sci. Adv.* **2**, e1500921 (2016).
82. Wong, B. G., Mancuso, C. P., Kiriakov, S., Bashor, C. J. & Khalil, A. S. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. *Nat. Biotechnol.* **36**, 614–623 (2018).
83. Zhong, Z. et al. Automated continuous evolution of proteins in vivo. *ACS Synth. Biol.* **9**, 1270–1276 (2020).
84. DeBenedictis, E. A. et al. Systematic molecular evolution enables robust biomolecule discovery. *Nat. Methods*. **19**, 55–64 (2022).
85. Marks, D. S. et al. Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE* **6**, e28766 (2011).
86. Stiffler, M. A. et al. Protein structure from experimental evolution. *Cell Syst.* **10**, 15–24.e5 (2020).
87. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
88. Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods*. **16**, 687–694 (2019).
89. Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J. & Arnold, F. H. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl Acad. Sci. USA* **116**, 8852–8858 (2019).
90. Carr, I. M. et al. Inferring relative proportions of DNA variants from sequencing electropherograms. *Bioinformatics* **25**, 3244–3250 (2009).
91. Shen, M. W., Zhao, K. T. & Liu, D. R. Reconstruction of evolving gene variants and fitness from short sequencing reads. *Nat. Chem. Biol.* **17**, 1188–1198 (2021).
92. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
93. Amarasinghe, S. L. et al. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 1–16 (2020).
94. Ravi, R. K., Walton, K. & Khosroheidari, M. in *Disease Gene Identification: Methods and Protocols* (ed. DiStefano, J. K.) 223–232 (Springer, 2018).
95. Stoler, N. & Nekrutenko, A. Sequencing error profiles of Illumina sequencing instruments. *NAR. Genomics Bioinform.* **3**, 1–9 (2021).
96. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, 1348–1365 (2021).
97. Karst, S. M. et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat. Methods*. **18**, 165–169 (2021).
98. Wenger, A. M. et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **37**, 1155–1162 (2019).
99. Zurek, P. J., Knyphausen, P., Neufeld, K., Pushpanath, A. & Hoffelder, F. UMI-linked consensus sequencing enables phylogenetic analysis of directed evolution. *Nat. Commun.* **11**, 6023 (2020).
100. Wilson, B. D., Eisenstein, M. & Soh, H. T. High-fidelity nanopore sequencing of ultra-short DNA targets. *Anal. Chem.* **91**, 6783–6789 (2019).
101. Murray, K. D. & Borevitz, J. O. Axe: rapid, competitive sequence read demultiplexing using a trie. *Bioinformatics* **34**, 3924–3925 (2018).
102. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
103. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
104. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods. Mol. Biol.* **1151**, 165–188 (2014).
105. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods*. **11**, 801–807 (2014).
106. Sirawaraporn, W., Sathitkul, T., Sirawaraporn, R., Yuthavong, Y. & Santi, D. V. Antifolate-resistant mutants of *Plasmodium falciparum* dihydrofolate reductase. *Biochemistry* **94**, 1124–1129 (1997).
107. Hankins, E. G., Warhurst, D. C. & Sibley, C. H. Novel alleles of the *Plasmodium falciparum* dhfr highly resistant to pyrimethamine and chlorocycloguanil, but not WR99210. *Mol. Biochem. Parasitol.* **117**, 91–102 (2001).
108. Long, M. et al. Directed evolution of ornithine cyclodeaminase using an EvolvR-based growth-coupling strategy for efficient biosynthesis of l-proline. *ACS Synth. Biol.* **9**, 1855–1863 (2020).

109. García-García, J. D. et al. Potential for applying continuous directed evolution to plant enzymes: an exploratory study. *Life* **10**, 1–16 (2020).
110. García-García, J. D. et al. Using continuous directed evolution to improve enzymes for plant applications. *Plant. Physiol.* **188**, 971–983 (2022).
111. Chatterjee, A. et al. *Saccharomyces cerevisiae* TH14p is a suicide thiamine thiazole synthase. *Nature* **478**, 542–546 (2011).
112. Joshi, J. et al. Structure and function of aerotolerant, multiple-turnover TH14 thiazole synthases. *Biochem. J.* **478**, 3265–3279 (2021).
113. Rix, G. et al. Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nat. Commun.* **11**, 5644 (2020). **This work demonstrates the use of OrthoRep to evolve in a scalable manner a large collection of highly diverse orthologues of an enzyme (TrpB), which was mined for promiscuous activities leading to the biosynthesis of valuable chemicals.**
114. Dunn, M. F. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. *Arch. Biochem. Biophys.* **519**, 154–166 (2012).
115. Buller, A. R. et al. Directed evolution of the tryptophan synthase β -subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl Acad. Sci. USA* **112**, 14599–14604 (2015).
116. Watkins-Dulaney, E., Straathof, S. & Arnold, F. Tryptophan synthase: biocatalyst extraordinaire. *ChemBioChem* **22**, 5–16 (2021).
117. Romney, D. K., Murciano-Calles, J., Wehrmüller, J. E. & Arnold, F. H. Unlocking reactivity of TrpB: a general biocatalytic platform for synthesis of tryptophan analogues. *J. Am. Chem. Soc.* **139**, 10769–10776 (2017).
118. Boville, C. E., Romney, D. K., Almhjell, P. J., Sieben, M. & Arnold, F. H. Improved synthesis of 4-cyanotryptophan and other tryptophan analogues in aqueous solvent using variants of TrpB from *Thermotoga maritima*. *J. Org. Chem.* **85**, 7447–7452 (2018).
119. Javanpour, A. A. & Liu, C. C. Evolving small-molecule biosensors with improved performance and reprogrammed ligand preference using OrthoRep. *ACS Synth. Biol.* **10**, 2705–2714 (2021).
120. Jensen, E. D. et al. Integrating continuous hypermutation with high-throughput screening for optimization of *cis,cis*-muonic acid production in yeast. *Microb. Biotechnol.* **14**, 2617–2626 (2021).
121. Wellner, A. et al. Rapid generation of potent antibodies by autonomous hypermutation in yeast. *Nat. Chem. Biol.* **17**, 1057–1064 (2021). **This work demonstrates the use of OrthoRep to drive the rapid evolution of custom antibodies displayed on the surface of yeast cells, including nanomolar-affinity nanobodies that bind and neutralize SARS-CoV-2.**
122. Pezo, V. et al. Noncanonical DNA polymerization by aminoadenine-based siphoviruses. *Science* **372**, 520–524 (2021).
123. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
124. Ling, X. et al. Improving the efficiency of precise genome editing with site-specific Cas9–oligonucleotide conjugates. *Sci. Adv.* **6**, 1–9 (2020).
125. Wang, C. et al. Microbial single-strand annealing proteins enable CRISPR gene-editing tools with improved knock-in efficiencies and reduced off-target effects. *Nucleic Acids Res.* **49**, 1–16 (2021).
126. Stevens, A. J. et al. Design of a split intein with exceptional protein splicing activity. *J. Am. Chem. Soc.* **138**, 2162–2165 (2016).
127. Mills, D. R., Peterson, R. L. & Spiegelman, S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl Acad. Sci. USA* **58**, 217–224 (1967).
128. Beaudry, A. A. & Joyce, G. F. Directed evolution of an RNA enzyme. *Science* **257**, 635–641 (1992).
129. Leman, J. K. et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods.* **17**, 665–680 (2020).
130. Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX — a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng.* **24**, 381–403 (2007).
131. Torrisi, M., Pollastri, G. & Le, Q. Deep learning methods in protein structure prediction. *Comput. Struct. Biotechnol. J.* **18**, 1301–1310 (2020).
132. Rollins, N. J. et al. Inferring protein 3D structure from deep mutation scans. *Nat. Genet.* **51**, 1170–1176 (2019).
133. Langan, R. A. et al. De novo design of bioactive protein switches. *Nature* **572**, 205–210 (2019).
134. Dou, J. et al. De novo design of a fluorescence-activating β -barrel. *Nature* **561**, 485–491 (2018).
135. Basanta, B. et al. An enumerative algorithm for de novo design of proteins with diverse pocket structures. *Proc. Natl Acad. Sci. USA* **117**, 22135–22145 (2020).
136. Anishchenko, I. et al. De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).
137. Hadadi, N., MohammadiPeyhani, H., Miskovic, L., Seijo, M. & Hatzimanikatis, V. Enzyme annotation for orphan and novel reactions using knowledge of substrate reactive sites. *Proc. Natl Acad. Sci. USA* **116**, 7298–7307 (2019).
138. Gumulya, Y. et al. Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* **1**, 878–888 (2018).
139. Xie, V. C., Pu, J., Metzger, B. P. H., Thornton, J. W. & Dickinson, B. C. Contingency and chance erase necessity in the experimental evolution of ancestral proteins. *eLife* **10**, 1–87 (2021).
140. Kaltenbach, M. et al. Evolution of chalcone isomerase from a noncatalytic ancestor article. *Nat. Chem. Biol.* **14**, 548–555 (2018).
141. Biebricher, C. K. & Eigen, M. The error threshold. *Virus Res.* **107**, 117–127 (2005).
142. Pu, J., Zinkus-Boltz, J. & Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nat. Chem. Biol.* **13**, 432–438 (2017).
143. Packer, M. S., Rees, H. A. & Liu, D. R. Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat. Commun.* **8**, 956 (2017).
144. Badran, A. H. et al. Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* **535**, 58–63 (2016).
145. Inamoto, I., Sheoran, I., Popa, S. C., Hussain, M. & Shin, J. A. Combining rational design and continuous evolution on minimalist proteins that target the E-box DNA site. *ACS Chem. Biol.* **16**, 35–44 (2021).
146. Wang, T., Badran, A. H., Huang, T. P. & Liu, D. R. Continuous directed evolution of proteins with improved soluble expression. *Nat. Chem. Biol.* **14**, 972–980 (2018).
147. Thuronyi, B. W. et al. Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* **37**, 1070–1079 (2019).
148. Berman, C. M. et al. An adaptable platform for directed evolution in human cells. *J. Am. Chem. Soc.* **140**, 18093–18103 (2018).
149. English, J. G. et al. VEGAS as a platform for facile directed evolution in mammalian cells. *Cell* **178**, 748–761.e17 (2019).

Acknowledgements

The authors thank members of their groups for insightful discussions. This work was funded by National Institutes of Health (NIH) National Institute of General Medical Sciences (NIGMS) 1R35GM139513 (C.C.L.); NIH NIGMS 1R35GM136354 (M.D.S.); MIT Robert J Silbey Fellowship (A.A.M.); MIT School of Science Fund for Future of Science (A.A.M.); US Department of Energy, Office of Science, Basic Energy Sciences under Award DE-SC0020153 (A.D.H.); Innovative Genomics Institute and Laboratory for Genomics Research (J.E.H., J.E.D. and D.V.S.); UC Berkeley Miller Basic Research Fellowship (O.Z.); NIH National Institute of Biomedical Imaging and Bioengineering (NIBIB) 1R01EB027793 (A.S.K.); Department of Defense (DoD) Vannevar Bush Faculty Fellowship N00014-20-1-2825 (A.S.K.); NIH NIGMS 1R01GM125887 (F.H.A.); and Ministerio de Ciencia e Innovación - Consejo Superior de Investigaciones Científicas (MICIN-CSIC) PTI + REC-EU SGL2103051 and EU Horizon 2020 research and innovation programme FET Open 965018-BIOCELLPHE (L.A.F.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or other funding agencies.

Author contributions

Introduction (R.S.M., G.R., A.A.M., M.D.S. and C.C.L.); Experimentation (R.S.M., G.R., A.A.M., B.A., D.S., H.C., J.E.H., Q.Z., J.E.D., D.V.S., F.C., S.K., L.A.F., M.D.S. and C.C.L.); Results (R.S.M., G.R. and C.C.L.); Applications (R.S.M., G.R., A.A.M., B.A., D.S., H.C., J.E.H., Q.Z., J.D.G.-G., Z.J.H., P.J.A., F.H.A., A.S.K., A.D.H., J.E.D., D.V.S., F.C., S.K., L.A.F., M.D.S. and C.C.L.); Reproducibility and data deposition (R.S.M., G.R. and C.C.L.); Limitations and optimizations (R.S.M., G.R., A.A.M., B.A., D.S., H.C., J.E.H., Q.Z., J.E.D., D.V.S., F.C., S.K., L.A.F., M.D.S. and C.C.L.); Outlook (R.S.M., G.R. and C.C.L.). R.S.M. and G.R. contributed equally and are listed in alphabetical order.

Competing interests

C.C.L. is a co-founder of K2 Biotechnologies, which applies continuous evolution to antibody engineering. All other authors declare no competing interests.

Peer review information

Nature Reviews Methods Primers thanks Jumi Shin, Chong Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1038/s43586-022-00119-5>.

RELATED LINKS

European Nucleotide Archive (ENA): <https://www.ebi.ac.uk/ena/browser/home>
 National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA): <https://www.ncbi.nlm.nih.gov/sra>
 NCBI BioProject: <https://www.ncbi.nlm.nih.gov/bioproject/>

© Springer Nature Limited 2022