

Review

Not peer-reviewed version

Crafting Genetic Diversity: Unlocking the Potential of Protein Evolution

[Vamsi Krishna Gali](#) , [Kang Lan Tee](#) ^{*} , [Tuck Seng Wong](#) ^{*}

Posted Date: 19 December 2023

doi: 10.20944/preprints202312.1401.v1

Keywords: Protein engineering; directed evolution; genetic diversity; random mutagenesis; molecular cloning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

Crafting Genetic Diversity: Unlocking the Potential of Protein Evolution

Vamsi Krishna Gali ¹, Kang Lan Tee ^{1,2,*} and Tuck Seng Wong ^{1,2,3,4,*}

¹ Department of Chemical & Biological Engineering, University of Sheffield, Sir Robert Hadfield Building, Mappin Street, Sheffield S1 3JD, United Kingdom;

² Evolutor Ltd, The Innovation Centre, 217 Portobello, Sheffield S1 4DP, United Kingdom;

³ National Center for Genetic Engineering & Biotechnology (BIOTEC), National Science & Technology Development Agency (NSTDA), 113 Thailand Science Park, Phahonyothin Road, Khlong Nueng, Khlong Luang, Pathum Thani 12120, Thailand;

⁴ School of Pharmacy, Bandung Institute of Technology, Bandung, West Java, Indonesia.

* Correspondence: k.tee@sheffield.ac.uk (K.L.T.); Tel: +44 114 222 7507 (K.L.T.); Fax: +44 114 222 7501 (K.L.T.); t.wong@sheffield.ac.uk (T.S.W.); Tel: +44 114 222 7591 (T.S.W.); Fax: +44 114 222 7501 (T.S.W.)

Abstract: Genetic diversity is the foundation of evolutionary resilience, adaptive potential, and the flourishing vitality of living organisms, serving as the cornerstone for robust ecosystems and the continuous evolution of life on Earth. The landscape of directed evolution, a powerful biotechnological tool inspired by natural evolutionary processes, has undergone a transformative shift propelled by innovative strategies for generating genetic diversity. This shift is fuelled by several factors, encompassing the utilization of advanced toolkits like CRISPR-Cas and base editors, enhanced comprehension of biological mechanisms, cost-effective custom oligo pool synthesis, and the seamless integration of artificial intelligence and automation. This comprehensive review looks into the myriad methodologies employed for constructing gene libraries, both *in vitro* and *in vivo*, categorized into three major classes: random mutagenesis, focused mutagenesis, and DNA recombination. The objectives of this review are threefold: firstly, to present a panoramic overview of recent advances in genetic diversity creation, secondly, to inspire novel ideas for further innovation in the genetic diversity generation, and thirdly, to provide a valuable resource for individuals entering the field of directed evolution.

Keywords: protein engineering; directed evolution; genetic diversity; random mutagenesis; molecular cloning

1. Introduction

The year 2018 marked a significant milestone in the field of protein engineering, as it witnessed the well-deserved recognition of Prof. Frances H. Arnold with the Nobel Prize in Chemistry [1,2]. This prestigious accolade honoured her ground-breaking contributions to directed protein evolution, particularly through the utilization of a simple yet powerful algorithm. Directed evolution involves iterative cycles of genetic diversity creation, followed by screening and selection (Figure 1). The far-reaching applications and profound impacts of directed evolution are evident in various domains, including but not limited to, the development of highly active or stable enzymes, the creation of novel enzymatic functions and chemistries, and the engineering of biopharmaceutical proteins.

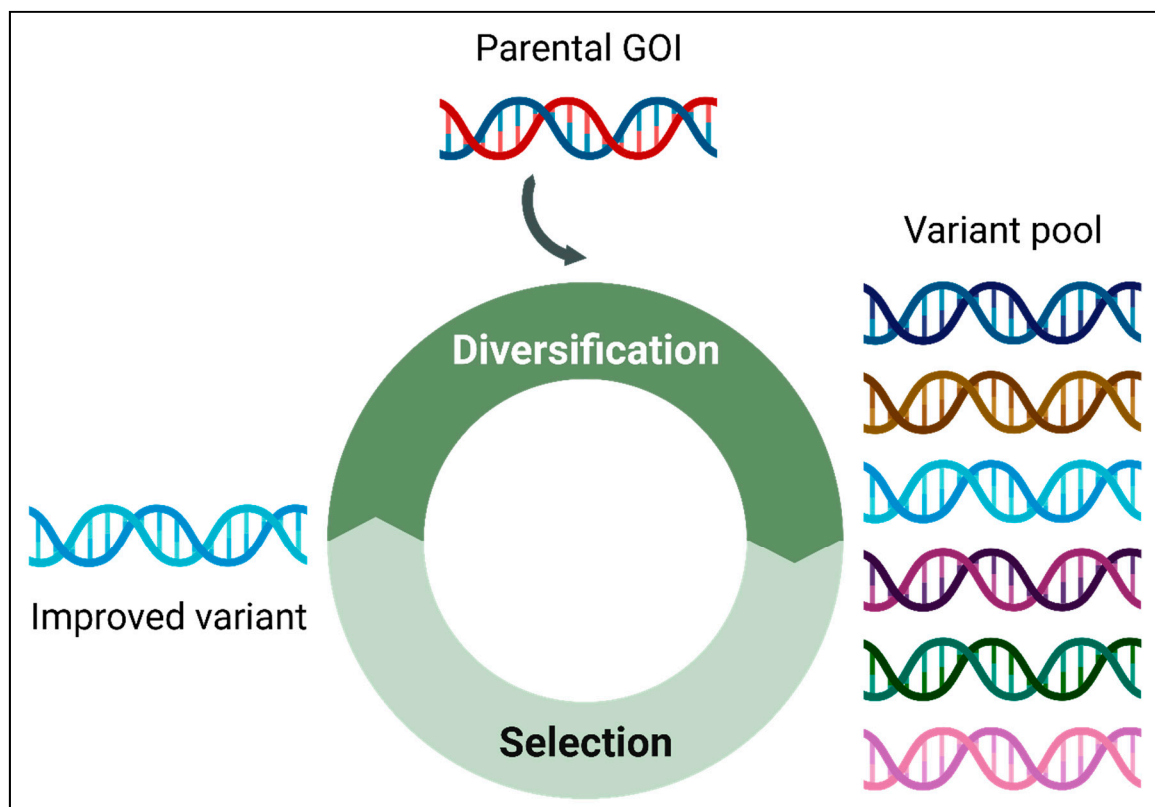


Figure 1. The directed evolution cycle. The parental gene of interest (GOI) undergoes mutagenesis to generate a diverse pool of genetic variants. This pool is then subjected to a selection process targeting the desired phenotype, enabling the identification of improved variant(s). This iterative cycle is repeated until the desired trait is successfully achieved.

At the core of protein evolution lies the essential prerequisite of genetic diversity. Indeed, without this diversity, the evolutionary process becomes untenable. In recognition of this pivotal aspect, this review article is dedicated to examining methodologies deliberately designed for the creation of genetic diversity, with a focus on publications from 2014 onward. This article serves as an extension to our prior reviews on the same subject in 2006 [3] and 2013 [4].

The methodologies for generating genetic diversity can be categorized into three main classes: random mutagenesis, focused mutagenesis, and DNA recombination (Figure 2). There exist hybrid methods that integrate features derived from two or more of these classes, showcasing the versatility and adaptability of current approaches. In random mutagenesis, mutations are randomly introduced into the target gene of interest (GOI). Conversely, focused mutagenesis involves the randomization or semi-randomization of selected amino acid residues within the target gene. In DNA recombination, a collection of homologous or non-homologous genes undergo random fragmentation and subsequent reassembly to form chimeric constructs. The fundamentals of directed evolution and genetic diversity creation have been comprehensively documented in our recently published protein engineering textbook [5]. Therefore, this review primarily explores the significant advancements achieved in the past decade.

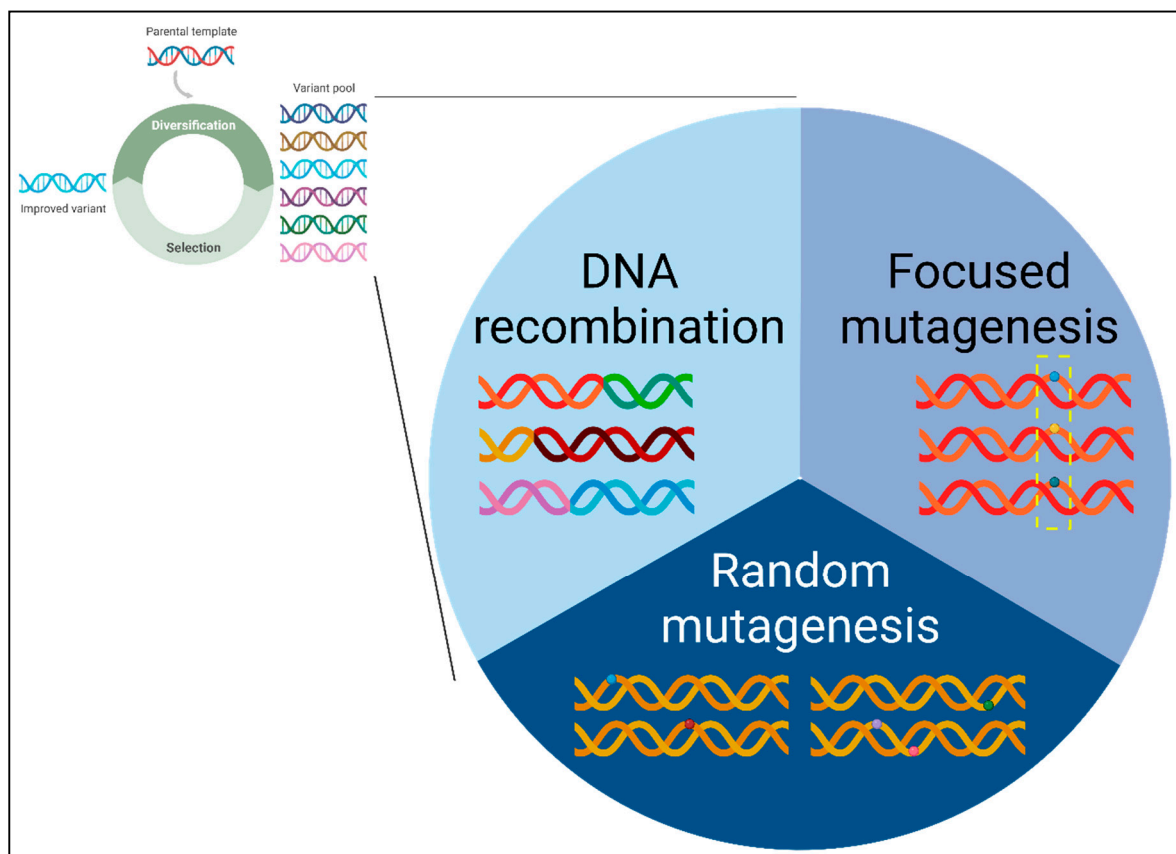


Figure 2. Classification of genetic diversity creation methods. The diverse methods for generating a genetically varied gene pool can be systematically categorized into three main classes: random mutagenesis, focused mutagenesis, and DNA recombination. Random mutagenesis involves the introduction of random mutations throughout the starting parental gene sequence. Focused mutagenesis targets mutations to specific pre-selected region(s) or amino acid residue(s) within the starting parental gene sequence. DNA recombination generates chimeric sequences by combining segments from a set of either homologous or non-homologous parental sequences.

This review commences by discussing the latest advancements in molecular cloning, a crucial initial step in the process of generating genetic diversity. Subsequent to this, we delve into the assessment of gene library quality. Building upon this foundation, we offer succinct summaries of innovative methodologies within each category of genetic diversity creation, emphasizing their individual merits and limitations.

2. Molecular cloning

DNA cloning serves as the cornerstone for mutagenesis experiments, involving the insertion of a target GOI or a gene library into a cloning/expression vector. Molecular cloning also plays an integral role in all synthetic biology projects at various stages. Conventionally, this process relies on the use of restriction endonucleases, which, while fundamental in molecular biology, present certain limitations. Drawbacks such as sequence specificity, time-consuming restriction digestion and ligation reactions, and the creation of undesirable 'scars' underscore the necessity for more versatile DNA cloning techniques. To address contemporary demands and challenges and facilitate rapid advancements, DNA assembly methods must be robust, cost-effective, swift, user-friendly, and highly scalable.

In this section, we present a comprehensive overview of the latest advancements in DNA cloning technology over the past decade. These innovations are categorized into *in vivo*, *in vitro*, modular DNA assembly, and automated DNA assembly methods.

2.1. *In vivo* cloning

In vivo cloning techniques (Table 1) leverage the cellular machinery of a host organism, such as bacteria or yeast, to incorporate a DNA fragment into a vector. Recent descriptions of these techniques often rely on DNA recombination or repair systems, such as homologous recombination (HR) and non-homologous end joining (NHEJ). The appeal of *in vivo* cloning techniques lies in their distinct advantages: they obviate the need for *in vitro* ligation reactions and are not constrained by restriction endonuclease sites. Introducing a short homologous sequence (30 – 40 bp) to any vector or DNA fragment through PCR facilitates recombinational cloning, underscoring the simplicity and versatility of this approach.

An underutilized yet progressively popular *in vivo* technique involves harnessing the RecA-independent recombination (RAIR) machinery [6] found in commonly employed laboratory bacterial strains, such as *Escherichia coli* DH5 α . RAIR is thought to be dependent on XthA, a 3'-to-5' exonuclease that resects the 3'-ends of linear DNA fragments introduced into *E. coli* cells, exposing the single-stranded 5'-overhangs [7]. Subsequently, the complementary single-stranded DNA ends hybridize with each other, and gaps are filled by DNA polymerase I and ligase LigA. Despite being originally described over three decades ago [8], the application of high-fidelity polymerases and low-cost oligonucleotides, coupled with insights into its molecular mechanisms [6,7,9–13], holds the promise of expanding its widespread adoption.

Traditionally, the budding yeast *Saccharomyces cerevisiae* has served as a platform for HR-based cloning [14–16], employing a shuttle vector to propagate plasmids in both yeast and *E. coli*. However, this approach limited the utility of yeast-based cloning exclusively to these two organisms. Recent breakthroughs in yeast recombination cloning [17–21] have transcended these limitations, allowing the technique to be applied universally for cloning multiple fragments of interest into any vector. Key innovation involves incorporating an origin of replication and a selection marker for yeast as a cassette or DNA fragment, which is co-inserted with other fragments for cloning. Assembled plasmids can then be selected in and purified from yeast, ready for transformation or transfection into desired host organisms. HR-based cloning in yeast can also be adapted for high-throughput DNA assembly and is amenable to automation [22].

While *S. cerevisiae* exhibits efficient HR machinery, its NHEJ activity is limited. By leveraging previous findings that the thermotolerant yeast *Kluyveromyces marxianus* possesses a highly efficient NHEJ pathway [23], an NHEJ-mediated cloning method was developed [24]. This method employs a functional marker selection system (*e.g.*, *ura3*) for the cloning of DNA fragments in *K. marxianus*. NHEJ-mediated cloning exploits the sequence-independent nature of NHEJ-based joining, allowing the cloning of DNA fragments without the need for homologous ends – a departure from the HR-based cloning approach.

CRasPy-Cloning harnesses the precision of Cas9 to cleave DNA at a user-specified locus, synergizing with the yeast's remarkably efficient homologous recombination [25]. This innovative approach enables the simultaneous cloning and engineering a bacterial genome in yeast. The successful application of CRasPy-Cloning was demonstrated through the cloning and engineering of the 0.816 Mbp genome of *Mycoplasma pneumoniae*, showcasing its potential for bacterial genome manipulation.

Most *in vivo* cloning methods utilize bacteria or yeasts as hosts, necessitating transformation and plasmid extraction procedures for the introduction of extracted plasmids into other organisms. A recent advancement in this domain is the Phage Enzyme-Assisted *In Vivo* DNA Assembly (PEDA) method [26]. Simultaneous expression of phage-derived T5 DNA exonuclease and T4 DNA ligase facilitates *in vivo* DNA assembly in a diverse range of microorganisms such as *Cupriavidus necator*, *Pseudomonas putida*, *Lactobacillus plantarum*, and *Yarrowia lipolytica*. Another cutting-edge technique, the Yeast Life Cycle (YLC) assembly method [27], capitalizes on CRISPR-Cas9 and yeast meiosis to iteratively assemble large DNA fragments. The advantage of the YLC method lies in bypassing challenging *in vitro* steps associated with handling and importing large DNA fragments into a host system.

Table 1. *In vivo* cloning methods developed in the past decade (HR, homologous recombination; NHEJ, non-homologous end joining).

<i>In vivo</i> cloning method	Method description	Reference(s)
Bacterial <i>in vivo</i> cloning (HR-based)	Depends on RecA-independent recombination (RAIR) pathway. DNA fragments generated via PCR or restriction enzyme digestion with overlapping homologous sequences can be used to directly transform bacteria.	[6,7,9–16]
Yeast <i>in vivo</i> cloning (HR-based)	Highly efficient HR pathway in yeast (<i>S. cerevisiae</i>) can be used for assembling multiple DNA fragments with homologous sequences directing the order of assembly.	[18–21]
Yeast <i>in vivo</i> cloning (NHEJ-based)	Highly efficient NHEJ pathway in thermotolerant yeast (<i>K. marxianus</i>) allows directly joining two DNA fragments without the need for homologous ends.	[24]
CR easPy-cloning	Combines ability of CRISPR-Cas9 and HR pathway of yeast to clone and edit large bacterial genomes at multiple loci.	[25]
Phage Enzyme-Assisted <i>In Vivo</i> DNA Assembly (PEDA) method	Simultaneous expression of exonuclease and ligase allows <i>in vivo</i> cloning in wide range of microorganisms.	[26]
Yeast Life Cycle (YLC) assembly method	Combines CRISPR-Cas9 and meiosis of yeast to iteratively assemble large DNA fragments.	[27]

2.2. *In vitro* cloning

Gibson assembly, Seamless Ligation Cloning Extract (SLiCE), Sequence and Ligation Independent Cloning (SLIC), and In-Fusion cloning are all *in vitro* techniques. Their underlying principle relies on the presence of homology regions at the ends of vector and insert DNA fragments. These methods have gained widespread popularity due to their ability to overcome limitations associated with restriction endonuclease-based cloning. Recent advancements have integrated the flexibility of Gibson assembly with the precision of CRISPR-Cas9, allowing for the targeting of double-strand breaks at any desired location. This is particularly advantageous in situations where PCR amplification of vector fragments is challenging, and restriction endonuclease sites are unavailable. Guide RNAs guide the Cas9 endonuclease to target any double-stranded DNA sequence, effectively linearizing the vector. The linearized vector can then be employed in Gibson assembly or other cloning techniques [28,29]. CRISPR-Cas9 has also been utilized to excise large DNA fragments (~100kb) from bacterial chromosomes, a process that may be difficult to achieve through PCR amplification. The cut fragments are subsequently cloned using Gibson assembly [30]. Variations of the SLiCE technique, such as Zero-Background Red α (ZeBR α), leverage the *ccdB* gene to eliminate any background from uncut or re-ligated vectors during cloning [31].

Other recent advances, such as T5 Exonuclease-Dependent Assembly (TEDA), Single 3'-exonuclease-based multifragment DNA assembly (SENAX), and T5 exonuclease-mediated low-temperature DNA cloning (TLTC) [32–34], operate on the principle of exonuclease-generated

overhangs within the homologous regions of vectors and DNA fragments. These methods offer a cost-effective alternative to Gibson assembly, as they only require the exonuclease enzyme for the *in vitro* reaction to create the necessary homologous overhangs. The subsequent steps of gap repair and DNA fragment ligation take place *in vivo* post-transformation into bacterial cells.

Alternatives to generating compatible overhangs are techniques such as Uracil-Specific Excision Reagent (USER) Cloning. In USER Cloning, vector and insert fragments with short overlapping homology regions are produced through PCR. The primers used in this process introduce a single deoxyuracil (dU) residue, subsequently cleaved by the USER enzyme. This cleavage results in 3' overhangs, facilitating the annealing of multiple DNA fragments. Although USER cloning has a longstanding history [35], advancements in the past decade, such as optimizing the melting temperature of annealing DNA fragments [36] and employing *in silico* design tools like AMUSER [37], have enhanced its efficacy. Similar techniques that introduce compatible overhangs through PCR include QuickStep-Cloning [38], where two parallel asymmetrical PCR reactions generate overhangs. An improvement on this method is PTO-QuickStep cloning [39,40], where phosphorothioate (PTO) bonds, introduced via primers, are processed by iodine cleavage to generate overhangs. Nicks in the plasmids resulting from both USER cloning and QuickStep-Cloning are sealed following transformation into bacterial cells. For scarless and sequence-independent DNA assembly, the method using thermostable exonuclease and ligase (DATEL) [41,42] presents another alternative. This technique employs a combination of Taq and Pfu DNA polymerases to cleave single-stranded DNA flaps generated during the annealing of DNA fragments with overlaps. The resulting nicks are then ligated using a heat-stable DNA ligase.

In summary, the fundamental principle underlying the latest advancements in *in vitro* DNA assembly (Table 2) is to streamline the process, ensuring simplicity, cost-effectiveness, and versatility. Leveraging overlapping DNA sequences among fragments to be assembled allows for the insertion of any DNA fragment into any chosen vector without being confined by the limitations of restriction endonuclease sites. While each discussed technique presents its unique advantages and limitations, the optimal choice for a particular DNA assembly project will depend on specific requirements and contextual factors.

Table 2. *In vitro* cloning methods developed in the past decade.

<i>In vitro</i> cloning method	Method description	Reference(s)
T5 Exonuclease-Dependent Assembly (TEDA)	Exonuclease generates homologous overhangs, gap repair and ligation completed <i>in vivo</i> .	[33]
Single 3'-exonuclease-based multifragment DNA assembly (SENAX)		[32]
T5 exonuclease-mediated low-temperature DNA cloning (TLTC)		[34]
Uracil-Specific Excision Reagent (USER)	Uracil-specific endonuclease generates homologous overhangs by digesting deoxyuracil introduced by primers.	[36,37]
PTO-QuickStep cloning	Phosphorothioate bonds introduced by primers are processed by iodine cleavage to generate homologous overhangs.	[39,40]

Scarless and sequence-independent DNA assembly method using thermostable exonuclease and ligase (DATEL)	Thermostable exonuclease generates homologous overhangs and thermostable ligase joins DNA fragments.	[41,42]
---	--	---------

2.3. Modular DNA assembly

The *in vivo* and *in vitro* cloning techniques outlined above present numerous advantages compared to traditional methods relying on restriction endonucleases. However, they often involve custom-designed primers and constructs tailored for specific experiments or purposes. A paradigm shift occurred with modular DNA assembly in synthetic biology. This approach revolves around using standardized DNA 'parts' to construct complex assemblies, which, in turn, can serve as 'parts' for even more intricate constructions. Once these components are generated, sequence-verified, and validated, they become reusable and shareable among researchers. This fosters standardization, ultimately leading to significant time and cost savings in the field.

Initially, modular DNA assembly relied on type IIP restriction enzymes, leading to the development of the BioBrick system for standardization. More recent techniques have shifted to the Golden Gate cloning technology, leveraging type IIS restriction enzymes. Unlike type IIP enzymes, which recognize and cut within palindromic DNA sequences, type IIS enzymes identify non-palindromic sequences and cleave DNA outside of the recognition site. This innovation has emerged as a potent tool for seamless cloning, allowing user-defined DNA sequences to serve as cutting sites and facilitating the precise assembly of DNA fragments in a predetermined manner.

Golden Gate assembly has been embraced in the creation of two extensively utilized hierarchical modular DNA cloning methods: MoClo and Golden Braid. These methodologies employ DNA 'parts' organized in different hierarchical levels, each level being more intricate than the one preceding it. Comprehensive reviews [43,44] provide invaluable resources for users seeking to initiate projects with these techniques. While these methods excel in seamlessly cloning multiple DNA fragments, their application has been constrained by limited vector choices – specifically, the requirement for destination vectors designed to contain one (or more) appropriately oriented pairs of recognition sites for a type IIS restriction enzyme. Recent advancements [45] have addressed this limitation by introducing type IIS restriction sites that are compatible with type IIP restriction sites, significantly broadening the range of vectors compatible with modular DNA assembly. A summary of other recently developed modular assembly techniques is provided in Table 3.

Table 3. Modular DNA assembly methods developed in the past 10 years.

Modular DNA assembly method	Method description	Reference(s)
Biopart Assembly Standard for Idempotent Cloning (BASIC)	Makes use of reusable linkers and parts. Orthogonal oligonucleotide linkers with single stranded overhangs are used to assemble DNA parts. Flexibility with the order of various DNA parts.	[46]
Modular Idempotent DNA Assembly System (MIDAS)	Requires three type IIS restriction enzymes and more complex than other modular DNA assemblies. Advantages include the ability to add new parts between existing parts rather than at the end.	[47]
MetClo Assembly	Controlling methylation (which cuts or blocks the recognition site) of a single type IIS recognition	[48]

	enzyme allows for a simpler hierarchical DNA assembly system.	
Start-Stop Assembly	3-bp overhangs corresponding to start and stop codons allow for scarless assembly of coding sequences.	[49,50]
PaperClip DNA Assembly	Unlike most other modular DNA assembly methods, does not require restriction enzymes. Four oligos per DNA part allow flexible ordering of DNA parts and reuse of oligos.	[51]

2.4. Automated DNA assembly

Amidst numerous breakthroughs, advancements, and refinements, a diverse array of *in vivo* and *in vitro* DNA cloning methods now caters to a spectrum of scenarios – from straightforward insertion of a single DNA fragment into a vector to intricate assemblies involving large and/or multiple DNA fragments. As experimental designs grow increasingly complex, the norm is now the generation of hundreds, if not thousands, of library sequences. Consequently, there arises a demand for high-throughput methods capable of assembling a large number of DNA sequences to generate libraries.

The integration of automation into these processes (Table 4) offers several advantages, including the elimination of manual handling errors, enhanced reliability and reproducibility, and significant time and cost savings. This shift toward high-throughput methods represents a pivotal stride in accommodating the evolving demands of sophisticated experimental designs with large-scale library creation. Liquid handling robots like Opentrons play a pivotal role in enabling this shift, offering benefits such as cost-effectiveness in hardware, open-source software accessibility, and user-friendly operation. These robotic systems prove invaluable for automating a diverse array of workflows, ranging from nucleic acid extraction to PCR and various DNA assembly techniques.

Table 4. Automated DNA assembly methods reported in the past decade.

Automated DNA assembly methods	Method description	Reference(s)
DNA assembly with BASIC on Opentrons (DNA-BOT)	Modular DNA assembly technique BASIC has been automated using robotic liquid handler Opentrons	[52]
Plasmid Maker	Combining the use of artificial restriction enzymes, custom software and robotic systems, an end-to-end system designed for automated DNA assembly.	[53]
AssemblyTron	Golden Gate and HR-dependent <i>in vivo</i> assemblies were automated using Opentrons	[54]

3. Genetic diversity creation

The primary objective of creating genetic diversity is to generate various protein variants, with the hope that a subset of these variants, albeit often a small fraction, will exhibit favourable phenotypes compared to their parent or wildtype protein. The introduction of mutations into the GOI is a common approach in mutagenesis to achieve this diversity. The outcome of protein variants is influenced by several factors. Firstly, considering that there are 61 codons encoding 20 amino acids, altering the codon usage or the GC content of the DNA sequence can impact the mutagenesis

outcome. This effect becomes particularly prominent when the mutational spectrum of the employed mutagenesis method is highly biased. Secondly, the organization of the genetic code imposes constraints on the mutagenesis outcome.

While protein engineers predominantly work with DNA, it's worth noting that RNA mutagenesis methods also exist. Fukuda *et al.* developed an RNA-mutagenesis method that leverages the intracellular RNA-editing mechanism [55]. In this approach, guide RNAs guide the editing enzyme, human adenosine deaminase acting on RNA (ADAR), inducing adenosine-to-inosine (A-to-I) mutations on RNA molecules. Other RNA-based mutagenesis strategies involve using Q β replicase to generate complex mRNA libraries [56]. However, compared to DNA, RNA is a transient molecule, making it more challenging to track genotype-phenotype linkage. The conversion of RNA to cDNA is typically required to identify mutations introduced onto RNA. In the subsequent sections, our focus will primarily be on DNA mutagenesis methods.

3.1. Quality of a gene library

It is strongly advisable to assess the quality of a gene library, especially when employing a mutagenesis method for the first time on the target gene sequence. This evaluation is commonly achieved by sequencing a small subset, typically ranging from 10 to 20 randomly selected 'variants'. If the gene library's quality is found to be suboptimal, proceeding to the expression of the protein library and subsequent selection/screening may not be prudent. The probability of identifying enhanced protein variants would be too minimal to justify the resources invested in such an undertaking.

Several key indicators are commonly employed to evaluate the quality of a gene library. Taking random DNA mutagenesis as an example, a high-quality gene library should exhibit the following characteristics:

- Mutations are precisely targeted to the GOI (i.e., no off-target mutations).
- Mutations are uniformly distributed along the entire GOI.
- All bases (A/T/G/C) experience mutations at the same frequency and are substituted with their three counterparts equally.
- The mutation frequency (number of errors per 1 kb of DNA) is not excessively high, preventing the predominance of non-functional protein variants.
- Duplicated sequences are avoided/eliminated.
- Wildtype sequences are absent in the gene library (i.e., no template carry-over).

Achieving an absolutely ideal gene library is practically elusive unless it exclusively comprises synthetic genes. The extent of deviation from this ideal serves as a practical measure of the gene library's quality. The Schwaneberg group conducted a comprehensive large-scale comparison among random mutagenesis libraries created using three approaches: error-prone polymerase chain reaction (epPCR) with low mutagenic conditions, epPCR with high mutagenic conditions, and Sequence Saturation Mutagenesis (SeSaM) [57–59]. After sequencing 1000 mutations for each library, the library quality was evaluated on both the DNA and protein levels [60,61]. The protein-level assessment primarily employed the Mutagenesis Assistant Program (MAP) [62–64]. The SeSaM library exhibits a preference for transversion mutations, contrasting with epPCR libraries that display a transition bias. Additionally, the SeSaM library demonstrates a significantly higher number of consecutive nucleotide substitutions. These characteristics result in a greater number and more diverse amino acid substitutions in the SeSaM library.

In directed evolution, preserving genotype-phenotype linkage is crucial for tracing an improved phenotype back to its genetic origins. During the library transformation process, transformants incorporating more than one plasmid could account for over 20% of the constructed library, thereby compromising the library quality. To mitigate the occurrence of multiple-plasmid transformants, it is possible to reduce their frequency by optimizing the amount of plasmid DNA used for transformation [65].

3.2. Random mutagenesis

Motivated by the technical simplicity of epPCR, several modified protocols have been developed. One adaptation is a modified epPCR protocol specifically optimized for small amplicons [66]. Another variation, known as Casting epPCR (cepPCR), directs random mutations to a specific region within the target GOI [66]. Furthermore, efforts are underway to extend epPCR methods beyond the conventional microbial hosts *E. coli* and *S. cerevisiae*, with attempts to optimize them for other microbial hosts. For example, in the transformation of *Bacillus subtilis* into a host for directed evolution, the epPCR product was fused with flanking regions and an antibiotic-resistant marker. This composite PCR product was then integrated into the chromosome through homologous recombination after transformation into the supercompetent cells of the *B. subtilis* strain SCK6 [67]. Additionally, a random mutagenesis protocol has been devised for the methylotrophic yeast *Pichia pastoris* [68]. This protocol involves the sequential amplification of plasmids using Phi29 DNA polymerase, encompassing error-prone rolling circle amplification (RCA) followed by multiple displacement amplification (MDA). Through these steps, it becomes feasible to obtain microgram amounts of plasmids for subsequent electroporation into *Pichia* cells, addressing a key challenge in employing *Pichia* for directed evolution.

In the following section, we will explore recent advancements in four key areas: (1) *In vivo* mutagenesis in *E. coli*, (2) virus-assisted mutagenesis, (3) random base editing, and (4) random insertion and deletion (InDel).

3.2.1. In vivo mutagenesis in *E. coli*

In vivo mutagenesis provides several distinct advantages over *in vitro* mutagenesis. This approach allows the integration of genetic diversity creation and selection, eliminating the need for a separate transformation step that often limits the size of gene libraries in *in vitro* mutagenesis. Additionally, it avoids the labour-intensive process associated with *in vitro* gene library creation.

Liu and colleague developed highly effective, inducible, broad-spectrum mutagenesis systems in *E. coli*, elevating mutation rates by over 320000 times compared to basal levels [69]. These plasmid systems rely on the induced and combinatorial expression of proteins associated with proofreading (*dnaQ926*), translesion synthesis (*umuD'*, *umuC*, *reacA730*), mismatch repair (*dam*, *seqA*), base excision (*ugi*, *cda1*), and base selection (*emrR*). Although efficient, this global mutagenesis strategy could introduce extensive mutations throughout the genome of the host, leading to undesirable issues such as toxicity, reduced library size, silencing of mutagenic plasmids, or the introduction of parasite variants into DNA libraries (mutations outside the target GOI that allow the host to circumvent the selection scheme).

MutaT7 is a targeted *in vivo* mutagenesis strategy that overcomes the challenges associated with global mutagenesis strategies [70]. It employs a DNA-damaging cytidine deaminase fused to a processive T7 RNA polymerase (T7RNAP), allowing continuous directed mutations to any DNA region downstream of a T7 promoter. To enhance mutation rates, eMutaT7 replaces the cytidine deaminase from rat APOBEC1 (rApo1) with *Petromyzon marinus* cytidine deaminase (PmCDA1), resulting in an increased mutation rate from 0.34 mutations/kb/day (MutaT7) to 4 mutations/kb/day [71]. The MutaT7 toolbox has recently been expanded further with the use of adenosine deaminase-T7 RNA polymerase fusion proteins [72], such as TadA8e fused to T7RNAP and TadA7.10 fused to T7RNAP. TadA8e and TadA7.10 are variants of *E. coli* tRNA adenosine deaminase, evolved to operate on DNA [73]. Despite its utility, the MutaT7 toolkit has limitations, including a restricted mutational spectrum, strand bias, and the necessity to prevent the repair of deoxyuridine (e.g., deleting uracil-DNA glycosylase in the host or using a uracil-DNA glycosylase inhibitor) for significant mutagenesis when using cytidine deaminase.

When employing a highly processive T7RNAP, mutations may extend beyond the intended target GOI. To confine the mutagenesis to a specific region, multiple copies of the T7 terminator are required, serving as a boundary to restrict mutagenesis [70]. Alternatively, in the T7-DIVA strategy [74], the catalytically dead Cas9 (dCas9), tethered to a custom-designed crRNA, acts as a 'roadblock' for the base deaminase-T7RNAP fusion proteins, effectively constraining the window of mutagenesis.

Instead of relying on base deaminase as a mutagenic agent and T7RNAP as a 'guide protein' (GP) to target mutations to specific loci, the EvolvR system offers continuous nucleotide diversification within a tunable window length at user-defined loci [75,76]. This is accomplished by directly generating mutations using engineered DNA polymerases (variant of *E. coli* PolII with reduced fidelity) targeted to specific loci through CRISPR-guided nickases (nCas9).

3.2.2. Virus-assisted mutagenesis

Viruses provide a distinctive avenue for designing rapid laboratory evolution experiments, capitalizing on their inherent capacity to evolve at a much faster pace than many living organisms [77]. This accelerated evolution is facilitated by their smaller genome size, which tolerates a high frequency of mutations, and a rapid rate of replication. These attributes present an excellent opportunity for the directed evolution of various biomolecules.

In the Viral Evolution of Genetically Actuating Sequences (VEGAS) method, the highly mutagenic RNA alphavirus Sindbis, belonging to the *Togaviridae* family and lacking known proof-reading capability, was employed to establish a mammalian-directed evolution system [78]. Estimates of RNA virus mutation frequencies range from 10^{-5} to 10^{-3} mutations per base replicated. The mutation rate of the Sindbis virus was quantified to be $1.0 \times 10^{-4} \pm 3.7 \times 10^{-5}$ mutations/base/hr. To create a robust directed evolution platform that capitalizes on the replicative and mutagenic potential of the Sindbis virus, artificial selective pressure must be applied. Each Sindbis viral particle requires 240 copies of the structural proteins E1, E2, and capsid to form a functional viral particle capable of maturing and propagating; without this envelope, the virus is unable to mature and propagate. By engineering restrictions on structural genome transcription, it is possible to apply a selective pressure to transgenic Sindbis virus carrying a GOI. The VEGAS system therefore allows for the simultaneous operation of viral mutagenesis, selection, and heredity. It's noteworthy that other reports on virus-assisted directed evolution, which didn't utilize viruses for creating genetic diversity, are not included in this review.

3.2.3. Random base editing

We have introduced the use of base editors (BE; *e.g.*, cytidine deaminase and adenosine deaminase) for random mutagenesis in *E. coli* in section 3.2.1. On first glance, this section on random base editing may seem redundant or repetitive. However, dedicating an additional section to random base editing is warranted for three reasons: (1) to discuss method variation, (2) to highlight the versatility of this approach, extending it to a wide range of organisms/cells beyond *E. coli*, and (3) to showcase its diverse applications. Indeed, the summary of methods in Table 5 emphasizes the popularity of this approach, further justifying a dedicated section on this topic.

Methods utilizing BEs for targeted random mutagenesis share a common chimeric protein design (Figure 3): a BE is tethered to a GP, with or without any accessory protein(s). The BE functions as the mutagenic agent by deaminating cytidine to uridine or adenosine to inosine. The GP directs the BE to specific gene loci for mutagenesis. Key variations among methods primarily reside in four areas: (1) the choice of BE, (2) the selection of GP, (3) the mechanisms of linkage between GP and BE, and (4) the utilization or absence of accessory protein(s).

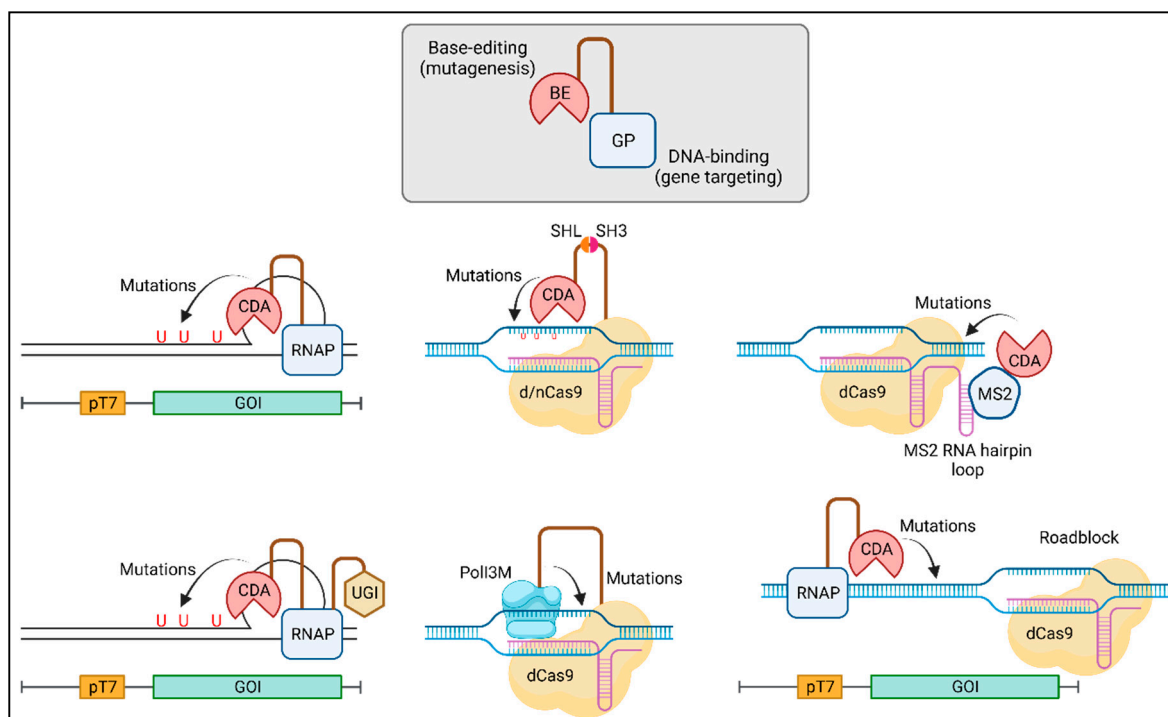


Figure 3. Targeted random mutagenesis using chimeric proteins comprising a base editor (BE) and a guide protein (GP), following the general BE-GP protein architecture. BE is the mutagenic agent, introducing random mutations through its base editing activity (e.g., cytidine and adenosine deamination). GP, with DNA-binding capability, guides or leads the BE to its target locus within the gene of interest (GOI) to effect mutagenesis. Typical BE choices include cytidine deaminase (CDA) or error-prone DNA polymerase (Pol). Frequently used GP candidates are T7 RNA polymerase (RNAP) or catalytically dead Cas9 (dCas9)/Cas9 nickase (nCas9). BE is tethered to GP via gene fusion or protein-protein or protein/RNA interactions through the utilization of the SRC homology domain 3 (SH3) and the MS2 bacteriophage coat protein. In some methods, an accessory protein such as uracil DNA glycosylase inhibitor (UGI) is required.

The most commonly employed BEs include apolipoprotein B mRNA editing enzyme catalytic subunit 1 (rAPOBEC1) from rat (*Rattus norvegicus*), cytidine deaminase (PmCDA1) from sea lamprey (*Petromyzon marinus*), human activation-induced cytidine deaminase (hAID), *E. coli* tRNA adenosine deaminase (TadA), and their respective variants. CRISPR-Cas and T7RNAP represent the most popular choices for GPs. While gene fusion stands out as the most straightforward method for connecting a BE to its GP, alternative strategies have proven effective, including the utilization of the SRC homology domain 3 (SH3) and the MS2 bacteriophage coat protein. To augment base editing efficiency, the inclusion of accessory proteins is frequently essential. Uracil DNA glycosylase inhibitor (UGI), an 83-residue protein derived from *Bacillus subtilis* bacteriophage PBS1, is extensively utilized for this purpose [79]. Additionally, the incorporation of mismatch and base excision proteins, Apn2p and Msh6p, has been shown to enhance editing efficiency.

With the BE-GP approach demonstrating growing maturity and proven success across diverse microbes and mammalian cells, we foresee ongoing developments and expect to witness adaptations of this method to additional microbial systems.

Table 5. Targeted and non-targeted (global) random base editing methods, reported in the past 10 years.

Method/first author's name	Guide protein (GP)	Base editor (BE)	Linkage between GP and BE	Organisms/cells validated	Ref

(year of publication)					
<i>Targeted mutagenesis</i>					
Nishida <i>et al.</i> (2016)	dCas9 ¹ or nCas9 ²	PmCDA1	Gene fusion or interaction between SH3 (SRC homology domain 3) and SHL (SH3 interaction ligand)	<i>S. cerevisiae</i> and CHO	[80]
CRISPR-X (2016)	dCas9 ¹	hAID* Δ^4	MS2 bacteriophage coat protein (MCP) binding to the MS2 RNA stem-loop	K-562 cell	[81]
TAM (2016)	dCas9 ¹	hAID, hAID CD ⁵ , hAID P182X ⁶ or hAID R190X ⁷	Gene fusion	K-562 cell and HEK293T cell	[82]
Komor <i>et al.</i> (2016)	dCas9 ¹ or nCas9 ²	hAID, rAPOBEC1, hAPOBEC3G or PmCDA1	Gene fusion	U2OS cell, HEK293T cell and HCC1954 cell	[83]
Gehrke <i>et al.</i> (2018)	nCas9 ²	rAPOBEC1 or engineered hAPOBEC3A	Gene fusion	U2OS cell and HEK293T	[84]
TRACE (2020)	T7 RNA polymerase	rAPOBEC1 or hAID* Δ^4	Gene fusion	HEK293T cell	[85]
TRIDENT (2021)	T7 RNA polymerase	PmCDA1 or yeTadA1.0 ⁸	Gene fusion	<i>S. cerevisiae</i>	[86]
Volke <i>et al.</i> (2022)	nCas9 ²	rAPOBEC1	Gene fusion	<i>P. putida</i> and <i>P. aeruginosa</i>	[87]
Skrekas <i>et al.</i> (2023)	dCas9 ¹	hAID* Δ^4 , TadA8e ⁹ or TadA8e V106W ¹⁰	Gene fusion	<i>S. cerevisiae</i>	[88]
CoMuTER (2023)	dCas3 ³	PmCDA1 or rAPOBEC1	Gene fusion	<i>S. cerevisiae</i>	[89]
<i>Global mutagenesis</i>					
Pan <i>et al.</i> (2021)		rAPOBEC1	N/A	<i>S. cerevisiae</i>	[90]

¹Catalytically dead Cas9, containing D10A and H840A mutations; ²Cas9 nickase, containing D10A mutation;³Nuclease-deficient Cas3, containing H74A and D75A mutations; ⁴Truncated AID, *i.e.*, AID Δ 196-198, containing mutations K10E, T82I and E156G; ⁵Truncated AID retaining only the catalytic domain, *i.e.*, AID Δ 94-

198; ⁶Truncated AID, *i.e.*, AID Δ 182-198; ⁷Truncated AID, *i.e.*, AID Δ 190-198; ⁸Variant of *E. coli* TadA, evolved in yeast; ⁹Variant of *E. coli* TadA; ¹⁰Variant of *E. coli* TadA.

3.2.4. Random insertion and deletion

Insertions and deletions in genomes occur naturally due to replication slippage or error-prone NHEJ of double-stranded breaks. While their utilization in protein engineering is infrequent due to their tendency to be highly deleterious, often resulting in frame-shift mutations that significantly alter the protein sequence or prematurely terminate translation, there is evidence suggesting potential benefits. Instances of altered protein functionality through insertions and deletions have been reported, such as the broadening of substrate specificity in β -lactamase [91] and the modification of coenzyme specificity in Rossmann fold enzymes [92]. These findings underscore the need for further exploration and investigation into the potential applications of this approach in protein engineering.

Methods for insertion and deletion in protein engineering have been developed over the last two decades, as comprehensively outlined in a recent review [93]. Some approaches leverage the higher slippage rates of certain polymerases to insert or delete 1 to 2 bases, inducing frame-shift mutations. Alternatively, other methods involve the fragmentation of GOI using DNase I or cleavage through exonucleases, endonucleases, or chemicals. Subsequently, nucleotides are added through processes like terminal deoxynucleotide transferase [94] or rolling circle amplification [95], incorporating random numbers or blocks of nucleotides. This results in libraries containing both in-frame and frame-shift variants.

Another set of methods relies on transposons to generate insertion and deletion libraries, and these approaches can prevent frame-shift occurrences through careful transposon sequence design [96–98]. Typically, transposons are designed with recognition sites for selected restriction enzymes, enabling cleavage and subsequent religation to generate the insertion or deletion of nucleotide triplets. A notable example is the recently developed TRIAD method that randomly inserts or deletes one to three nucleotide triplets [99]. This method randomly inserts engineered mini-Mu transposons, defining the location of the insertion/deletion event. For deletion, the recognition site for the type II restriction enzyme MlyI is designed into both ends of the transposon. After MlyI digestion and religation, 3bp are deleted. Repeating this process using MlyI custom cassettes enables longer deletions of up to 9 bp. For insertion, an asymmetric transposon with NotI and MlyI restriction sites is used. Following restriction digest, a cassette carrying up to three NNN triplets can be inserted. The TRIAD method has been applied to evolve arylesterase activity in a phosphotriesterase [99] and enhance antibody affinity [100].

3.3. Focused mutagenesis

Advancement in focused mutagenesis is marked by five noteworthy trends. These trends encompass (1) a transition from site-directed mutagenesis to multi-site directed mutagenesis and even massive mutagenesis, (2) the incorporation of CRISPR/Cas9 in focused mutagenesis, mirroring trends seen in cloning and random mutagenesis, (3) the formulation of strategies to minimize library size, (4) the development of computational tools for automated oligo design, and (5) a shift from column-synthesized oligos to microarray-synthesized oligos for protein engineering. The subsequent sections will discuss each of these trends individually.

3.3.1. Multi-site directed mutagenesis

Significant progress in multi-site directed mutagenesis has predominantly emerged in methodologies utilizing single-stranded DNA (ssDNA) templates, as summarized in Table 6. A key example is Nicking Mutagenesis [101], leveraging a nicking enzyme (Nt.BbvCI) and exonucleases (ExoIII, ExoI) for the preparation of ssDNA templates. This approach involves a pool of mutagenic primers defining the targeted mutagenesis sites, which anneal to the ssDNA and undergo isothermal assembly. Following template strand removal, the complementary mutant strand is synthesized via PCR. Continuous protocol enhancements have been reported to bolster efficiency (Table 6). In

contrast, the SLUPT method generates the ssDNA template by eliminating the phosphorylated strand in the linearly amplified GOI [102].

Another group of methods opts for the creation of gene fragments bearing the desired mutations, which are subsequently combined in a sequential assembly using techniques like megaprimer PCR, overlap extension PCR, or Golden Gate assembly. Drawing inspiration from genome recombineering methods such as MAGE, Higgins *et al.* devised a plasmid recombineering method for in vivo multi-site directed mutagenesis [103].

Another development involves leveraging solid-phase gene synthesis technologies to design libraries [104,105]. This approach minimizes the risk of stop codons and unintended mutations during PCR, thereby reducing screening efforts [104]. Notably, Öling *et al.* achieved an impressive 161 multisite mutations using this methodology [106].

Table 6. Multi-site directed mutagenesis reported in the last decade.

Method/first author (year of publication)	Method description and mutagenic agent	Template	Maximum mutated sites per gene	Ref
<i>ssDNA template</i>				
Nicking Mutagenesis (2016)	<ul style="list-style-type: none"> • ssDNA template generated with nicking enzyme (Nt.BbvCI) and exonucleases (ExoIII, ExoI). • Phosphorylated mutagenic primers annealed to template and extended by PCR using Phusion polymerase and Taq DNA ligase. • Template ssDNA removed with nicking enzyme (Nb.BbvCI) and exonucleases. • Complementary mutant strand synthesized by PCR. 	Plasmid ssDNA	7	[101]
Darwin Assembly (2018)	<ul style="list-style-type: none"> • ssDNA template created with nicking enzyme (Nt.BbvCI) and exonuclease. • Phosphorylated mutagenic primers annealed to template and extended by PCR using Q5 polymerase and Taq DNA ligase. • Outnest primers used for PCR amplification of assembled fragment. • Template removal may or may not be required depending on the complexity of construct. • Library cloned into vector using Golden Gate. 	Plasmid ssDNA	19	[107]

Nicking Mutagenesis (2021)	<ul style="list-style-type: none"> • Modified from Nicking Mutagenesis (2016). • Improved library coverage by using longer mutagenic primers and quick annealing conditions 	Plasmid ssDNA	15	[108]
SLUPT (2021)	<ul style="list-style-type: none"> • GOI amplification with dUTP and one phosphorylated primer. • ssDNA template created by degradation of phosphorylated strand using lambda exonuclease. • Phosphorylated mutagenic primers annealed to template and extended by PCR using Phusion polymerase and Taq DNA ligase. • Template ssDNA removal using uracil DNA glycosylase. • Library cloned into vector by restriction digest and ligation. 	Linear GOI ssDNA	17	[102]
SUNi Mutagenesis (2023)	<ul style="list-style-type: none"> • Derived from Nicking Mutagenesis. • Longer mutagenic primers with a 5'-G/C to improve efficiency. • Codon design at mutagenic site to minimize parent. 	Plasmid ssDNA	n.r*	[109]
<i>Assembly of mutated gene fragments</i>				
Combinatorial Codon mutagenesis (CCM) (2014, 2017)	<ul style="list-style-type: none"> • Megaprimers generated by one-pot amplification with mutagenic primers and an end primer. • Random combinations of mutations generated by PCR reassembly of GOI with mutant megaprimers and two end primers. • Template removed by gel extraction of assembled GOI variants. • Assembled GOI library used as template for further rounds of megaprimer generation and reassembly to accumulate mutations per gene. • Library cloned into vector by restriction digest and ligation. 	Plasmid dsDNA	7	[110,111]

Chung <i>et al.</i> (2017)	<ul style="list-style-type: none"> • Megaprimers generated by one-pot or multi-pot amplification with mutagenic primers and an end primer. • Random combinations of mutations generated by PCR reassembly of GOI with mutant megaprimers and two end primers. • Library cloned by Gibson assembly. 	Linear GOI dsDNA	14	[112]
Golden Mutagenesis (2019)	<ul style="list-style-type: none"> • Fragments generated using mutagenic primers with Type IIS recognition sites. • Template removed by gel extraction of fragments. • Fragments assembled with vector via Golden gate. • Web-tool for primer design and sequencing result analysis. 	Plasmid dsDNA	5	[113]
Hejlesen <i>et al.</i> (2020)	<ul style="list-style-type: none"> • Fragments with overlapping regions generated by PCR amplification with mutagenic primers. • Prolonged overlap extension PCR created linear repetitive multimers of the template. • Linear multimers transformed and circularised in bacteria. 	Plasmid dsDNA	3	[114]
<i>In vivo method</i>				
Plasmid recombineering (2017)	<ul style="list-style-type: none"> • <i>In vivo</i> method. • Plasmid and mutagenic primers electroporated into a recombineering strain. • Repeat recombineering using variant plasmids as template to increase sites of mutations. 	Plasmid dsDNA	2	[103]

*n.r. not reported.

3.3.2. CRISPR/Cas9-mediated mutagenesis

She *et al.* introduced a PCR-free, two-step *In vitro* CRISPR/Cas9-mediated Mutagenic (ICM) system designed for both single-site and multi-site directed mutagenesis [115]. In the first step, site-specific plasmid digestion is achieved by employing a complex of Cas9 with specific single guide RNA (sgRNA), followed by degradation with T5 exonuclease to create a 15-nucleotide homologous region. Subsequently, in step 2, primers containing the desired mutations are annealed to generate

double-stranded DNA fragments, which are then ligated into the linearized plasmid. A distinct advantage of employing a PCR-free approach is the attainment of greater genetic diversity, attributed to the elimination of biases introduced by PCR, such as preferential primer binding to the template. Anticipating ongoing advancements, we foresee more methodologies integrating the use of CRISPR/Cas9 in the near future.

3.3.3. The numbers game

In focused mutagenesis, with the escalation of target sites, the library size (number of permutations) undergoes exponential expansion. When combined with oversampling to ensure certain degree of diversity coverage, this exerts significant strain on library screening, making the navigation of the entire diversity space practically challenging – a scenario often denoted as permutation explosion. Commonly, three strategies are employed to tackle this issue: the divide-and-conquer approach, the codon randomization strategy, and the use of machine learning (ML).

GeneORator [116] employs a divide-and-conquer approach. In this method, target sites are partitioned into subsets, such as subsets A and B. A permutation library is established for each subset, generating libraries A and B. Both libraries A and B are then independently screened for the desired phenotype. Another notable divide-and-conquer strategy is the iterative saturation mutagenesis (ISM) pioneered by Reetz [117], taking the approach one step further. Beneficial variants from each subset act as templates to randomize the other subset, leading to the creation of new libraries AB or BA for the subsequent round of screening. A potential limitation of the divide-and-conquer approach is that it may not capture some pairwise and higher-order effects, known as epistasis, between residues in the gene libraries.

M-ISM, a variant of the ISM method [118], employs a computational tool to design a 'small-intelligent' library for each subset. The term 'small-intelligent' denotes a minimal gene library size devoid of inherent amino acid biases, stop codons, or rare codons specific to the protein expression host [119]. Rather than sampling each target site with 19 amino acid alphabets, it is feasible to obtain improved protein variants using reduced amino acid alphabets, potentially even with just 1 or 2 alphabets [120].

Degenerate Codon Optimization for Informed Libraries (DeCOIL) is a recently introduced ML method that optimizes degenerate codon libraries to sample protein variants likely to exhibit both high fitness and high diversity within the sequence search space [121]. Like all ML approaches, DeCOIL necessitates a training library to develop its ML model.

3.3.4. Automated oligo design

Mutation Maker [122] and GeneGenie [123] serve as oligo design platforms for protein engineering. Taking Mutation Maker as an example, the platform designs oligos for various applications, including site-scanning saturation mutagenesis, multi-site directed mutagenesis, and PCR-based accurate gene synthesis. The designed oligos can be assembled through overlap extension PCR, with the program optimizing melting temperature (T_m) and codon choice. The use of degenerate oligos remains popular in focused mutagenesis. Web-based applications like CodonGenie [124] facilitate the selection of degenerate codons. For high-throughput or massively parallel protein engineering, automated oligo design proves useful in minimizing errors and expediting experimental design.

3.3.5. Oligo pool (oPool) for cost-effective library construction

Rather than relying on oligonucleotides with degenerate codons synthesized through solid-phase phosphoramidite chemistry in a column-based synthesis, there is a growing preference for using microarray-synthesized oligonucleotides or oligo pools (oPools) [125]. oPools are mixes of thousands of individually designed polynucleotides, each up to 350 bases in length.

Individually designed oligos offer several distinct advantages, including the removal of redundancy from the gene library, extensive coverage of mutational space, precise selection of codons

optimized for protein expression in a chosen host, and facilitation of studies such as deep mutational scanning (DMS) [126] that necessitate large oligonucleotide pools. Programmed Allelic Series (PALS) is a technique devised for massively parallel single amino acid mutagenesis [127]. It achieves this by integrating a low-cost oPool with overlap-extension mutagenesis. oPool has been demonstrated to yield higher quality gene libraries compared to degenerate oligos [128].

While oPools show significant promise for protein engineering, it is essential to acknowledge certain challenges [125]. First, although the number of individual user-defined oligo sequences in a pool is large, their individual concentrations are low. Second, in focused mutagenesis applications, some oligos might preferentially bind to the template. Exacerbated by low concentrations, certain mutations may not be present in the final constructed gene library. Third, as the length of the oligos in the pool increases, the percentage of truncated molecules also rises, further diminishing the expected concentration of full-length molecules. Fourth, the error rates for oPools are typically higher than those for column-synthesized oligos.

Irrespective of the method chosen for creating a gene library, effective target site identification is crucial for focused mutagenesis. A recent comprehensive review by the Dalby group offers an excellent overview of methodologies for identifying target sites based on sequence or structural information [129]. Additionally, the review provides a valuable list of computational tools designed for target site identification.

3.4. DNA recombination

While considerable strides have been made in random mutagenesis and focused mutagenesis categories, advancements in the DNA recombination category have been relatively modest. An article sought to tackle the challenge of recombining protein modules from distant parents with minimal disruption at crossover sites. To overcome this hurdle, an approach called key motif-directed recombination was introduced [130]. Members of the same protein superfamily often share common structural or sequence motifs. Validated through the creation of α/β -hydrolase chimeras, this method strategically conducted recombination at key sequence motif regions. These chimeras retained their biological functions and exhibited desirable properties.

4. Applications of genetic diversity creation and future prospects

Directed evolution has firmly established itself as a powerful tool in biotechnology, diversifying its applications from evolving single enzymes or proteins to fascinating areas such as metabolic pathway engineering, organismal engineering, viral engineering, and the engineering of molecular biology tools. In the section below, we provide recent examples of applications beyond engineering a single protein.

Biofuels have emerged as a promising alternative to fossil fuels in sustainable energy solutions. Agricultural waste, a rich source of cellulose, provides a valuable feedstock for biofuel production [131]. Cellobiose, an intermediate in the conversion of cellulose into glucose monomers for fermentative processes, is crucial for efficient biofuel production. Directed evolution has played a key role in engineering two essential proteins in the cellobiose utilization pathway in *S. cerevisiae*: β -glucosidase and cellodextrin transporter [132,133]. These advancements have significantly increased ethanol production, contributing to the development of a more sustainable biofuel industry.

In section 3.2.3 above, we extensively covered the use of the 'BE-GP' protein architecture for targeted random mutagenesis in a wide range of biological systems. By excluding the GP, the method transforms into a potent global mutagenesis approach suitable for adaptive laboratory evolution (ALE). Pan *et al.* showcased the application of BE for ALE of *S. cerevisiae*, enhancing resistance to isobutanol and acetate, and boosting the production of β -carotene [90].

In a recent study, directed evolution was leveraged to enhance the transduction capabilities of adeno-associated viruses (AAVs) [134]. Directed evolution of the AAV capsid protein resulted in variants exhibiting significantly improved transduction efficiencies. This breakthrough facilitated the genetic manipulation of microglia, specialized immune cells in the central nervous system, with unprecedented ease. Given the implication of microglial dysfunction in neurodegenerative disorders

and brain cancers, this advancement opens avenues for a more profound understanding of their functions and potential therapeutic interventions.

In genome editing and targeting, the landscape has been revolutionized by the emergence of CRISPR-Cas9. The widely utilized Cas9 nuclease, originating from *Streptococcus pyogenes*, encounters challenges attributed to its substantial size, constraining its effectiveness in delivery into cells. A smaller alternative, the Cas9 orthologue from *Campylobacter Jejuni* (CjCas9), presents an appealing solution [135]. Nevertheless, the intricate PAM sequence recognized by CjCas9 imposes limitations on its versatility. Through successive rounds of directed evolution, a variant of CjCas9 named evoCjCas9 was found, boasting an altered PAM recognition sequence that is ten times more prevalent in the genome than the canonical PAM sequence [136]. Beyond this, evoCjCas9 demonstrates increased nuclease activity compared to its wildtype counterpart, thereby broadening the horizons of CRISPR-Cas9 technology. Throughout this review, CRISPR-Cas9 has been consistently emphasized for its role in cloning and genetic diversity creation. It's truly inspiring to witness the 'ripple effect' where proteins crafted through protein engineering contribute to the continual expansion of the capabilities of protein engineering itself.

The convergence of automation and the emergence of artificial intelligence (AI) is poised to bring about a transformative shift in the domain of directed evolution, promising substantial improvements in scale, efficiency, and precision of experiments. Automation stands to simplify labour-intensive processes like library construction, screening, and the identification of desired variants. AI tools, coupled with the capability of analysing extensive datasets, have the potential to steer experimental and variant design, expediting the discovery process.

Embarking upon this transformative era necessitates the unlocking of untapped biological potential. At the heart of this unlocking process lies the art and science of creating genetic diversity. It is in the diverse tapestry of our genetic landscape that the blueprint for the future of the bioeconomy is etched. Through the strategic exploration and manipulation of genetic diversity, we forge a path into an era where the frontiers of bioengineering unfold boundlessly, promising novel discoveries and innovations that will shape the landscape of tomorrow.

References

1. Arnold FH (2019) *Angew Chem Int Ed Engl* 58: 14420. doi:10.1002/anie.201907729
2. Fasan R, Jennifer Kan SB, Zhao H (2019) *ACS Catal* 9: 9775. doi:10.1021/acscatal.9b02737
3. Wong TS, Zhurina D, Schwaneberg U (2006) *Comb Chem High Throughput Screen* 9: 271. doi:10.2174/138620706776843192
4. Tee KL, Wong TS (2013) *Biotechnol Adv* 31: 1707. doi:10.1016/j.biotechadv.2013.08.021
5. Wong TS, Tee KL (2020) *A practical guide to protein engineering*. Springer Cham
6. Watson JF, Garcia-Nafria J (2019) *J Biol Chem* 294: 15271. doi:10.1074/jbc.REV119.009109
7. Nozaki S, Niki H (2019) *J Bacteriol* 201. doi:10.1128/JB.00660-18
8. Jones DH, Howard BH (1991) *Biotechniques* 10: 62.
9. Garcia-Nafria J, Watson JF, Greger IH (2016) *Sci Rep* 6: 27459. doi:10.1038/srep27459
10. Huang F, Spangler JR, Huang AY (2017) *PLoS One* 12: e0183974. doi:10.1371/journal.pone.0183974
11. Jacobus AP, Gross J (2015) *PLoS One* 10: e0119221. doi:10.1371/journal.pone.0119221
12. Kostylev M, Otwell AE, Richardson RE, Suzuki Y (2015) *PLoS One* 10: e0137466. doi:10.1371/journal.pone.0137466
13. Wang Y, Liu Y, Chen J, Tang MJ, Zhang SL, Wei LN, Li CH, Wei DB (2015) *Genet Mol Res* 14: 12306. doi:10.4238/2015.October.9.19
14. Chen X, Yuan H, He W, Hu X, Lu H, Li Y (2005) *Sci China C Life Sci* 48: 330. doi:10.1360/04yc0040
15. Gibson DG (2009) *Nucleic Acids Res* 37: 6984. doi:10.1093/nar/gkp687
16. Oldenburg KR, Vo KT, Michaelis S, Paddon C (1997) *Nucleic Acids Res* 25: 451. doi:10.1093/nar/25.2.451
17. Ding W, Weng H, Jin P, Du G, Chen J, Kang Z (2017) *Bioengineered* 8: 296. doi:10.1080/21655979.2017.1308986
18. Joska TM, Mashruwala A, Boyd JM, Belden WJ (2014) *J Microbiol Methods* 100: 46. doi:10.1016/j.mimet.2013.11.013

19. Kuijpers NG, Solis-Escalante D, Bosman L, van den Broek M, Pronk JT, Daran JM, Daran-Lapujade P (2013) *Microb Cell Fact* 12: 47. doi:10.1186/1475-2859-12-47
20. Mashruwala AA, Boyd JM (2016) *Methods Mol Biol* 1373: 33. doi:10.1007/7651_2015_275
21. van Leeuwen J, Andrews B, Boone C, Tan G (2015) *Cold Spring Harb Protoc* 2015: pdb prot085100. doi:10.1101/pdb.prot085100
22. Ip K, Yadin R, George KW (2020) *Methods Mol Biol* 2205: 79. doi:10.1007/978-1-0716-0908-8_5
23. Abdel-Banat BM, Nonklang S, Hoshida H, Akada R (2010) *Yeast* 27: 29. doi:10.1002/yea.1729
24. Hoshida H, Murakami N, Suzuki A, Tamura R, Asakawa J, Abdel-Banat BM, Nonklang S, Nakamura M, Akada R (2014) *Yeast* 31: 29. doi:10.1002/yea.2993
25. Ruiz E, Talenton V, Dubrana MP, Guesdon G, Lluch-Senar M, Salin F, Sirand-Pugnet P, Arfi Y, Lartigue C (2019) *ACS Synth Biol* 8: 2547. doi:10.1021/acssynbio.9b00224
26. Pang Q, Ma S, Han H, Jin X, Liu X, Su T, Qi Q (2022) *ACS Synth Biol* 11: 1477. doi:10.1021/acssynbio.1c00529
27. He B, Ma Y, Tian F, Zhao GR, Wu Y, Yuan YJ (2023) *Nucleic Acids Res* 51: 8283. doi:10.1093/nar/gkad599
28. Jeong YK, Yu J, Bae S (2019) *Scientific Reports* 9. doi:10.1038/s41598-019-41356-1
29. Wang JW, Wang A, Li K, Wang B, Jin S, Reiser M, Lockey RF (2015) *Biotechniques* 58: 161. doi:10.2144/000114261
30. Jiang W, Zhao X, Gabrieli T, Lou C, Ebenstein Y, Zhu TF (2015) *Nat Commun* 6: 8101. doi:10.1038/ncomms9101
31. Richter D, Bayer K, Toesko T, Schuster S (2019) *Sci Rep* 9: 2980. doi:10.1038/s41598-019-39768-0
32. Dao VL, Chan SA, Zhang JY, Ngo RKJ, Poh CL (2022) *Scientific Reports* 12. doi:10.1038/s41598-022-07878-x
33. Xia Y, Li K, Li J, Wang T, Gu L, Xun L (2019) *Nucleic Acids Res* 47: e15. doi:10.1093/nar/gky1169
34. Yu F, Li X, Wang F, Liu Y, Zhai C, Li WQ, Ma LX, Chen WP (2023) *Frontiers in Bioengineering and Biotechnology* 11. doi:10.3389/fbioe.2023.1167534
35. Nisson PE, Rashtchian A, Watkins PC (1991) *PCR Methods Appl* 1: 120. doi:10.1101/gr.1.2.120
36. Cavaleiro AM, Kim SH, Seppala S, Nielsen MT, Norholm MH (2015) *ACS Synth Biol* 4: 1042. doi:10.1021/acssynbio.5b00113
37. Genee HJ, Bonde MT, Bagger FO, Jespersen JB, Sommer MO, Wernersson R, Olsen LR (2015) *ACS Synth Biol* 4: 342. doi:10.1021/sb500194z
38. Jajnesniak P, Wong TS (2015) *J Biol Eng* 9: 15. doi:10.1186/s13036-015-0010-3
39. Jajnesniak P, Tee KL, Wong TS (2019) *Int J Mol Sci* 20. doi:10.3390/ijms20163908
40. Jajnesniak P, Tee KL, Wong TS (2022) *Methods Mol Biol* 2461: 123. doi:10.1007/978-1-0716-2152-3_8
41. Jin P, Ding W, Du G, Chen J, Kang Z (2016) *ACS Synth Biol* 5: 1028. doi:10.1021/acssynbio.6b00078
42. Kang Z, Ding W, Jin P, Du G, Chen J (2018) *Methods Mol Biol* 1772: 421. doi:10.1007/978-1-4939-7795-6_24
43. Bird JE, Marles-Wright J, Giachino A (2022) *ACS Synth Biol* 11: 3551. doi:10.1021/acssynbio.2c00355
44. Marillonnet S, Grutzner R (2020) *Curr Protoc Mol Biol* 130: e115. doi:10.1002/cpmb.115
45. Sorida M, Bonasio R (2023) *Cell Reports Methods* 3. doi:10.1016/j.crmeth.2023.100564
46. Storch M, Casini A, Mackrow B, Fleming T, Trewhitt H, Ellis T, Baldwin GS (2015) *ACS Synth Biol* 4: 781. doi:10.1021/sb500356d
47. van Dolleweerd CJ, Kessans SA, Van de Bittner KC, Bustamante LY, Bundela R, Scott B, Nicholson MJ, Parker EJ (2018) *Acs Synthetic Biology* 7: 1018. doi:10.1021/acssynbio.7b00363
48. Lin D, O'Callaghan CA (2018) *Nucleic Acids Res* 46: e113. doi:10.1093/nar/gky596
49. Taylor GM, Heap JT (2020) *Methods Mol Biol* 2205: 219. doi:10.1007/978-1-0716-0908-8_13
50. Taylor GM, Mordaka PM, Heap JT (2019) *Nucleic Acids Res* 47: e17. doi:10.1093/nar/gky1182
51. Trubitsyna M, Honsbein A, Jayachandran U, Elfick A, French CE (2020) *Methods Mol Biol* 2205: 161. doi:10.1007/978-1-0716-0908-8_10
52. Storch M, Haines MC, Baldwin GS (2020) *Synthetic Biology* 5. doi:10.1093/synbio/ysaa010
53. Enghiad B, Xue P, Singh N, Boob AG, Shi CY, Petrov VA, Liu R, Peri SS, Lane ST, Gaither ED, Zhao HM (2022) *Nature Communications* 13. doi:10.1038/s41467-022-30355-y
54. Bryant JA, Kellinger M, Longmire C, Miller R, Wright RC (2023) *Synthetic Biology* 8. doi:10.1093/synbio/ysac032
55. Fukuda M, Umeno H, Nose K, Nishitarumizu A, Noguchi R, Nakagawa H (2017) *Sci Rep* 7: 41478. doi:10.1038/srep41478

56. Kopsidas G, Carman RK, Stutt EL, Raicevic A, Roberts AS, Siomos MA, Dobric N, Pontes-Braz L, Coia G (2007) *BMC Biotechnol* 7: 18. doi:10.1186/1472-6750-7-18
57. Wong TS, Roccatano D, Loakes D, Tee KL, Schenk A, Hauer B, Schwaneberg U (2008) *Biotechnol J* 3: 74. doi:10.1002/biot.200700193
58. Wong TS, Tee KL, Hauer B, Schwaneberg U (2004) *Nucleic Acids Res* 32: e26. doi:10.1093/nar/gnh028
59. Wong TS, Tee KL, Hauer B, Schwaneberg U (2005) *Anal Biochem* 341: 187. doi:10.1016/j.ab.2005.03.023
60. Zhao J, Frauenkron-Machedjou VJ, Kardashliev T, Ruff AJ, Zhu L, Bocola M, Schwaneberg U (2017) *Appl Microbiol Biotechnol* 101: 3177. doi:10.1007/s00253-016-8035-1
61. Zhao J, Kardashliev T, Joelle Ruff A, Bocola M, Schwaneberg U (2014) *Biotechnol Bioeng* 111: 2380. doi:10.1002/bit.25302
62. Verma R, Wong TS, Schwaneberg U, Roccatano D (2014) *Methods Mol Biol* 1179: 279. doi:10.1007/978-1-4939-1053-3_19
63. Wong TS, Roccatano D, Schwaneberg U (2007) *Biotechnol J* 2: 133. doi:10.1002/biot.200600201
64. Wong TS, Roccatano D, Zacharias M, Schwaneberg U (2006) *J Mol Biol* 355: 858. doi:10.1016/j.jmb.2005.10.082
65. Li H, Li J, Jin R, Chen W, Liang C, Wu J, Jin JM, Tang SY (2018) *Biotechnol Lett* 40: 1101. doi:10.1007/s10529-018-2559-7
66. Yang J, Ruff AJ, Arlt M, Schwaneberg U (2017) *Biotechnol Bioeng* 114: 1921. doi:10.1002/bit.26327
67. Ye B, Li Y, Tao Q, Yao X, Cheng M, Yan X (2020) *Front Microbiol* 11: 570280. doi:10.3389/fmicb.2020.570280
68. Tachioka M, Sugimoto N, Nakamura A, Sunagawa N, Ishida T, Uchiyama T, Igarashi K, Samejima M (2016) *Biotechnol Biofuels* 9: 199. doi:10.1186/s13068-016-0613-z
69. Badran AH, Liu DR (2015) *Nat Commun* 6: 8425. doi:10.1038/ncomms9425
70. Moore CL, Papa LJ, 3rd, Shoulders MD (2018) *J Am Chem Soc* 140: 11560. doi:10.1021/jacs.8b04001
71. Park H, Kim S (2021) *Nucleic Acids Res* 49: e32. doi:10.1093/nar/gkaa1231
72. Mengiste AA, Wilson RH, Weissman RF, Papa Iii LJ, Hendel SJ, Moore CL, Butty VL, Shoulders MD (2023) *Nucleic Acids Res* 51: e31. doi:10.1093/nar/gkad003
73. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR (2017) *Nature* 551: 464. doi:10.1038/nature24644
74. Alvarez B, Mencia M, de Lorenzo V, Fernandez LA (2020) *Nat Commun* 11: 6436. doi:10.1038/s41467-020-20230-z
75. Halperin SO, Tou CJ, Wong EB, Modavi C, Schaffer DV, Dueber JE (2018) *Nature* 560: 248. doi:10.1038/s41586-018-0384-8
76. Sadanand S (2018) *Nat Biotechnol* 36: 819. doi:10.1038/nbt.4247
77. Jewel D, Pham Q, Chatterjee A (2023) *Curr Opin Chem Biol* 76: 102375. doi:10.1016/j.cbpa.2023.102375
78. English JG, Olsen RHJ, Lansu K, Patel M, White K, Cockrell AS, Singh D, Strachan RT, Wacker D, Roth BL (2019) *Cell* 178: 748. doi:10.1016/j.cell.2019.05.051
79. Wang L, Xue W, Yan L, Li X, Wei J, Chen M, Wu J, Yang B, Yang L, Chen J (2017) *Cell Res* 27: 1289. doi:10.1038/cr.2017.111
80. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, Shimatani Z, Kondo A (2016) *Science* 353. doi:10.1126/science.aaf8729
81. Hess GT, Fresard L, Han K, Lee CH, Li A, Cimprich KA, Montgomery SB, Bassik MC (2016) *Nat Methods* 13: 1036. doi:10.1038/nmeth.4038
82. Ma Y, Zhang J, Yin W, Zhang Z, Song Y, Chang X (2016) *Nat Methods* 13: 1029. doi:10.1038/nmeth.4027
83. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR (2016) *Nature* 533: 420. doi:10.1038/nature17946
84. Gehrke JM, Cervantes O, Clement MK, Wu Y, Zeng J, Bauer DE, Pinello L, Joung JK (2018) *Nat Biotechnol* 36: 977. doi:10.1038/nbt.4199
85. Chen H, Liu S, Padula S, Lesman D, Griswold K, Lin A, Zhao T, Marshall JL, Chen F (2020) *Nat Biotechnol* 38: 165. doi:10.1038/s41587-019-0331-8
86. Cravens A, Jamil OK, Kong D, Sockolosky JT, Smolke CD (2021) *Nat Commun* 12: 1579. doi:10.1038/s41467-021-21876-z
87. Volke DC, Martino RA, Kozaeva E, Smania AM, Nickel PI (2022) *Nat Commun* 13: 3026. doi:10.1038/s41467-022-30780-z
88. Skrekas C, Limeta A, Siewers V, David F (2023) *ACS Synth Biol* 12: 2278. doi:10.1021/acssynbio.2c00690

89. Zimmermann A, Prieto-Vivas JE, Cautereels C, Gorkovskiy A, Steensels J, Van de Peer Y, Verstrepen KJ (2023) *Nat Commun* 14: 3389. doi:10.1038/s41467-023-39087-z
90. Pan Y, Xia S, Dong C, Pan H, Cai J, Huang L, Xu Z, Lian J (2021) *ACS Synth Biol* 10: 2440. doi:10.1021/acssynbio.1c00217
91. Hwang J, Cho KH, Song H, Yi H, Kim HS (2014) *Antimicrob Agents Chemother* 58: 6265. doi:10.1128/AAC.02648-14
92. Toledo-Patiño S, Pascarelli S, Uechi G-i, Laurino P (2022) *bioRxiv*: 2022.05.16.491946. doi:10.1101/2022.05.16.491946
93. Savino S, Desmet T, Franceus J (2022) *Biotechnol Adv* 60: 108010. doi:10.1016/j.biotechadv.2022.108010
94. Fujii R, Kitaoka M, Hayashi K (2006) *Nucleic Acids Res* 34: e30. doi:10.1093/nar/gnj032
95. Kipnis Y, Dellus-Gur E, Tawfik DS (2012) *Protein Eng Des Sel* 25: 437. doi:10.1093/protein/gzs023
96. Liu SS, Wei X, Ji Q, Xin X, Jiang B, Liu J (2016) *J Biotechnol* 227: 27. doi:10.1016/j.jbiotec.2016.03.038
97. Hayes F, Hallet B (2000) *Trends Microbiol* 8: 571. doi:10.1016/s0966-842x(00)01857-6
98. Jones DD (2005) *Nucleic Acids Res* 33: e80. doi:10.1093/nar/gni077
99. Emond S, Petek M, Kay EJ, Heames B, Devenish SRA, Tokuriki N, Hollfelder F (2020) *Nat Commun* 11: 3469. doi:10.1038/s41467-020-17061-3
100. Skamaki K, Emond S, Chodorge M, Andrews J, Rees DG, Cannon D, Popovic B, Buchanan A, Minter RR, Hollfelder F (2020) *Proc Natl Acad Sci U S A* 117: 27307. doi:10.1073/pnas.2002954117
101. Wrenbeck EE, Klesmith JR, Stapleton JA, Adeniran A, Tyo KE, Whitehead TA (2016) *Nat Methods* 13: 928. doi:10.1038/nmeth.4029
102. Meinke G, Dalda N, Brigham BS, Bohm A (2021) *Synth Biol (Oxf)* 6: ysaa030. doi:10.1093/synbio/ysaa030
103. Higgins SA, Ouonkap SVY, Savage DF (2017) *ACS Synth Biol* 6: 1825. doi:10.1021/acssynbio.7b00112
104. Hoebenreich S, Zilly FE, Acevedo-Rocha CG, Zilly M, Reetz MT (2015) *ACS Synth Biol* 4: 317. doi:10.1021/sb5002399
105. Li A, Sun Z, Reetz MT (2018) *Chembiochem* 19: 2023. doi:10.1002/cbic.201800339
106. Oling D, Lawenius L, Shaw W, Clark S, Kettleborough R, Ellis T, Larsson N, Wigglesworth M (2018) *ACS Synth Biol* 7: 2317. doi:10.1021/acssynbio.8b00118
107. Cozens C, Pinheiro VB (2018) *Nucleic Acids Res* 46: e51. doi:10.1093/nar/gky067
108. Kirby MB, Medina-Cucurella AV, Baumer ZT, Whitehead TA (2021) *Protein Eng Des Sel* 34. doi:10.1093/protein/gzab017
109. Mighell TL, Toledano I, Lehner B (2023) *PLoS One* 18: e0288158. doi:10.1371/journal.pone.0288158
110. Belsare KD, Andorfer MC, Cardenas FS, Chael JR, Park HJ, Lewis JC (2017) *ACS Synth Biol* 6: 416. doi:10.1021/acssynbio.6b00297
111. Bloom JD (2014) *Mol Biol Evol* 31: 2753. doi:10.1093/molbev/msu220
112. Chung DH, Potter SC, Tanomrat AC, Ravikumar KM, Toney MD (2017) *Protein Eng Des Sel* 30: 347. doi:10.1093/protein/gzx013
113. Pullmann P, Ulpinnis C, Marillonnet S, Gruetzner R, Neumann S, Weissenborn MJ (2019) *Sci Rep* 9: 10932. doi:10.1038/s41598-019-47376-1
114. Hejlesen R, Fuchtbauer EM (2020) *Biotechniques* 68: 345. doi:10.2144/btn-2019-0104
115. She W, Ni J, Shui K, Wang F, He R, Xue J, Reetz MT, Li A, Ma L (2018) *ACS Synth Biol* 7: 2236. doi:10.1021/acssynbio.8b00245
116. Currin A, Kwok J, Sadler JC, Bell EL, Swainston N, Ababi M, Day P, Turner NJ, Kell DB (2019) *ACS Synth Biol* 8: 1371. doi:10.1021/acssynbio.9b00063
117. Reetz MT, Carballeira JD (2007) *Nat Protoc* 2: 891. doi:10.1038/nprot.2007.72
118. Wang X, Zheng K, Zheng H, Nie H, Yang Z, Tang L (2014) *J Biotechnol* 192 Pt A: 102. doi:10.1016/j.jbiotec.2014.10.023
119. Tang L, Gao H, Zhu X, Wang X, Zhou M, Jiang R (2012) *Biotechniques* 52: 149. doi:10.2144/000113820
120. Sun Z, Lonsdale R, Li G, Reetz MT (2016) *Chembiochem* 17: 1865. doi:10.1002/cbic.201600296
121. Yang J, Ducharme J, Johnston KE, Li FZ, Yue Y, Arnold FH (2023) *ACS Synth Biol* 12: 2444. doi:10.1021/acssynbio.3c00301
122. Hiraga K, Mejzlik P, Marcisin M, Vostrosablin N, Gromek A, Arnold J, Wiewiora S, Svarba R, Prihoda D, Clarova K, Klempir O, Navratil J, Tupa O, Vazquez-Otero A, Walas MW, Holy L, Spale M, Kotowski J, Dzamba D, Temesi G, Russell JH, Marshall NM, Murphy GS, Bitton DA (2021) *ACS Synth Biol* 10: 357. doi:10.1021/acssynbio.0c00542

123. Swainston N, Currin A, Day PJ, Kell DB (2014) *Nucleic Acids Res* 42: W395. doi:10.1093/nar/gku336
124. Swainston N, Currin A, Green L, Breitling R, Day PJ, Kell DB (2017) *PeerJ Comput. Sci.* 3: e120. doi:10.7717/peerj-cs.120
125. Kuiper BP, Prins RC, Billerbeck S (2022) *Chembiochem* 23: e202100507. doi:10.1002/cbic.202100507
126. Fowler DM, Fields S (2014) *Nat Methods* 11: 801. doi:10.1038/nmeth.3027
127. Kitzman JO, Starita LM, Lo RS, Fields S, Shendure J (2015) *Nat Methods* 12: 203. doi:10.1038/nmeth.3223
128. Medina-Cucurella AV, Steiner PJ, Faber MS, Beltran J, Borelli AN, Kirby MB, Cutler SR, Whitehead TA (2019) *Protein Eng Des Sel* 32: 41. doi:10.1093/protein/gzz013
129. Yu H, Ma S, Li Y, Dalby PA (2022) *Biotechnol Adv* 56: 107926. doi:10.1016/j.biotechadv.2022.107926
130. Zhou X, Gao L, Yang G, Liu D, Bai A, Li B, Deng Z, Feng Y (2015) *Chembiochem* 16: 455. doi:10.1002/cbic.201402456
131. Kumar JA, Sathish S, Prabu D, Renita AA, Saravanan A, Deivayanai VC, Anish M, Jayaprabakar J, Baigenzhenov O, Hosseini-Bandegharai A (2023) *Chemosphere* 331: 138680. doi:10.1016/j.chemosphere.2023.138680
132. Hu ML, Zha J, He LW, Lv YJ, Shen MH, Zhong C, Li BZ, Yuan YJ (2016) *Front Microbiol* 7: 241. doi:10.3389/fmicb.2016.00241
133. Kim H, Oh EJ, Lane ST, Lee WH, Cate JHD, Jin YS (2018) *J Biotechnol* 275: 53. doi:10.1016/j.jbiotec.2018.04.008
134. Lin R, Zhou Y, Yan T, Wang R, Li H, Wu Z, Zhang X, Zhou X, Zhao F, Zhang L, Li Y, Luo M (2022) *Nat Methods* 19: 976. doi:10.1038/s41592-022-01547-7
135. Kim E, Koo T, Park SW, Kim D, Kim K, Cho HY, Song DW, Lee KJ, Jung MH, Kim S, Kim JH, Kim JH, Kim JS (2017) *Nat Commun* 8: 14500. doi:10.1038/ncomms14500
136. Schmidheini L, Mathis N, Marquart KF, Rothgangl T, Kissling L, Bock D, Chanez C, Wang JP, Jinek M, Schwank G (2023) *Nat Chem Biol.* doi:10.1038/s41589-023-01427-x

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.