Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Modern methods for laboratory diversification of biomolecules

Sinisa Bratulic and Ahmed H Badran

Genetic variation fuels Darwinian evolution, yet spontaneous mutation rates are maintained at low levels to ensure cellular viability. Low mutation rates preclude the exhaustive exploration of sequence space for protein evolution and genome engineering applications, prompting scientists to develop methods for efficient and targeted diversification of nucleic acid sequences. Directed evolution of biomolecules relies upon the generation of unbiased genetic diversity to discover variants with desirable properties, whereas genome-engineering applications require selective modifications on a genomic scale with minimal off-targets. Here, we review the current toolkit of mutagenesis strategies employed in directed evolution and genome engineering. These state-of-the-art methods enable facile modifications and improvements of single genes, multicomponent pathways, and whole genomes for basic and applied research, while simultaneously paving the way for genome editing therapeutic interventions.

Address
Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

Corresponding author: Badran, Ahmed H (ahbadran@broadinstitute.org)

Current Opinion in Chemical Biology 2017, **41**:50–60

This review comes from a themed issue on **Mechanistic biology**

Edited by **Gregory A. Weiss**

For a complete overview see the Issue and the Editorial

Available online 2nd November 2017

http://dx.doi.org/10.1016/j.cbpa.2017.10.010

Introduction

Naturally occurring biomolecules are products of evolution in service of the survival of an organism, but they often lack the catalytic efficiency, specificity, or stability necessary for industrial or therapeutic applications. To improve these properties, proteins and other biopolymers are subjected to rounds of directed evolution, a powerful and flexible scheme to systematically endow biomolecules with desirable traits. This approach can employ a variety of mutagenesis strategies to modulate the frequency, distribution, and spectrum of mutations to explore biomolecule sequence landscapes, and applies selections or screens to identify and assess improved variants. The choice of diversification strategy is critical to the success of a directed evolution campaign, as the sequence landscape of a standard protein or biopolymer is typically too vast to be exhaustively searched [1]. Where *a priori* information is limited, unbiased and random *in vitro* [2,3] or *in vivo* [4–7] mutagenesis methods have successfully generated libraries of variants with improved or novel functionalities. Alternatively, bioinformatics, structural, or biochemical information can be leveraged to comprehensively explore a portion of the variant landscape by focusing mutagenesis on functionally relevant sites [8,9]. Finally, newly identified beneficial mutations can be isolated or integrated into single sequences using *in vitro* [10,11] or *in vivo* [12•,13] recombination methods. Recently, laboratory evolution has shifted to techniques that directly couple the diversification and assessment steps, providing the basis for continuous *in vivo* evolution strategies [14] that streamline previously lengthy experiments and minimize the need for human intervention.

Whereas these methods have proven crucial for studying structure-function relationships of single macromolecules [15], functional genomic screening and genome engineering applications typically require unbiased methods for *in situ* genome modification. Early methods to produce strain libraries relied on treatments with chemical mutagens/stressors [16] or transposon mutagenesis [17], and later integrated targeted methods employing homologous recombination capabilities, or *recombineering* (recombination-mediated genetic engineering) [18]. These approaches have been applied in both eukaryotes [19] and prokaryotes [20], and extended to enable *in vivo* continuous genome engineering [21••,22]. The recent discovery of CRISPR-Cas9 systems has reshaped this field, owing to their effectiveness as programmable nucleases or DNA-binding domains, and enabling novel, comparatively facile strategies for targeted diversification in cells [23].

In this review we focus on novel approaches for the diversification of biomolecules and generation of variant libraries, which underlie the application of evolutionary principles in molecular biology research and engineering. We first discuss untargeted mutagenesis methods that are commonly applied in generating diversity for directed protein evolution, and highlight novel methods that have been developed over the past decade. We transition to more targeted mechanisms of creating diversity, and address recent advances in targeted genome modifications with an emphasis on the latest developments in CRISPR-based systems.

Diversification methods for directed protein evolution and functional studies

Random mutagenesis

Natural mutation rates are low ($\sim 10^{-9}$ mutations/nucleotide/generation [24]) and, therefore, inappropriate for diversification of nucleic acid sequences in a laboratory setting. Laboratory evolution of biomolecules critically depends on elevated mutation rates for the discovery of improved or novel activities, mirroring biological principles that exist naturally (e.g. somatic hypermutation is employed to generate substantial antibody diversity [25]). Early protein evolution efforts catalyzed the development of approaches to increase mutation rates in a sequence-independent fashion to facilitate the unbiased construction of large and diverse gene libraries, spearheaded by techniques like error-prone PCR (epPCR). PCR protocols can be modified to reduce the fidelity of the reaction by modulating buffer composition [26] and dNTP ratios [3], introducing nucleoside analogues [2], using proofreading-deficient polymerases [27,28], or treating oligonucleotides with chemical mutagens [29]. While these various approaches increase overall mutation rates, the distribution of specific base changes that are generated can limit the chemical diversity of the resultant libraries. This distribution, called the mutational spectrum, cumulatively describes the efficiency and bias of sequence space exploration by a mutagenesis method. Despite widespread implementation, epPCR suffers from a bias in mutational spectrum (Table 1) to predominately incorporate transitions (A↔G or T↔C), yielding libraries enriched in synonymous mutations or conservative nonsynonymous mutations given the redundancy and assignments of the 64 natural codons. The sequence saturation mutagenesis (SeSaM) [30•] method was developed to specifically address this bias (Table 1), where the promiscuous base-pairing nucleotide inosine is enzymatically incorporated in the variant library and later replaced with canonical nucleotides through standard PCR. A recent improvement, SeSam-Tv-II [31••], increases the likelihood of consecutive mutations, especially double transversions (A/G↔C/T), thereby improving library quality and generating variants that are typically inaccessible by conventional epPCR [31••].

While epPCR-based methods introduce genetic variation primarily through point mutations, insertion and/or deletion (indels) of codons can also have considerable consequences on biomolecule function. These types of variants can be readily accessed using complementary methods such as TRINS [32], which incorporates short tandem repeats generated by rolling circle amplification into a target sequence. The resultant diversity is, however, limited to short sequence duplications rather than truly random insertions, with a significant fraction of the diversified pool encoding frameshifting insertions that can limit downstream discovery efforts (Table 1). It is possible to access in-frame deletions of multiple codons through Mu transposon mutagenesis [33], where the gene of interest bridges a TAT periplasm-directing signal and TEM-1 β-lactamase, ensuring that only an in-frame transposition event creates a functional TAT-β-lactamase product.

Conversely, approaches that do not rely on PCR can simplify library generation by minimizing researcher intervention. In error-prone rolling circle amplification (epRCA) [34], isothermal amplification and mutagenesis are coupled, and RCA-generated libraries can be directly transformed into *Escherichia coli* without further processing by restriction and ligation reactions (Table 1). RCA can also be combined with Kunkel mutagenesis [35] in selective RCA (sRCA) mutagenesis [36]. Plasmids are first produced from an *ung*$^-$ *dut*$^-$ *E. coli* strain to undergo non-specific uridylation (dT→dU), and subsequently amplified by PCR using mutagenic primers. Treatment with uracil-DNA glycosylase (UDG) creates abasic sites in the uracil-containing template, leaving only the mutagenized product for amplification by RCA. The sRCA approach increases effective library sizes and improves the mutagenesis efficiency by eliminating the non-mutated background sequences (Table 1).

Compared to *in vitro* mechanisms that require discreet manipulations to achieve the desired mutagenesis, *in vivo* mutagenesis approaches take inspiration from naturally occurring cellular, error-prone replication machinery. Early attempts at *in vivo* mutagenesis were inspired by SOS response and relied on its components [25], where mutator strains [4] enabled elevated mutagenesis *in vivo*, but to date these approaches have lacked mechanisms to control the resultant mutation rates and spectra (Table 1). In instances where unbiased, whole organism mutagenesis is desirable (e.g. for genome, plasmid, and viral evolution), a mutagenesis plasmid (MP) system encoding inducible dominant mutator alleles can be employed across variable *E. coli* strains. This *in vivo* mutagenesis approach was developed to enable control over a broad dynamic range of mutation rates concomitant with an unbiased mutational spectrum (Figure 1a) [6].

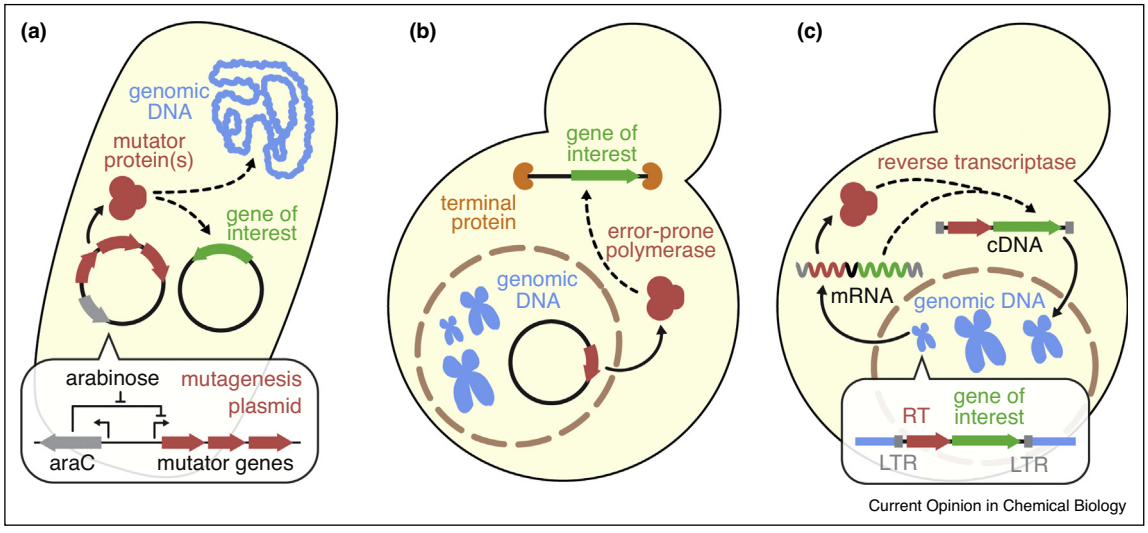
Traditional *in vivo* mutagenesis approaches can also indiscriminately mutagenize the host genome and accessory gene sequences beyond the target locus. This intrinsically limits the mutagenesis rate and may catalyze the appearance of undesirable 'cheaters' in selections or screens. To overcome this limitation, an error-prone Pol I polymerase system was developed to preferentially mutagenize plasmids bearing ColE1 and related origins of replication in *E. coli* over chromosomal sequences [5]. Similarly, a system based on an orthogonal plasmid and an error-prone polymerase pair was recently developed for *in vivo* continuous evolution in yeast [37••]. The orthogonality of mutagenesis in this approach is a significant improvement over previous *in vivo* methods (Table 1),

Table 1

Comparison of strategies for library diversification

Strategy		Method	Advantage	Disadvantage
Random mutagenesis	*In vitro*	epPCR [2,3,26–28]	Easy to implement, high mutation rates, no prior information required	Biased mutational spectra, laborious cloning procedures
		SeSaM, SeSaM-TV-II [30•,31••]	Unbiased mutational spectra, transversions and consecutive mutations accessible	Laborious, technically demanding
		epRCA [34], sRCA [36]	Easy to implement, efficient, random hexamers instead of primers, isothermal amplification, products can be directly transformed	Mutations introduced throughout the entire plasmid, plasmid multimers
		TRINS [32]	Expansion of sequence space sampling through insertion of short repeats	Out-of-frame insertions
	In vivo	Mutator strains [4]	Easy to implement	Low mutation rates, off-target mutagenesis
		Mutagenesis plasmid [6]	Inducible, broad mutational spectrum, portable across strains	Off-target mutagenesis
		Orthogonal plasmid/DNA polymerase [37••]	Mutagenesis constrained to the plasmid	>50 copies of the linear plasmid per cell (could affect selection strategies)
		Ty1 retrotransposon mutagenesis [7]	Mutagenesis constrained to the retrotransposon, capable of evolving longer sequences (up to 5 kb), broad host range (*K. lactis*, and different *S. cerevisiae* strains), large library sizes	Possible integration to off-target loci, relatively low mutation rate
Focused mutagenesis	*In vitro*	Cassette-mutagenesis [8,9]	Comprehensive sampling of variation for a small number of sites	Requires *a priori* structural/biochemical information
		PFunkel [42]	Multiple-site saturation possible	Technically demanding
		OmniChange [41]	Efficient saturation of up to five sites, omits restriction enzymes and ligases	Requires phosphorothiolated primers, library sizes not large enough for full saturation of five or more sites
		TrimerDimer [44]	Possible to saturate up to 4 consecutive positions, reduces screening effort by eliminating STOP and redundant codons	Technically demanding
		Small-intelligent libraries [45]	Eliminates STOP and redundant codons from libraries, flexible	Experimental distribution of amino-acid substitutions differs from the theoretical, limited primer length/ distance between saturated sites
		Systemic allelic series construction through reversibly terminated inosine [43•]	Facile generation of deep-mutational scanning libraries	Technically demanding
		Mutagenesis through massively parallel DNA synthesis [47••,49]	Low cost, single volume method to create libraries with one mutation per sequence	Technically demanding, limited length of the target sequence
Recombination	*In vitro*	DNA shuffling [10]	Rapid improvement of enzyme properties, flexible, robust shuffling of pre-existing variants	Homology requirement, laborious, high background
		STeP [11]	Easy to implement, single-tube reaction	Homology requirement, high background
		Template-change PCR [57]	Mimics *in vivo* recombination through single crossovers	Low mosaicism, homology requirement
		RACHITT [58]	Highly diverse libraries, low sequence identity required	Technically demanding and laborious (preparation of ssDNA templates)
		NRR [59]	Independent of homology	High number of inactive clones (insertions, deletions, rearrangements)
	In vivo	MORPHING [13]	Allows combinatorial targeted mutagenesis and assembly with a flexible number of crossovers	Laborious, requires ~50 bp overhangs for *in vivo* assembly
		Heritable recombination in yeast [12•]	Large library sizes, repeated diversification through recombination and rearrangement of beneficial mutations	*A priori* knowledge required for mutagenic cassette design, *in vitro* library generation

Figure 1



Methods for *in vivo* random mutagenesis. **(a)** A mutagenesis plasmid with arabinose-inducible mutator genes for *in vivo* directed evolution in *E. coli* [6]. This system enables tunable whole-genome mutagenesis, as well as the mutagenesis of sequences carried on plasmid or phage vectors. **(b)** An orthogonal system for mutagenesis in *S. cerevisiae* [37••]. The method uses a linear plasmid capped with P1 terminal proteins that recruit a specialized error-prone TP-DNA polymerase. Because the P1-capped plasmid is located in the cytoplasm, the host genome is not subjected to mutagenesis. **(c)** An *in vivo* mutagenesis system based on a synthetic Ty1 retrotransposon in *S. cerevisiae* [7]. The retrotransposon element, flanked by long terminal repeats (LTRs), is first transcribed, then reverse transcribed in an error-prone fashion, and finally re-integrated into the stable genomic locus.

as mutations predominantly accumulate in a targeted sequence (Figure 1b). Yet such directed approaches need not rely on error-prone polymerases. TagTEAM [38] is an innovative approach that uses a fusion protein encoding the yeast 3-methyladenine DNA glycosylase MAG1 and the tetR DNA-binding domain to enhance mutagenesis to regions containing arrays of tetO sites.

Beyond such artificial systems, biological systems can naturally exploit elevated mutagenesis to generate diversity and ensure that at least a fraction of the population can overcome challenges imposed by the surrounding environment. For example, error-prone reverse transcription, which drives natural variation during retroviral replication, plays critical roles in facilitating viral escape from immune system surveillance and the discovery of antiviral drug resistance [39]. To extend this capability to laboratory evolution, a novel retrotransposon-dependent *in vivo* mutagenesis strategy was developed to enable robust continuous evolution schemes [7]. This strategy implements the yeast retrotransposon Ty1 to generate gene libraries, where the evolving target sequence is transcribed using host machinery, and then reverse transcribed by the error-prone Ty1 enzyme (Figure 1c, Table 1). The mutagenized cDNA is integrated into a stable genomic locus, at which point the biomolecule variant can be assessed by functional selection or screen. Related mutagenesis systems are also known occur naturally, such as in the *Bordetella bronchiseptica* bacteriophage retroelement, which is comprised of an error-prone reverse transcriptase that selectively mutates codons of a key surface protein that determines host specificity [40]. This natural phage-display system has been coopted for use in the laboratory to evolve high-affinity T4 lysozyme binders, relying exclusively on the natural diversity generated by the retroelement [40].

Focused mutagenesis

In order to effectively explore the vast sequence landscape of the average protein, structural and biochemical information can be leveraged to confine genetic variation to one or multiple sites implicated in a property of interest. Typically, the selected positions are randomized using degenerate oligonucleotide codon sequences (NNN, NNB, NNK, etc.) [8]. Whereas improved variants may be discovered in libraries where a single position has been subjected to saturation mutagenesis, it is often necessary to investigate multiple sites in concert to access significant improvements in the desired activity. While site-directed mutagenesis methods can, in principle, access more than ten sites in a gene of interest [9] in a single reaction, the efficient simultaneous evaluation of saturation mutagenesis at multiple sites across a gene requires a different approach to library construction. One such method, OmniChange [41], facilitates site-saturation mutagenesis at up to five codons independent of their location in the target sequence. This method relies on a single PCR step with phosphorothioate primers to

generate modified DNA fragments, which are combinatorially assembled and transformed without requiring enzyme-based cloning steps (Table 1).

Site-directed mutagenesis strategies can be extended to generate comprehensive libraries for studying sequence-function relationships. PFunkel [42] combines traditional Kunkel mutagenesis with a *Pfu* polymerase and thermostable ligase to mutagenize a DNA template using oligo libraries. In one application, PFunkel afforded a comprehensive codon mutagenesis library for the full-length TEM-1 gene (287 codons). A novel strategy for creating systematic allelic series (SAS) [43•] relies on reversible termination during linear amplification of a template using a capped inosine triphosphate (rtITP). Terminated amplification products are then isolated, and rtITP is non-enzymatically liberated. Finally, terminated products are extended and amplified in a PCR reaction that introduces dNTPs at sites where inosines were incorporated. This enables the unbiased introduction of only a single mutation per molecule following amplification by PCR (Table 1).

Complementary approaches to improve focused mutagenesis library quality rely on the elimination of redundant and premature stop codons. One approach to mitigate codon redundancy, TrimerDimer [44], introduces codons into degenerate primers via chemical synthesis and can focus site-saturation mutagenesis at multiple consecutive positions. While this reduces the total number of requisite primers for site saturation mutagenesis, TrimerDrimer might not be readily accessible to the majority of researchers (Table 1). An alternative approach integrates bioinformatics analyses to design optimized sets of degenerate primers, yielding primer mixtures that contain exactly one codon per amino acid [45]. These primer mixtures can be used in conjunction with a preferred mutagenesis method to generate 'small-intelligent' mutagenesis libraries where each protein variant is represented by a single gene, thereby minimizing genotypic redundancy and reducing unnecessary screening load (Table 1).

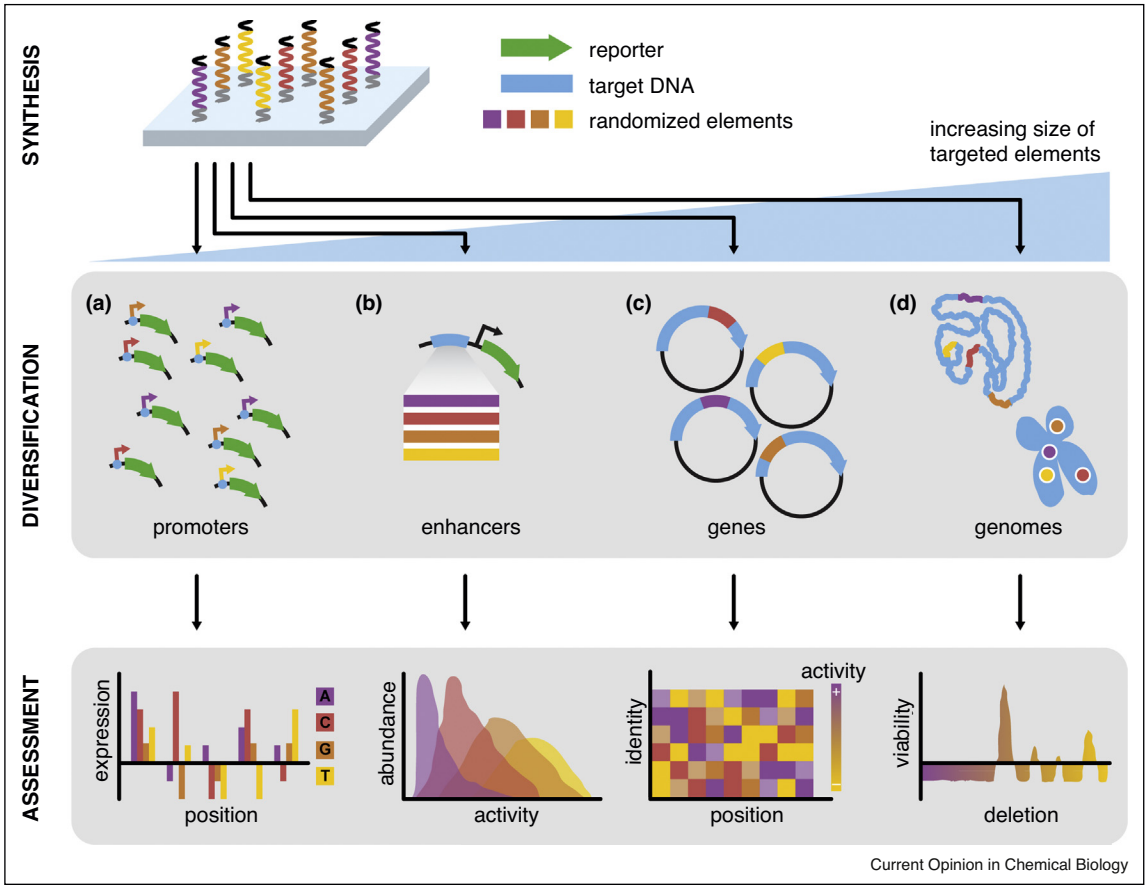
Emerging technologies for massively parallel DNA synthesis [46] have extended these approaches to functionally probe large libraries of sequence variants (Table 1). It is now possible to implement deep mutagenesis scanning of every position in a given sequence through a single experiment. In one demonstration, synthetic saturation mutagenesis [47••] afforded a library of 10^9 members covering every possible nucleotide variant of a 35 base pair promoter sequence (Figure 2a). Such microarray-based DNA synthesis can also be extended to assess putative regulatory elements in genomic and episomal contexts using massively parallel reporter assays (Figure 2b) [48,49]. Furthermore, deep mutational scanning is not restricted to short regulatory sequences, where microarray-based synthesis of mutagenic primers has been used to create libraries containing >93% of every missense mutation in the transcription factors Gal4 and p53 (Figure 2c) [50]. Finally, libraries of guide RNA pairs (see the 'CRISPR-Cas9 based methods' section) can be readily synthesized using this technology. In one implementation, a tiling deletion scan (ScanDel) [49] approach was used to interrogate loss-of-function mutations in the *HPRT1* genomic locus (Figure 2d). Similar approaches have been used to investigate the functional effects of variation in promoters, enhancers, splice sites, and proteins with single-nucleotide resolution [51].

Recombination

Recombination enables the combinatorial rearrangement of mutations between closely related nucleic acid sequences. While its origins and its impact on diversity and adaptation are still a topic of investigation [52], recombination has been observed in numerous biological contexts, ranging from mating type switching in *Saccharomyces cerevisiae* [53] to antigenic variation that enables pathogens to evade host immune responses [54]. In the laboratory, recombination counters the loss of heterogeneity that regularly occurs during the assessment stage of directed evolution schemes, where competing beneficial mutations can potentially drive each other to extinction. It is worth mentioning that even when recombination is not explicitly sought or desired in a diversification strategy, some experimental approaches, like epPCR [55] or *in vivo* mutagenesis [56], can inadvertently lead to recombination.

Furthermore, recombination techniques can be harnessed to create novel combinations of beneficial and/or epistatic mutations. The first *in vitro* recombination protocol, DNA shuffling [10], involved fragmentation and subsequent PCR-based reassembly of random DNA fragments (Figure 3a). Subsequent methods simplified DNA shuffling by eliminating the laborious DNA fragmentation step (Table 1) [10]. A simple PCR-based staggered extension and template switching [11] or template-change PCR [57] allows facile generation of chimeric gene sequences. A major limitation of traditional DNA shuffling methods has been the requirement for high sequence similarity between library members (Table 1). Subsequent *in vitro* recombination approaches that reduce [58] or completely eliminate [59] the sequence similarity requirement have been devised to overcome this drawback. For example, RACHITT [58] bypasses the requirement for high homology between parental fragments by employing transient DNA templates which serve as hybridization scaffolds prior to full length chimera assembly. This approach generates chimeric recombination products with multiple crossover points and virtually no parental background sequences, but requires extensive preparation of ssDNA template preparation for scaffolding (Table 1).

Figure 2



Microarray-based saturation mutagenesis methods. DNA synthesis on microarrays allows for rapid, cost-effective production of saturation mutagenesis libraries, enabling the production of extensive libraries of **(a)** promoters [47••], **(b)** enhancers [48], **(c)** genes [49], and even (**d**) whole genomes [50]. In many instances, libraries of near-complete coverage of single mutant [47••] or single codon [49] sequence space facilitate the robust interrogation of sequence-function relationships. Following functional assessment of the generated variants using one of a host of *in vitro* or *in vivo* assays, high-throughput sequencing can inform fitness levels of the generated mutants in exquisite detail.

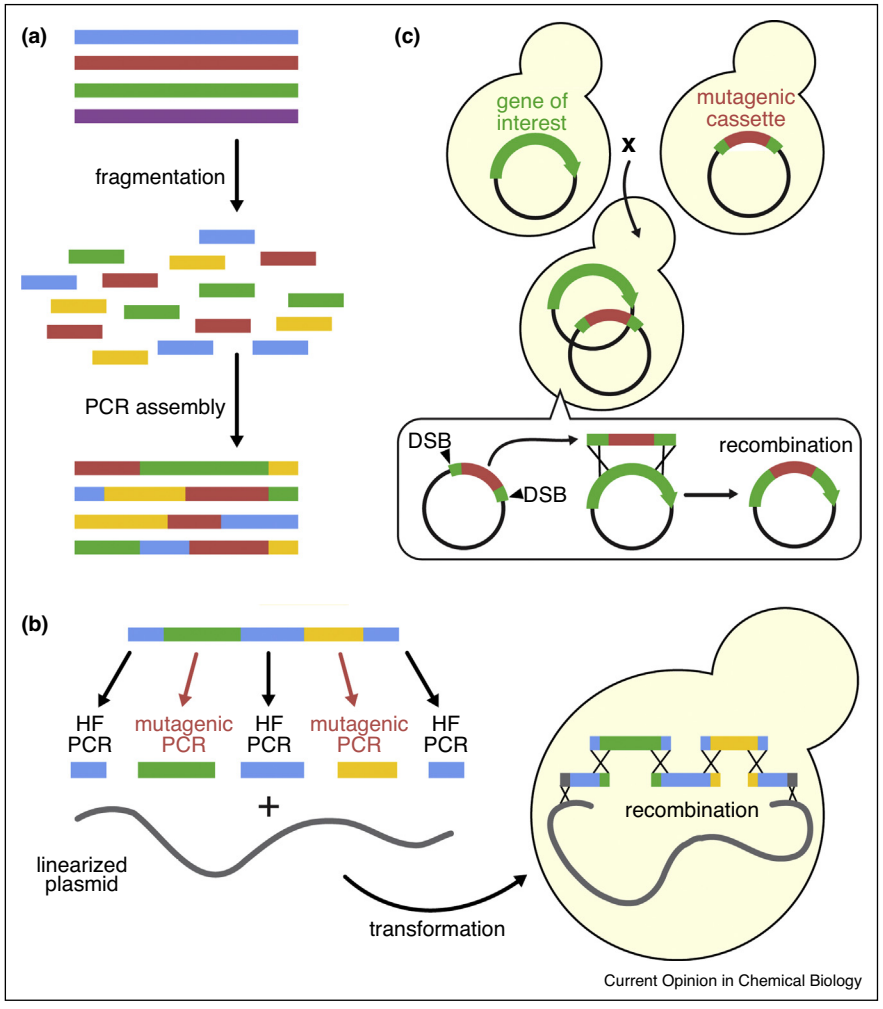
Current recombination approaches are shifting from *in vitro* to *in vivo* methods that enable continuous evolution. Homologous recombination (HR) natively occurs with high-efficiency in *S. cerevisiae*, where PCR-diversified fragments can be assembled into chimeric variants by Mutagenic Organized Recombination Process by Homologous *IN vivo* Grouping (MORPHING) [13] (Figure 3b). It should be noted that prior bioinformatics, structural, or biochemical analysis can be leveraged in this method to target diversification to elements that may be tolerant to modification and limit changes to alternate regions (Table 1). It is also possible to leverage HR in diversification by mutagenesis, as was recently demonstrated using a heritable recombination system in *S. cerevisiae* [12•], generating libraries with up to 10^{13} diverse sequences (Figure 3c). This approach relies on the *in situ* production of double-stranded donor templates that have been mutagenized to saturation at defined sites. Similar techniques could also be implemented to enable library recombination in prokaryotes using *in vivo* assembly (IVA) [60], which was developed for insertions, deletions, and mutagenesis in *E. coli*.

Mutagenesis methods for genome engineering and genome-wide screening

Recombineering methods

Recombineering, genome engineering using recombinase enzymes, has been the method of choice for modifying bacterial and yeast genomes over the past two decades. Bacterial recombineering exploits components derived from λ phage to integrate nucleic acids into the host genome [20], and facilitates the introduction of point mutations, deletions, insertions, and rearrangements into genomic or extrachromosomal DNA [18,20]. Conversely, recombineering in *S. cerevisiae* relies on the endogenous machinery for HR. A highly efficient ssDNA-mediated recombineering approach has been successfully applied

Figure 3



Methods for *in vitro* and *in vivo* recombination. **(a)** DNA shuffling [10], an *in vitro* recombination method, relies on fragmenting a pool of parental sequences, and uses PCR to assemble full-length chimeras. **(b)** *MORPHING* [13] is a method that combines *in vitro* mutagenesis and *in vivo* recombination in yeast for enzyme evolution. **(c)** An *in vivo* recombination method [12•] in yeast that uses homologous recombination for cassette mutagenesis of a target sequence.

to *S. cerevisiae* [61], enabling modification and assembly of gene pathways.

Efforts to scale up recombineering to the entire bacterial genome motivated the development of multiplex automated genome engineering (MAGE) [21••], an efficient method for genome-wide ssDNA-mediated allelic replacements. MAGE has been successfully integrated with conjugative assembly genome engineering (CAGE) to replace all 314 TAG stop codons in *E. coli* with the synonymous TAA codon [62•]. This recoded organism was later used as a chassis to generate a library of chromosomally integrated aminoacyl-tRNA synthetases, and select for variants with enhanced ability for multi-site incorporation of nonstandard amino acids [63]. However, MAGE requires hosts deficient in methyl-directed mismatch repair (MMR), often resulting in substantial off-target mutagenesis [62•]. A recently developed method called portMAGE reduces off-target mutagenesis by using a temperature-inducible MMR mutant, thereby temporally limiting mutagenesis [22]. Furthermore, portMAGE is transferable across hosts, permitting the genome modification of diverse bacterial species.

While *E. coli* is a preferred bacterial host for many laboratory applications, the utility of recombineering approaches has been realized beyond this workhorse bacterium. Using an iterative recombineering approach called SIRCAS (stepwise integration of rolling circle amplified segments) [64], the *Salmonella typhimurium* genome was extensively modified in the largest genomic recoding effort to date, enabling 1557 synonymous

leucine substitutions. In addition to whole genome mutagenesis through recombineering, the flexibility of this oligo-based approach was recently demonstrated in strategy that extended its applications to directed protein evolution [65]. As a proof of concept, the authors used plasmid recombineering to subject the 110-residue iLOV protein to saturation mutagenesis with near complete coverage (99.8% of all mutations), uncovering many variants with improved thermostability.

CRISPR-Cas9 based methods

Numerous bacterial and archaeal species carry adaptive immunity mechanisms based on clustered regularly interspaced short palindromic repeats (CRISPR) and associated Cas proteins to protect against foreign nucleic acids [66]. The elucidation of CRISPR-Cas biology has rendered this system convenient for engineering of prokaryotic [67], archaeal [68], and eukaryotic [23] genomes. The simplest and most commonly used implementation of CRISPR-Cas systems consists of only two components: an endonuclease (e.g. Cas9) capable of generating double-stranded breaks (DSBs) in DNA targeted by a single-guide RNA (sgRNA) [23,66,69•]. Cas9-mediated DSBs are repaired via non-homologous end-joining (NHEJ), typically resulting in small indels and inactivation of the target sequence. In addition, it has been used to introduce large genomic deletions and catalyze chromosomal rearrangements [70].

Homology-directed repair (HDR) pathways can alternatively address DSBs following Cas9-mediated cleavage, whereby a donor DNA molecule encoding homologous sequences flanking the targeted site is used as a template. HDR enables more precise control over outcomes of Cas9-mediated genome editing, and has been successfully used to introduce modifications in genomes of human [69•] and other eukaryotic [23] cells. However, early attempts to use CRISPR-Cas systems for genome engineering revealed two limitations to this approach: insufficient Cas9 specificity for the target sequence, leading to off-target mutations [71], and poor efficiency of HDR repair relative to that of NHEJ [69•]. Yet the ability of Cas9 and homologous nucleases to generate DSBs in an sgRNA-defined manner *in situ* has propelled the recombineering field forward by eliminating three limitations of the native recombineering methodology. First, Cas9 nuclease can be used for counter-selection against cheaters (through cleavage of undesirable variants) [72•]. Second, CRISPR-Cas9-mediated recombineering can be used to efficiently generate donor templates [72•]. Finally, unlike early recombineering protocols, Cas9-assisted recombineering can modify genomes in a scarless fashion [73], obviating potentially adverse effects of genomic scars.

CRISPR-Cas systems have been extended to enable forward genetic screening to comprehensively annotate functional genomic elements. A genome-scale CRISPR-Cas knockout screen targeting more than 18,000 human genes [74••] revealed genes essential for cell viability of melanoma and stem cell lines, and identified genes whose loss is implicated in resistance to vemurafenib. In a more focused approach, a similar study probed 291 human genes to discover host genes responsible for susceptibility to diphtheria and anthrax toxins [75].

To expand the possible modifications accessible using CRISPR-Cas9, additional enzymes have been covalently appended to a nuclease-deficient Cas9 (dCas9), exploiting its RNA-programmable DNA-binding activity to recruit or directly present fused enzymes to the targeted nucleic acid [23]. These tools have been extended to the realm of directed evolution through engineered deaminase-Cas9 systems to evolve known variants of GFP [76], identify novel mutants conferring resistance to bortezomib [76] and imatinib [77], and to enable robust RNA-programmed recombination through a recombinase-dCas9 fusion [78].

Conclusion

Diversification methods for directed evolution and genome modification have rapidly improved over the past decade. Random mutagenesis methods can now reliably generate chemically diverse libraries with near-ideal mutational spectra and at high mutational frequencies. Novel focused mutagenesis methods readily achieve simultaneous saturation of numerous noncontiguous sites. Advances in DNA synthesis now allow for the exhaustive study of sequence-function relationships in the single-mutant neighborhood of a protein-coding sequence. For applications that require exploration of a broader sequence landscape, recombination methods can provide libraries with a defined number of cross-overs, independent of sequence similarity between parental sequences. And continuous *in vivo* methods have provided novel avenues of biomolecule evolution that seek to limit researcher intervention between rounds of diversification and assessment.

While the early strain and genome modification methods paralleled the development of gene diversification strategies, the field was hampered by the lack of generalized tools for high-throughput *in vivo* targeted modification. This changed when recombineering set in motion the extension of directed evolution approaches to genome-scale modification for prokaryotes and yeast. However, it was not until the discovery of RNA-guided endonucleases that genomes of other eukaryotes could be modified with equal ease. Recent advances exemplify the wide applicability of CRISPR-Cas9 as programmable DNA-binding domains to engineer tools for targeted mutagenesis and synthetic biology directly in the native genomic context. The staggering rate at which new approaches are developed in this field undoubtedly ensures that

CRISPR-Cas9 will become a routine genome editing approach, not just in basic science and industrial biotechnology, but in applied biomedical research as well.

Acknowledgements

This work was generously supported by the Broad Institute of MIT & Harvard and the National Institutes of Health Early Investigator Award DP5-OD-024590.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Ruff AJ, Dennig A, Schwaneberg U: **To get what we aim for – progress in diversity generation methods**. *FEBS J* 2013, **280**:2961-2978.
2. Zaccolo M, Williams DM, Brown DM, Gherardi E: **An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues**. *J Mol Biol* 1996, **255**: 589-603.
3. Cadwell RC, Joyce GF: **Randomization of genes by PCR mutagenesis**. *Genome Res* 1992, **2**:28-33.
4. Greener A, Callahan M, Jerpseth B: **An efficient random mutagenesis technique using an *E. coli* mutator strain**. *Mol Biotechnol* 1997, **7**:189-195.
5. Camps M, Naukkarinen J, Johnson BP, Loeb LA: **Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I**. *Proc Natl Acad Sci U S A* 2003, **100**:9727-9732.
6. Badran AH, Liu DR: **Development of potent in vivo mutagenesis plasmids with broad mutational spectra**. *Nat Commun* 2015, **6**:8425.
7. Crook N, Abatemarco J, Sun J, Wagner JM, Schmitz A, Alper HS: **In vivo continuous evolution of genes and pathways in yeast**. *Nat Commun* 2016, **7**:13051.
8. Wells JA, Vasser M, Powers DB: **Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites**. *Gene* 1985, **34**:315-323.
9. Seyfang A, Huaqian Jin J: **Multiple site-directed mutagenesis of more than 10 sites simultaneously and in a single round**. *Anal Biochem* 2004, **324**:285-291.
10. Stemmer WPC: **Rapid evolution of a protein in vitro by DNA shuffling**. *Nature* 1994, **370**:389-391.
11. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH: **Molecular evolution by staggered extension process (StEP) in vitro recombination**. *Nat Biotechnol* 1998, **16**:258-261.
12. • Romanini DW, Peralta-Yahya P, Mondol V, Cornish VW: **A heritable recombination system for synthetic Darwinian evolution in yeast**. *ACS Synth Biol* 2012, **1**:602-609.

This study describes a system for *in vivo* recombination and mutagenesis that can be used for continuous evolution of enzymes in yeast.

13. Gonzalez-Perez D, Molina-Espeja P, Garcia-Ruiz E, Alcalde M: **Mutagenic Organized Recombination Process by Homologous In vivo Grouping (MORPHING) for directed enzyme evolution**. *PLOS ONE* 2014, **9**:1-12.
14. Badran AH, Liu DR: **In vivo continuous directed evolution**. *Curr Opin Chem Biol* 2015, **24**:1-10.
15. Dean AM, Thornton JW: **Mechanistic approaches to the study of evolution: the functional synthesis**. *Nat Rev Genet* 2007, **8**:675-688.
16. Dragosits M, Mattanovich D: **Adaptive laboratory evolution – principles and applications for biotechnology**. *Microb Cell Factories* 2013, **12**:64.
17. Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC: **Global transposon mutagenesis and a minimal Mycoplasma genome**. *Science* 1999, **286**:2165-2169.
18. Sharan SK, Thomason LC, Kuznetsov SG, Court DL: **Recombineering: a homologous recombination-based method of genetic engineering**. *Nat Protoc* 2009, **4**:206-223.
19. Orr-Weaver TL, Szostak JW, Rothstein RJ: **Genetic applications of yeast transformation with linear and gapped plasmids**. *Methods Enzymol* 1983, **101**:228-245.
20. Datsenko KA, Wanner BL: **One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products**. *Proc Natl Acad Sci USA* 2000, **97**:6640-6645.
21. •• Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM: **Programming cells by multiplex genome engineering and accelerated evolution**. *Nature* 2009, **460**: 894-898.

The authors describe Multiplex Automated Genome Engineering (MAGE), a streamlined recombineering methodology for recursive genome modification in *E. coli* across multiple genomic loci.

22. Nyerges A, Csorgo B, Nagy I, Balint B, Bihari P, Lazar V, Apjok G, Umenhoffer K, Bogos B, Posfai G *et al.*: **A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species**. *Proc Natl Acad Sci U S A* 2016, **113**:2502-2507.
23. Komor AC, Badran AH, Liu DR: **CRISPR-based technologies for the manipulation of eukaryotic genomes**. *Cell* 2016, **168**:1-17.
24. Lynch M: **Evolution of the mutation rate**. *Trends Genet* 2010, **26**:345-352.
25. Goodman MF: **Better living with hyper-mutation**. *Environ Mol Mutagen* 2016, **57**:421-434.
26. Lin-Goerke JL, Robbins DJ, Burczak JD: **PCR-based random mutagenesis using manganese and reduced dNTP concentration**. *BioTechniques* 1997, **23**:409-412.
27. Mondon P, Grand D, Souyris N, Emond S, Bouayadi K, Kharrat H: In *Mutagen: A Random Mutagenesis Method Providing a Complementary Diversity Generated by Human Error-Prone DNA Polymerases*. Edited by Braman J. Humana Press; 2010:373-386.
28. Vanhercke T, Ampe C, Tirry L, Denolf P: **Reducing mutational bias in random protein libraries**. *Anal Biochem* 2005, **339**:9-14.
29. Lai YP, Huang J, Wang LF, Li J, Wu ZR: **A new approach to random mutagenesis in vitro**. *Biotechnol Bioeng* 2004, **86**: 622-627.
30. • Wong TS, Tee KL, Hauer B, Schwaneberg U: **Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution**. *Nucleic Acids Res* 2004, **32**:e26.

See annotation to Ref. [31••].

31. •• Mundhada H, Marienhagen J, Scacioc A, Schenk A, Roccatano D, Schwaneberg U: **SeSaM-Tv-II generates a protein sequence space that is unobtainable by epPCR**. *ChemBioChem* 2011, **12**:1595-1601.

This study describes a random mutagenesis method capable of yielding a balanced mutational spectrum, thereby eliminating biases inherent in PCR-based mutagenesis protocols and increasing the chemical diversity of gene libraries.

32. Kipnis Y, Dellus-Gur E, Tawfik DS: **TRINS: a method for gene modification by randomized tandem repeat insertions**. *Protein Eng Des Sel* 2012, **25**:437-444.
33. Liu SS, Wei X, Ji Q, Xin X, Jiang B, Liu J: **A facile and efficient transposon mutagenesis method for generation of multi-codon deletions in protein sequences**. *J Biotechnol* 2016, **227**:27-34.
34. Fujii R, Kitaoka M, Hayashi K: **One-step random mutagenesis by error-prone rolling circle amplification**. *Nucleic Acids Res* 2004, **32**:e145.
35. Kunkel TA: **Rapid and efficient site-specific mutagenesis without phenotypic selection**. *Proc Natl Acad Sci U S A* 1985, **82**:488-492.

36. Huovinen T, Brockmann EC, Akter S, Perez-Gamarra S, Ylä-Pelto J, Liu Y, Lamminmäki U: **Primer extension mutagenesis powered by selective rolling circle amplification**. *PLoS ONE* 2012, **7**.

37. •• Ravikumar A, Arrieta A, Liu CC: **An orthogonal DNA replication system in yeast**. *Nat Chem Biol* 2014, **10**:175-177.
The authors developed an error-prone orthogonal DNA replication system for continuous evolution in the yeast *S. cerevisiae*. This method obviates off-target mutagenesis of the host genome by selectively replicating a plasmid bearing the evolving gene of interest in an error-prone fashion.

38. Finney-Manchester SP, Maheshri N: **Harnessing mutagenic homologous recombination for targeted mutagenesis in vivo by TaGTEAM**. *Nucleic Acids Res* 2013, **41**:e99.

39. Cuevas JM, Geller R, Garijo R, Lopez-Aldeguer J, Sanjuan R: **Extremely high mutation rate of HIV-1 in vivo**. *PLoS Biol* 2015, **13**:e1002251.

40. Yuan TZ, Overstreet CM, Moody IS, Weiss GA: **Protein engineering with biosynthesized libraries from *Bordetella bronchiseptica* bacteriophage**. *PLoS ONE* 2013, **8**:e55617.

41. Dennig A, Shivange AV, Marienhagen J, Schwaneberg U: **Omnichange: the sequence independent method for simultaneous site-saturation of five codons**. *PLoS ONE* 2011, **6**.

42. Firnberg E, Ostermeier M: **PFunkel: efficient, expansive, user-defined mutagenesis**. *PLoS ONE* 2012, **7**:e52031.

43. • Haller G, Alvarado D, McCall K, Mitra RD, Dobbs MB, Gurnett CA: **Massively parallel single-nucleotide mutagenesis using reversibly terminated inosine**. *Nat Methods* 2016, **13**:923-925.
A mutagenesis method based on using rtITP to generate arrays of alleles with single-nucleotide mutations for comprehensive investigation of mutational effects in a gene of interest.

44. Gaytán P, Contreras-Zambrano C, Ortiz-Alvarado M, Morales-Pablos A, Yáñez J: **TrimerDimer: an oligonucleotide-based saturation mutagenesis approach that removes redundant and stop codons**. *Nucleic Acids Res* 2009, **37**.

45. Tang L, Gao H, Zhu X, Wang X, Zhou M, Jiang R: **Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers**. *BioTechniques* 2012, **52**:149-158.

46. Cleary MA, Kilian K, Wang Y, Bradshaw J, Cavet G, Ge W, Kulkarni A, Paddison PJ, Chang K, Sheth N *et al.*: **Production of complex nucleic acid libraries using highly parallel in situ oligonucleotide synthesis**. *Nat Methods* 2004, **1**:241-248.

47. •• Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J: **High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis**. *Nat Biotechnol* 2009, **27**:1173-1175.
The authors describe a microarray-based method for multiplexed analysis of mutational effects in non-coding elements through comprehensive assessment of all possible single-nucleotide mutations in prokaryotic and eukaryotic promoters.

48. Inoue F, Kircher M, Martin B, Cooper GM, Witten DM, McManus MT, Ahituv N, Shendure J: **A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity**. *Genome Res* 2017, **27**:38-52.

49. Gasperini M, Findlay GM, McKenna A, Milbank JH, Lee C, Zhang MD, Cusanovich DA, Shendure J: **CRISPR/Cas9-mediated scanning for regulatory elements required for HPRT1 expression via thousands of large, programmed genomic deletions**. *Am J Hum Genet* 2017, **101**:192-205.

50. Kitzman JO, Starita LM, Lo RS, Fields S, Shendure J: **Massively parallel single-amino-acid mutagenesis**. *Nat Methods* 2015, **12**:203-206.

51. Gasperini M, Starita L, Shendure J: **The power of multiplexed functional analysis of genetic variants**. *Nat Protoc* 2016, **11**:1782-1787.

52. Pesce D, Lehman N, de Visser JA: **Sex in a test tube: testing the benefits of in vitro recombination**. *Philos Trans R Soc Lond B Biol Sci* 2016, **371**.

53. Haber JE: **Mating-type genes and MAT switching in *Saccharomyces cerevisiae***. *Genetics* 2012, **191**:33-64.

54. Vink C, Rudenko G, Seifert HS: **Microbial antigenic variation mediated by homologous DNA recombination**. *FEMS Microbiol Rev* 2012, **36**:917-948.

55. Yu W, Rusterholtz KJ, Krummel AT, Lehman N: **Detection of high levels of recombination generated during PCR amplification of RNA templates**. *Biotechniques* 2006, **40**:499-507.

56. Badran AH, Guzov VM, Huai Q, Kemp MM, Vishwanath P, Kain W, Nance AM, Evdokimov A, Moshiri F, Turner KH *et al.*: **Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance**. *Nature* 2016, **533**:58-63.

57. Shu C, Zhou J, Crickmore N, Li X, Song F, Liang G, He K, Huang D, Zhang J: **In vitro template-change PCR to create single crossover libraries: a case study with *B. thuringiensis* Cry2A toxins**. *Sci Rep* 2016, **6**:23536.

58. Coco WM, Levinson WE, Crist MJ, Hektor HJ, Darzins A, Pienkos PT, Squires CH, Monticello DJ: **DNA shuffling method for generating highly recombined genes and evolved enzymes**. *Nat Biotechnol* 2001, **19**:354-359.

59. Bittker JA, Le BV, Liu JM, Liu DR: **Directed evolution of protein enzymes using nonhomologous random recombination**. *Proc Natl Acad Sci U S A* 2004, **101**:7011-7016.

60. García-Nafría J, Watson JF, Greger IH: **IVA cloning: a single-tube universal cloning system exploiting bacterial in vivo assembly**. *Sci Rep* 2016, **6**:27459.

61. DiCarlo JE, Conley AJ, Penttilä M, Jäntti J, Wang HH, Church GM: **Yeast oligo-mediated genome engineering (YOGE)**. *ACS Synth Biol* 2013, **2**:741-749.

62. • Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB *et al.*: **Precise manipulation of chromosomes in vivo enables genome-wide codon replacement**. *Science* 2011, **333**:348-353.
This study describes the integration of multiplex automated genome engineering (MAGE) and conjugative assembly genome engineering (CAGE) methods to replace all TAG codons with synonymous TAA stop codons in a single *E. coli* genome.

63. Amiram M, Haimovich AD, Fan C, Wang Y-S, Aerni H-R, Ntai I, Moonan DW, Ma NJ, Rovner AJ, Hong SH *et al.*: **Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids**. *Nat Biotechnol* 2015, **33**:1272-1279.

64. Lau YH, Stirling F, Kuo J, Karrenbelt MAP, Chan YA, Riesselman A, Horton CA, Schafer E, Lips D, Weinstock MT *et al.*: **Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA**. *Nucleic Acids Res* 2017, **45**:6971-6980.

65. Higgins SA, Ouonkap SVY, Savage DF: **Rapid and programmable protein mutagenesis using plasmid recombineering**. *ACS Synth Biol* 2017 http://dx.doi.org/10.1021/acssynbio.7b00112.

66. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: **A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity**. *Science* 2012, **337**:816-821.

67. Mougiakos I, Bosma EF, de Vos WM, van Kranenburg R, van der Oost J: **Next generation prokaryotic engineering: the CRISPR-Cas toolkit**. *Trends Biotechnol* 2016, **34**:575-587.

68. Nayak DD, Metcalf WW: **Cas9-mediated genome editing in the methanogenic archaeon *Methanosarcina acetivorans***. *Proc Natl Acad Sci U S A* 2017, **114**:201618596.

69. • Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM: **RNA-guided human genome engineering via Cas9**. *Science* 2013, **339**:823-826.
The authors showcase that RNA-programmable Cas9 can be used for efficient editing of mammalian genomes.

70. Xiao A, Wang Z, Hu Y, Wu Y, Luo Z, Yang Z, Zu Y, Li W, Huang P, Tong X *et al.*: **Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish**. *Nucleic Acids Res* 2013, **41**:1-11.

71. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK: **Improving CRISPR-Cas nuclease specificity using truncated guide RNAs**. *Nat Biotechnol* 2014, **32**:279-284.

72. • Oh JH, Van Pijkeren JP: **CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri***. *Nucleic Acids Res* 2014, **42**:1-11.
The authors combine CRISPR-Cas with ssDNA recombineering to increase the efficiency of genome editing in lactic acid bacteria.

73. Reisch CR, Prather KLJ: **The no-SCAR (scarless Cas9 assisted recombineering) system for genome editing in *Escherichia coli***. *Sci Rep* 2015, **5**:15096.

74. •• Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, Heckl D, Ebert BL, Root DE, Doench JG *et al.*: **Genome-scale CRISPR-Cas9 knockout screening in human cells**. *Science* 2014, **343**:84-87.
The authors establish a method for genome-scale functional screening using a lentiviral sgRNA library targeting more than 18,000 human genes.

75. Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, Wei W: **High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells**. *Nature* 2014, **509**:487-491.

76. Hess GT, Frésard L, Han K, Lee CH, Li A, Cimprich KA, Montgomery SB, Bassik MC: **Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells**. *Nat Methods* 2016, **13**:1036-1042.

77. Ma Y, Zhang J, Yin W, Zhang Z, Song Y, Chang X: **Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells**. *Nat Methods* 2016, **13**:1-9.

78. Chaikind B, Bessen JL, Thompson DB, Hu JH, Liu DR: **A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells**. *Nucleic Acids Res* 2016, **44**:9758-9770.