Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Targeted mutagenesis of multiple chromosomal regions in microbes

Bálint Csörgő[1,2], Akos Nyerges[3,4] and Csaba Pál[3]

Directed evolution allows the effective engineering of proteins, biosynthetic pathways, and cellular functions. Traditional plasmid-based methods generally subject one or occasionally multiple genes-of-interest to mutagenesis, require time-consuming manual interventions, and the genes that are subjected to mutagenesis are outside of their native genomic context. Other methods mutagenize the whole genome unselectively which may distort the outcome. Recent recombineering- and CRISPR-based technologies radically change this field by allowing exceedingly high mutation rates at multiple, predefined loci in their native genomic context. In this review, we focus on recent technologies that potentially allow accelerated tunable mutagenesis at multiple genomic loci in the native genomic context of these target sequences. These technologies will be compared by four main criteria, including the scale of mutagenesis, portability to multiple microbial species, off-target mutagenesis, and cost-effectiveness. Finally, we discuss how these technical advances open new avenues in basic research and biotechnology.

Addresses
[1] Department of Microbiology and Immunology, University of California, San Francisco, 94143, San Francisco, CA, USA
[2] Genome Biology Unit, European Molecular Biology Laboratory, 69117, Heidelberg, Germany
[3] Synthetic and Systems Biology Unit, Biological Research Centre, 6726, Szeged, Hungary
[4] Department of Genetics, Harvard Medical School, 02115, Boston, MA, USA

Corresponding authors:
Csörgő, Bálint (Balint.Csorgo@ucsf.edu), Pál, Csaba (cpal@brc.hu)

Current Opinion in Microbiology 2020, **57**:22–30

This review comes from a themed issue on **Microbial systems biology**

Edited by **Athanasios (Nassos) Typas** and **Gene-Wei Li**

For a complete overview see the Issue and the Editorial

Available online 26th June 2020

https://doi.org/10.1016/j.mib.2020.05.010

Introduction

On a sufficiently long timescale with a large enough population size, biological evolution can produce myriad intricate solutions to various selective pressures. Over time, the best performing genetic variants are continuously selected resulting in highly specialized gene products with optimal properties. Humans have long sought to speed up and control this process to produce whole organisms or specific biomolecules with desired traits [1]. With the advent and continuing advancement of molecular biological techniques, efforts to direct evolution have greatly increased in specificity, capable of targeting single genes within organisms [2,3]. The concurrent development of highly efficient methods for the screening of gene variant libraries [4,5] has allowed for the isolation of a range of enzymes with improved or completely novel functions [6].

The most comprehensive approach for achieving these objectives requires saturation mutagenesis, that is, the ability to generate and screen all possible amino acid variants and their combinations at as many positions of a protein as possible. Although a variety of techniques have long existed capable of generating gene variant libraries towards this goal, recent years have seen the development of more refined mutagenesis technologies with increased targeting precision, increased ranges of attainable mutation rates, and decreased biases in mutational spectra. We briefly summarize these most recent advances and their related applications, focusing on tools developed and employed in microbial systems. These technologies can be broadly categorized based on whether the mutagenized target DNA is on an extrachromosomal element or on genomic DNA. While the former is more amenable to a wide variety of highly precise strategies, which allow for true saturation, the latter allows probing the effects of mutations in their truly native contexts by coupling the mutation generation and variant selection steps. Finally, we highlight recent approaches that are overcoming these limitations of mutagenesis of user-defined chromosomal segments and offer new possibilities for both fundamental evolutionary biology questions, as well as industrial applications.

Extrachromosomal mutagenesis

Extrachromosomal mutagenesis methods have the inherent ability of focusing the generation of genetic variants to a specified segment of DNA, allowing for saturation studies of selected regions of interest. However, these libraries are generated separately from the functional selection process and require labor-intensive cloning and transformation steps that often present a limit to the final number of variants that are screened. The most long-standing such method has been error-prone PCR [7], which makes use of the low fidelity of DNA polymerases under certain conditions. This approach has long been

used for numerous protein engineering applications to attain novel variants for example, with new catalytic activities [8], improved stability [9], or novel binding capabilities [10]. Drawbacks of error-prone PCR methods include relatively low per base mutation rates and inherently biased mutational spectra making it impossible to achieve saturation [11]. Improvements to overcome these limitations have been made in techniques such as 'sequence saturation mutagenesis' where a universal base is inserted throughout the target sequence [12] and also in 'casting error-prone PCR', where target sequences are divided up into smaller fragments [13], making for higher levels of mutational coverage.

A more targeted PCR-based approach that allows for true saturation of selected positions but generates variants of much shorter sequences, is site saturation mutagenesis (SSM), which utilizes synthetic oligonucleotides carrying one or more degenerate codon (such as NNK). To increase efficiency of this approach, prior identification of key residues of the given gene-product through phylogenetic analysis of homologous proteins is usually performed, and regions deemed important for functionality are then targeted. Numerous variations of this technique exist, the most common being QuikChange mutagenesis [14] where overlapping oligonucleotides carrying the degenerate codons are used to amplify the target sequence from a plasmid. Recent variations of SSM include nicking mutagenesis [15] and mutagenesis with reversibly terminated deoxyinosine triphosphates [16], both of which allowed for comprehensive saturation libraries of the active sites of the *bla* gene in *Escherichia coli* encoding TEM-1 β; as well as a two-step PCR strategy applicable to difficult-to-randomize genes [17]. SSM has numerous applications, including the engineering of protein binding and selectivity [18], increased enzymatic activity [19], or enhanced therapeutic efficacy [20]. Most recently, SSM was utilized for the mutagenesis of phage tail fiber residues, limiting bacterial phage resistance, thereby increasing efficiency of phage therapy [21].

Recent advances in DNA synthesis capabilities have allowed the massively parallel *in vitro* generation of gene variant libraries by high-throughput oligo synthesis [22]. Although such oligos are limited in size (~350 nucleotide maximal length), this has enabled complete saturation mutagenesis of short genes, including a tRNA gene in yeast [23]. Recently, tiling of multiple (19 in this case) such oligonucleotide libraries allowed for the generation of the complete first-order fitness landscape of a much larger adeno-associated virus capsid gene [24]. The high cost of DNA synthesis has been an obstacle in generating large gene variant libraries in this fashion, however techniques such as DropSynth, an emulsion-based DNA synthesis method [25•] hold promise in making this approach more attainable.

Several methodologies also exist for mutating episomal target genes in a continuous manner, which has the advantage of not requiring prior *in vitro* synthesis of variants of PCR-based or synthesis-based approaches. Many of these techniques utilize error-prone (EP) variants of DNA polymerases for replicating plasmid DNA leading to mutagenesis of the target sequence. This was originally achieved by transforming the vector into a mutator host strain with EP DNA polymerase enzymes and defective/deleted mismatch-repair systems such as *E. coli* strain XL1-Red [26]. However, the systematically high mutation rates of such strains eventually lead to deleterious effects in the cell, slowing growth and making the cells difficult to transform. A more targeted approach utilizes an EP DNA polymerase I (Pol I) enzyme to mutagenize the cargo of a Pol I-dependent plasmid [27]. A more refined version of this principle was developed recently in yeast termed OrthoRep, where the replication of a plasmid carrying the targeted DNA sequence is fully dependent upon an engineered orthologous EP DNA polymerase that otherwise does not replicate the host genome or other plasmids [28••]. This system allowed the generation of a detailed fitness landscape of the malarial dihydrofolate reductase against the anti-malarial drug pyrimethamine. An entirely different approach utilizes bacteriophage and their lifecycles as vessels for mutagenizing a target gene. In phage-assisted continuous evolution (PACE), propagation of the M13 bacteriophage relies on bacterial production of the pIII infectivity protein, which in turn is dependent on functional library variants encoded within the phage. In this fashion, genes of interest can be encoded on M13 and continuously mutagenized to rapidly generate saturation libraries. PACE was recently employed to evolve *Bacillus thuringiensis* δ-endotoxin variants able to target previously resistant insect pests [29], to generate a variety of proteins with improved soluble expression [30], and to evolve Cas9 variants with altered PAM specificity and higher precision [31].

Genomic strategies

Targeting chromosomal DNA sequences for mutagenesis has the advantages of eliminating labor-intensive cloning and PCR steps and coupling the variant generation and selection steps, all while maintaining the native genetic context of the target. The first approaches aiming to generate chromosomal gene variant libraries utilized various DNA damaging forces or compounds affecting the entire chromosome of an organism. A range of both physical (e.g. ultraviolet irradiation, gamma rays) and chemical (e.g. ethyl methanesulfonate, nitrous acid) mutagens induce mutations at random sequences throughout the genome [32] and can be utilized for genome-wide gene inactivation screens.

Chemical mutagenesis protocols are conceptually simple and broadly applicable, but hey have associated health hazards, and the associated mutational spectra are generally biased. In a similar vein, a mutagenesis plasmid (MP) approach has been developed, where selected dominant mutator genes are expressed from a vector in the bacteria of interest, leading to a systematic increase in mutation rates with less bias in the mutational spectra than physical and chemical approaches [33]. Overall, mutator strains and chemical mutagenesis do not require specification of the genomic regions relevant for the selected phenotype: they increase overall bacterial genomic mutation rate. As a consequence, they cannot be focused to specific regions for in-depth saturation studies, and result in deleterious off-target effects.

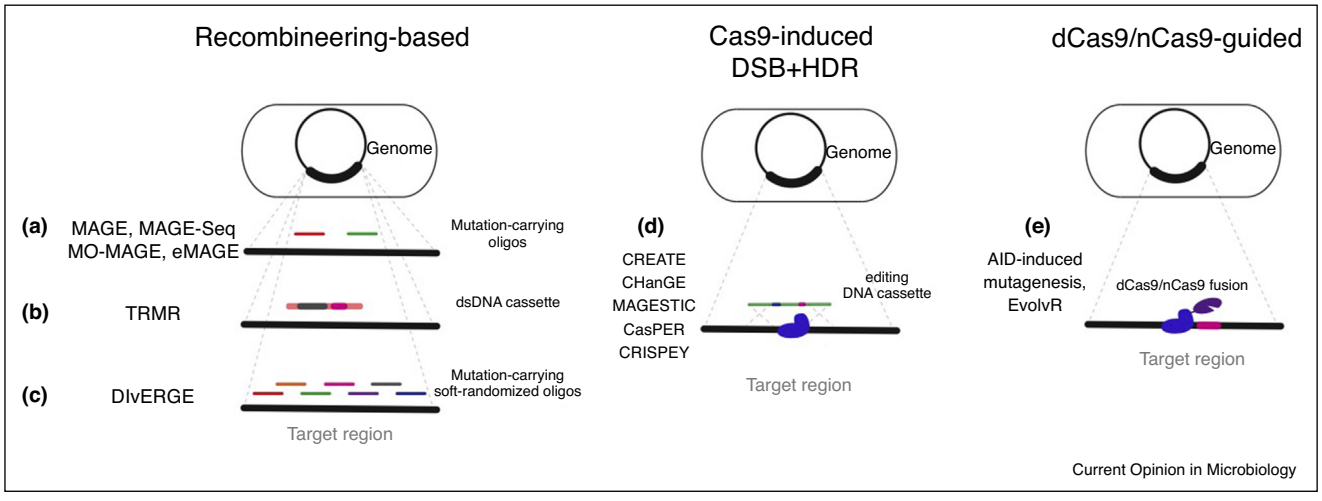
Alternatively, synthetic constructs can be genomically integrated to achieve targeted mutagenesis in a continuous fashion. In one approach dubbed *in vivo* continuous evolution (ICE), retroelements are constructed in yeast to encode a targeted gene of interest which undergoes EP reverse-transcription and genomic integration in a continuous fashion [34]. In another approach, an array of specific sites can be genomically integrated next to a selected region of interest to which a glycosylase enzyme is recruited that is capable of mutagenizing a ~20 kb genomic region [35]. These approaches solve the problem of localizing mutagenesis within the genome, however they require extensive prior construct development resulting in considerable modifications to the native contexts.

Targeted mutagenesis of multiple genomic loci

Recent years have seen the development of a number of diverse strategies that all aim to combine the high precision of extrachromosomal mutagenesis with genomic targeting for the saturation mutagenesis of specified genomic sequences within their native contexts. Two key technologies have enabled these advances: the development and optimization of single-stranded oligonucleotide-based recombineering methods [36], and the advent of CRISPR-Cas genome engineering technologies [37]. These approaches allow for unprecedented precision in the targeted modification of microbial genomes, and, through various strategies, can be adapted to mutagenize distinct chromosomal regions (Figure 1 and Table 1).

Recombineering-based approaches rely on the annealing of synthetic single-stranded oligonucleotides to the lagging strands at open replication forks. This process requires specific single-stranded DNA annealing proteins (e.g. the phage λ's Red Beta protein for *E. coli* [38] or other RecT variants [39]) to work at a high efficiency in a given organism. In a landmark paper, recombineering was developed to introduce multiple mutations across the genome in a process called multiplex automated genome engineering (MAGE) [40]. 20 separate oligonucleotides containing degenerate ribosome binding site (RBS) sequences were

Figure 1



Schematic representation of microbial genome editing methods capable of targeted saturation genome-mutagenesis. Recombineering-based approaches rely on single-stranded DNA oligonucleotide-stranded or double-stranded DNA cassette-mediated homologous recombination. **(a)** MAGE, MAGE-Seq, MO-MAGE, and eMAGE utilize single-stranded DNA oligonucleotides that carry user-defined mutations and incorporate those into the genomic target. **(b)** TMMR achieves the same outcome by recombining selectable dsDNA cassettes, that carry the desired modification, into the target. **(c)** DIvERGE uses partially overlapping DNA oligonucleotides that carry randomly distributed random point-mutations along their entire length to perform mutagenesis at the target region. **(d)** Methods relying on Cas9-induced double-stranded breaks plus homologous recombination (Cas9-induced DSB + HDR) exploit the lethal effect of CRISPR-Cas9-induced DSBs to select the integration of an editing DNA cassette that is carrying the modification-of-interest. **(e)** Catalytically impaired Cas9- (dead(d)- or nicking(n)-) guided methods exploit Cas9's ability to sequence-specifically recognize the target sequence and bring it to the proximity of a Cas9-fused mutator enzyme and thus introduce desired mutations. See section '*Targeted mutagenesis of multiple genomic loci*' for an extended description of each method.

Table 1

Efficacy and costs of targeted mutagenesis methods

Basis of technology	Method	Targeting window, efficiency	Applicable species	Off-target effects	Cost
Recombineering-based	MAGE [40], MAGE-seq [43]	Up to 30 nucleotides using a single oligo or hundreds of nucleotides (e.g. 219 of essential gene *infA*) in parallel using multiple oligos (1 per 1–2 saturated codons)	Optimized for *E. coli* and available in multiple *Gammaproteobacteria*	High because MMR deficient strain required for high efficiency, but the use of inducible dominant-negative MMR variant (pORTMAGE) can eliminate off-target effects	Cost-effective, however each oligonucleotide can only mutagenize a target up to 30 bp
	TRMR [41]	Thousands of nucleotides in parallel using multiple oligos	Optimized for *E. coli*	Low	High cost, due to the neccesity of high-throughput DNA synthesis
	MO-MAGE [44]	Thousands of nucleotides in parallel using multiple oligos	Optimized for *E. coli*	High, MMR deficient strain required for high efficiency	High cost, due to the neccesity of high-throughput DNA synthesis
	Eukaryotic MAGE [48]	Hundreds of nucleotides in parallel using multiple oligos, constrained by requirement for replication fork, rarer in eukaryotes	Optimized for *S. cerevisiae*	High, MMR deficient strain required for high efficiency	Cost-effective, however each oligonucleotide can only mutagenize a target up to 30 bp
	DIvERGE [47]	Thousands of nucleotides in parallel using multiple oligos	Optimized for *E. coli*, applicable to a range of *Enterobacteriacae*	Undetectable due to usage of inducible dominant-negative MMR variant	Cost-effective, each oligonucleotide can mutagenize a target up to 72 bp
Cas9-induced DSB, HDR	CREATE [53]	Thousands of nucleotides in parallel using multiple repair cassettes	Optimized for *E. coli*, applicable to *S. cerevisiae*	Not examined in-depth, expected to be low	High cost, due to the neccesity of high-throughput DNA synthesis
	CRISPR library [55], CHAnGE [56], MAGESTIC [57•], CRISPR variant libraries [58•], CasPER [59••], CRISPEY [60]	Thousands of nucleotides in parallel using multiple repair cassettes	Optimized for *S. cerevisiae*	Low, Cas9-mediated targeting showed high specificity	High cost, due to the necessity of high-throughput DNA synthesis
dCas9/nCas9-guided	AID-induced mutagenesis [61–63]	Maximum of 2 parallel targets demonstrated, mutagenesis limited to ~100 nucleotides surrounding PAM-constrained target site, high bias in mutational spectra	*E. coli*, *S. cerevisiae*, human cells	Potentially high	Moderate, due to the necessity of plasmid construction before mutagenesis
	EvolvR [64]	Targeting of 2 parallel targets demonstrated, mutagenesis limited to 50–350 nucleotides in vicinity of PAM-constrained target site with declining mutagenesis with increased distance	Optimized for *E. coli*	High, low-fidelity DNA polymerase raises background mutation rate over 100-fold	Moderate, due to the necessity of plasmid construction before mutagenesis

simultaneously targeted to various genes all involved in the biosynthesis of lycopene, leading to a fivefold increase in the production of this industrially relevant isoprenoid compound in only three days. In a method termed trackable multiplex recombineering, the MAGE approach was further refined to include barcodes within each oligonucleotide to allow for massively parallel mutagenesis of multiple genomic regions and the subsequent identification of modified sequences that resulted in improved phenotypes of interest [41]. This approach allowed for the mutagenesis of the RBSs of close to all genes in *E. coli* to modify their expression levels allowing improved growth in various environments [42]. MAGE can be employed in a highly focused manner, as synthesizing a library of oligonucleotides each carrying a degenerate codon of the same gene, allowed for the saturation codon mutagenesis of the essential gene *infA* in *E. coli* [43]. Measuring the fitness of each individual variant, combined with amplicon deep sequencing, enabled the in-depth analysis of the effects of codon usage across an entire gene.

In order to scale up the mutagenizing capabilities of MAGE to allow for potential saturation of extended genomic targets and enhanced multiplexability, microarray-oligonucleotide (MO)-MAGE was developed, where the mutation-inducing oligos are synthesized from microarray chips, allowing for parallel synthesis of large (>55 000) libraries [44]. Alternatively, the introduction of exogenous oligos to generate variants may be circumvented through a retroelement-based approach where a mutagenic T7 RNA polymerase enzyme generates variants of a sequence encoded on a retroelement in a continuous manner [45]. A specialized reverse transcriptase ultimately generates variants of single-stranded DNA which then edits the target sequence through ssDNA-recombineering.

A key drawback of MAGE-based recombineering approaches is the requirement of a mismatch repair (MMR)-deficient host for high efficiency mutagenesis. This leads to a high background mutation rate, leading to several off-target mutations, potentially confounding the phenotypic effects of saturation mutagenesis of the targeted region. One solution to this obstacle is the utilization of counter-selection markers such as the *tetA-sacB* system [46] or a system employing *ccdB* [47]. Through a two-step recombination process, the counter-selectable markers are integrated at the genomic site of interest, which is subsequently targeted using mutagenizing oligos. Counter-selection allows enrichment of cells which have incorporated the mutagenizing oligos all without the requirement of MMR inactivation. Alternatively, a simplified approach dubbed portable, plasmid-based MAGE (pORTMAGE) was developed, which utilizes inducible expression of a dominant negative MMR protein allele to achieve high efficiency recombineering while eliminating off-target effects [48]. Building on this advance, it became possible to specifically target extended genomic regions for saturation mutagenesis without any detectable off-target effects. This was achieved in a method called directed evolution with random genomic mutations (DIvERGE), which utilizes pools of oligonucleotides synthesized using a soft-randomization protocol (where the alternative nucleotides are spiked in at low (0.5–2%) amounts) at each nucleotide position [49••]. Such a synthesis approach significantly reduces the oligonucleotide costs of other methods such as MO-MAGE. The tiling of such 90mer oligos allows for the coverage of entire chromosomal genes for saturation mutagenesis. DIvERGE simultaneously targets multiple, user-defined regions, up to 10 s of kilobases in total, and has broad, controllable mutagenesis spectra for each nucleotide position [49••]. Importantly, DIvERGE is applicable to a range of bacterial host species without the need for prior genomic modification and off-target mutagenesis rate is expected to be very low [49••]. DIvERGE was utilized to perform simultaneous combinatorial saturation mutagenesis of the 4 genes (a total of 9.5 kb) encoding the target proteins of the antibiotics ciprofloxacin and gepotidacin [39,49••], while saturation mutagenesis of the target gene for the drug trimethoprim resulted in combinations of 5 mutations showing a >3900-fold increase in drug resistance [49••]. Overall, recombineering-based approaches now allow for the most extensive and controllable mutagenesis of multiple chromosomal regions in microbes, opening entirely new possibilities for future applications (see future perspectives).

Despite these unmatched capabilities, recombineering-based approaches do have some inherent limitations. Recombineering relies on active replication forks within the target cell, meaning the slower division time of eukaryotes makes the approach less efficient [50]. Also, ssDNA annealing proteins are not universal in their efficiencies in diverse bacterial organisms, meaning specific systems have to be optimized for different species [39, 51–53,68]. Finally, it generally relies on the *in vitro* synthesis of oligonucleotides to generate diversity. The advent of CRISPR-Cas-based gene editing technologies has offered solutions to some of these limitations. Double-stranded breaks of chromosomal DNA greatly enhance the recombination frequency of introduced homologous templates. Repurposed CRISPR-based systems (generally employing Cas9) can specifically cleave a genomic sequence of interest, leading to a vast improvement in the frequency of edited microbial cells [54]. Combining this capability with large-scale oligonucleotide synthesis led to the development of CRISPR-enabled trackable genome engineering (CREATE), which utilizes pools of 10^4–10^6 barcoded oligos to achieve genomic mutagenesis at chromosomal sites in bacteria [55]. A number of similar approaches were recently developed in yeast [56,57•,58•,59••,60,61], demonstrating the expanded potential of CRISPR-based targeted mutagenesis approaches in eukaryotes (see Box 1 for specific applications of these technologies).

All of these CRISPR-based methods require the prior synthesis of large pools of DNA oligonucleotides which serve as the editing templates for gene variation generation. Fusing various mutagenizing enzymes to a catalytically inactive version of Cas9 (dCas9) allows for their targeted localization within the genome, allowing for highly specific mutagenesis. One such approach utilizes fusion [62] or recruitment [63] of activation-induced cytidine deaminase (AID) to dCas9 to generate targeted mutagenesis specified by the single guide RNA (sgRNA) sequences. Using multiple sgRNAs allowed for tiling of longer mutagenized sequences and was used to identify drug resistance mutations against various cancer therapeutics in mammalian cells [62,63]. A similar approach fused AID to zinc-finger and

Box 1 Current applications of CRISPR-based mutagenizing technologies

CRISPR-enabled trackable genome engineering (CREATE) combines the genome editing capabilities of CRISPR-Cas9 with large-scale DNA oligo synthesis to achieve targeted chromosomal mutagenesis in bacteria [55]. This approach was used to saturate all codons of the *folA* drug-target gene in *E. coli*, and identify all resistance conferring individual mutations. CREATE can be used in multiplex, and was used to target 50 000 genomic sites to select for variants with improved tolerance to temperature and to the industrial solvents furfural and acetate [55]. CREATE can also be performed iteratively, generating combinations of thousands of mutations to achieve 60-fold improvement in the production of the industrially important chemical 3-hydroxypropionic acid [66]. The technique has also been used for the parallel mutagenesis of 19 genes involved in lysine metabolism in *E. coli*, identifying determinants capable of increasing production of the metabolite [67]. Building upon the basic principles laid down by CREATE, several methods have been recently developed to expand these capabilities to eukaryotes as well. These approaches have allowed the genomic integration of large libraries of variants and have enabled a wide variety of applications, including: determining the functional consequences of premature-termination codons at various locations within all annotated essential genes [57•], the saturation mutagenesis of a 29 amino acid region of the Siz1 protein for increased tolerance to the growth-inhibitor furfural [58•], the saturation editing of the essential gene SEC14 and identification of amino acids critical for chemical inhibition of lipid signaling [59••], the generation of a set of tiling deletion mutants for characterization of the SGS1 DNA helicase enzyme [60], the generation and screening of combinations of mutations in two key enzymes of the mevalonate pathway resulting in improved isoprenoid production [61], and studying the fitness consequences of 16 006 natural genetic variants through a retroelement-based approach to generate variation [62].

transcription activator-like effector proteins to achieve targeted variant generation in *E. coli* [64]. Finally, a CRISPR-guided Cas9 nickase was recently used to guide an engineered EP nick-translating DNA polymerase to specific genomic target sites, raising mutation rate by 3–4 orders of magnitude compared to background levels [65••]. This system, termed EvolvR is capable of generating all single substitutions in a 60-nucleotide window after 16 hours in 1 μl of saturated culture. Notwithstanding certain limitations of existing CRISPR-guided targeted genomic mutagenesis tools such as biases in mutational spectra, potential off-target effects, limited targeting window size, and an increased background mutation rate in the case of EvolvR, these technologies hold great promise in potential applications going forward.

Future perspectives for *in vivo* chromosomal saturation mutagenesis

The technologies currently allowing for the most controlled and complete mutagenesis of chromosomal sequences of interest (such as DIvERGE [49••], CREATE [55], and EvolvR [65••]) will open new doors in what is possible in directed evolution. Broadly, examples of these future applications include: (1) Targeted mutagenesis along the full length of multiple genes within a genome. This will allow the engineering of novel cellular functions involving multiple proteins, such as evolving novel metabolic functions from complex pathways. (2) Metabolic engineering in previously underutilized species. These above techniques can be adapted to a range of bacteria, including those with untapped metabolic potential resulting in optimization of novel industrially relevant pathways. (3) Saturation mutagenesis of multiple genes, allowing the directed evolution of multiprotein complexes. Improvement of complex traits often requires co-evolution of interacting amino acids coded at distinct loci, whose mutations provide no benefits individually. (4) Forecasting the dynamics of resistance evolution to novel antimicrobial drugs. System-wide mutagenesis affecting gene expression levels will aid in identifying primary drug targets and mechanisms of action. Once these are identified, saturation mutagenesis of the encoding genes will allow detailed fitness landscapes in the presence of a given drug. (5) Optimization of *in vitro* synthesized DNA constructs. *In vitro* constructed DNA elements encoding for example, biosynthetic pathways, genetic circuits, or entire genomic segments often lack clear design principles thus leading to suboptimal performance. High-throughput variant generation of the constructs will lead to rapid optimization. Finally, (6) fundamental evolutionary biology questions, such as the conservation of epistatic effects between related species or the phenotypic effects of varying codon usage in different species could be studied in greatly enhanced detail.

In summary, the last several years have seen great strides in the ability to generate genetic variant libraries capable of saturation of selected sequences. Many of these techniques can complement each other and depending on the studied organism, the level of specificity, targeting window size, and level of saturation, the ideal strategy can be chosen for a range of diverse applications.

Conflict of interest statement

The authors declare competing financial interest. B.C., A.N., and C.P. are listed as inventors on patent application related to DIvERGE (PCT/EP2017/082574 (WO2018108987) Mutagenizing Intracellular Nucleic Acids).

Acknowledgements

B.C. is supported by the Eötvös National Scholarship of Hungary and a Marie Skłodowska-Curie Actions Individual Global Fellowship (number 844093) of the Horizon 2020 Research Program of the European Commission. A.N. was supported by a Long-Term Fellowship (number ALTF 160-2019) from EMBO. C.P is supported by the European Research Council H2020-ERC-2014-CoG 648364 – Resistance Evolution; 'Célzott Lendület' Programme of the Hungarian Academy of Sciences LP-2017–10/2017; 'Élvonal' KKP 126506, and GINOP-2.3.2–15–2016–00014 (EVOMER).

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. Doebley JF, Gaut BS, Smith BD: **The molecular genetics of crop domestication**. *Cell* 2006, **127**:1309-1321.
2. Romero PA, Arnold FH: **Exploring protein fitness landscapes by directed evolution**. *Nat Rev Mol Cell Biol* 2009, **10**:866-876.
3. Simon AJ, d'Oelsnitz S, Ellington AD: **Synthetic evolution**. *Nat Biotechnol* 2019, **37**:730-743.
4. Bershtein S, Tawfik DS: **Advances in laboratory evolution of enzymes**. *Curr Opin Chem Biol* 2008, **12**:151-158.
5. Packer MS, Liu DR: **Methods for the directed evolution of proteins**. *Nat Rev Genet* 2015, **16**:379-394.
6. Turner NJ: **Directed evolution drives the next generation of biocatalysts**. *Nat Chem Biol* 2009, **5**:567-573.
7. Leung DW, Chen E, Goeddel DV, Leung DW, Goeddel DV: **A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction**. *Technique* 1989, **1**:11-15.
8. Seelig B, Szostak JW: **Selection and evolution of enzymes from a partially randomized non-catalytic scaffold**. *Nature* 2007, **448**:828-831.
9. Bordes F, Tarquis L, Nicaud J-M, Marty A: **Isolation of a thermostable variant of Lip2 lipase from *Yarrowia lipolytica* by directed evolution and deeper insight into the denaturation mechanisms involved**. *J Biotechnol* 2011, **156**:117-124.
10. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, Gonzales APW, Li Z, Peterson RT, Yeh J-RJ *et al.*: **Engineered CRISPR-Cas9 nucleases with altered PAM specificities**. *Nature* 2015, **523**:481-485.
11. Patrick WM, Firth AE, Blackburn JM: **User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries**. *Protein Eng* 2003, **16**:451-457.
12. Wong TS, Tee KL, Hauer B, Schwaneberg U: **Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution**. *Nucleic Acids Res* 2004, **32**:e26.
13. Yang J, Ruff AJ, Arlt M, Schwaneberg U: **Casting epPCR (cepPCR): a simple random mutagenesis method to generate high quality mutant libraries**. *Biotechnol Bioeng* 2017, **114**:1921-1927.
14. Hogrefe HH, Cline J, Youngblood GL, Allen RM: **Creating randomized amino acid libraries with the QuikChange multi site-directed mutagenesis kit**. *BioTechniques* 2002, **33**:1158-1160 1162, 1164–1165.
15. Wrenbeck EE, Klesmith JR, Stapleton JA, Adeniran A, Tyo KEJ, Whitehead TA: **Plasmid-based one-pot saturation mutagenesis**. *Nat Methods* 2016, **13**:928-930.
16. 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**:923-924.
17. Li A, Acevedo-Rocha CG, Reetz MT: **Boosting the efficiency of site-saturation mutagenesis for a difficult-to-randomize gene by a two-step PCR strategy**. *Appl Microbiol Biotechnol* 2018, **102**:6095-6103.
18. van der Meer J-Y, Poddar H, Baas B-J, Miao Y, Rahimi M, Kunzendorf A, van Merkerk R, Tepper PG, Geertsema EM, Thunnissen A-MWH *et al.*: **Using mutability landscapes of a promiscuous tautomerase to guide the engineering of enantioselective Michaelases**. *Nat Commun* 2016, **7**:10911.
19. Kleinstiver BP, Sousa AA, Walton RT, Tak YE, Hsu JY, Clement K, Welch MM, Horng JE, Malagon-Lopez J, Scarfò I *et al.*: **Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing**. *Nat Biotechnol* 2019, **37**:276-282.
20. Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, Mattos CD, Myers CA, Kamisetty H, Blair P, Wilson IA *et al.*: **Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing**. *Nat Biotechnol* 2012, **30**:543-548.
21. Yehl K, Lemire S, Yang AC, Ando H, Mimee M, Torres MDT, de la Fuente-Nunez C, Lu TK: **Engineering phage host-range and suppressing bacterial resistance through phage tail fiber mutagenesis**. *Cell* 2019, **179**:459-469.e9.
22. Rocklin GJ, Chidyausiku TM, Goreshnik I, Ford A, Houliston S, Lemak A, Carter L, Ravichandran R, Mulligan VK, Chevalier A *et al.*: **Global analysis of protein folding using massively parallel design, synthesis, and testing**. *Science* 2017, **357**:168-175.
23. Li C, Qian W, Maclean CJ, Zhang J: **The fitness landscape of a tRNA gene**. *Science* 2016, **352**:837-840.
24. Ogden PJ, Kelsic ED, Sinai S, Church GM: **Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design**. *Science* 2019, **366**:1139-1143.
25. • Plesa C, Sidore AM, Lubock NB, Zhang D, Kosuri S: **Multiplexed gene synthesis in emulsions for exploring protein functional landscapes**. *Science* 2018, **359**:343-347.
This work introduces DropSynth, a scalable, low-cost, droplet-based method to construct 1000s of gene-length assemblies simultaneously. The authors applied DropSynth to successfully build more than 7000 synthetic, phylogenetically diverse homologs of two essential genes of *E. coli* and tested their functionality *in vivo*.
26. Greener A, Callahan M, Jerpseth B: **An efficient random mutagenesis technique using an *E. coli* mutator strain**. *Mol Biotechnol* 1997, **7**:189-195.
27. Camps M, Naukkarinen J, Johnson BP, Loeb LA: **Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I**. *PNAS* 2003, **100**:9727-9732.
28. •• Ravikumar A, Arzumanyan GA, Obadi MKA, Javanpour AA, Liu CC: **Scalable, continuous evolution of genes at mutation rates above genomic error thresholds**. *Cell* 2018, **175**:1946-1957.e13.
The authors describe OrthoRep, an orthogonal, highly error-prone DNA polymerase-plasmid pair in *Saccharomyces cerevisiae* that stably mutates plasmid-borne sequences up to 100 000-fold faster than the host genome. Using OrthoRep, the authors simultaneously evolved antimalarial drug-resistant dihydrofolate reductases in 90 replicates.
29. 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.
30. Wang T, Badran AH, Huang TP, Liu DR: **Continuous directed evolution of proteins with improved soluble expression**. *Nat Chem Biol* 2018, **14**:972-980.
31. Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, Zeina CM, Gao X, Rees HA, Lin Z *et al.*: **Evolved Cas9 variants with broad PAM compatibility and high DNA specificity**. *Nature* 2018, **556**:57-63.
32. Kodym A, Afza R: **Physical and chemical mutagenesis**. *Methods Mol Biol* 2003, **236**:189-204.
33. Badran AH, Liu DR: **Development of potent in vivo mutagenesis plasmids with broad mutational spectra**. *Nat Commun* 2015, **6**:8425.
34. 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**.
35. Finney-Manchester SP, Maheshri N: **Harnessing mutagenic homologous recombination for targeted mutagenesis in vivo by TaGTEAM**. *Nucleic Acids Res* 2013, **41** e99–e99.
36. Thomason LC, Sawitzke JA, Li X, Costantino N, Court DL: **Recombineering: genetic engineering in bacteria using homologous recombination: recombineering**. In *Current Protocols in Molecular Biology*. Edited by Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. John Wiley & Sons, Inc.; 2014. 1.16.1-1.16.39.

37. Doudna JA, Charpentier E: **The new frontier of genome engineering with CRISPR-Cas9**. *Science* 2014, **346**:1258096.

38. Murphy KC: **Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli***. *J Bacteriol* 1998, **180**:2063-2071.

39. Wannier TM, Nyerges A, Kuchwara HM, Czikkely M, Balogh D, Filsinger GT, Borders NC, Gregg CJ, Lajoie MJ, Rios X, Pál C, Church GM: **Improved bacterial recombineering by parallelized protein discovery**. *Proc Natl Acad Sci* 2020 http://dx.doi.org/10.1073/pnas.2001588117.

40. 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.

41. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LBA, Gill RT: **Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides**. *Nat Biotechnol* 2010, **28**:856-862.

42. Sandoval NR, Kim JYH, Glebes TY, Reeder PJ, Aucoin HR, Warner JR, Gill RT: **Strategy for directing combinatorial genome engineering in *Escherichia coli***. *Proc Natl Acad Sci U S A* 2012, **109**:10540-10545.

43. Kelsic ED, Chung H, Cohen N, Park J, Wang HH, Kishony R: **RNA structural determinants of optimal codons revealed by MAGE-Seq**. *Cell Syst* 2016, **3**:563-571.e6.

44. Bonde MT, Kosuri S, Genee HJ, Sarup-Lytzen K, Church GM, Sommer MOA, Wang HH: **Direct mutagenesis of thousands of genomic targets using microarray-derived oligonucleotides**. *ACS Synth Biol* 2014, **4**:17-22.

45. Simon AJ, Morrow BR, Ellington AD: **Retroelement-based genome editing and evolution**. *ACS Synth Biol* 2018, **7**:2600-2611.

46. Li X, Thomason LC, Sawitzke JA, Costantino N, Court DL: **Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in *Escherichia coli***. *Nucleic Acids Res* 2013, **41**:e204.

47. Wang H, Bian X, Xia L, Ding X, Müller R, Zhang Y, Fu J, Stewart AF: **Improved seamless mutagenesis by recombineering using ccdB for counterselection**. *Nucleic Acids Res* 2014, **42**:e37.

48. Nyerges Á, Csörgő B, Nagy I, Bálint B, Bihari P, Lázár V, Apjok G, Umenhoffer K, Bogos B, Pósfai G *et al.*: **A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species**. *PNAS* 2016, **113**:2502-2507.

49. •• Nyerges Á, Csörgő B, Draskovits G, Kintses B, Szili P, Ferenc G, Révész T, Ari E, Nagy I, Bálint B *et al.*: **Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance**. *PNAS* 2018, **115**:E5726-E5735.

DIvERGE is presented, which utilizes pools of partially overlapping soft-randomized oligonucleotides to cover multiple chromosomal targets up to tens of kilobases and perform saturation mutagenesis in multiple bacterial species. DIvERGE was utilized to perform simultaneous saturation mutagenesis at the four target-encoding genes of the antibiotic ciprofloxacin and gepotidacin, encompassing a total of 9.5 kb, while saturation mutagenesis of the drug-target of trimethoprim resulted in combinations of five mutations showing >3900-fold increase in drug resistance.

50. Barbieri EM, Muir P, Akhuetie-Oni BO, Yellman CM, Isaacs FJ: **Precise editing at DNA replication forks enables multiplex genome engineering in eukaryotes**. *Cell* 2017, **171**:1453-1467.e113.

51. Datta S, Costantino N, Zhou X, Court DL: **Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages**. *PNAS* 2008, **105**:1626-1631.

52. Van Pijkeren J-P, Neoh KM, Sirias D, Findley AS, Britton RA: **Exploring optimization parameters to increase ssDNA recombineering in *Lactococcus lactis* and *Lactobacillus reuteri***. *Bioengineered* 2012, **3**:209-217.

53. Aparicio T, Nyerges A, Nagy I, Pal C, Martínez-García E, de Lorenzo V: **Mismatch repair hierarchy of *Pseudomonas putida* revealed by mutagenic ssDNA recombineering of the pyrF gene**. *Environ Microbiol* 2020, **22**:45-58.

54. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA: **RNA-guided editing of bacterial genomes using CRISPR-Cas systems**. *Nat Biotech* 2013, **31**:233-239.

55. Garst AD, Bassalo MC, Pines G, Lynch SA, Halweg-Edwards AL, Liu R, Liang L, Wang Z, Zeitoun R, Alexander WG *et al.*: **Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering**. *Nat Biotechnol* 2017, **35**:48-55.

56. Sadhu MJ, Bloom JS, Day L, Siegel JJ, Kosuri S, Kruglyak L: **Highly parallel genome variant engineering with CRISPR–Cas9**. *Nat Genet* 2018, **50**:510-514.

57. • Bao Z, HamediRad M, Xue P, Xiao H, Tasan I, Chao R, Liang J, Zhao H: **Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision**. *Nat Biotechnol* 2018, **36**:505-508.

The authors report a CRISPR-mediated and homology directed-repair-mediated genome engineering method, termed CHAnGE, that enables the engineering of tens of thousands of user-defined genomic modifications in *S. cerevisiae* on a precise and trackable manner. By exploiting the single-nucleotide precision of CHAnGE, the authors performed saturation mutagenesis at a 29 amino acid long region of the Siz1 protein for increased tolerance to the growth-inhibitor furfural.

58. • Roy KR, Smith JD, Vonesch SC, Lin G, Tu CS, Lederer AR, Chu A, Suresh S, Nguyen M, Horecka J *et al.*: **Multiplexed precision genome editing with trackable genomic barcodes in yeast**. *Nat Biotechnol* 2018, **36**:512-520.

The authors describe a CRISPR-mediated and homology directed-repair-mediated genome engineering method with trackable barcodes, termed MAGESTIC, in *S. cerevisiae*. Compared to similar CRISPR-mediated methods, MAGESTIC simultaneously increases editing efficiency by recruiting the editing-donor DNA to the CRISPR recognition site using a LexA–Fkh1p fusion protein. By exploiting this advance, the authors performed saturation editing of the essential gene SEC14 in its native genomic context and identified amino acids critical for lipid signaling.

59. •• Guo X, Chavez A, Tung A, Chan Y, Kaas C, Yin Y, Cecchi R, Garnier SL, Kelsic ED, Schubert M *et al.*: **High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR–Cas9 in yeast**. *Nat Biotechnol* 2018, **36**:540-546.

This work introduces a CRISPR-mediated and homology directed-repair-mediated genome editing approach for generating pools of trackable mutants with up to 100% efficiency in *S. cerevisiae*. The authors demonstrate the utility of their method via the saturation mutagenesis of SGS1 and the point-mutagenesis of highly conserved residues of the same protein. Furthermore, they also generated a genome-wide knock-out library targeting 315 previously poorly characterized small open reading frames (smORFs) throughout the yeast genome and assessed their fitness effect under multiple environmental conditions.

60. Jakočiūnas T, Pedersen LE, Lis AV, Jensen MK, Keasling JD: **CasPER, a method for directed evolution in genomic contexts using mutagenesis and CRISPR/Cas9**. *Metab Eng* 2018, **48**:288-296.

61. Sharon E, Chen S-AA, Khosla NM, Smith JD, Pritchard JK, Fraser HB: **Functional genetic variants revealed by massively parallel precise genome editing**. *Cell* 2018, **175**:544-557.e16.

62. 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**:1029-1035.

63. 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.

64. Yang L, Briggs AW, Chew WL, Mali P, Guell M, Aach J, Goodman DB, Cox D, Kan Y, Lesha E *et al.*: **Engineering and optimising deaminase fusions for genome editing**. *Nat Commun* 2016, **7**:1-12.

65. •• Halperin SO, Tou CJ, Wong EB, Modavi C, Schaffer DV, Dueber JE: **CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window**. *Nature* 2018, **560**:248-252.

The authors describe EvolvR, that introduces semirandom mutations in a small region at the vicinity of any genomic or plasmid site that can be targeted by CRISPR/Cas9. By fusing a nicking variant of Cas9 (nCas9) to an DNA polymerase I, the EvolvR system recruits the increased mutagenesis activity of the error-prone polymerase to the Cas9 recognition site. The authors also apply EvolvR to select mutations that confer antibiotic resistance in plasmid and at genomic target sites.

66. Liu R, Liang L, Choudhury A, Bassalo MC, Garst AD, Tarasava K, Gill RT: **Iterative genome editing of *Escherichia coli* for 3-hydroxypropionic acid production**. *Metab Eng* 2018, **47**:303-313.

67. Bassalo MC, Garst AD, Choudhury A, Grau WC, Oh EJ, Spindler E, Lipscomb T, Gill RT: **Deep scanning lysine metabolism in *Escherichia coli***. *Mol Syst Biol* 2018, **14**:e8371.

68. Aparicio T, Nyerges A, Martínez-García E, de Lorenzo V: **High-efficiency multi-site genomic editing (HEMSE) of *Pseudomonas putida* through thermoinducible ssDNA recombineering**. *Science* 2020:100946 http://dx.doi.org/10.1016/j.isci.2020.100946.