

## 随机突变文库构建与筛选研究进展

陈珏<sup>1,2</sup>, 黄佳敏<sup>1,2</sup>, 燕天鹤<sup>1,2</sup>, 彭晓宇<sup>1,2</sup>, 林峻<sup>1,2,3</sup>

1 福州大学 应用基因组学研究所, 福建 福州 350108

2 福州大学 生物科学与工程学院, 福建 福州 350108

3 福州大学 福建省海洋酶工程重点实验室, 福建 福州 350108

陈珏, 黄佳敏, 燕天鹤, 等. 随机突变文库构建与筛选研究进展. 生物工程学报, 2021, 37(1): 163-177.

Chen J, Huang JM, Yan TH, et al. Progress in the construction and screening of random mutation library. Chin J Biotech, 2021, 37(1): 163-177.

**摘要:** 定向进化是一个循环过程, 在构建多样化基因序列和筛选功能基因变体之间交替进行。该技术目前已被广泛应用于 DNA 序列、基因功能和蛋白质结构的优化和分析。定向进化包括随机基因文库的生成、基因在合适宿主中的表达和突变文库的筛选。构建基因文库的关键是库容量和突变多样性, 而筛选变体的关键是高灵敏度和高通量。文中讨论了定向进化技术的最新进展, 这些新技术极大地加速和简化了传统的定向进化过程, 为定向进化的发展注入了强劲动力。

**关键词:** 定向进化, 突变文库, 随机突变, 高通量筛选, 半理性设计

## Progress in the construction and screening of random mutation library

Jue Chen<sup>1,2</sup>, Jiamin Huang<sup>1,2</sup>, Tianhe Yan<sup>1,2</sup>, Xiaoyu Peng<sup>1,2</sup>, and Jun Lin<sup>1,2,3</sup>

1 Institute of Applied Genomics, Fuzhou University, Fuzhou 350108, Fujian, China

2 College of Biological Science and Engineering, Fuzhou University, Fuzhou 350108, Fujian, China

3 Fujian Key Laboratory of Marine Enzyme Engineering, Fuzhou University, Fuzhou 350108, Fujian, China

**Abstract:** Directed evolution is a cyclic process that alternates between constructing different genes and screening functional gene variants. It has been widely used in optimization and analysis of DNA sequence, gene function and protein structure. It includes random gene libraries construction, gene expression in suitable hosts and mutant libraries screening. The key to construct gene library is the storage capacity and mutation diversity, to screen is high sensitivity and high throughput.

**Received:** May 9, 2020; **Accepted:** July 29, 2020

**Supported by:** National Natural Science Foundation of China (No. 31301537), Fujian Science and Technology Department Support Program (No. 2020Y0003), Joint Funds for the Innovation of Science and Technology of Fujian Province, China (No. 2018Y9105), Young Teachers Education Research Project, Department of Education, Fujian, China (No. JAT170080).

**Corresponding author:** Jun Lin. Tel: +86-591-22863805; Fax: +86-591-22866273; E-mail: jun@fzu.edu.cn

国家自然科学基金 (No. 31301537), 福建省科技厅引导性项目 (No. 2020Y0003), 福建省科技创新联合资金项目 (No. 2018Y9105), 福建省中青年教育科研项目 (No. JAT170080) 资助。

网络出版时间: 2020-08-12

网络出版地址: <https://kns.cnki.net/kcms/detail/11.1998.Q.20200810.1704.003.html>

This review discusses the latest advances in directed evolution. These new technologies greatly accelerate and simplify the traditional directional evolution process and promote the development of directed evolution.

**Keywords:** directed evolution, mutant library, random mutation, high throughput screening, semi-rational design

定向进化 (Directed evolution) 是一种广泛应用的工程策略, 通过反复突变和选择来提高目标分子的稳定性或生化功能。通过人工模拟进化历程, 将多样化的基因文库翻译成相应的基因产物文库, 根据基因型和表型之间的对应关系来筛选功能变体<sup>[1]</sup>。定向进化已成功用于蛋白质工程, 其优点是: 在未知蛋白质结构和作用机制时<sup>[2]</sup>, 能在现有的蛋白质 (天然或经过工程改造) 中引入突变, 然后筛选出具有增强活性或另一种所需性状的子代蛋白质, 持续进化以达到目标性能水平。定向进化利用连续的突变和筛选获得目标 DNA, 进而得到对应的蛋白质, 如改良的抗体<sup>[3]</sup>、更高效的<sup>[4]</sup>或活性增强的突变蛋白<sup>[5]</sup>, 这些生物分子可用于基础生物学和生物技术研究。目前, 实验室的定向进化如图 1 所示, 已被广泛用于优化或改变单个基因或基因产物的活性。

定向进化的成功主要取决于两个方面: 多样性突变和高通量筛选<sup>[6]</sup>。随着分子生物学技术的发展, 研究人员开发出各种各样的突变方法扩大突变文库。扩大突变文库提高了获取目标分子的

机率, 但也增加了突变文库的筛选难度和分析成本。定向进化策略的关键因素之一是开发合适的选择和筛选技术, 这将大大增加从大型突变文库中获取具有所需特性突变体的机会, 并减少时间和成本。高通量筛选技术简化了传统筛选过程, 减少了时间和成本, 更易检测到目标性能变体。因此, 采用定向进化技术, 不仅需要考虑如何构建突变体文库, 还要考虑如何在高通量条件下筛选突变体<sup>[7-8]</sup>。定向进化的主要局限性是必须开发一种高通量筛选方法, 蛋白质半理性设计旨在构建小而精的突变体文库, 比其他的随机方法更高效。

本文总结了近年来发表的最新的定向进化方法。用于生成突变文库的传统技术如易错 PCR 和 DNA 改组在不断完善中得到广泛应用。重点介绍了正交体内连续进化和 3 种 CRISPR 介导的定向进化平台: EvolvR、CRISPR-X 和 CasPER, 还提及了近年来快速发展的蛋白质半理性设计策略。此外, 介绍了近年来出现的高通量筛选方法, 这些方法共同加速了定向进化领域的发展。

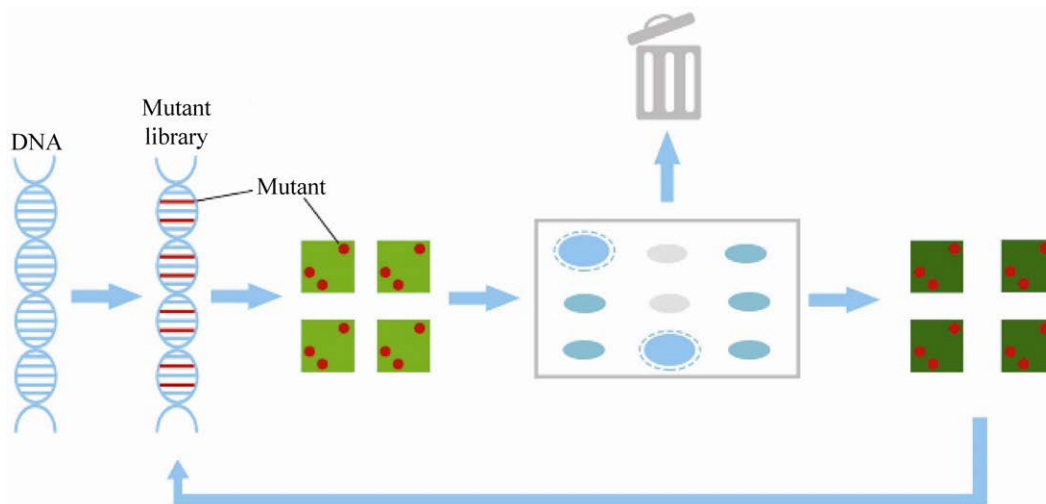


图 1 实验室定向进化中的关键步骤

Fig. 1 Key steps in the laboratory directed evolution.

## 1 突变文库构建策略

在自然界中, DNA 的遗传突变源于在基因组复制过程中引入错误, 紫外线或活性化学物质对 DNA 产生破坏。在实验室中, 选择合适的突变策略是成功构建大型 DNA 文库的第一步。突变文库的质量对定向进化起着决定性作用。不同的突变策略应该在多样的 DNA 文库产生以及在突变蛋白结构和功能研究上有广泛应用。用于生成突变文库的传统技术包括易错 PCR 和 DNA 改组至今仍在沿用, 而新型突变技术的出现简化了定向进化的进程。表 1 总结和比较了本文论述的几种突变文库构建策略。

### 1.1 传统突变技术及其进展

易错 PCR (Error-prone PCR) 是一种应用广泛的体外随机突变技术。Leung 等<sup>[9]</sup>研究表明 DNA 聚合酶在某些条件下保真度降低, 可在目的基因中随机引入核苷酸突变。McCullum 等<sup>[10]</sup>将易错 PCR 方法用于优化起始进化的蛋白质, 以改善蛋白质的稳定性、溶解性以及结合亲和力。Arnold 等<sup>[11]</sup>使用易错 PCR 将随机突变引入了枯草杆菌蛋白酶 E。Mao 等<sup>[12]</sup>采用易错 PCR 法, 突变具有草甘膦抗性的植物内源基因 EPSPS, 得到 5 种突变体 ELs-EPSPS, 将抗草甘膦的能力提高 5.4 倍。

之后, Shao 等<sup>[13]</sup>开发了原位易错 PCR (is-epPCR) 法, 可通过多轮 is-epPCR 获得一个序列的所有突变, 以进一步提高构建定向进化突变文库的效率。该方法具有以下优点: 1) 封闭环形 PCR 产物可以直接转化大肠杆菌感受态细胞, 且易于被其他抗生素筛选; 2) 在表达质粒中对可变 DNA 进行一步易错扩增, 建立突变文库并进行筛选; 3) 通过多轮 is-epPCR 可以在序列中获得目标突变的积累。

DNA 改组 (DNA shuffling) 也被广泛用于体外突变文库的产生。DNA 改组是由 W.P.C Stemmer 发明的一种同源基因体外重组的方法<sup>[14-15]</sup>。最初, 该技术用于随机重组单个基因的各种突变体, 如单基因改组。随后, DNA 改组也能有效重组不同微生物菌株和物种的同源基因, 如家族改组 (Family shuffling)<sup>[16]</sup>。Arnold 等<sup>[17]</sup>开发了一种简单且可重复的方法, 他们利用核酸内切酶 V 产生 DNA 片段, 成功实现了 DNA 改组, 其重组效率与 DNase I 相当。Ouyang 等<sup>[18]</sup>开发了一种新型的 DNA 改组方法, 以构建重组大肠杆菌菌株中的随机突变文库。漆酶 CotA 在大肠杆菌中的异源表达只有较低的催化效率, 不能满足工业要求, 而鉴定出的突变体 5E29 催化活性是原来的 1.21 倍。

表 1 突变文库构建策略的比对

Table 1 Comparison of mutant library construction strategies

Mutation method	Evolution strategy	<i>In vivo</i> or <i>in vitro</i>	Targeting	Source of mutation	References
Error-prone PCR	Random point mutations introduce genetic diversity through incomplete mutation response	<i>In vitro</i>	Single or multiple genomic loci	Random point mutation	[9-10]
DNA shuffling	Sequence diversity of recombinant genes based on homology	<i>In vitro</i>	Multiple genomic loci	Homologous recombination	[14-15]
OrthoRep	Error-prone replication target gene	<i>In vivo</i>	Orthogonal mutant plasmid	<i>In vivo</i> error-prone orthogonal replicase	[22]
EvolvR	Error-prone repair of single-strand gap produced by nCas9	<i>In vivo</i>	Target sequence	Error-prone DNA polymerase <i>in vivo</i>	[33]
CRISPR-X	dCas9 targets highly active cytidine deaminase AID	<i>In vivo</i>	Multiple genomic loci	dCas9 targeted AID	[34]
CasPER	Error-prone PCR and Cas9 binding mediate mutant library	<i>In vitro</i>	Single or multiple genomic loci	Error-prone PCR and <i>in vitro</i> evolution of Cas9	[35]

## 1.2 正交体内连续进化

聚合酶 (Polymerases) 校对和复制后错配修复是确保保真度的关键因素。研究人员为此开展了大量的 DNAPs (DNA polymerases) 工程工作, 得到了高度易错的正交 DNAPs 变体。在真核生物中, 酵母是用于定向进化的优秀宿主菌, 它能快速维持较大的种群规模。Herr 等<sup>[19]</sup>研究发现酵母能够承受 1 000 倍的突变率。Gunge 等<sup>[20]</sup>在酿酒酵母 *Saccharomyces cerevisiae* 中开发了一种特殊的 DNA 复制系统, 该系统由一对与基因组复制正交的 DNAP/质粒对组成, 工程化的正交 DNAP 只复制正交质粒。正交复制系统是体内连续进化的平台, 可以独立于宿主进行调控<sup>[21]</sup>。

以 Chang C. Liu 为代表的研究团队<sup>[22]</sup>在酵母中开发了一种高度易错的正交复制系统 (OrthoRep), 独立于宿主的体内连续进化系统, 其突变率比体内宿主基因组高 10 万倍。OrthoRep 由一个含有目标基因的末端蛋白 (Terminal proteins, TP) 质粒和另一个含有所有必需基因的质粒组成。DNAP 靶向 TP 质粒, 实现了正交 TP-DNA 聚合酶 (DNAP) 自主复制过程中的靶向突变, 从而导致目标质粒的快速突变。OrthoRep 基于一种 DNA 聚合酶 (TP-DNAP1), 它是能在酿酒酵母中自主复制的线性、高拷贝性细胞质 DNA 质粒 p1。在 p1 上编码的基因连续突变速度比基因组上编码基因的突变速度快约 100 000 倍, 而染色体基因则保持其自然突变率为每个碱基  $10^{-10}$  个替换 (s.p.b)。酵母细胞质定位的 *pGKL1*/TP-DNAP1 质粒/DNA 聚合酶对形成了一个正交的 DNA 复制系统。在不影响基因组复制的情况下, 其突变率显著提高, 从而支持了体内连续进化。Ravikumar 等<sup>[22]</sup>利用 OrthoRep 在 90 个独立复制实验中进化出耐药性疟疾二氢叶酸还原酶 (Drug-resistant malarial dihydrofolate reductases, DHFRs), 为生物分子和细胞功能的常规、高通量进化提供了一种新范例, 并证明其在耐药性研究方面的效用。

在现有的进化实验中, OrthoRep 能保持超过 300 代的高突变率。然而, OrthoRep 使用一种特殊的转录系统, 其组成部分 (如启动子) 限制了 OrthoRep 编码基因的表达强度。Ravikumar 等<sup>[22]</sup>试图增强 OrthoRep 的表达水平, 使其在蛋白质进化中更具通用性; 他们从之前的 OrthoRep 连续进化实验中筛选启动子突变体, 将这些启动子与基因编码的 3'poly(A) 尾巴结合。Zhong 等<sup>[23]</sup>开发了一组 OrthoRep 基因表达盒, 使 OrthoRep 编码基因超出酿酒酵母内源基因可以达到的表达水平。具体来说, 各种启动子突变以及经过基因编码的 poly(A) 尾巴增加了 OrthoRep 上编码基因的表达水平。表达水平在传代过程中稳定提高, 并在多个基因和具有不同突变率的 OrthoRep 系统中保持一致。这些表达上的改进扩展了 OrthoRep 的应用范围, 进一步将 OrthoRep 的适用性扩展到蛋白质体内连续进化。

Arzumanyan 等<sup>[24]</sup>开发了 *pGKL2*/TP-DNAP2 质粒/DNA 聚合酶对, 形成了第 2 个正交复制系统; 他们构建了 p2 质粒和 TP-DNAP2 所需的相关遗传技术, 并证明 *pGKL2* 可以编码和表达自定义基因, 可以工程化易错的 TP-DNAP2s。TP-DNAP2 复制的 *pGKL2* 既与酿酒酵母基因组复制正交, 又与 TP-DNAP1 复制的 *pGKL1* 正交, 从而在同一细胞中形成 2 个相互正交的 DNA 复制系统。这样, 一对正交复制系统可以调节 2 个聚合酶的错配, 实现多基因在体内以不同的突变率进化。这 2 个相互正交的 DNA 复制系统为合成生物学的新应用奠定了基础。2019 年, Javanpour 等<sup>[25]</sup>继续研究和扩展了 OrthoRep 的通用性, 这将促进 OrthoRep 在定向进化中的广泛应用; 他们利用 CRISPR/Cas9 加速 p1 的遗传操作, 显示至少 22 kb 的遗传物质可以很容易地在 p1 上编码, 这表明 OrthoRep 是一种高效通用的酵母系统, 广泛适用于体内目标基因连续进化。

### 1.3 CRISPR/Cas 介导的定向进化

从化脓性链球菌 *Streptococcus pyogenes* 中获得的 II 型 CRISPR 系统, 又称 CRISPR/Cas9 系统。CRISPR/Cas9 可在使用者选择的 DNA 序列上产生一个位点特异性双链断裂 (DSB)。这种 DSB 可以通过精确的同源定向修复 (HDR) 或易出错的非同源末端连接 (NHEJ) 修复机制进行修复, 达到定点敲除某种基因的目的。化脓性链球菌 Cas9 (SpCas9) 需要原型间隔区相邻基序 (PAM) 序列 (NGG) 用于靶识别, 从而限制可靶向的基因组位点<sup>[26]</sup>。

CRISPR 技术的不断发展, 为定向进化的研究带来了机遇和挑战。CRISPR/Cas9 已可用于靶向几乎任何基因组位点并执行定向进化。CRISPR/Cas 定向进化 (CDE) 平台采用 CRISPR/Cas9 在目的基因所有可能的编码序列位点产生 DSB, CRISPR/Cas9 要在 NGG PAM 序列进行切割。CDE 模仿达尔文进化 (图 2), 该方法高效简便且高度灵活。CDE 通过设计多个具有不同目标序列的 sgRNA 同时进行多位点编辑<sup>[27]</sup>, 而在新的靶位点只需要设计新的 sgRNA, 这极大地简化了基因编辑过程并增加了目标位点的选择性<sup>[28]</sup>。CDE 平台正成为植物科学研究的有力工具<sup>[29-31]</sup>。Butt 等<sup>[32]</sup>使用 CDE 进化水稻 SF3B1 剪接体蛋白, 展示了 CRISPR/Cas9

在定向进化上的应用; 他们设计 sgRNA 靶向文库, 克隆到二元载体中, 然后转化到根瘤土壤杆菌中培养, 在选择压力下再生以加速进化。

为了提高进化能力, 可以通过多种策略来扩大目标突变范围。科学家已开发多种 CRISPR/Cas 介导的定向进化平台, 包括 EvolvR<sup>[33]</sup>、CRISPR-X<sup>[34]</sup>和 CasPER<sup>[35]</sup>, 能在未来的 CDE 平台中发挥作用。

#### 1.3.1 EvolvR

传统的连续定向进化要求严格的培养条件<sup>[36]</sup>且需在用户定义的范围内突变目标核苷酸<sup>[37]</sup>。目前, 还没有一种高效的方法可以连续地使目标区域的所有核苷酸多样化。新的正向遗传学需要能够在目标区域不断地使所有核苷酸多样化的靶向突变体。

Halperin 等<sup>[33]</sup>开发了一种名为 EvolvR (图 3) 的新型诱变工具。使用 EvolvR 在大肠杆菌中靶向 2 个基因: 核糖体蛋白亚基 *rspE* 和 *rspL*。两者都有已知的突变, 分别针对大观霉素和链霉素的耐药性。EvolvR 分别靶向 *rspE* 和 *rspL*, 同时进化这两个基因, 以此产生对大观霉素和链霉素耐药的菌株。此外, 研究发现 *rspE* 的新突变也赋予了大观霉素抗性, 这证实 EvolvR 可用于产生新的蛋白质变体。

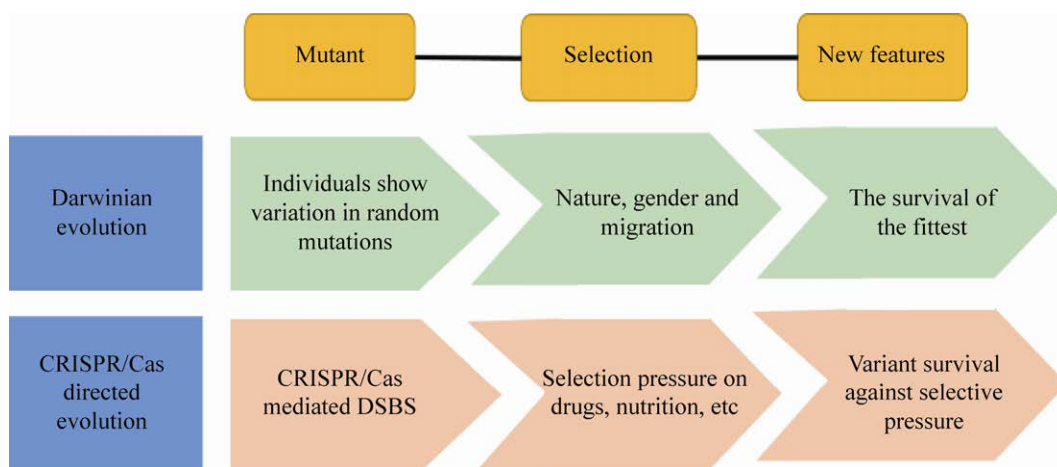


图 2 CRISPR/Cas 定向进化模仿了达尔文进化

Fig. 2 CRISPR/Cas directed evolution mimics Darwinian evolution.

EvolvR 系统巧妙地将 CRISPR/Cas9 的特异性与易错 DNA 聚合酶的突变能力结合。它通过用户定义的 20 nt 间隔序列的向导 RNA (gRNA) 来实现精准定位。EvolvR 有两个重要元件, 包括变体 Cas9 (nickase Cas9 或 nCas9)<sup>[38]</sup>, 其中 RuvC 核酸酶结构域发生了突变, 仅发生单链断裂。还有 1 个直接与 nCas9 融合的易错聚合酶 (Pol I 3M)<sup>[39]</sup>。Pol I 3M 包含 3 个点突变, 这些突变会增加其错误率并消除其校对活性。研究证明了这两种元件的融合使 Pol I 3M 的诱变活性指向 nCas9 靶向基因组位点, 而不是基因组中任何切口 DNA。

从机制上讲, nCas9 首先定位目标基因组位点, 并在与 gRNA 互补的单链 DNA 上产生 1 个切口 (图 3A)。nCas9 分离后 (图 3B), Pol I 3M 在该切口处结合 (图 3C)。Pol I 3M 聚合 1 条新的 DNA 链, 偶尔引入错误碱基, 并置换旧的 DNA 链 (图 3D)。在解离前, Pol I 3M 的瓣状内切核酸酶结构域切割原来的置换链, 留下了 1 个可连接的缺口 (图 3E)。Pol I 3M 解离后, 在宿主细胞内完成缺口的修复工作 (图 3F)。在任何可被 CRISPR/Cas9 靶向的基因组或质粒位点下游的小区域引入半随机突变, EvolvR 可以重新定位并连续进化相同的位点。

在第 1 次迭代中, EvolvR 在目标位点的突变率是野生型大肠杆菌的 24 500 倍。为了进一步提高靶向突变率, 将 3 个点突变 (K848A, K1003A, R1060A) 引入到 nCas9 (增强的 nCas9, enCas9)<sup>[40]</sup> 中以提高其解离率, 增加了 Pol I 3M 与 DNA 结合的机会, 从而提高了目标位点的突变率。此外, 考虑到 Pol I 3M 的持续合成能力 (15–20 nt) 是一个限制因素, 将 Pol I 3M 与 T7 噬菌体的硫氧还蛋白结合域 (Thioredoxin-binding domain, TBD) 融合<sup>[41]</sup>, 用于提高 Pol I 的持续合成能力。基于这些改进, 研究人员利用 EvolvR 测试到的突变率

是野生型的 212 000 倍, 且合成长度达 56 nt。

EvolvR 有很大的发展空间, 拓展 PAM 范围是其能被更广泛应用的一个关键, 可借鉴扩大 CRISPR 靶向范围的研究。Hu 等<sup>[42]</sup>利用一种噬菌体辅助的连续进化系统 (Phage-assisted continuous evolution, PACE) 进化出一种能够识别多种 PAM 序列的 SpCas9 变体 (xCas9)。xCas9 可以广泛识别多种 PAM 序列 (包括 NG、GAA 和 GAT), 这支持了它在人类细胞中的应用, 包括靶向转录激活, 核酸酶介导的基因破坏以及胞苷和腺嘌呤碱基编辑。xCas9 具有更广泛的 PAM 兼容性, 但与 SpCas9 相比, 它具有更大的 DNA 特异性。在 NGG 目标位点, 它的脱靶全基因组活性较低, 当使用非 NGG PAM 作为靶标时, 其脱靶活性较低。这一发现扩大了 CRISPR 系统 DNA 的靶向范围, 并证明在 Cas9 的编辑效率, PAM 兼容性和 DNA 特异性之间没有必要的联系。Anders 等<sup>[43]</sup>开发了一种 SpCas9 突变体 (SpCas9-NG), 用于识别 NG 而不是 NGG。先前的研究表明, SpCas9 的 Arg1 333 和 Arg1 335 分别识别 NGG-PAM 中第 2 和第 3 个碱基。通过消除 Arg1 335 和第 3 个 G 之间的碱基特异性相互作用来减少 PAM 约束。新引入的非碱基特异性相互作用弥补了与第 3 个 G 的碱基特异性相互作用的损失, 从而能够识别 NG PAM。研究者发现 SpCas9-NG 在 NG PAM 靶点诱导插入缺失。此外, SpCas9-NG 与活化诱导的胞苷脱氨酶 (Activation-induced cytidine deaminase, AID) 的融合可以介导 NG PAM 在靶位点产生 C-T 的转换。

EvolvR 系统已成为 CRISPR 工具箱中的一种最新的独创性工具。EvolvR 可以在任何细菌菌株中使用, 但目前只测试了其在大肠杆菌中的用途, 可能需要进一步优化其在致病性或肠道共生细菌中的用途。最后, EvolvR 能否应用于真核细胞尚待证实, 其复杂的基因组结构和 DNA 复制/修复途径可能需要进一步的工程优化<sup>[44]</sup>。

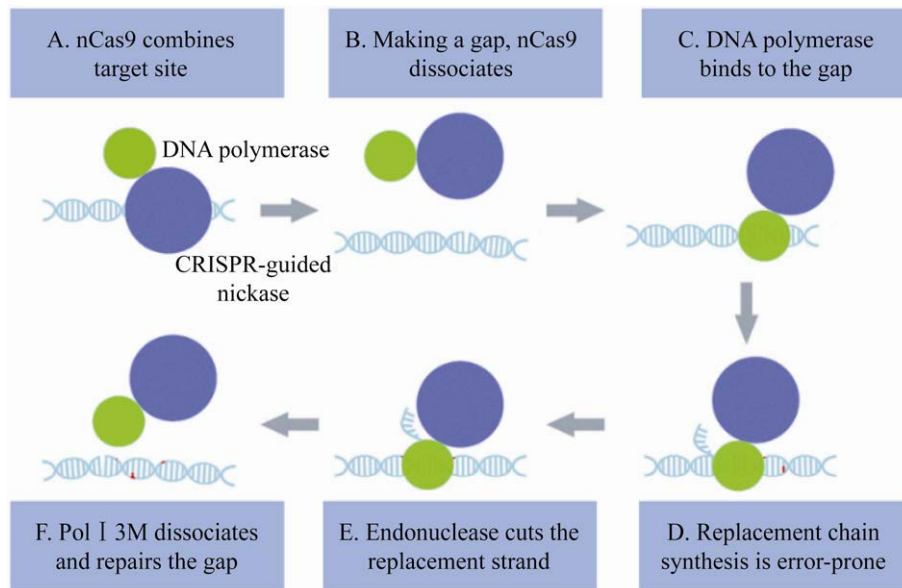


图3 EvolvR 突变过程示意图

Fig. 3 EvolvR mutation process diagram.

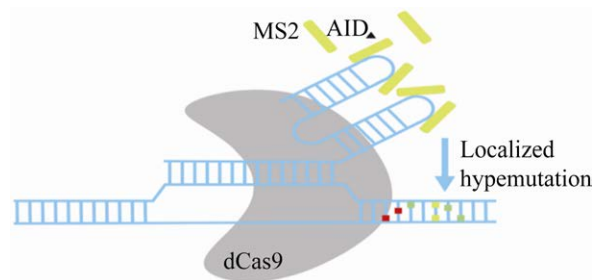
### 1.3.2 CRISPR-X

定向进化需要引入特定大小的 DNA 文库。随着 CRISPR/Cas9 系统的出现, 活性 Cas9 突变能有效引入插入和缺失, 但可能会破坏功能元件<sup>[45]</sup>, 使蛋白功能失活。而催化失活的 Cas9 (dCas9)<sup>[28]</sup>能将功能蛋白定位到特定的基因组位点, 这扩展了活性 Cas9 进行蛋白质工程的研究范围。

Hess 等开发的 CRISPR-X<sup>[34]</sup>, 通过 dCas9 靶向高活性胞苷脱氨酶 AID<sup>[46]</sup>以诱导局部、多样的点突变 (图 4)。活性 Cas9 引入突变主要通过产生插入和缺失, 或者通过 Cas9 切割后的同源重组生成寡核苷酸突变库。相比之下, CRISPR-X 产生局部序列多样化, 可同时突变多个基因组位点, 特异性诱变内源性靶标。测试后发现, 只要简单地电穿孔一个 sgRNA 和靶向 AID, GFP 就可以很容易地进化成 EGFP<sup>[34]</sup>。利用该技术成功突变了 GFP 并筛选出光谱转移变体 EGFP, 同时利用一种活性显著增加的 AID 变体可以突变转录起始位点上游和下游的内源性位点。此外, 对癌症治

疗药物硼替佐米的靶点 PSMB5 进行突变, 揭示已知的和新型的耐药机制, 并为 PSMB5 的功能研究和未来的药物研发提供参考<sup>[34]</sup>。

通过电穿孔引入 CRISPR-X, 同时靶向多个 sgRNAs, 可以诱发多个基因组位点突变。CRISPR-X 是一种强大的工具, 允许在 CRISPR 靶点引入不同的点突变, 连续突变和进化出新的蛋白质功能。可以设想, 该系统能使两种相互作用的内源性表达蛋白共同进化, 并为在自然环境下突变选择活性增强的抗体和酶蛋白提供一种简化的策略。

图4 CRISPR-X 技术原理示意图<sup>[34]</sup>Fig. 4 Schematic of the CRISPR-X technology<sup>[34]</sup>.

### 1.3.3 CasPER

同源重组能高效地靶向和整合异源 DNA 片段。在酵母中能将同源末端的线性片段整合到染色体<sup>[47-48]</sup>。通过 Cas9 能将较大的 DNA 片段组装并整合到酿酒酵母基因组中。但目前,大多数 CRISPR/Cas9 定向进化方法倾向于整合相对较小(80–120 bp)的 DNA 片段<sup>[49-50]</sup>,可能需要耗时构建供体突变文库或合成相对昂贵的多样化寡核苷酸 DNA<sup>[51]</sup>。在此基础上, Jakočiūnas 等扩展了现有的 CRISPR 技术,采用易错 PCR 和 Cas9 介导的基因组整合方法,将大型供体突变文库整合到单个或多个基因组位点,成功率达到 98%–99%<sup>[35]</sup>。该方法称为 CasPER,是 Cas9 介导蛋白质进化反应的简称。

CasPER 包括以下主要步骤(图 5):首先,选择合适的 DNA 突变片段。其次,在目标靶点设计 gRNA。之后,在宿主基因组中合成或扩增缺乏 PAM 的 DNA 模板,通过多轮易错 PCR (epPCR) 进行突变。epPCR 利用 Cas9 介导双链断裂 (DSB) 上下游基因组位点同源的 52 bp 引物产生突变文库。最后,通过电穿孔将突变的线性片段与携带 gRNA 的质粒共同转化到目标受体中。

CasPER 在同源基因组中能促进数百个核苷

酸的有效突变。通过多种方式将大型突变 DNA 片段整合到一起,其效率几乎达到 100%,从而产生数百万个突变体。研究员利用 CRISPR/Cas9 在基因组靶点引入 DSB,并通过天然同源定向修复整合 300–600 bp 线性 DNA 供体,提高酶在自然环境中的定向进化效率。该方法适用于以相同的效率靶向单基因组和双基因组基因座,使基因组整合子的突变分布保持均匀。应用 CasPER 对酿酒酵母甲羟戊酸途径中的 2 种关键代谢酶进行定向进化,可将异戊二烯产量提高 11 倍。

CasPER 通过同源重组支持异源 DNA 基因组整合,可以在供体整个片段上实现多重基因组工程,从而实现高效率和高通量的进化研究。CasPER 已在酵母中得到应用,未来它还能够应用于其他的具有高效的同源重组机制的有机体。

### 1.4 蛋白质半理性设计

理论上,定向进化首先通过随机突变和体外重组产生分子多样性,然后通过高通量筛选或选择来鉴定突变体。并不是所有的酶活性都适用高通量筛选方法,也不是所有的筛选方法都易于实现。定向进化存在局限性,即使文库量达数百万个,对一个蛋白质来说,仍然只能从庞大的文库

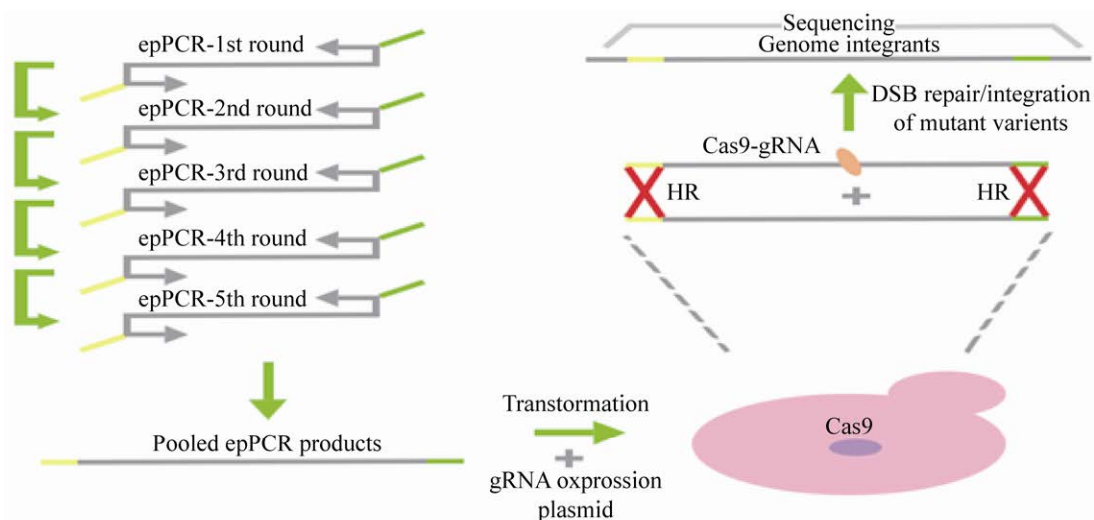


图 5 CasPER 技术原理示意图<sup>[35]</sup>

Fig. 5 Schematic of the CasPER process technology<sup>[35]</sup>.



中抽取小部分样本。实验方法的偏差性和遗传密码的简并性进一步限制了文库的设计<sup>[52-54]</sup>。研究人员不是通过构建更大的文库和进行更多的筛选或选择来解决这些问题,而是通过超越传统的定向进化。Lutz 等<sup>[55]</sup>开发出更小、更高质量文库的新策略,形成半理性设计 (Semi-rational design) 策略。半理性设计主要借助生物信息学方法,基于蛋白结构、功能、序列同源性以及计算预测法来预先选择合适的目标位置,再以有限的氨基酸多样性进行集中突变。该方法理性选取多个氨基酸残基进行突变,选用有效的密码子,通过构建高质量突变体文库,有针对性地对蛋白质进行改造<sup>[56]</sup>。

为了减少定向进化过程中的筛选工作, O'Maille 等依据结构和序列信息开发了一些半理性“文库设计”策略,如基于结构的组合蛋白质工程 (Structure-based combinatorial protein engineering, SCOPE)<sup>[57]</sup>、组合位点饱和突变 (Combinatorial site-saturation mutagenesis, CSSM)<sup>[58]</sup>及其他方法,旨在生成更小和功能更丰富的突变库。随着计算机技术的发展,半理性设计可以进一步缩小筛选的突变库容量,在尽可能小的文库中,产生更多有益突变。半理性设计越来越多地被用于设计“小而精”的突变体文库<sup>[59-60]</sup>。除了基于序列和结构的设计策略外,量子化学 (Quantum mechanics, QM) 和分子动力学模拟 (Molecular dynamics, MD) 计算以及机器学习算法能有效地探索氨基酸取代对蛋白质结构和稳定性影响。当针对选择性和活性的进化时,首选组合活性位点饱和突变 (Combinatorial active-site saturation test, CAST) 及迭代饱和突变 (Iterative saturation mutagenesis, ISM) 策略,这

2 种方法都在结合口袋附近的位点进行饱和突变 (Saturation mutagenesis, SM)<sup>[61]</sup>。随着半理性设计策略的快速发展, Fox 等<sup>[62]</sup>基于统计学方法开发的 ProSAR (Protein sequence activity relationships) 和 Gonzalez 等<sup>[63]</sup>基于序列同源性开发的 MORPHING (Mutagenic organized recombination process by homologous *in vivo* grouping) 工具也广泛用于蛋白酶的设计和改造。Chen 等<sup>[64]</sup>采用 3 种半理性策略, CSSM 和 2 种基于结构的计算设计算法 Corbit<sup>[65]</sup> 和 CRAM 构建了野生型 P450-BM3 血红素结构域的突变文库,获得了丙烷和乙烷羟基化的 P450-BM3 变体。在半理性设计的基础上,采用蛋白质工程技术, Jamil 等<sup>[66]</sup>成功地改善了单胞菌 PpADI (*Pseudomonas plecoglossicida*) 的精氨酸脱亚氨酶 (Arginine deiminase, ADI) 的疏水性以提高其结构稳定性和催化效率。Shin 等<sup>[67]</sup>首次采用蛋白半理性设计改善人参皂苷的  $\beta$ -糖苷酶 ( $\beta$ -glycosidase) 水解活性。这些研究表明,半理性设计使研究人员加强对酶的结构和功能理解。

## 2 突变体的选择和筛选方法

随着分子生物学的发展,已开发出多种突变方法来引入足够的遗传多样性。定向进化实验的主要限制是从庞大的突变体文库中鉴定出目标变体。常规的突变文库筛选方法无法适应较大突变体文库的筛选,人们开发了新的筛选和选择方法。成功的定向进化需要高通量的筛选或选择方法,这大大增加了获得所需特性的机会,并减少了时间和成本。表 2 给出了一些定向进化实验中文库选择和筛选方法的比较,包括适用范围和优缺点。

表 2 选择和筛选方法的比较

Table 2 Comparison of selection and screening methods

Strategy	Library quantity	Advantages	Disadvantages	References
Phage Display	$10^9$	Multiple proteins can be displayed	Not suitable for selecting multiple enzymatic properties; not suitable for selecting eukaryotic proteins	[68-70]
PACE	$10^9$	Fast and high throughput	Need phage	[42]
FACS	$>10^7$	High sensitivity and extremely high throughput	The expression level of fluorescent protein is related to enzyme activity	[80-81]
FADS	$>10^9$	Large library	Need fluorescent display	[82]

## 2.1 噬菌体展示技术

定向进化联合体外筛选技术是一种获取具有所需功能特性的核酸和蛋白质的强有力方法。噬菌体展示技术由 Smith 教授<sup>[68]</sup>于 1985 年首次引入,它是一种生产多肽(多达  $10^{10}$  种)和蛋白质以及筛选特定功能分子非常有效的方法<sup>[69-70]</sup>。1990 年, McCafferty 等<sup>[71]</sup>将噬菌体展示 (Phage display) 技术成功地从多肽的展示与筛选扩展到抗体的展示与筛选,人类从此进入了不经过免疫也能获得特异性抗体的时代。噬菌体展示具有高通量、快速、易用性等特点,可用于改造人类抗体<sup>[72-73]</sup>。截至 2016 年初,利用噬菌体展示研发的 6 种人类抗体已获批准并用于治疗<sup>[74]</sup>。2018 年,“多肽和抗体的噬菌体展示”技术荣获诺贝尔化学奖。

噬菌体展示技术与与时俱进。Frei 等<sup>[75]</sup>总结了噬菌体文库的合成路径,强调设计人性化的抗体文库和组合扫描诱变文库的重要性。Cung 等<sup>[76]</sup>还开发了一种微流体芯片,该芯片能够单轮识别无需细菌感染的高亲和力噬菌体展示肽,应用领域覆盖了蛋白质、肽和分子相互作用的基础研究,以及全自动蛋白质组学筛选。

## 2.2 噬菌体辅助连续进化

对基因反复突变和功能性选择,定向进化取得了非凡的成功。现有的定向进化方法费力耗时难以扩展,需要一个能在体内连续进化的平台。噬菌体辅助连续进化 (Phage-assisted continuous evolution, PACE) 系统<sup>[36]</sup>将目的蛋白与噬菌体生命周期耦联,通过自动化连续培养实现目标分子连续进化。噬菌体生命周期短,仅 10 min 即可产生子代噬菌体。该系统利用 M13 噬菌体的增殖能力,依赖于诱导蛋白质 III (pIII 蛋白) 的特性,构建了一个包含目的基因而 pIII 蛋白缺陷型的选择噬菌体。与其他系统不同,目标基因在 M13 噬菌体中编码,而突变(致突变质粒)和选择性(辅助质粒)因子则位于大肠杆菌中。当噬菌体感染大肠杆菌时,大肠杆菌突变质粒触发噬菌体的突变,仅释放

那些诱导蛋白质 III (pIII) 产生的突变噬菌体,从而实现筛选。PACE 系统一天中可以进行多代噬菌体选择性复制和数十轮进化,无需人工干预。

PACE 代表了生命系统中许多蛋白质和核酸的整合,快速生成具有新活性的生物分子。PACE 已经用于优化一个扩展的 PAM SpCas9 突变体 (xCas9)<sup>[42]</sup>。PACE 可在 8 d 执行多达 200 轮的蛋白质进化,为多种蛋白质或酶的改造提供了可能。David Liu 研究团队<sup>[42]</sup>利用 PACE 成功进化了 T7 RNA 聚合酶,该酶识别独特的启动子。PACE 在一周内进化出 3 种目标活性的酶,这 3 种酶活性均超过其野生型 T7 RNAP,提高了实验室进化速度。

David Liu 研究团队不断优化 PACE,开发了可溶性表达噬菌体辅助的连续进化 (Soluble expression phage-assisted continuous evolution, SE-PACE)<sup>[77]</sup>。SE-PACE 是一种快速进化蛋白质的系统,以快速增加其可溶性表达。实验表明,在短短 3 d 内,SE-PACE 进化出了几个抗体片段,表达量提高了 7 倍,同时保持了其结合活性。另外还开发了一种具有独立活性的 SE-PACE 来纠正麦芽糖结合蛋白 (Maltose-binding protein, MBP) 的折叠缺陷变体,并进化出具有改进表达特性的真核细胞胞苷脱氨酶 APOBEC1 变体。当这些进化的 APOBEC1 变体取代野生型胞苷脱氨酶时,可以改善 Cas9 衍生碱基编辑的表达活性。这些结果表明,SE-PACE 可以应用于多种蛋白质并快速提高它们的可溶性表达。

2019 年,David Liu 研究团队又开发了一种利用噬菌体辅助连续进化的碱基编辑器 (Phage-assisted continuous evolution of base editors, BE-PACE),以提高其编辑效率和目标序列兼容性<sup>[78]</sup>。碱基编辑器利用 DNA 修饰酶和催化缺陷的 CRISPR 蛋白精确地引入点突变。迄今为止,已经开发出两类碱基编辑器:胞嘧啶碱基编辑器 (CBE) 可将 C•G 转换为 T•A,腺嘌呤碱基编辑器 (ABE) 可将 A•T 转换为 G•C<sup>[79]</sup>。BE-PACE 每天可进行数十代的突变、选择和复制,从而产生具

有提高目标序列兼容性和更高活性的碱基编辑器。该团队利用 BE-PACE 系统,克服传统 CBE 的靶序列限制的缺陷,进化产生了 3 种类型的胞嘧啶碱基编辑器。进化的 EvoAPOBEC1-BE4max 与野生 APOBEC1 脱氨酶相比,在 GC 背景下编辑胞嘧啶的效率提高 26 倍,并且在所有测试序列上都能有效编辑。该团队还进化出一种 CDA1 脱氨酶的 CBE (CBE-EvoCDA1-BE4max) 变体,对于较难编辑的靶位点有较高的编辑效率。BE-PACE 系统不但可以丰富碱基编辑性能,还可以扩大碱基编辑的应用范围,为碱基编辑系统的优化提供强大工具。这些发现为碱基编辑的快速发展建立了一个系统,并为其使用和改进提供帮助。

### 2.3 流式细胞荧光分选技术

定向进化的主要瓶颈是缺乏快速和通用的高通量筛选 (High-throughput screening, HTS) 方法。流式细胞荧光分选 (Fluorescence activated cell sorting, FACS) 能筛选分离出活性显著改善、底物特异性改变的酶变体<sup>[80]</sup>。FACS 因其高灵敏度和高分析量 (每天多达  $10^8$  个突变体) 的特性从而成为筛选突变文库的强大工具。Tan 等<sup>[81]</sup>开发了一种荧光活化细胞分选系统,建立了首个针对岩藻糖基化酶的单细胞高通量筛选技术平台,能够对这些酶进行高通量筛选 ( $>10^7$  个突变体/h),并设计一个配套策略来快速定量评估单细胞高通量筛选效率;在这项研究中,他们设计和合成了几个用于岩藻糖基化反应的荧光标记底物衍生物。这些岩藻糖基化酶产生了易于保留在细胞中的荧光寡糖产物,从而促进一种新的基于流式细胞仪的岩藻糖基化酶定向进化系统的开发。研究证明这一系统可作为一种通用的筛选工具,用于  $\alpha$ -1,3-岩藻糖基转移酶、 $\alpha$ -1,2-岩藻糖基转移酶,以及岩藻糖苷酶的转糖苷活性等糖基化反应的定向进化。FACS 提高了筛选通量,但并非所有酶都适用 FACS。今后,如何保持基因型和表型之间的联系是 FACS 要面临的主要挑战。

### 2.4 液滴微流控技术

Baret 等<sup>[82]</sup>将 FACS 技术与液滴技术相结合,开发了一种荧光活化液滴分选 (Fluorescence activated droplet sorting, FADS) 系统,可以高通量筛选含氟底物的酶库。Agresti 等<sup>[83]</sup>提出了一种通用的超高通量筛选平台,该平台基于液滴的微流控技术,改变了筛选的规模和速度。该技术已经成为酶定向进化的有力工具,利用超高通量进行初始纯化选择以去除无活性的突变体,以每秒数千次的速度进行分类,确定了辣根过氧化物酶 (Horseradish peroxidase, HRP) 的突变体,该突变体的催化速率比亲本快 10 倍以上。与最先进的机器人筛选系统相比,该系统能够在 10 h 内筛选出  $10^8$  个 HRP 突变体,使用的试剂总量小于 150  $\mu$ L,整个检测速度提高了 1 000 倍,成本降低了约 100 万倍。

筛选庞大的突变文库以实现酶对映体选择性的定向进化一直极具挑战性。以前,改进酶对映体选择性是在微量滴定板上进行的,并且此类方法的通量被限制为每天少于  $10^4$  个菌落<sup>[84]</sup>。基于超高通量液滴的微流控技术可显著提高筛选菌株的能力。之后, Ma 等<sup>[85]</sup>提出“超高通量双通道微流液滴筛选 (Dual-channel microfluidic droplet screening, DMDS)”技术,可每天筛选多达  $10^7$  种酶变体。使用该系统筛选酯酶的对映选择性能获得性质优良的消炎药洛芬烯的对映异构体,已从 500 万个突变体中鉴定出一种对映选择性增加 700 倍的突变体,证明了该筛选平台可用于复杂酶特性 (如区域选择性、化学选择性和对映选择性) 的定向进化。

在以前的研究工作中, Ma 等基于微滴反应器建立了酶的超高通量筛选系统。筛选速度可以高达  $10^8$  个克隆/d,比常规筛选高出 1 000 倍以上,同时大大减少了昂贵的荧光底物的消耗。然而,由于许多商业荧光底物不能适当地保留在微滴反应器中,导致荧光信号在微反应之间扩散,降低了筛选精度和效率,限制了系统的应用范围。Ma 等<sup>[86]</sup>提出了一种酶底物设计新策略,称作“荧光液滴捕获 (Fluorescence droplet entrapment, FDE)”。利用

2 种不同的底物工程策略, 设计了几种 FDE 底物, 用于检测磷酸三酯酶、羧酸酯酶和糖苷酶活性。

### 3 总结与展望

传统突变技术易错 PCR 和 DNA 改组不断引入创新, 而新技术的涌现为定向进化的未来发展提供了一个新视角。本文主要介绍以下几个新的突变技术: 1) 正交 DNA 聚合酶-质粒 OrthoRep 是一种简化的连续进化系统, 加速了定向进化的发展。它能够使目标基因在宿主酵母繁殖时快速稳定地突变和进化。2) 传统的定向进化通常不可能在多个位点进行独立的突变, 而 CRISPR 介导的定向进化可以利用不同的 sgRNAs 定位多个位点, 扩大了突变文库, 并将其应用于动物、植物、微生物等不同物种中, 具有广阔的发展前景。3) 蛋白质半理性设计快速发展, 有针对性地构建高质量的小型突变文库, 进一步降低了筛选工作量。

稳定的高通量筛选方法至关重要。如何使高通量筛选方法适用更广泛的酶功能, 是未来研究的重点。科学家们还需不断地改进现有的技术手段, 研发和优化更高效的文库构建和筛选方法, 为定向进化创造更加诱人的前景。

### REFERENCES

- [1] Packer MS, Liu DR. Methods for the directed evolution of proteins. *Nat Rev Genet*, 2015, 16(7): 379-394.
- [2] Tobin MB, Gustafsson C, Huisman GW. Directed evolution: the 'rational' basis for 'irrational' design. *Curr Opin Struct Biol*, 2000, 10(4): 421-427.
- [3] Doerner A, Rhiel L, Zielonka S, et al. Therapeutic antibody engineering by high efficiency cell screening. *FEBS Lett*, 2014, 588(2): 278-287.
- [4] Bornscheuer UT, Huisman GW, Kazlauskas RJ, et al. Engineering the third wave of biocatalysis. *Nature*, 2012, 485(7397): 185-194.
- [5] Soskine M, Tawfik DS. Mutational effects and the evolution of new protein functions. *Nat Rev Genet*, 2010, 11(8): 572-582.
- [6] Wong TS, Roccatano D, Zacharias M, et al. A statistical analysis of random mutagenesis methods used for directed protein evolution. *J Mol Biol*, 2006, 355(4): 858-871.
- [7] Wong TS, Tee KL, Hauer B, et al. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res*, 2004, 32(3): e26.
- [8] Körfer G, Pitzler C, Vojcic L, et al. *In vitro* flow cytometry-based screening platform for cellulase engineering. *Sci Rep*, 2016, 6: 26128.
- [9] Leung DW, Chen E, Goeddel DV. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique*, 1989, 1: 11-15.
- [10] McCullum EO, Williams BAR, Zhang JL, et al. Random mutagenesis by error-prone PCR//Braman J, Ed. *In vitro* Mutagenesis Protocols. Totowa, NJ: Humana Press, 2010, 634: 103-109.
- [11] Chen K, Arnold FH. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc Natl Acad Sci USA*, 1993, 90(12): 5618-5622.
- [12] Mao CJ, Xie HJ, Chen SG, et al. Error-prone PCR mutation of *Ls-EPSPS* gene from *Liriope spicata* conferring to its enhanced glyphosate-resistance. *Pestic Biochem Physiol*, 2017, 141: 90-95.
- [13] Shao WL, Ma KS, Le YL, et al. Development and use of a novel random mutagenesis method: *in situ* error-prone PCR (*is-epPCR*)//Reeves A, Ed. *In vitro* Mutagenesis. New York, NY: Humana Press, 2017: 497-506.
- [14] Yuan L, Kurek I, English J, et al. Laboratory-directed protein evolution. *Microbiol Mol Biol Rev*, 2005, 69(3): 373-392.
- [15] Porter JL, Rusli RA, Ollis DL. Directed evolution of enzymes for industrial biocatalysis. *ChemBioChem*, 2016, 17(3): 197-203.
- [16] Patel PH, Loeb LA. DNA polymerase active site is highly mutable: evolutionary consequences. *Proc Natl Acad Sci USA*, 2000, 97(10): 5095-5100.
- [17] Chen KQ, Arnold FH. Enzyme engineering for nonaqueous solvents: random mutagenesis to enhance activity of subtilisin E in polar organic

- media. Bio/Technology, 1991, 9(11): 1073-1077.
- [18] Ouyang FJ, Zhao M. Enhanced catalytic efficiency of CotA-laccase by DNA shuffling. *Bioengineered*, 2019, 10(1): 182-189.
- [19] Herr AJ, Ogawa M, Lawrence NA, et al. Mutator suppression and escape from replication error-induced extinction in yeast. *PLoS Genet*, 2011, 7(10): e1002282.
- [20] Gunge N, Sakaguchi K. Intergeneric transfer of deoxyribonucleic acid killer plasmids, pGK11 and pGK12, from *Kluyveromyces lactis* into *Saccharomyces cerevisiae* by cell fusion. *J Bacteriol*, 1981, 147(1): 155-160.
- [21] Ravikumar A, Arrieta A, Liu CC. An orthogonal DNA replication system in yeast. *Nat Chem Biol*, 2014, 10(3): 175-177.
- [22] Ravikumar A, Arzumanyan GA, Obadi MKA, et al. Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell*, 2018, 175(7): 1946-1957.e13.
- [23] Zhong ZW, Ravikumar A, Liu CC. Tunable expression systems for orthogonal DNA replication. *ACS Synth Biol*, 2018, 7(12): 2930-2934.
- [24] Arzumanyan GA, Gabriel KN, Ravikumar A, et al. Mutually orthogonal DNA replication systems *in vivo*. *ACS Synth Biol*, 2018, 7(7): 1722-1729.
- [25] Javanpour AA, Liu CC. Genetic compatibility and extensibility of orthogonal replication. *ACS Synth Biol*, 2019, 8(6): 1249-1256.
- [26] Sternberg SH, Redding S, Jinek M, et al. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 2014, 507(7490): 62-67.
- [27] Ma XL, Zhang QY, Zhu QL, et al. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol Plant*, 2015, 8(8): 1274-1284.
- [28] Gaj T, Gersbach CA, Barbas CR. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*, 2013, 31(7): 397-405.
- [29] Zhang YL, Ma XL, Xie XR, et al. CRISPR/Cas9-based genome editing in plants. *Prog Mol Biol Transl Sci*, 2017, 149: 133-150.
- [30] Feng ZY, Zhang BT, Ding WN, et al. Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res*, 2013, 23(10): 1229-1232.
- [31] Lu YM, Zhu JK. Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol Plant*, 2017, 10(3): 523-525.
- [32] Butt H, Eid A, Momin AA, et al. CRISPR directed evolution of the spliceosome for resistance to splicing inhibitors. *Genome Biol*, 2019, 20: 73.
- [33] Halperin SO, Tou CJ, Wong EB, et al. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature*, 2018, 560(7717): 248-252.
- [34] Hess GT, Frésard L, Han K, et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Methods*, 2016, 13(12): 1036-1042.
- [35] Jakočiūnas T, Pedersen LE, Lis AV, et al. CasPER, a method for directed evolution in genomic contexts using mutagenesis and CRISPR/Cas9. *Metab Eng*, 2018, 48: 288-296.
- [36] Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. *Nature*, 2011, 472(7344): 499-503.
- [37] Ma YQ, Zhang JY, Yin WJ, et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat Methods*, 2016, 13(12): 1029-1035.
- [38] Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, 337(6096): 816-821.
- [39] Camps M, Naukkarinen J, Johnson BP, et al. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proc Natl Acad Sci USA*, 2003, 100(17): 9727-9732.
- [40] Slaymaker IM, Gao LY, Zetsche B, et al. Rationally engineered Cas9 nucleases with improved specificity. *Science*, 2016, 351(6268): 84-88.
- [41] Wang Y, Prosen DE, Mei L, et al. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Res*, 2004, 32(3): 1197-1207.
- [42] Hu JH, Miller SM, Geurts MH, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, 2018, 556(7699): 57-63.
- [43] Anders C, Niewoehner O, Duerst A, et al. Structural

- basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 2014, 513(7519): 569-573.
- [44] Nishimasu H, Shi X, Ishiguro S, et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*, 2018, 361(6408): 1259-1262.
- [45] Korkmaz G, Lopes R, Ugalde AP, et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol*, 2016, 34(2): 192-198.
- [46] Ma YW, Yu L, Zhang X, et al. Highly efficient and precise base editing by engineered dCas9-guide tRNA adenosine deaminase in rats. *Cell Discov*, 2018, 4: 39.
- [47] Orr-Weaver TL, Szostak JW, Rothstein RJ. Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci USA*, 78(10): 6354-6358.
- [48] Szostak JW, Orr-Weaver TL, Rothstein RJ, et al. The double-strand-break repair model for recombination. *Cell*, 1983, 33(1): 25-35.
- [49] Guo XG, Chavez A, Tung A, et al. High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR-Cas9 in yeast. *Nat Biotechnol*, 2018, 36(6): 540-546.
- [50] Barbieri EM, Muir P, Akhuetie-Oni BO, et al. Precise editing at DNA replication forks enables multiplex genome engineering in eukaryotes. *Cell*, 2017, 171(6): 1453-1467.e13.
- [51] Nyerges Á, Csörgő B, Draskovits G, et al. Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *Proc Natl Acad Sci USA*, 2018, 115(25): E5726-E5735.
- [52] Wong TS, Zhurina D, Schwaneberg U. The diversity challenge in directed protein evolution. *Comb Chem High Throughput Screen*, 2006, 9(4): 271-288.
- [53] 曲戈, 赵晶, 郑平, 等. 定向进化技术的最新进展. *生物工程学报*, 2018, 34(1): 1-11.  
Qu G, Zhao J, Zheng P, et al. Recent advances in directed evolution. *Chin J Biotech*, 2018, 34(1): 1-11 (in Chinese).
- [54] 曲戈, 张锟, 蒋迎迎, 等. 2018 诺贝尔化学奖: 酶定向进化与噬菌体展示技术. *生物学杂志*, 2019, 36(1): 1-6, 19.  
Qu G, Zhang K, Jiang YY, et al. The Nobel Prize in Chemistry 2018: the directed evolution of enzymes and the phage display technologies. *J Biol*, 2019, 36(1): 1-6, 19 (in Chinese).
- [55] Lutz S. Beyond directed evolution-semi-rational protein engineering and design. *Curr Opin Biotechnol*, 2010, 21(6): 734-743.
- [56] Cheng F, Zhu LL, Schwaneberg U. Directed evolution 2.0: improving and deciphering enzyme properties. *Chem Commun (Camb)*, 2015, 51(48): 9760-9772.
- [57] O'Maille PE, Bakhtina M, Tsai MD. Structure-based combinatorial protein engineering (SCOPE). *J Mol Biol*, 2002, 321(4): 677-691.
- [58] Widersten M, Hansson LO, Tronstad L, et al. Use of phage display and transition-state analogs to select enzyme variants with altered catalytic properties: glutathione transferase as an example. *Methods Enzymol*, 2000, 328: 389-404.
- [59] 曲戈, 朱彤, 蒋迎迎, 等. 蛋白质工程: 从定向进化到计算设计. *生物工程学报*, 2019, 35(10): 1843-1856.  
Qu G, Zhu T, Jiang YY, et al. Protein engineering: from directed evolution to computational design. *Chin J Biotech*, 2019, 35(10): 1843-1856 (in Chinese).
- [60] Lassila JK. Conformational diversity and computational enzyme design. *Curr Opin Chem Biol*, 2010, 14(5): 676-682.
- [61] Qu G, Li AT, Acevedo-Rocha CG, et al. The crucial role of methodology development in directed evolution of selective enzymes. *Angew Chem Int Ed Engl*, 2020, 59, DOI: 10.1002/anie.201901491.
- [62] Fox RJ, Davis SC, Mundorff EC, et al. Improving catalytic function by ProSAR-driven enzyme evolution. *Nat Biotechnol*, 2007, 25(3): 338-344.
- [63] Gonzalez-Perez D, Molina-Espeja P, Garcia-Ruiz E, et al. Mutagenic organized recombination process by homologous in vivo grouping (MORPHING) for directed enzyme evolution. *PLoS ONE*, 2014, 9(3): e90919.
- [64] Chen MMY, Snow CD, Vizcarra CL, et al. Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. *Protein Eng Des Sel*, 2012, 25(4):

- 171-178.
- [65] Treynor TP, Vizcarra CL, Nedelcu D, et al. Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. *Proc Natl Acad Sci USA*, 2007, 104(1): 48-53.
- [66] Jamil S, Liu MH, Liu YM, et al. Hydrophobic mutagenesis and semi-rational engineering of arginine deiminase for markedly enhanced stability and catalytic efficiency. *Appl Biochem Biotechnol*, 2015, 176(5): 1335-1350.
- [67] Shin KC, Choi HY, Seo MJ, et al. Improved conversion of ginsenoside Rb<sub>1</sub> to compound K by semi-rational design of *Sulfolobus solfataricus*  $\beta$ -glycosidase. *AMB Express*, 2017, 7: 186.
- [68] Smith GP. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 1985, 228(4705): 1315-1317.
- [69] Bradbury A. Recent advances in phage display: the report of the Phage Club first meeting. *Immunotechnology*, 1997, 3(3): 227-231.
- [70] Siegel DL. Research and clinical applications of antibody phage display in transfusion medicine. *Transfus Med Rev*, 2001, 15(1): 35-52.
- [71] McCafferty J, Griffiths AD, Winter G, et al. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*, 1990, 348(6301): 552-554.
- [72] Smothers JF, Henikoff S, Carter P. Tech. Sight. Phage display. Affinity selection from biological libraries. *Science*, 2002, 298(5593): 621-622.
- [73] Azzazy HME, Highsmith WE Jr. Phage display technology: clinical applications and recent innovations. *Clin Biochem*, 2002, 35(6): 425-445.
- [74] Frenzel A, Schirrmann T, Hust M. Phage display-derived human antibodies in clinical development and therapy. *mAbs*, 2016, 8(7): 1177-1194.
- [75] Frei JC, Lai JR. Protein and antibody engineering by phage display. *Methods Enzymol*, 2016, 580: 45-87.
- [76] Cung K, Slater RL, Cui Y, et al. Rapid, multiplexed microfluidic phage display. *Lab Chip*, 2012, 12(3): 562-565.
- [77] Wang TN, Badran AH, Huang TP, et al. Continuous directed evolution of proteins with improved soluble expression. *Nat Chem Biol*, 2018, 14(10): 972-980.
- [78] Thuronyi BW, Koblan LW, Levy JM, et al. Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat Biotechnol*, 2019, 37(9): 1070-1079.
- [79] Rees HA, Liu DR. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet*, 2018, 19(12): 770-788.
- [80] Yang GY, Withers SG. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *Chem Bio Chem*, 2009, 10(17): 2704-2715.
- [81] Tan YM, Zhang Y, Han YB, et al. Directed evolution of an  $\alpha$ 1,3-fucosyltransferase using a single-cell ultrahigh-throughput screening method. *Sci Adv*, 2019, 5(10): w8451.
- [82] Baret JC, Miller OJ, Taly V, et al. Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. *Lab Chip*, 2009, 9(13): 1850-1858.
- [83] Agresti JJ, Antipov E, Abate AR, et al. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci USA*, 2010, 107(9): 4004-4009.
- [84] Liebeton K, Zonta A, Schimossek K, et al. Directed evolution of an enantioselective lipase. *Chem Biol*, 2000, 7(9): 709-718.
- [85] Ma FQ, Chung MT, Yao YB, et al. Efficient molecular evolution to generate enantioselective enzymes using a dual-channel microfluidic droplet screening platform. *Nat Commun*, 2018, 9: 1030.
- [86] Ma FQ, Fischer M, Han Y, et al. Substrate engineering enabling fluorescence droplet entrapment for IVC-FACS-based Ultrahigh-throughput screening. *Anal Chem*, 2016, 88(17): 8587-8595.

(本文责编 陈宏宇)