合成生物学使能技术的研究进展

李雷, 姜卫红, 覃重军, 薛小莉*, 芦银华*

中国科学院上海生命科学研究院植物生理生态研究所, 中国科学院合成生物学重点实验室, 上海 200032

* 联系人, E-mail: yhlu@sibs.ac.cn; xlxue@sibs.ac.cn

收稿日期: 2015-03-15; 接受日期: 2015-05-26

国家重点基础研究发展计划(批准号: 2011CBA00801, 2012CB721102)、国家高技术研究发展计划(批准号: 2012AA021703)、国家自然科学基金(批准号: 31200059)和国家自然科学基金委员会-中国科学院学科发展战略研究合作项目(批准号: L1222037)资助
doi: 10.1360/N052015-00048

摘要 作为一门拥有巨大潜力的新兴工程学科, 合成生物学的发展主要得益于各种使能技术(enabling technology)的创新开发与应用. 从基本功能元件的构建与标准化, 到高通量的微芯片基因合成技术与各种尺度(从 bp 至 Mb)的 DNA 拼接组装方法, 再到强大的基因组编辑工具, 在过去十几年里合成生物学使能技术取得了长足的进步. 同时, 新颖的使能技术也为遗传学、癌症治疗、疾病监测以及生物制造等领域提供了优秀的研究工具, 促进了多个学科的发展. 如果将这些使能技术作为“配件工具”, 那么相对应的“主体设备”——底盘细胞也因工具的不断創新得到了快速发展. 微生物最小基因组的分析以及对基因组的连续删减优化, 为构建一个具有可预测、可控制表型的优良底盘细胞奠定了基础. 为促进基于细胞疗法的人类疾病治疗, 哺乳动物细胞作为底盘细胞也正在开发中. 本文对合成生物学使能技术的最新发展进行了深入总结和梳理, 探讨了这些使能技术在合成生物学乃至整个生命科学研究中的应用及其重要意义.

关键词合成生物学
使能技术
生物元件
基因合成
DNA 组装
基因组编辑
底盘细胞

合成生物学是一门新兴的学科, 利用工程与计算的原理重新设计生命系统, 使其具备新的特征, 例如, 通过设计改造微生物以产生新的化合物, 治疗人类疾病或解决环境问题等^[1,2]. 以拨动开关(toggle switch)与合成振荡器(synthetic oscillator)为标志, 合成生物学在 2000 年正式诞生^[3,4]. 虽然只有短暂的十几年, 该学科已经取得了长足的进展, 其中具有代表性的事件包括: (i) 各种高阶遗传环路的构建, 如边缘探测器(edge-detection circuit)与多细胞模式形成(multicellular pattern formation)等^[5,6]; (ii) 青蒿素前体青蒿酸在微生物中的高效合成^[7]; (iii) 嵌合抗原受体 T 细胞癌症免疫疗法的开发^[8,9]. 在此过程中, 作为一门拥有巨大潜力的工程学科, 合成生物学创新

发展了一系列优秀的使能技术.

首先, 为实现在生命系统的理性设计, 生物元件的标准化与模块化一直是合成生物学家追求的目标^[10]. 从启动子、核糖体结合位点(ribosome binding site, RBS)到终止子, 目前均实现了强度跨度大、梯度密集的元件库的构建, 并且其强度可进行软件预测, 为后续生物合成途径的代谢流平衡优化与复杂遗传环路的设计等提供了坚实的保障^[11-14]. 其次, DNA 合成与组装技术作为合成生物学最基础的工具, 其发展很大程度上决定着人们对生命系统的改造能力. 如今, 基因合成借助于芯片技术与下一代高通量测序平台, 已从传统的柱式合成(column-based synthesis)发展为高通量、高保真的芯片合成(array-based

引用格式: 李雷, 姜卫红, 覃重军, 等. 合成生物学使能技术的研究进展. 中国科学: 生命科学, 2015, 45: 950-968

Li L, Jiang W H, Qin Z J, et al. Recent advances in the enabling technologies for synthetic biology. SCIENTIA SINICA Vitae, 2015, 45: 950-968, doi: 10.1360/N052015-00048

synthesis)^[15-17]. 同时, 不同尺度(从 bp 至 Mb)的高效 DNA 组装方法也相继被开发^[18], 尤其是 Gibson 等人^[19]的等温一步法和 Noskov 等人^[20]发展的酿酒酵母(*Saccharomyces cerevisiae*)体内组装. 最后, 对生命系统的改造已经从单基因的操作, 进入了对整个基因组设计改造的时代^[21]. 无论是对遗传物质的直接修改还是转录或翻译层面上的整体优化, 全局扰动的理念产生了各种优秀的基因组编辑工具, 如多元自动化基因组工程(multiplex automated genome engineering, MAGE)^[22]、全局转录工程(global transcription machinery engineering, gTME)^[23,24]与 CRISPR/Cas9(cluster regularly interspaced short palindromic repeats/CRISPR-associated endonuclease)基因组编辑技术^[25]等. 这些使能技术的创新发展也促进了理想底盘细胞的构建, 从而为不同遗传环路与生物合成途径等提供了适配的反应系统(图 1).

合成生物学使能技术的快速发展, 不仅为人们生命现象及其本质的认识提供了便捷、高效的研究工具, 同时也极大地促进了包括生物制造、癌症治疗、疫苗研发以及疾病或环境污染物监测等多个领域的迅速发展. 例如, 通过改造蓝藻(*Cyanobacteria*)直接利用温室气体二氧化碳生产化工品^[26], 基于 DNA 折纸技术与基因环路开发的埃博拉病毒检测试纸^[27], 基于嵌合抗原受体的癌症免疫疗法^[8]以及感应重金属砷的生物传感器的开发^[28]等.

1 生物元件的鉴定与标准化

基于“自下而上”的合成生物学建构理念和对生

图1 合成生物学使能技术概要(网络版彩图)

命系统精确控制的目标, 生物元件(包括调控元件与结构元件)的挖掘与标准化是合成生物学最基础的环节之一. 由此, 各种合成生物学元件被不断挖掘, 并建立了相应的标准化数据库, 如 iGEM Registry 与 SynBERC registry 等(<http://parts.igem.org> 与 <https://registry.synberc.org>). 2013 年, Kahl 和 Endy^[29]对合成生物学使能技术的应用情况进行调研时发现, 很多标准化的元件已被普遍使用, 如 BBa_J23100 系列启动子、BBa_B0015 双向终止子与 BBa_115010 Cph1 感光受体与 EnvZ 的嵌合蛋白等. 从启动子、RBS、基因间隔区到终止子, 这些调控元件在不同的层面上控制着基因的表达水平, 对其进行系统鉴定与标准化将显著提高人们对反应系统的调控能力. 同时, 作为生化反应的实体——结构元件的开发, 如合成途径关键酶的鉴定^[7]与嵌合光受体的构建^[30]等, 也将进一步拓展改造生命系统的潜能.

2009 年, Purnick 和 Weiss^[31]对于合成生物学调控网络构建的基本功能元件进行了梳理, 总结于下表 1 中. 本文主要对启动子、RBS 与终止子的鉴定与设计进行了总结.

1.1 启动子

作为控制基因表达最重要的一环, 启动子是最早被合成生物学家鉴定与标准化的元件. 2005 年由 Alper 等人^[11]、Siegl 等人^[32]、Blazeck 和 Alper^[33]及 Tornøe 等人^[34]引入启动子库以精细调控代谢途径的理念后, 多种宿主细胞(如大肠杆菌(*Escherichia coli*)、酵母以及放线菌(*Actinobacteria*)等)启动子库相继被开发. 除了构建启动子文库外, 近几年来, 关于启动子设计也取得了很大的进步.

(1) 通过调节染色质结构(如修饰核小体结合序列)改变启动子强度. 通常, 人们一般利用易错 PCR 方式, 随机突变启动子核心区域或其间隔区序列以改变启动子的强度. 2014 年, Curran 等人^[35]发现, 在酵母中通过调控启动子区附近核小体的占有量也可影响启动子强度, 最高获得了 16 倍的强度差异. 但由于实用性较小, 这种设计主要体现为理念的新颖性, 进一步扩展了人们改造启动子的策略.

(2) 绝缘启动子的开发. 虽然启动子强度可利用报告基因进行精确定量, 但当更换不同的宿主或控制不同基因表达时, 由于不同转录因子的干扰或不同 mRNA 二级结构的形成, 其标定的强度往往就不可靠了. 2011 年, Davis 等人^[36]在原核启动子-35 区

表 1 用于构建合成调控网络的基本元件^[31]

元件	范例	功能
转录水平		
组成型启动子	<i>lacIq</i> , <i>CMV</i> , <i>EF1</i> , <i>UBC</i> , <i>SV40</i> , <i>T7</i> , <i>sp6</i> , <i>RSV36</i> 与 <i>U6</i>	总是起始转录
调控区	<i>tetO</i> , <i>lacO</i> , <i>cuO</i> , <i>ara</i> , <i>EST</i> , <i>glnA</i> , <i>Or</i> , <i>UASG</i> , <i>NRI</i> , <i>gal4</i> 与 <i>rhl box</i>	抑制子和激活子结合位点
诱导型与组织特异型启动子	<i>ara</i> , <i>stress</i> , <i>trc</i> , <i>nitric oxide</i> , <i>FOS</i> , <i>ethano</i> , <i>lac</i> , <i>gal</i> , <i>rhl</i> , <i>lux</i> , <i>FUS1</i> , <i>FIG1</i> , <i>ste5</i> , <i>ade2</i> , <i>lys21</i> , <i>fdhH</i> , <i>TEF</i> , <i>SSRE</i> , <i>sal</i> , <i>glnK</i> , <i>cyc1</i> 与 <i>CAG</i>	通过诱导或细胞状态控制启动子的转录
细胞命运调节子	<i>GATA</i> 因子, <i>MYOD</i> 与 <i>NGN1</i>	控制细胞分化
翻译水平		
RNAi	逻辑功能与 RNAi 抑制子	构建遗传开关、逻辑门
核糖调节子(Riboregulators)	配体控制的核酶	调控开关
RBS	RBS 计算器	控制翻译
翻译后水平		
磷酸化级联反应	酵母磷酸化途径与 MAPK 信号传递骨架	调制遗传环路的输出
蛋白受体设计	TNT 受体, <i>ACTR</i> 与 <i>EST</i> 受体	控制检测阈值和不同功能域的組合
蛋白降解	<i>Ssra</i> 标签与富含脯氨酸、谷氨酸、丝氨酸或苏氨酸的小肽	以不同速率控制蛋白降解
位置信号	<i>NLS</i> , <i>NES</i> 与 <i>mito</i>	控制细胞核或线粒体蛋白的输入/输出
多细胞水平		
细胞间通讯	<i>AHL</i> , <i>ADH/乙醛</i> , <i>NRL/乙酸</i> 与 <i>CRE1/IRT4</i>	利用可扩散的小分子控制细胞间的通讯
荧光报告系统	<i>EGFP</i> , <i>EYFP</i> , <i>ECFP</i> , <i>LacZ</i> , <i>DsRed</i> 与 <i>ZsYellow</i>	表达水平的检测
抗生素	<i>amp</i> , <i>chlor</i> , <i>kan</i> , <i>bla</i> , <i>puro</i> , <i>bleo</i> 和 <i>neo</i>	不同细胞系的筛选

与-10 区临近区域处(-105 至+55 bp), 添加了一段特殊的绝缘序列, 巧妙地规避了上述两个缺点, 提高了启动子库的通用性及其强度的预测能力。

(3) 采用人工神经网络模型直接预测启动子强度。随着越来越多启动子文库的建立, 人们对启动子序列如何影响其强度有了更深的认识。因此, 借助具有自我学习、归纳能力的人工神经网络系统, 并运用计算机辅助设计, 可以直接获得所需强度的启动子序列, 其准确度达到 98%^[37]。

(4) 动态响应启动子的挖掘。组成型启动子文库建立后, 人们也构建了一系列诱导性启动子的文库。虽然理论上这些文库可以实现对代谢途径的精细调控, 但耗时耗力, 而且如果使用诱导剂也会增加工业生产成本。2013 年, Keasling 研究组^[38]借助动态响应的理念, 利用 DNA 芯片直接筛选响应青蒿二萜合成途径毒性中间产物的启动子, 在此基础上构建了正、负反馈回路。该策略不仅简化了人为调节的过程, 而且青蒿二萜产量也比采用组成型或诱导性启动子时有显著提高。

1.2 RBS 与其他翻译调控元件

虽然转录水平的控制十分重要, 但对翻译效率

的准确预测有助于优化重组蛋白的表达或目标代谢产物的高效合成。2009 年, Voigt 研究组^[12]利用热动力学模型, 开发出一种稳健的 RBS 计算器(RBS calculator), 实现了对 RBS 强度的有理设计, 其动态范围达到 100000 倍。同时, 当采用不同遗传材料时, 可准确预测出相同 RBS 对应的不同蛋白表达水平。由于 RBS 序列一般较短, 可利用 PCR 的方式便捷地引入至目标基因上游, 相比启动子的更换, 可以实现对基因表达强度的更加快速地调节。

最近, 一种新颖的控制翻译效率的方法被开发——两个氨基酸寡肽形成的上游开放阅读框(upstream open reading frames, uORF)直接抑制下游基因的翻译^[39]。当更换不同氨基酸时, 其抑制效率可发生改变, 从而构建了不同抑制强度的翻译调控元件。在作者检测的所有哺乳动物细胞系中, uORF 的调控效率均可实现较为准确的预测。

1.3 终止子

在构建复杂的遗传环路或多基因代谢途径时, 为避免不同转录单元的相互影响, 终止子, 尤其是强终止子的使用显得尤为重要。同时, 序列多样性终止子的开发, 也可以避免同源重组的发生, 进而提高代

谢途径或遗传环路的结构稳定性. 2013 年, Endy 与 Voigt 两个研究组^[13,14]先后报道了大肠杆菌终止子库的筛选与鉴定. 在 Voigt 的工作中, 作者先测量了大肠杆菌中全部注释的 227 个天然终止子强度, 在分析颈环结构(stem-loop)与尿嘧啶碱基束(U-tract)如何影响终止子强度的基础上, 重新构建了 265 个人工终止子库, 获得了一系列转录终止效果显著的终止子元件^[14].

综上所述, 目前已完成了多种不同来源的启动子、RBS 与终止子元件库的构建, 并基本实现了运用软件进行强度预测, 从而使大家对于生命系统的控制变得更加精确, 减少了试错的过程. 当然, 除了上述 3 种经典的调控元件外, 越来越多新颖的元件不断被开发, 如核酸开关(riboswitch)^[40]、核酸调节子(ribo regulator)^[41]和可调控基因间序列(tunable intergenic regions, TIGRs)^[42]等顺式作用元件, 以及转录激活样效应子(transcription activated-like effector, TALE)^[43]、CRISPR/Cas9 系统^[44]和倒转重组酶^[45]等反式元件. 丰富多样的功能元件的挖掘与标准化, 将加快人们对遗传环路、代谢途径乃至整个基因组进行更为简便、高效的设计与优化.

2 DNA 合成与组装技术

DNA 合成与组装技术是合成生物学最基础的工

具, 在实现了对生物元件的鉴定、标准化后, 功能模块、高阶系统以及完整生命体的构建, 都依赖于这些技术的发展. 尤其在代谢工程领域, 外源基因的密码子优化与代谢途径流量的快速调试, 都迫切地需要廉价的 DNA 合成技术与高效、便捷的组装方法. 2010 年, 人造生命蕈状支原体“Synthia”的诞生, 更是得益于酵母体内拼接等微生物基因组尺度的 DNA 组装方法的开发^[46]. 最近几年, 由于芯片技术与高通量测序的引入, 自动化、高通量、高保真的基因合成技术开始出现, 虽然还未实现商业化, 但发展十分迅速. 同时, 不同尺度的体内、体外 DNA 组装方法也相继诞生, 基本满足了不同类型的反应系统的构建(表 2).

2.1 DNA 合成

DNA 化学合成, 主要分为寡核苷酸合成(oligo synthesis)与基因合成(gene synthesis). 寡核苷酸合成一般采用固相亚磷酸胺三酯法(phosphoramidite chemistry), 以核苷酸单体为原料, 经过脱保护、偶联、封闭与氧化 4 步循环反应而得到^[47,48]. 这种方法经过后期的优化, 以 1/200 的错误率最长可合成 200~300 nt 的寡核苷酸序列^[17]. 20 世纪 90 年代早期, 芯片技术的引入实现了寡核苷酸的高通量合成^[58,59]. 但是由于“脱嘌呤”现象与“边缘效应”的影响, 其合成序列的准确性受到了挑战. 不过, 在 Agilent 开发的基于喷墨打印技术的芯片合成与 CustomArray 开发

表 2 常见 DNA 合成与组装方法

合成产物	大小	方法	优势	劣势	应用范围
寡核苷酸	~200 bp	固相亚磷酸胺三酯法	商业化、自动化	基于柱式反应, 合成成本较高	引物合成与基因合成主要来源 ^[47,48]
基因	~10 ³ bp	连接酶链式反应	组装错误率低, 特异性比 PCR 更好	不允许寡核苷酸间存在缺口	含重复序列的基因合成 ^[49]
途径或环路	~10 ⁴ bp	聚合酶循环拼接	允许寡核苷酸间存在缺口	PCR 可能引入突变	基因合成常规方法 ^[16,50]
		Biobrick/Bglbrick	标准化组装	产生疤痕, 耗时较长	标准元件库与小的遗传环路的构建 ^[51,52]
		环式聚合酶延伸克隆	无疤痕, 仅需 Phusion 聚合酶	PCR 可能引入突变	复杂组合元件库的克隆 ^[53]
		Golden gate 克隆	无疤痕, 无需引入同源臂	耗时	重组基因建库与 TALEs 构建 ^[54]
		连接酶循环反应	无疤痕, 一次可组装 20 个小片段	设计较为繁琐	小的遗传环路构建 ^[55]
基因组	~10 ⁵⁻⁶ bp	连接与序列非依赖克隆	标准化组装, 仅需 T4 DNA 聚合酶	产生疤痕, 设计较为繁琐	重组质粒与多蛋白复合物的构建 ^[56]
		Gibson 等温一步法	组装效率高, 组装尺度范围广(10 ² ~10 ⁵ bp), 一次可组装多个片段	组装酶系价格偏高	不同尺度代谢途径和小的基因组组装 ^[19]
		酵母体内拼接	一次可组装多个片段, 组装尺度大	相对体外组装较为耗时	微生物基因组的组装 ^[46,57]

的基于半导体电化学酸合成后, 出错率得到了显著的改善, 基本与柱式合成相当^[60,61]。同时, 其单次合成数量却是柱式合成的 100~10000 倍, 因此极大地降低了 DNA 合成成本。

寡核苷酸的合成主要用于常规引物的合成, 也是基因合成的主要来源。基因合成借助连接酶链式反应(ligase chain reaction, LCR)或聚合酶循环组装(polymerase cycling assembly, PCA), 利用具有重叠区的寡核苷酸逐步组装获得。但当需要大规模基因合成时, 例如, 基因组的组装时, 由于其传统方法通量低、成本高, 其应用受到了限制。因此, 借助芯片合成与原位组装的理念, 最近几年已迅速开发出了多种高通量的基因合成技术。

在应用芯片技术进行基因合成时, 相对寡核苷酸的合成, 主要有两个问题: (i) 虽然一次合成寡核苷酸数量很多, 但每一条寡核苷酸的浓度却很低, 不能满足基因合成的需要; (ii) 寡核苷酸序列可能存在相似性, 如高 GC 含量的序列, 导致基因错误组装。针对这些问题, 2004 年, 田敬东等人^[62]提出了解决方案, 他们首先通过扩增获得高浓度的寡核苷酸序列, 满足了浓度的需求; 其次, 应用杂交原理直接筛选正确的寡核苷酸序列用于基因的合成, 从而实现了基于芯片的基因合成。但是, 由于这种方法是基于杂交筛选的原理, 很难实现规模化, 而且要求寡核苷酸序列具有很好的正交性, 因此一直没有得到广泛运用。不过后来, 他们利用喷墨合成器(ink-jet synthesizer)与隔离微孔(microwells), 一次可进行 10000 条寡核苷酸以上的合成, 铺平了商业化的道路^[16]。

在基因合成中, 无论是寡核苷酸的合成还是后期的组装, 都有可能产生错误。尤其是当需要高通量、自动化的合成时, 高保真的出发片段将显得尤为重要。目前, 主要采用两种方法对出错率加以控制: (i) 借助酶直接修复出错位点, 如商业化的酶系 ErrASE, 包含了外切酶、内切酶与解离酶等, 成功地减少了出错的频率^[15]; (ii) 利用下一代高通量测序技术, 直接从源头解决问题, 边合成边测序, 这也是将来高保真、高通量基因合成商业化的必然趋势^[63]。

2.2 不同尺度的 DNA 组装方法

为了保证合成基因的序列准确性, DNA 合成长度一般不超过 5 kb, 更大尺度 DNA 分子的合成将借助于最近发展的 DNA 组装方法。除了传统的分子克

隆, 目前已诞生了依赖于连接、聚合或同源重组原理等多种体内、体外 DNA 组装新方法。其中, 体外 DNA 组装方法的代表有 Gibson 等温一步法^[19]、序列与连接非依赖性克隆(sequence and ligation-independent cloning, SLIC)^[56]、连接酶拼接法(ligase cycling reaction, LCR)^[55]与 Golden gate 克隆^[54]等; 体内主要是酵母转化偶联重组(transformation-associated recombination, TAR)介导的 DNA 拼接^[20,57]。这些方法的开发, 使得 DNA 组装无论是尺度还是效率上, 均有了质的飞跃。本文主要介绍两类被广泛使用的 DNA 组装方法——Gibson 等温一步法和酵母体内拼接方法。

2008 年, Venter 研究组^[57]完成了 580 kb 生殖道支原体(*Mycoplasma genitalium*)的人工构建, 紧接着又于 2010 年实现了人造生命“Synthia”(蕈状支原体, *Mycoplasma mycoides*, 1.08 M)的重头合成^[46], 掀起了合成生物学新的浪潮。这些突破性的进展, 除了依靠高通量、低成本的基因合成技术外, 主要得益于 Gibson 组装方法与酵母体内拼接方法的开发。Gibson 组装方法中最具代表性的方法是 Gibson 等温一步法。该方法主要依靠一组酶系: T5 外切酶、Phusion DNA 聚合酶与 Taq DNA 连接酶。在片段与载体、片段与片段间引入重叠区后, 利用外切酶产生黏性末端, 聚合酶与连接酶加以修复, 从而实现多个片段的一次性无痕拼接, 其目前组装的最大尺度为 580 kb^[19]。

酿酒酵母由于自身具有高效的同源重组能力, 早在 2002 年就由 Noskov 等人^[20]提出了酵母转化偶联重组的理念, 后期被广泛地用于各种尺度 DNA 组装, 实现了生物元件库、遗传环路以及微生物基因组的快速构建。例如, 除了上述支原体的人工构建外, Gibson 等人^[64]又尝试了在酵母体内一次性组装 38 个寡核苷酸序列的目标。赵惠民研究组^[65,66]也在同年开发了类似的组装方法——DNA 装配器(DNA assembler), 实现了 D-木糖代谢途径与玉米黄质合成途径的快速构建, 也成功用于组合转录工程, 快速对纤维素降解途径进行了流平衡优化。

3 基因组编辑工具

在分子生物学革命之前, 对于生命的改造一直依赖于随机诱变(random mutagenesis)与选择育种(selective breeding), 如抗生素高产菌株的物理或(和)

化学诱变与农作物、家畜的驯化等. 虽然那时对于“生命是什么”的认识还很缺乏, 但这些例子都涉及到了对生命整个基因组的修饰. 后来, 转座子的发现提供了一种新的工具——利用转座子随机插入的特点, 可直接筛选与重要表型相关的功能基因. 事实上, 随着人类基因组测序的完成, 人们不再局限于对单个基因的操作, 为对基因组水平的有序设计, 各种高效、新颖的基因组编辑工具应运而生. 本节从 DNA、RNA 与蛋白质 3 种不同生物大分子层面开发获得的基因组编辑工具进行简要阐述.

3.1 基因组工程

DNA 合成与组装技术的迅速发展, 促进了基因组工程(genome engineering)的诞生. 无论是单一基因的靶向敲除, 还是基因组的多重编辑, 都有赖于高效的基因组操作工具. 此外, 这些工具的创新应用, 也促进了遗传模型的构建、致病基因的筛选与微生物代谢工程等研究工作.

(1) 靶向基因组工程(targeted genome engineering). 靶向基因组编辑工具主要有两类: (i) 重组酶, 如环化重组酶(cyclization recombination enzyme, Cre)与翻转酶(flipase, Flp)等; (ii) 基于模块化元件构建的核酶或转录调节子, 如锌指蛋白(zinc-finger protein, ZFP)、转录激活样效应子(TALE)和 CRISPR/Cas9 系统等(图 2A). 重组酶是一类催化两个高度特异位点间 DNA 序列剪切反应的核酶, 如 Cre-*loxP* 系统. 当两个 *loxP* 位点方向相同时, Cre 将位点间的 DNA 序列敲除, 而方向相反时, 这将序列进行倒转. Cre 重组酶来源于噬菌体 P1, 已被广泛开发用于多种生物的基因敲除. 但这仅是对少数基因进行敲除, 无法实现大规模、多个基因的同时编辑. 不过, 在 2014 年, *Science* 杂志报道的一篇利用 Cre-*loxP* 系统进行酵母染色体简化的工作, 突破了这一瓶颈^[67]. 文章主要利用“自上而下”的理念来构建基因组精简的酵母染色体. 在合成每个基因时, 两侧都引入了 *loxP* 位点, 然后组装成完整的染色体. 通过控制 Cre 的表达量与多次删减, 利用菌株自适应选择, 可迅速将大量非必需基因剔除. Cre-*loxP* 系统的灵活应用, 避免了人们对必需基因认识的不足, 加快了基因组精简优化的过程, 也实现了合成生物学“自下而上”合成与“自上而下”删减 2 种策略的巧妙结合.

然而, 在采用重组酶进行基因组编辑时, 需要将

特定的识别位点引入, 这增加了操作的复杂性. 最近, 一系列模块化的生物元件被开发, 这些元件可以人为设计直接靶向结合目标序列, 不再需要引入特定识别位点. 首先被发现的是 ZFP. ZFP 是一类由小的锌指模块串联而成的 DNA 结合蛋白, 每 3 个模块可以特定识别一个碱基^[68]. ZFP 可以连接其他效应结构域, 如核酶 Fox I 与转录激活结构域 VP64 等, 然后靶向编辑目标序列, 其中应用最为广泛的是锌指核酶(zinc-finger nucleases, ZFNs). 但由于其识别特异性较差导致明显的脱靶效应, 模块组装时操作复杂, 成本过高以及蛋白本身对细胞具有一定的毒害, ZFNs 现在基本不再使用, 替代这种早期技术的是转录激活样效应子核酶(transcriptional activator-like effector nucleases, TALENs)与 CRISPR/Cas9 系统.

转录激活样效应子(TALEs), 是第二类新型的模块化元件, 来源于黄单胞菌属(*Xanthomonas*), 它借助细菌三型分泌系统侵染植物, 干扰宿主的生长或引发疾病^[25,43]. 天然 TALEs 蛋白的 N 端为蛋白分泌和转位结构域, C 端包含核定位序列与转录激活结构域. 中间即是可人为根据靶序列设计的 DNA 结合结构域. 为什么可人为设计呢? 主要原因是这个结构域是由多个几乎完全相同的基序(motif)组装而成, 每个基序由 33~35 个氨基酸组成, 识别一个特定碱基. 在不同基序中仅 12 与 13 位氨基酸有差异(称为重复可变的双残基, repeat variable di-residues, RVDs). 由此, 任何序列(几十 bp 长度)均可相应设计出特异识别的 TALE 蛋白, 实现蛋白与 DNA 的靶向结合. TALE 蛋白也被开发用于基因的敲除、激活或抑制, 其中研究最为广泛的是 TALENs. 其主要的缺点是组装比较繁琐, 较难实现规模化.

最近一两年, 一类强大的模块化编辑工具——CRISPR/Cas 系统被发现. CRISPR/Cas 是一类来源于细菌或古菌的天然防御系统, 抵抗外源 DNA 的入侵^[69]. 其中 II 型 CRISPR/Cas 系统由于其 Cas 蛋白(如 Cas9)为单一模块, 不像 I 型与 III 型 Cas 蛋白为多亚基组成的复合物, 其应用最为普遍. 当外源质粒或病毒进入细胞后, CRISPR 会转录形成一个顺式激活 RNA(trans-activating crRNA, tracrRNA)和系列反式 RNA(CRISPR RNA, crRNA), 两者形成功能性的双链 RNA 二级结构, 直接结合外源 DNA, 并引导 Cas9 核酶进行剪切, 从而加以破坏. 2012 年, Jinek 等人^[70]发现, tracrRNA 与 crRNA 可直接嵌合形成一个双链

RNA 分子(guide RNA, gRNA)发挥功能. 除了对靶基因序列切割位点毗邻基序(protospacer adjacent motif, PAM)有简单的要求外(如来自酿脓链球菌(*Streptococcus pyogenes*)的 Cas9 核酶靶向识别的 PAM 结构需要三核苷酸 NGG 序列), 可人为设计 gRNA 编辑基因组任意的位点. 唯一不足的是由于 Cas9 蛋白自身具有核酶功能, 仅需一个 gRNA 就可识别并切割 DNA, 相对需要一对 TALENs 才能发挥 DNA 切割活性而言, 识别特异性较差, 容易导致脱靶效应. 不过, Tsai 等人^[71]在突变 Cas9 蛋白(保留 DNA 结合活性, 丧失 DNA 切割活力)后添加外源核酸内切酶 *Fox I*, 通过模仿 TALENs, 提高了该系统的靶向特异性. 后来, 多个研究组^[72-74]发现通过适当缩短 gRNA 与 DNA 碱基配对的长度或抑制 Cas9 非同源末端连接的活性, 都能有效降低该系统的脱靶效应, 推进了其临床应用的步伐. Boettcher 和 McManus^[75]在最新一期综述中, 也详细对比了 TALENs 与 CRISPR/Cas9 系统的优缺点, 可指导研究者选择最佳的基因组研究工具.

由于 CRISPR/Cas9 系统中模块化元件十分简单, 无需像 TALE 蛋白进行繁琐组装, 其后应用一发而不可收. 除了构建转录激活子、抑制子和靶向荧光定位蛋白外(失活的 Cas9 蛋白分别与激活域、抑制结构域或荧光蛋白结合), 最重要的是直接用于基因突变或敲除. 除了在微生物中实现靶向编辑外, 重要的农作物如水稻(*Oryza sativa*)、小麦(*Triticum aestivum*)等^[76], 模型动物如小鼠(*Mus musculus*)、果蝇(*Drosophila melanogaster*)和线虫(*Caenorhabditis elegans*)等^[77-79], 都实现了内源基因的定点修饰, 其构建过程简单, 一步可获得纯合突变体. 同时, CRISPR/Cas 系统也实现了同时对 5 个基因进行定点突变, 其效率达到 80%^[80]. 如今, 该方法又被开发用于遗传病的基因治疗, 如阻止杜氏肌营养不良小鼠模型中的肌肉退化等^[81]. 最后, CRISPR/Cas 系统还被用于人类细胞大规模功能基因的筛选, 如白喉毒素与炭疽毒素耐药基因鉴定^[82,83]. 继 2012 年 TALE 成为 *Science* 年度十大科学突破后, 2013 年, CRISPR/Cas 系统又被评为年度十大科学突破, 基因组编辑工具对其他领域正产生着日趋重要的影响.

(2) 多重基因组工程(multiplexed genome engineering). 2009 年, 一种多元自动化基因组工程(multiplex automated genome engineering, MAGE)的方法

提供了一类不同于靶向核酶的高通量基因组编辑工具(图 2B)^[22]. 该方法从 DNA 复制的视角, 利用单链寡核苷酸(single-stranded oligos, ssDNA; 30~100 bp)模拟冈崎片段, 直接修饰基因组多个位点. 由于细胞复制时间短、寡核苷酸易高通量合成等优势, 该方法实现了对基因组的快速进化. 作者利用 MAGE 方法, 对番茄红素合成途径 24 个基因的转录水平进行了同时优化, 使其产量提高了 5 倍. 后来, 作者又利用该方法将大肠杆菌 TAG 终止密码子部分替换为 TAA, 再借助接合组装基因组改造技术(conjugative assembly genome engineering, CAGE), 创建了一株 TAG 密码子完全被剔除的大肠杆菌^[84,85]. 目前, MAGE 技术已被应用于某些革兰氏阴性菌、阳性菌和哺乳动物细胞等^[86-88].

3.2 sRNA 或 RNAi 介导的编辑工具

除了直接对 DNA 进行基因组编辑外, Lee 研究组^[89]在 2013 年开发了一种基于 sRNA 的转录后修饰元件, 以期对多个靶基因实现同时抑制(图 2C). 作者先尝试了酪氨酸合成途径旁路中 4 个基因的组合抑制, 又采用 130 个合成 sRNA 鉴定了可显著提高戊二胺合成的关键基因, 发现 *murE* 的抑制可以使产量提高 55%. 依据相同的原理, 赵惠民研究组^[90]在酵母中开发了一类基于 RNA 干扰的基因组进化方法(RNAi-assisted genome evolution, RAGE), 通过 3 轮迭代的 RNAi 筛选, 作者鉴定了 3 个敲除靶点, 通过组合敲除提高了细胞耐受能力. 这些方法加快了重要功能基因的筛选, 其应用前景值得期待.

3.3 全局转录工程

生命进化过程中产生了很多全局转录调控因子, 如 σ 因子、TATA 盒结合蛋白与 CRP(cAMP-response protein)蛋白等. 如果对这些蛋白进行随机突变, 可能产生多种复杂表型, 为筛选具有优良性状的菌株奠定了基础. 2006 年, Stephannopoulos 研究组^[24]开发了一种全局转录机器工程(global transcription machinery engineering, gTME), 实现了大规模基因表达重编程的目标(图 2D). 作者对酿酒酵母 TATA 盒结合蛋白进行了随机突变, 然后直接筛选耐受高浓度葡萄糖与乙醇的新型酵母, 使乙醇产量提高了 50%. 这种优秀的思路后来被广泛普及, 应用于多种全局调控蛋白的改造, 包括上述提及的 ZFP 与 TALE 蛋白,

图2 基因组编辑工具(网络版彩图)

A: 靶向核酶介导的基因组修饰技术; B: 多元自动化基因组工程; C: sRNA 介导的转录组工程; D: 全局转录工程

实现了耐药菌株的构建或功能基因的大规模筛选等^[91,92].

4 底盘细胞

如果将生物元件、DNA 合成与组装技术以及基因组编辑工具作为“配件工具”，那么其“主体设备”——底盘细胞的选择与优化，对于目标产物的高效合成与复杂遗传环路的建立将起到事半功倍的效果。构建理想底盘细胞的主要理念是在保持宿主细胞具有基本的自我复制与代谢能力的基础上，进行必要的功能精简优化，以获得正交性好、稳健性高、普适性强的合成生物学操作平台^[93]。目前，构建微生物底盘细胞的主要策略是基因组的精简，主要研究对象包括大肠杆菌、链霉菌(*Streptomyces* sp.)与酵母等^[94-96]。同时，为了满足更多复杂产物的合成与基因治疗的目标，哺乳动物细胞作为底盘也正在发展中。理想底盘细胞的构建不仅能够显著提高各种精细化工品、可再生能源与天然药物的合成能力，而且可以加深人们对生命系统的进化、稳健性与信息编码等本质问题的认识。

4.1 微生物最小基因组的分析与鉴定

生命系统存在高度的冗余性，对其遗传物质进行有理简化，将提高对细胞的可控力，优化细胞对底物、能量的利用效率。由于微生物基因组较小，结构与调控相对简单，所以成为构建最小基因组底盘细胞的主要对象。

(1) 天然进化的小基因组微生物。在已完成基因组测序的物种中，有一些天然进化的小基因组微生物(表 3)。这些微生物为了适应独特的生态环境，如在动物、植物和昆虫体内共生、寄生，以及在深海极端环境等，其基因组常要丢失大量生存所非必需的基因、缺失许多代谢途径基因簇、删除重复序列、减少非编码区的序列、增加编码蛋白的基因密度等^[97]。人体内寄生的生殖道支原体是可以获得纯培养、已知具有最小基因组的微生物，也是最早完成测序的几个微生物基因组之一^[98]。其 580 kb 的基因组有 477 种蛋白编码区，携带了所有生存的关键基因，包括复制、转录、翻译、蛋白折叠，以及有限的能量代谢相关的基因，如合成维生素、核酸前体物、能量供应和用于人体寄生的物质。在木虱(*Psyllids*)的叶柄瘿中分离到细胞内共生的一种细菌的微生物 *Carsonella ruddii*。该细菌的基因组大小仅 159 kb，是已知内共

表 3 天然小基因组的微生物^[97]

代谢模型	生活方式	物种	环境	基因组(kb)	基因数	G+C 含量(%)	细胞直径(mm)
A. 异养型	A. 1. 共生型	生殖道支原体 (<i>Mycoplasma genitalium</i>)	人类细胞 1	580	477	31.7	0.3
		骑行纳古菌 (<i>Nanoarchaeum equitans</i>)	古菌(<i>Ignicoccus</i> sp.)	490	536	31.6	0.4
		蚜虫巴克纳氏菌 (<i>Buchnera aphidicola</i>)	雪松(<i>Cinara cedri</i>)	420	362	20.1	3.0
		昆虫寄生菌(<i>Carsonella ruddii</i>)	木虱(<i>Pachypsylla venusta</i>)	159	182	16.0	3.0
		叶蝉内共生菌(<i>Sulcia muelleri</i>)	草翅叶蝉 (<i>Homalodisca coagulate</i>)	245	227	22.4	30.0
	A. 2. 自由生活	遍在远洋杆菌 (<i>Pelagibacter ubique</i>)	海水	1308	1354	29.7	0.3
B. 自养型	B. 1. 共生型	蛤类共生菌(<i>Ruthia magnifica</i>)	壮丽伴溢蛤 (<i>Calypptogena magnifica</i>)	1200	976	34.0	1.0
		蛤类共生菌 (<i>Vesicomysocius okutanii</i>)	深海蛤类 <i>Calypptogena okutanii</i>	1000	937	31.6	1.0
	B. 2. 自由生活型	海洋原绿球藻 (<i>Prochlorococcus marinus</i>)	海水	1657	1717	30.8	0.6

生细菌中最小的基因组^[99]。高密度的基因组包含相互重叠的基因, 预测编码 182 种蛋白, 缺少了许多独立生存的必需基因。基因组中超过 1/2 的基因用于编码蛋白翻译和氨基酸代谢, 也有少量 tRNA 基因, 但是缺少一些关键的氨基酸代谢途径。在深海热液口分离到的一株纳古菌(*Nanoarchaeum equitans*)与另一株古菌 *Ignicoccus* sp. 共生。该纳古菌的基因组大小为 490 kb, 预测编码 536 种蛋白, 含有完整而高度简化的代谢途径^[100]。为了适应高温环境和共生生活, 基因组保留了 DNA 修复和重组相关的基因。

在海洋环境中蓝细菌是最丰富的微生物, 特别是原绿球藻属(*Prochlorococcus*)。该属海洋原绿球藻(*P. marinus*)菌株具有较小的基因组(1.7~2.7 Mb), 是已知可以光合自养的小基因组微生物。比较分析不同的原绿球藻菌株的基因组发现, 它们会因不同环境发生适应性的变化。这些基因组编码了完整的代谢途径, 包括 Calvin-Benson 循环固定碳、利用 CO₂ 合成细胞构架的电子传递链, 而氮代谢、转运系统、DNA 修复系统等相对于其他淡水蓝细菌而言则简化或丢失^[101]。遍在远洋杆菌(*Pelagibacter ubique*)基因组大小为 1.308 Mb, 编码 1354 种蛋白, 可以合成所有的氨基酸和大部分的辅因子。基因组几乎没有冗余或无功能的基因, 也没有假基因或噬菌体基因, 基

因之间的非编码序列也很短, 这些现象可能反映了基因组最小化的结果^[102]。

(2) 生物信息学分析最小的微生物基因组。除了分析天然的小基因组微生物外, 研究者们还通过对细胞生长与繁殖所必需基因的理论分析, 以及利用比较基因组分析可能的“最小微生物基因组”。通过重构 DNA、RNA 和蛋白质合成途径, Forster 和 Church^[103]认为最小细胞是仅需依赖于小分子底物就能自我复制的, 因此只需要包含基因组复制、转录、RNA 处理及修饰、翻译及翻译后修饰等几个生物化学亚系统。他们提出了理论最小的基因组为 113 kb, 编码 151 个基因, 包括 38 个 RNA 及 113 个蛋白编码基因。这些基因多来源于大肠杆菌, 但是出于最简化考虑, 它们使用了噬菌体的复制和翻译机器。

最小基因组的预测可以用比较基因组的方法, 即通过寻找不同微生物基因组中的保守基因来确定生存必需的基因集合。Mushegian 和 Koonin^[104]最早利用比较基因组的方法分析了基因组分别为 1.83 Mb 和 0.58 Mbp 的革兰氏阴性菌流感嗜血杆菌(*Haemophilus influenzae*)以及革兰氏阳性菌生殖道支原体。他们认为在这两个进化距离比较远的物种中保守的基因极有可能是生存必需的, 由此提出了 256 个保守的蛋白编码基因组成的最小的基因集合。然

而, 有些生存必需基因的功能可能被非同源的基因替代, 这些基因在比较基因组的分析中会被遗漏. 随着参与比较的基因组数目的增加, 被认为是生存必需的基因大量减少. 当 8 个基因组参加比较时, 保守的最小基因集合已经降低为 156 个^[105]. 此外, 约 1/3 的生存必需基因的功能是未知的, 这也大大影响了比较基因组方法对生存必需基因预测的准确性^[103].

为了得到更接近最小的细菌基因组的预测, Gil 等人^[106]综合了比较基因组分析及多方面的实验数据, 提出了最小基因集合为 206 个基因, 包括了完整的 DNA 和 RNA 代谢、蛋白修饰、折叠和转运、能量和中间产物代谢等, 此外还预测了 8 个具有保守序列的、假定的必需基因. 但是, 该分析还是遗漏了一些重要生理过程的基因, 如参与染色体复制起始的基因 *dnaA*, 以及关键调控基因 *seqA* 和 *had*, 辅助 *dnaB* 功能的 *dnaC* 基因等.

(3) 实验推测最小的微生物基因组. 除了理论分析和推测最小微生物基因组外, 在不同的微生物中还有大量基因组水平的系统实验, 以验证和进一步推断生存必需基因. 在基因组范围大规模的生存必需基因的鉴定包括遗传足迹法检测插入失活基因、单基因失活, 以及对文献中相关实验数据的整合等. 这些基因组系统突变的研究结果提示, 细菌基因组约 90% 和真核生物基因组约 80% 均可以被敲除失活, 而细胞仍然可以生长繁殖^[107]. 表 4 显示, 即使是天然已经很小基因组的生殖道支原体, 通过基因组系统的转座子突变, 仍然有 100 个基因可以被敲除失活, 而全部 43 个 RNA 编码基因是不能被敲除的, 从而使生殖道支原体基因组的必需基因数目减少为 382 个. 对于天然基因组较大的微生物, 例如, 枯草芽孢杆菌 (*Bacillus subtilis*), 通过系统地失活基因, 在约 4100 个编码基因中, 只有 192 个不能被突变失活, 加上预测的其他 79 个基因, 共推测有 271 个基因为生存必需基因^[108].

大肠杆菌基因组必需基因数目的推测结合了理论分析以及多种系统的基因组实验结果. Gerdes 等人^[109]用遗传足迹法(genetic footprinting)对大肠杆菌 K-12 菌株 MG1655 基因组进行了 Tn-5 转座子插入突变, 共鉴定出 620 个生存必需基因. 但是, 基因组中存在转座子不容易插入的位点(insertion cold spot), 使得该方法容易将存在不易插入位点的或基因长度较短而未被失活的生存非必需基因错误评估为生存必需

基因. 此外, 该方法鉴定的是对大肠杆菌生长重要的基因, 并非严格意义的生存必需基因. 有些生存非必需基因的失活可能会使菌株生长变慢而在一个群体中失去生存优势, 这类基因的失活不易被检测到而被误认为是生存必需基因. 相比而言, Baba 等人^[110]通过单基因敲除查找生存必需基因更为严谨. 他们共得到 3985 个单基因敲除的大肠杆菌突变菌株, 称为 Keio collection; 而对 303 个基因无法得到敲除突变株, 因此, 这些基因被认为是实验条件下的生存必需基因. 然而, 不管是哪种方法, 通过失活基因来鉴定生存必需基因都会遗漏基因组中存在的执行生存必需功能的冗余基因. 对于这些基因, 失活或敲除一个基因并不影响另一个基因发挥功能, 因而不会被鉴定为生存必需基因.

在 EcoGene 数据库(http://www.ecogene.org/old/topic.php?topic_id=5)中, 其他研究者根据 Baba 等人发表的生存必需基因结果进行了不断地更新. 在 EcoGene 中, 只有在所有测试过的生长条件中都为生存必需的基因才被定义为必需基因. 例如, 在不同温度、氧浓度、盐浓度或添加了不同的营养物质后可以得到单一的基因敲除菌株的话, 则该基因被认为是生存非必需基因. 此外, 因基因组其他位置的抑制突变(suppressor mutation)使有些生存必需基因可以得到敲除菌株, 则该基因仍应被列为生存必需基因. 据此, 共列出了 299 个生存必需基因. 此外, PEC(the Profiling of *Escherichia coli* chromosome)数据库(<http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp>)还搜集了不同实验数据来源的大肠杆菌生存必需基因的信息. 在这个数据库中大肠杆菌 K-12 菌株 MG1655 的生存必需基因为 302 个, 而 K-12 菌株 W3110 的生存必需基因为 300 个. 由于不同的研究中培养和实验条件可能会不一致, PEC 数据库中生存必需基因的信息也只能作为一个参考.

4.2 微生物基因组的大规模删减

通过对基因组中的生存非必需的 DNA 进行大片段删减的测试, 并将可敲除的区域合并到一个菌株中, 这样逐步构建一系列基因组减小化的菌株, 最终得到基因组最小化的菌株. 基因组删减的关键在于建立完善的大片段无痕敲除策略, 其主要运用的原理是同源重组、位点特异性重组与双链断裂修复重组等. 目前, 大肠杆菌、枯草芽孢杆菌、阿维链霉菌

(*Streptomyces avermitilis*)与酵母等少数模式微生物,其基因组均实现了不同程度简化,其中大肠杆菌精简幅度最高,为38.9%^[94-96,111]。

最小基因组底盘细胞的构建,一方面加深了人们对于生命起源、进化与代谢调控的认识,另一方面也为微生物代谢工程提供了更稳健、更适配的底盘细胞,发展前景值得期待。

(1) 基因组删减技术的发展。基因组删减技术包括精确敲除和随机敲除。随机敲除是利用转座子随机插入和切离介导的重组造成染色体DNA片段的缺失。但是,随机敲除在后续的缺失突变体筛选和鉴定方面比较麻烦,而且无法控制敲除片段的大小和位置。为了快速地删减微生物基因组,科学家们发展了不同的精确打靶技术,其中包括位点特异性重组、自杀质粒介导的同源重组和Lambda噬菌体Red系统介导的线性DNA重组^[112]。

位点特异性重组介导的基因组敲除利用了Cre重组酶/*loxP*或Flp重组酶/*FRT*(Flp识别位点)位点特异重组系统^[113]。Cre(Flp)是位点特异性整合酶,能介导2个*loxP(FRT)*位点之间的DNA片段的整合。在目的敲除区域两端按照同一方向插入34 bp的*loxP(FRT)*

位点,在整合酶的作用下发生分子内整合而使2个位点之间的DNA片段缺失。该方法需要先在染色体目标敲除区域两端插入两个*loxP(FRT)*位点,步骤比较繁琐,并且会在染色体上留下一个*loxP(FRT)*“疤痕”,对后续的敲除工作产生干扰,因此不太适合于基因组的连续敲除。

自杀质粒介导的敲除方法利用了不能在宿主菌中进行游离复制的质粒,在选择抗性标记的情况下,被迫与宿主染色体之间进行同源重组,整合到染色体上^[114]。质粒还带有与目的敲除区域两端同源的片段,通过选择*sacB*(蔗糖敏感)或*rpsL*(链霉素敏感)等反筛标记(counter selection marker)促进第二次同源重组^[115]。还可以利用I-Sce I内切酶介导的双链断裂促进重组^[114],使得质粒骨架被切除下来,根据同源重组的位置不同可以得到野生型或目标区域敲除的菌株。由于该方法在每次敲除实验都需要先构建特殊的基因打靶载体,因此过程比较繁琐。

目前,在基因敲除方法中应用最广泛的是Lambda噬菌体的RecET重组系统(图3)。该系统介导的线性DNA重组系统由于直接利用了PCR扩增的线性DNA片段,避免了繁琐的载体构建过程。Red系统

步骤1. PCR扩增FRT位点之间的抗性基因

步骤2. 转化导入至含有λ Red重组酶的菌株

步骤3. 筛选具有抗生素抗性的转化子

步骤4. 利用FLP表达质粒消除抗性基因

图3 一个简单的基因中断方法^[116]

H1 和 H2: 同源区; P1 和 P2: PCR 引物

只需要线性 DNA 上含有大于 40 bp 与染色体目的敲除区域两端同源的序列即可发生有效的同源重组^[116]. 在 Red 系统中, Gamma 蛋白抑制了大肠杆菌中的 RecBCD 核酸酶活性, 使线性 DNA 片段不被降解; Exo 外切酶从 5' 到 3' 方向切割线性 DNA 使其产生单链末端; Beta 单链 DNA 结合蛋白则促进了线性 DNA 与染色体的重组. 通过选择线性 DNA 片段上的抗性标记即可筛选到目的区域敲除的菌株. 抗性标记的去除可以利用直接识别 2 个 *FRT* 位点而可以删除中间区域的 Flp 重组酶或 I-Sce I 内切酶介导的双链断裂同源重组.

(2) 大肠杆菌基因组删减的进展. 2001 年, 日本开展了最小基因组工厂(minimal genome factory)项目, 包括最小基因组大肠杆菌的构建^[117]. Hashimoto 等人^[115]利用基因组无痕敲除方法, 将大肠杆菌 MG1655 基因组中不含生存必需基因的大于 40 kb 的区域进行了平行敲除, 再利用 P1 噬菌体转导将这些敲除累积在一个菌株中, 这样得到了 $\Delta 16$ 菌株. 该菌株基因组从 4.6 Mb 减小至 3.2 Mb(29.7%). 但是, 随着敲除区域的增加, 敲除菌株生长变得缓慢, $\Delta 16$ 菌株生长代时达到 45 min, 而野生型菌株的生长代时仅为 26 min. 此外, 这些生长非必需区域的敲除使得细胞大小和染色体复制与分配都受到了影响. 在此基础上, 他们继续将基因组删减, 在 2011 年得到了基因组减小至 2.89 Mb 的 $\Delta 33$ 菌株, 基因组减小了 38.9%, 这是目前已知的基因组最小的大肠杆菌菌株^[94]. 值得一提的是, Kato 和 Hashimoto^[118]报道了覆盖大肠杆菌 MG1655 的整个基因组范围的敲除信息, 包括对染色体上不同区域进行的大片段敲除、中等长度及小片段的敲除, 以及回补的区域等信息. 这些信息对于大肠杆菌最小基因组的删减和研究来说非常宝贵.

从应用的角度出发, 最小基因组不一定意味着最小的生存必需基因集合, 而是对工业应用有利的合适的基因集合. Shigenobu 等人^[119]通过比较大肠杆菌与基因组大小仅为 0.64 Mb, 且被认为与大肠杆菌有共同祖先的蚜虫(*Acyrtosiphon pisum*)共生菌巴克纳氏菌(*Buchnera*), 列出了大肠杆菌特有的基因作为候选的敲除基因^[121]. 为了得到健壮生长的大肠杆菌菌株, 他们保留了对生长重要的基因, 在此基础上, 对于 10 个生长非必需基因的区域进行了连续敲除, 并检测这些敲除菌株对生长是否有影响. 将对生长

没有影响的敲除区域再富集在一个菌株, 构建了 MGF-01 菌株. 该菌株基因组减小至 3.6 Mbp, 基因组减小了 22%. MGF-01 菌株在生长对数期的生长速率与野生型菌株一样, 但在野生型菌株进入生长静止期后该菌株还能继续保持生长, 最后的细胞密度为野生型的 1.5 倍, 因此比野生型菌株更具生长优势. 以该菌为生产宿主, 其 *L*-苏氨酸的产量比野生型高 2.4 倍^[120]. 在 MGF-01 的基础上将基因组继续删减至 2.98 Mbp, 获得菌株 DGF-298^[121]. 该菌株比野生型菌株生长更快, 最后的细胞密度也更高, 将是未来工业应用的良好宿主.

为了构建合适的底盘细胞, Posfai 等人^[93]利用比较基因组学的方法分析了在大肠杆菌 MG1655 基因组中的存在、而在其他几株大肠杆菌基因组中不存在的区域, 确定了 100 处可以敲除的区域, 总长度占基因组的 20%. 据此, 他们连续敲除了 43 处生存非必需基因, 包括重组或转移元件以及隐蔽的毒力因子等, 将基因组减小了 15%. 敲除后的菌株生长未受到影响, 而且具有更高的转化效率. 其中, 他们构建的菌株 MDS42(基因组减小了 14.3%)已有商业出售(<http://www.scarabgenomics.com/default.aspx>). MDS42 由于去掉了大量的生存非必需基因, 提高了作为商业菌株的生产力. 同样的遗传改造在野生型菌株和 MDS42 菌株中, 后者生产 *L*-苏氨酸的产量增加了约 80%^[122]. 由于删除了所有移动的遗传元件, MDS42 的进化频率降低了, 有利于不稳定的质粒和基因的保存和传代, 比较适合作为合成生物学研究和应用的底盘生物^[123]. 在此基础上, 通过敲除压力诱导的易于出错的 DNA 聚合酶, 使得 MDS42 的基因组更加稳定, 即使是一些生长抑制的生物分子元件也能稳定保持, 使该菌株成为更理想的宿主^[124].

(3) 其他微生物基因组删减的进展. 除了大规模删减大肠杆菌的基因组, 在完成模式微生物和一些有应用价值的微生物的基因组测序与分析后, 人们也开始对这些基因组进行删减工作. 如枯草芽孢杆菌具有强大的产生蛋白和分泌到细胞外的能力, 在工业上是重要的酶生产细胞工厂. 其基因组大小约 4.2 Mb, 预测编码 4245 个蛋白, 基因组敲除实验推测只有 271 个生存必需基因. 通过连续敲除基因组上 17 处非生长必需区段, 获得了基因组删减 1 Mb 的突变株 MG1M^[125]. 与野生型菌株 168 相比, 突变株 MG1M 产生纤维素酶和蛋白酶的能力不变, 但是生

表 4 实验推测的不同微生物基因组的必需基因数目^{a)[107]}

物种	N_{PCG}^*	N_E^{**}	培养基	方法
生殖道支原体 (<i>Mycoplasma genitalium</i>)	482	382 (79)	丰富培养基	随机转座子插入
肺炎支原体(<i>Mycoplasma pneumoniae</i>)	782	405 (52)	丰富培养基	随机转座子插入
新凶手弗朗西斯氏菌 (<i>Francisella novicida</i>)	1574	396 (25)	丰富培养基	随机转座子插入
流感嗜血杆菌(<i>Haemophilus influenzae</i>)	1737	670 (38)	丰富培养基	随机转座子插入
金黄色葡萄球菌(<i>Staphylococcus aureus</i>)	2632	351 (13)	丰富培养基	随机转座子插入
谷氨酸棒状杆菌 (<i>Corynebacterium glutamicum</i>)	2993	658 (22)	丰富培养基	随机转座子插入
贝氏不动杆菌(<i>Acinetobacter baylyi</i>)	3197	499 (16)	基本培养基	基因靶向敲除
霍乱弧菌(<i>Vibrio cholerae</i>)	3887	789 (20)	丰富培养基	随机转座子插入
结核分枝杆菌(<i>Mycobacterium tuberculosis</i>)	3918	614 (16)	丰富培养基	随机转座子插入
枯草杆菌(<i>Bacillus subtilis</i>)	4245	271 (6.4)	丰富培养基	靶向质粒插入
大肠杆菌(<i>Escherichia coli</i>)	4288	299 (8.0)	丰富培养基	基因靶向敲除
鼠伤寒沙门氏菌(<i>Salmonella typhimurium</i>)	4425	257 (5.8)	丰富培养基	随机转座子插入
铜绿假单胞菌(<i>Pseudomonas aeruginosa</i>)	5565	335 (6.0)	丰富培养基	随机转座子插入
粟酒裂殖酵母 (<i>Schizosaccharomyces pombe</i>)	4914	1260 (26)	丰富培养基	基因靶向敲除
酿酒酵母(<i>Saccharomyces cerevisiae</i>)	5797	1105 (19)	丰富培养基	基因靶向敲除

a) *: 蛋白编码基因的总数是根据截止至 2011 年 2 月的最新的文献或特殊的网页来统计; **: 生存必需基因的数目是根据表中所引文献确定的

长稍慢. 获得的另一个突变株 MBG874 的基因组删减了 874 kb^[126]. 与野生型菌株相比, 该突变株的基因表达调控网络发生了变化, 产生纤维素酶和蛋白酶的水平显著提高. 进一步研究发现, 缺失了精氨酸降解相关的区域(*rocDEF-rocR*)是产酶水平提高的原因^[127].

链霉菌是重要的抗生素产生菌. 对 9 Mb 的除虫链霉菌(*S. avermitilis*)基因组进行了连续删减, 获得了基因组减少 1.4 Mb 的突变株^[95]. 在多个突变株中导入了其他链霉菌来源的 4 种抗生素生物合成基因簇后, 与原始产生菌相比, 突变株可以更多地产生链霉素(streptomycin)和头霉素(cephamycin)C. 在模式链霉菌-天蓝色链霉菌(*S. coelicolor*)中, 连续敲除了放线紫红素(actinorhodin)、十一烷基灵菌红素(prodigiosin)、肌酸磷酸激酶(creatine phosphokinase, CPK)和钙-依赖抗生素(calcium-dependent antibiotic, CDA)等 4 种内源的抗生素生物合成基因簇, 并在 *rpoB* 和 *rpsL* 基因上引入了点突变. 在突变株中重新导入放线紫红素生物合成基因簇或异源的氯霉素

(chloramphenicol)和 Congocidine 生物合成基因簇, 发现突变株比原始产生菌产生更多的抗生素^[128]. Zhou 等人^[129]还获得了连续敲除天蓝色链霉菌基因组中全部 10 种聚酮和聚肽类抗生素生物合成基因簇, 以及近端粒 900 kb 的序列的菌株(基因组共删减 1.22 Mb). 在 2 个突变株中放线紫红素的产量比野生型提高了, 但是在 3 个突变株中几乎不产生放线紫红素.

4.3 从微生物到哺乳动物细胞

虽然微生物底盘细胞被广泛地应用, 但是依旧存在不可避免的缺陷, 例如, 来源于哺乳动物的单克隆抗体或其他蛋白的表达, 就可能导致其活性丧失. 微生物缺乏复杂的翻译后修饰系统, 如糖基化等, 可能导致蛋白的错误折叠, 而合成的蛋白药物也可能引起免疫排斥. 因此, 对哺乳动物细胞进行改造十分重要, 这也将促进基于细胞疗法的疾病治疗.

哺乳动物细胞除了用于人源蛋白表达与简单遗传环路(逻辑门与记忆系统等)构建外, 目前最成功的

应用是基于嵌合抗原受体(chimeric antigen receptors, CARs)的 T 细胞免疫疗法开发^[130]. 2011 年, Porter 等人^[8]将靶向慢性淋巴性白血病(chronic lymphoid leukaemia)特异性免疫抗原的单链抗体 V 区直接与 T 细胞 ξ 链融合, 从而利用自身免疫细胞直接靶向消灭癌细胞. 后来, 该方法也在多种实体瘤中得到了证实, 其应用前景巨大^[131,132].

由于哺乳动物细胞染色质结构的高度复杂性, 如大量增强子(enhancers)、沉默子(silencers)与绝缘子(insulators)的存在, 导致了外源基因整合表达水平很难准确预测. 构建遗传背景清晰的人工底盘染色体(artificial chassis chromosome)是解决方案之一, 目前基于基因组删减策略也获得了一定程度简化的染色体, 但未删减的部分依旧影响外源基因的表达, 因此还需要进一步的精简与优化^[133,134].

5 总结与展望

随着合成生物学的迅速发展, 各种高效、简便的使能技术相继诞生. 不同尺度的 DNA 合成组装方法

与高通量测序技术作为最基础的工具, 辅助靶向或全局基因组编辑工具, 为元件库、基因线路或生物合成途径、高阶遗传系统乃至底盘细胞的构建, 奠定了坚实基础. (i) 高通量测序技术确保了大规模 DNA 合成过程中序列的准确性, 高效便捷的 DNA 组装方法与精细元件库的配合使用, 可快速优化生物合成途径提高目标产物产量, 以及靶向基因组编辑工具可加快基因组精简的底盘细胞构建等. 同时, 更多优秀的使能技术如 DNA 折纸技术与单分子测序技术的出现, 也将为合成生物学及其他相关学科提供更好的服务. 由于生命系统的高度复杂性, 对其改造一般都需经过“设计-构建-检测-分析”的过程. 随着合成生物学的进步, 对于生命系统的预测能力将极大提高, 从而可减少这种繁琐的试错过程. (ii) 虽然对于生命系统已有了多种设计、改造与优化的方法, 包括基于染色质结构、转录、转录后与翻译层面等开发的多种使能技术, 然而对于蛋白质翻译后修饰工具的开发还十分欠缺. (iii) 加强和促进不同合成生物学科学团队间的合作, 分享各种新型使能技术将有助于整个合成生物学学科的发展.

参考文献

- Way J C, Collins J J, Keasling J D, et al. Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell*, 2014, 157: 151–161
- Cameron D E, Bashor C J, Collins J J. A brief history of synthetic biology. *Nat Rev Microbiol*, 2014, 12: 381–390
- Gardner T S, Cantor C R, Collins J J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 2000, 403: 339–342
- Elowitz M B, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*, 2000, 403: 335–338
- Tabor J J, Salis H M, Simpson Z B, et al. A synthetic genetic edge detection program. *Cell*, 2009, 137: 1272–1281
- Liu C L, Fu X F, Liu L L, et al. Sequential establishment of stripe patterns in an expanding cell population. *Science*, 2011, 334: 238–241
- Paddon C J, Westfall P J, Pitera D J, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 2013, 496: 528–532
- Porter D L, Levine B L, Kalos M, et al. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. *N Eng J Med*, 2011, 365: 725–733
- Grupp S A, Kalos M, Barrett D, et al. Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. *N Eng J Med*, 2013, 368: 1509–1518
- Cheng A A, Lu T K. Synthetic biology: an emerging engineering discipline. *Annu Rev Biomed Eng*, 2012, 14: 155–178
- Alper H, Fischer C, Nevoigt E, et al. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci USA*, 2005, 102: 12678–12683
- Salis H M, Mirsky E A, Voigt C A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol*, 2009, 27: 946–950
- Cambray G, Guimaraes J C, Mutalik V K, et al. Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res*, 2013, 41: 5139–5148
- Chen Y J, Liu P, Nielsen A A, et al. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat Methods*, 2013, 10: 659–664

- 15 Kosuri S, Eroshenko N, LeProust E M, et al. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat Biotechnol*, 2010, 28: 1295–1299
- 16 Quan J Y, Saaem I, Tang N, et al. Parallel on-chip gene synthesis and application to optimization of protein expression. *Nat Biotechnol*, 2011, 29: 449–452
- 17 Kosuri S, Church G M. Large-scale *de novo* DNA synthesis: technologies and applications. *Nat Methods*, 2014, 11: 499–507
- 18 Ma S Y, Tang N, Tian J D. DNA synthesis, assembly and applications in synthetic biology. *Curr Opin Chem Biol*, 2012, 16: 260–267
- 19 Gibson D G, Young L, Chuang R Y, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, 2009, 6: 343–345
- 20 Noskov V, Kouprina N, Leem S H, et al. A genetic system for direct selection of gene-positive clones during recombinational cloning in yeast. *Nucleic Acids Res*, 2002, 30: E8
- 21 Esvelt K M, Wang H H. Genome-scale engineering for systems and synthetic biology. *Mol Syst Biol*, 2013, 9: 641
- 22 Wang H H, Isaacs F J, Carr P A, et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 2009, 460: 894–898
- 23 Alper H, Stephanopoulos G. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng*, 2007, 9: 258–267
- 24 Alper H, Moxley J, Nevoigt E, et al. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science*, 2006, 314: 1565–1568
- 25 Gaj T, Gersbach C A, Barbas C F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*, 2013, 31: 397–405
- 26 Atsumi S, Higashide W, Liao J C. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat Biotechnol*, 2009, 27: 1177–1180
- 27 Pardee K, Green A A, Ferrante T, et al. Paper-based synthetic gene networks. *Cell*, 2014, 159: 940–954
- 28 Merulla D, Buffi N, Beggah S, et al. Bioreporters and biosensors for arsenic detection. *Biotechnological solutions for a world-wide pollution problem. Curr Opin Biotechnol*, 2013, 24: 534–541
- 29 Kahl L J, Endy D. A survey of enabling technologies in synthetic biology. *J Biol Eng*, 2013, 7: 13
- 30 Levskaya A, Chevalier A A, Tabor J J, et al. Engineering *Escherichia coli* to see light—these smart bacteria ‘photograph’ a light pattern as a high-definition chemical image. *Nature*, 2005, 438: 441–442
- 31 Purnick P E, Weiss R. The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol*, 2009, 10: 410–422
- 32 Siegl T, Tokovenko B, Myronovskiy M, et al. Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. *Metab Eng*, 2013, 19: 98–106
- 33 Blazeck J, Alper H S. Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol J*, 2013, 8: 46–58
- 34 Tornøe J, Kusk P, Johansen T E, et al. Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. *Gene*, 2002, 297: 21–32
- 35 Curran K A, Crook N C, Karim A S, et al. Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat Commun*, 2014, 5: 4002
- 36 Davis J H, Rubin A J, Sauer R T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res*, 2011, 39: 1131–1141
- 37 Meng H L, Wang J F, Xiong Z Q, et al. Quantitative design of regulatory elements based on high-precision strength prediction using artificial neural network. *PLoS One*, 2013, 8: e60288
- 38 Dahl R H, Zhang F, Alonso-Gutierrez J, et al. Engineering dynamic pathway regulation using stress-response promoters. *Nat Biotechnol*, 2013, 31: 1039–1046
- 39 Ferreira J P, Overton K W, Wang C L. Tuning gene expression with synthetic upstream open reading frames. *Proc Natl Acad Sci USA*, 2013, 110: 11284–11289
- 40 Isaacs F J, Dwyer D J, Collins J J. RNA synthetic biology. *Nat Biotechnol*, 2006, 24: 545–554
- 41 Callura J M, Dwyer D J, Isaacs F J, et al. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc Natl Acad Sci USA*, 2010, 107: 15898–15903
- 42 Pfleger B F, Pitera D J, D Smolke C, et al. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol*, 2006, 24: 1027–1032

- 43 Garg A, Lohmueller J J, Silver P A, et al. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res*, 2012, 40: 7584–7595
- 44 Jiang W Y, Bikard D, Cox D, et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*, 2013, 31: 233–239
- 45 Bonnet J, Yin P, Ortiz M E, et al. Amplifying genetic logic gates. *Science*, 2013, 340: 599–603
- 46 Gibson D G, Glass J I, Lartigue C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 2010, 329: 52–56
- 47 Caruthers M H, Barone A D, Beaucage S L, et al. Chemical synthesis of deoxyoligonucleotides by the phosphoramidite method. *Method Enzymol*, 1987, 154: 287–313
- 48 Caruthers M H. Gene synthesis machines: DNA chemistry and its uses. *Science*, 1985, 230: 281–285
- 49 Tian J D, Ma K S, Saaem I. Advancing high-throughput gene synthesis technology. *Mol Biosyst*, 2009, 5: 714–722
- 50 Czar M J, Anderson J C, Bader J S, et al. Gene synthesis demystified. *Trends Biotechnol*, 2009, 27: 63–72
- 51 Anderson J C, Dueber J E, Leguia M, et al. BglBricks: a flexible standard for biological part assembly. *J Biol Eng*, 2010, 4: 1
- 52 Shetty R P, Endy D, Knight T F Jr. Engineering BioBrick vectors from BioBrick parts. *J Biol Eng*, 2008, 2: 5
- 53 Quan J, Tian J. Circular polymerase extension cloning of complex gene libraries and pathways. *Plos one*, 2009, 4: e6441
- 54 Engler C, Gruetznier R, Kandzia R, et al. Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS One*, 2009, 4: e5553
- 55 de Kok S, Stanton L H, Slaby T, et al. Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth Biol*, 2014, 3: 97–106
- 56 Schmid-Burgk J L, Xie Z, Frank S, et al. Rapid hierarchical assembly of medium-size DNA cassettes. *Nucleic Acids Res*, 2012, 40: e92
- 57 Gibson D G, Benders G A, Andrews-Pfannkoch C, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, 2008, 319: 1215–1220
- 58 Pease A C, Solas D, Sullivan E J, et al. Light-generated oligonucleotide arrays for rapid DNA-sequence analysis. *Proc Natl Acad Sci USA*, 1994, 91: 5022–5026
- 59 Fodor S P, Read J L, Pirrung M C, et al. Light-directed, spatially addressable parallel chemical synthesis. *Science*, 1991, 251: 767–773
- 60 Ghindilis A L, Smith M W, Schwarzkopf K R, et al. CombiMatrix oligonucleotide arrays: genotyping and gene expression assays employing electrochemical detection. *Biosens Bioelectron*, 2007, 22: 1853–1860
- 61 Hughes T R, Mao M, Jones A R, et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol*, 2001, 19: 342–347
- 62 Tian J D, Gong H, Sheng N J, et al. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*, 2004, 432: 1050–1054
- 63 Schwartz J J, Lee C, Shendure J. Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules. *Nat Methods*, 2012, 9: 913–915
- 64 Noskov V N, Karas B J, Young L, et al. Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS Synth Biol*, 2012, 1: 267–273
- 65 Du J, Yuan Y B, Si T, et al. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res*, 2012, 40: e142
- 66 Shao Z, Zhao H, Zhao H. DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res*, 2009, 37: e16
- 67 Annaluru N, Muller H, Mitchell L A, et al. Total synthesis of a functional designer eukaryotic chromosome. *Science*, 2014, 344: 55–58
- 68 Khalil A S, Lu T K, Bashor C J, et al. A synthetic biology framework for programming eukaryotic transcription functions. *Cell*, 2012, 150: 647–658
- 69 Sorek R, Lawrence C M, Wiedenheft B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem*, 2013, 82: 237–266
- 70 Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, 337: 816–821
- 71 Tsai S Q, Wyvekens N, Khayter C, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*, 2014, 32: 569–576
- 72 Fu Y F, Sander J D, Reyon D, et al. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*, 2014, 32: 279–284
- 73 Chu V T, Weber T, Wefers B, et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in

- mammalian cells. *Nat Biotechnol*, 2015, 33: 543–548
- 74 Maruyama T, Dougan S K, Truttmann M C, et al. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol*, 2015, 33: 538–542
- 75 Boettcher M, McManus M T. Choosing the right tool for the job: RNAi, TALEN, or CRISPR. *Mol Cell*, 2015, 58: 575–585
- 76 Shan Q W, Wang Y P, Li J, et al. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol*, 2013, 31: 686–688
- 77 Wang H Y, Yang H, Shivalila C S, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, 2013, 153: 910–918
- 78 Gratz S J, Cummings A M, Nguyen J N, et al. Genome engineering of drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics*, 2013, 194: 1029–1035
- 79 Friedland A E, Tzur Y B, Esvelt K M, et al. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat Methods*, 2013, 10: 741–743
- 80 Cong L, Ran F A, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 2013, 339: 819–823
- 81 Long C Z, McAnally J R, Shelton J M, et al. Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science*, 2014, 345: 1184–1188
- 82 Wang T, Wei J J, Sabatini D M, et al. Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, 2014, 343: 80–84
- 83 Shalem O, Sanjana N E, Hartenian E, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, 2014, 343: 84–87
- 84 Lajoie M J, Rovner A J, Goodman D B, et al. Genomically recoded organisms expand biological functions. *Science*, 2013, 342: 357–360
- 85 Isaacs F J, Carr P A, Wang H H, et al. Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. *Science*, 2011, 333: 348–353
- 86 Rios X, Briggs A W, Christodoulou D, et al. Stable gene targeting in human cells using single-strand oligonucleotides with modified bases. *PLoS One*, 2012, 7: e36697
- 87 Swingle B, Markel E, Costantino N, et al. Oligonucleotide recombination in Gram-negative bacteria. *Mol Microbiol*, 2010, 75: 138–148
- 88 van Pijkeren J P, Britton R A. High efficiency recombineering in lactic acid bacteria. *Nucleic Acids Res*, 2012, 40: e76
- 89 Na D, Yoo S M, Chung H, et al. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat Biotechnol*, 2013, 31: 170–174
- 90 Si T, Luo Y, Bao Z, et al. RNAi-assisted genome evolution in *Saccharomyces cerevisiae* for complex phenotype engineering. *ACS Synth Biol*, 2014, 4: 283–291
- 91 Park K S, Lee D K, Lee H, et al. Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. *Nat Biotechnol*, 2003, 21: 1208–1214
- 92 Zhang H, Li J, Hou S, et al. Engineered TAL effector modulators for the large-scale gain-of-function screening. *Nucleic Acids Res*, 2014, 42: e114
- 93 Posfai G, Plunkett G, Feher T, et al. Emergent properties of reduced-genome *Escherichia coli*. *Science*, 2006, 312: 1044–1046
- 94 Iwade Y, Honda H, Sato H, et al. Oxidative stress sensitivity of engineered *Escherichia coli* cells with a reduced genome. *FEMS Microbiol Lett*, 2011, 322: 25–33
- 95 Komatsua M, Uchiyama T, Omura S, et al. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc Natl Acad Sci USA*, 2010, 107: 2646–2651
- 96 Murakami K, Tao E, Ito Y, et al. Large scale deletions in the *Saccharomyces cerevisiae* genome create strains with altered regulation of carbon metabolism. *Appl Microbiol Biotechnol*, 2007, 75: 589–597
- 97 Moya A, Gil R, Latorre A, et al. Toward minimal bacterial cells: evolution vs. design. *FEMS Microbiol Rev*, 2009, 33: 225–235
- 98 Fraser C M, Gocayne J D, White O, et al. The minimal gene complement of *Mycoplasma genitalium*. *Science*, 1995, 270: 397–403
- 99 Nakabachi A, Yamashita A, Toh H, et al. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science*, 2006, 314: 267
- 100 Das S, Paul S, Bag S K, et al. Analysis of *Nanoarchaeum equitans* genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. *BMC Genomics*, 2006, 7: 186
- 101 Dufresne A, Garczarek L, Partensky F. Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol*, 2005, 6: R14
- 102 Dewall M T, Cheng D W. The minimal genome: a metabolic and environmental comparison. *Brief Funct Genomics*, 2011, 10: 312–315
- 103 Forster A C, Church G M. Towards synthesis of a minimal cell. *Mol Syst Biol*, 2006, 2: 45
- 104 Mushegian A R, Koonin E V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad*

- Sci USA, 1996, 93: 10268–10273
- 105 Klasson L, Andersson S G. Evolution of minimal-gene-sets in host-dependent bacteria. Trends Microbiol, 2004, 12: 37–43
- 106 Gil R, Silva F J, Pereto J, et al. Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev, 2004, 68: 518–537
- 107 Korona R. Gene dispensability. Curr Opin Biotechnol, 2011, 22: 547–551
- 108 Kobayashi K, Ehrlich S D, Albertini A, et al. Essential *Bacillus subtilis* genes. Proc Natl Acad Sci USA, 2003, 100: 4678–4683
- 109 Gerdes S Y, Scholle M D, Campbell J W, et al. Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. J Bacteriol, 2003, 185: 5673–5684
- 110 Baba T, Ara T, Hasegawa M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol, 2006, 2: 2006–2008
- 111 Westers H, Dorenbos R, van Dijl J M, et al. Genome engineering reveals large dispensable regions in *Bacillus subtilis*. Mol Biol Evol, 2003, 20: 2076–2090
- 112 Feher T, Papp B, Pal C, et al. Systematic genome reductions: theoretical and experimental approaches. Chem Rev, 2007, 107: 3498–34513
- 113 Yu B J, Kang K H, Lee J H, et al. Rapid and efficient construction of markerless deletions in the *Escherichia coli* genome. Nucleic Acids Res, 2008, 36: e84
- 114 Posfai G, Kolisnychenko V, Bereczki Z, et al. Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome. Nucleic Acids Res, 1999, 27: 4409–4415
- 115 Hashimoto M, Ichimura T, Mizoguchi H, et al. Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. Mol Microbiol, 2005, 55: 137–149
- 116 Datsenko K A, Wanner B L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA, 2000, 97: 6640–66405
- 117 Mizoguchi H, Mori H, Fujio T. *Escherichia coli* minimum genome factory. Biotechnol Appl Biochem, 2007, 46: 157–167
- 118 Kato J, Hashimoto M. Construction of consecutive deletions of the *Escherichia coli* chromosome. Mol Syst Biol, 2007, 3: 132
- 119 Shigenobu S, Watanabe H, Hattori M, et al. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. Nature, 2000, 407: 81–86
- 120 Mizoguchi H, Sawano Y, Kato J, et al. Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome. DNA Res, 2008, 15: 277–284
- 121 Hirokawa Y, Kawano H, Tanaka-Masuda K, et al. Genetic manipulations restored the growth fitness of reduced-genome *Escherichia coli*. J Biosci Bioeng, 2013, 116: 52–58
- 122 Lee J H, Sung B H, Kim M S, et al. Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. Microb Cell Fact, 2009, 8: 2
- 123 Umenhoffer K, Feher T, Baliko G, et al. Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. Microb Cell Fact, 2010, 9: 38
- 124 Csorgo B, Feher T, Timar E, et al. Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. Microb Cell Fact, 2012, 11: 11
- 125 Ara K, Ozaki K, Nakamura K, et al. *Bacillus* minimum genome factory: effective utilization of microbial genome information. Biotechnol Appl Biochem, 2007, 46: 169–178
- 126 Morimoto T, Kadoya R, Endo K, et al. Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. DNA Res, 2008, 15: 73–81
- 127 Manabe K, Kageyama Y, Morimoto T, et al. Combined effect of improved cell yield and increased specific productivity enhances recombinant enzyme production in genome-reduced *Bacillus subtilis* strain MGB874. Appl Environ Microbiol, 2011, 77: 8370–8381
- 128 Gomez-Escribano J P, Bibb M J. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol, 2011, 4: 207–215
- 129 Zhou M, Jing X, Xie P, et al. Sequential deletion of all the polyketide synthase and nonribosomal peptide synthetase biosynthetic gene clusters and a 900-kb subtelomeric sequence of the linear chromosome of *Streptomyces coelicolor*. FEMS Microbiol Lett, 2012, 333: 169–179
- 130 Wieland M, Fussenegger M. Engineering molecular circuits using synthetic biology in mammalian cells. Annu Rev Chem Biomol Eng, 2012, 3: 209–234
- 131 Krebs K, Bottinger N, Huang L R, et al. T Cells expressing a chimeric antigen receptor that binds hepatitis B virus envelope proteins control virus replication in mice. Gastroenterology, 2013, 145: 456–465

- 132 Kloss C C, Condomines M, Cartellieri M, et al. Combinatorial antigen recognition with balanced signaling promotes selective tumor eradication by engineered T cells. *Nat Biotechnol*, 2013, 31: 71–75
- 133 Bergmann J H, Martins N M, Larionov V, et al. HACKing the centromere chromatin code: insights from human artificial chromosomes. *Chromosome Res*, 2012, 20: 505–519
- 134 Ren X Y, Tahimic C G, Katoh M, et al. Human artificial chromosome vectors meet stem cells-new prospects for gene delivery. *Stem Cell Rev*, 2006, 2: 43–50

Recent Advances in the Enabling Technologies for Synthetic Biology

LI Lei, JIANG WeiHong, QIN ZhongJun, XUE XiaoLi & LU YinHua

CAS Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences (CAS), Shanghai 200032, China

As a new discipline with great potential, synthetic biology has rapidly developed mainly owing to the innovation and application of various enabling technologies. From the construction and standardization of biological parts to high-throughput microchip-based gene synthesis, convenient DNA assembly methods at different scales, and high-efficiency genome-editing tools, great progress has been made in the enabling technologies for synthetic biology. Furthermore, advances in numerous other disciplines such as genetics, cancer therapy, disease surveillance, and biomanufacturing are driven by these cutting-edge engineering techniques. If these aforementioned operating tools are viewed as accessories, the corresponding main equipment, the cellular chassis, has also been greatly developed in recent years. Bioinformatic analysis and large-scale deletions in microbial genomes provide a strong basis for constructing predictable and controllable chassis strains. In addition, mammalian cells have been optimized to construct excellent chassis cells for the treatment of human diseases based on cell therapy. This review discusses the currently available versatile enabling technologies as well as their increasing application and significance in life sciences research.

synthetic biology, enabling technology, biological parts, gene synthesis, DNA assembly, genome editing, chassis

doi: 10.1360/N052015-00048