

## 特约评述

DOI: 10.12211/2096-8280.2020-013

## 基因组的“读-改-写”技术

王会<sup>1,2</sup>, 戴俊彪<sup>1,2</sup>, 罗周卿<sup>2</sup>

(<sup>1</sup> 深圳大学生命与海洋科学学院, 广东 深圳 518055; <sup>2</sup> 中国科学院深圳先进技术研究院, 深圳合成生物学创新研究院, 合成基因组学研究中心, 广东省合成基因组学重点实验室, 深圳合成基因组学重点实验室, 广东 深圳 518055)

**摘要:** 基因组是生命系统的指令中枢, 对基因组的研究是生命科学的核心内容, 基因组研究相关技术的开发是深化对基因组序列和功能认识的重要推动力量。通过基因组测序获取基因组全序列, 通过人工诱变、定点编辑研究基因组局部序列的功能与调控, 通过对基因组的从头设计与化学再造实现对生命性状的定制, 是基因组研究的三个不同层面。从一代测序到三代测序, 基因组“读”技术极大地降低了成本和难度, 提升了速度和精准度, 引领着复杂基因组、大型基因组从草图走向完成图时代。通过人工诱变、定点编辑等技术可以改变野生型基因组的局部序列, 研究基因组序列的功能与调控。从人工诱变到定点编辑, 从ZFN到CRISPR, 基因组“改”技术在效率、适用对象和简便性上有了显著的提高, 为“基因型-表型”研究提供了有力工具, 精准编辑、高通量编辑逐步走向应用。通过对基因组的从头设计与化学再造, 书写人工基因组, 可以获得对基因组全局的系统认识, 实现对生命性状的定制。从病毒基因组合成、细菌基因组合成到酵母基因组合成, 再到国际基因组写计划, 基因组“写”技术在适用对象上不断拓展, 人工设计、化学再造正成为复杂生物学问题研究和已有性状优化、新性状引入的一把利器。本文主要综述了基因组测序(读)、基因组编辑(改)和基因组合成(写)技术的发展历程、各自的特征、目前的研究进展及在基因组研究方面的一些应用, 并对近期相关技术的可能突破点进行了总结和展望。“读-改-写”技术互为支撑, 推动基因组研究在致知和致用领域两面开花。

**关键词:** 合成生物学; 功能基因组; 基因组测序; 基因组编辑; 合成基因组学

**中图分类号:** Q5      **文献标志码:** A

## Reading, editing, and writing techniques for genome research

WANG Hui<sup>1,2</sup>, DAI Junbiao<sup>1,2</sup>, LUO Zhouqing<sup>2</sup>

(<sup>1</sup> College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518055, Guangdong, China; <sup>2</sup> Guangdong Provincial Key Laboratory of Synthetic Genomics, Shenzhen Key Laboratory of Synthetic Genomics, Center for Synthetic Genomics, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, Guangdong, China)

**Abstract:** Genome carries the entire genetic information of life. Genome-related researches are ultimate fundamentals for life sciences. Technological development in genomic researches has deepened our understanding of genomes and their function. Obtaining genome sequences through sequencing, studying their function and regulation through editing

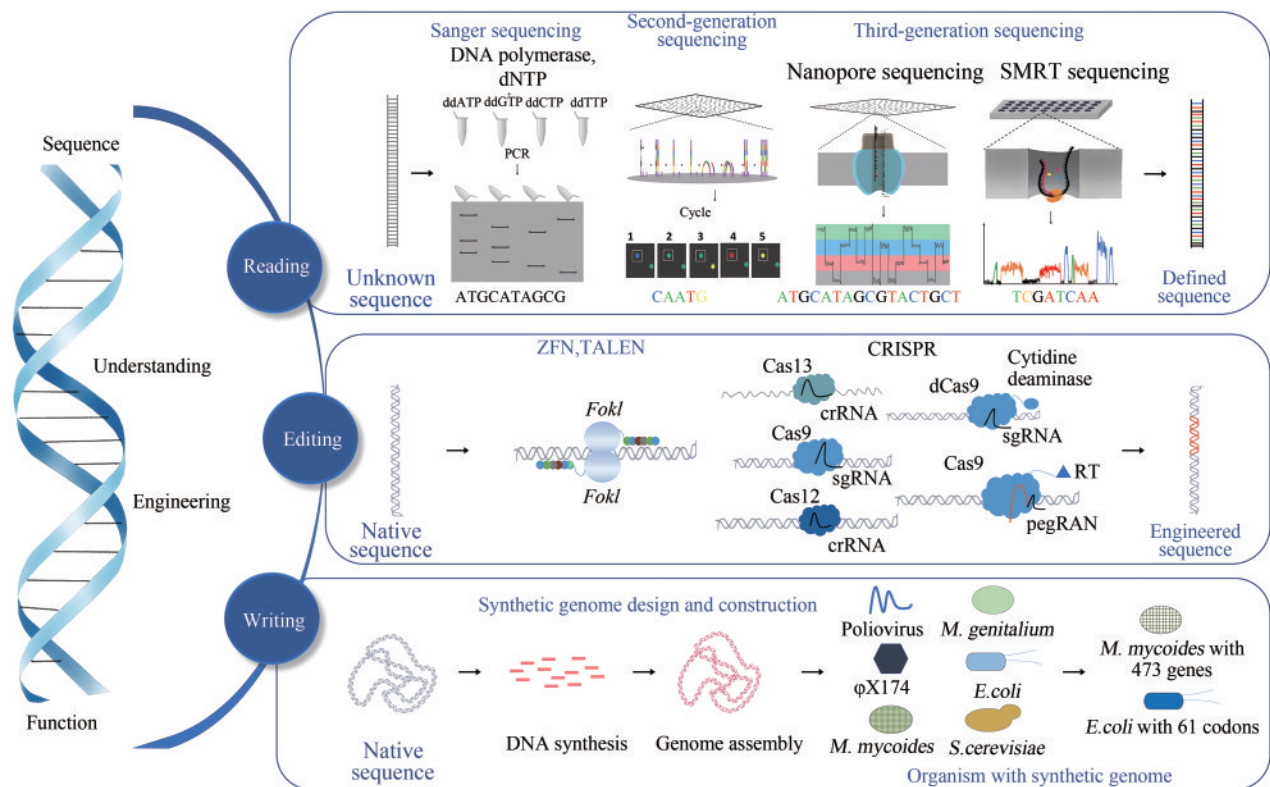
收稿日期: 2020-02-29 修回日期: 2020-04-19

基金项目: 国家重点研发计划“合成生物学”重点专项(2018YFA0900100); 国家自然科学基金(31725002, 31800069和31800082); 深圳合成基因组学重点实验室项目(ZDSYS201802061806209); 广东省合成基因组学重点实验室(2019B030301006)和深圳市科技计划(KQTD20180413181837372)

引用本文: 王会, 戴俊彪, 罗周卿. 基因组的“读-改-写”技术[J]. 合成生物学, 2020, 1(5): 503-515

Citation: WANG Hui, DAI Junbiao, LUO Zhouqing. Reading, editing, and writing techniques for genome research[J]. Synthetic Biology Journal, 2020, 1(5): 503-515

and creating customer designed genomes through synthesis are three important aspects of genome research. From the first-generation sequencing to the third-generation sequencing, the "reading" technology has greatly reduced the cost and difficulty, while improved the speed, enabling the production of complete genomic information for complex and large genomes. From random mutagenesis to site-specific genome editing and, from ZFN to CRISPR, the genome "editing" technology has improved significantly in efficiency, applicability, and simplicity, providing a wealth of materials to dissect "genotype-phenotype" relationship. Accurate editing and high-throughput editing are moving towards applications in various areas. From viral genome, bacterial genome to yeast genome, and ultimately to human genome, synthetic genomics has moved from simple organisms to many complex organisms. Precise, fast and low-cost synthesis technologies are important for the development of synthetic genomics. This article reviews the histories, features, present status, and applications of technologies for genome sequencing (reading), genome editing (editing) and genome synthesizing (writing). The potential breakthrough of these technologies in the near future is also summarized and prospected. The ability to read, edit and write a genome has been and will continue to advance not only our understanding but also better utilization of living systems.



**Keywords:** synthetic biology; functional genomics; genome sequencing; genome editing; synthetic genomics

基因组指的是生命体所携带的遗传物质的总和，如人类基因组包括细胞核和线粒体所含有的全部遗传物质<sup>[1]</sup>。遗传物质主要可分为DNA和RNA两种。自然界中绝大部分生物的遗传物质为DNA，只有少部分生物如RNA病毒，它们的遗传物质为RNA。基因组可分为编码区和非编码区。编码区能够通过转录和翻译过程制造蛋白质，

执行相应的功能；非编码区虽然不会被翻译，但是对于基因组遗传信息的表达也是必需的。在生命的进化过程中，基因组的组成与结构也在不断演化。不同种类生物的基因组在遗传信息的存储量、编码区和非编码区在基因组中所占的比例等方面都存在着较大差异（表1）<sup>[2]</sup>。

对于基因组序列及其功能的解析一直是基因

**表 1** 不同生物基因组中含有的编码序列与非编码序列的比较

**Tab. 1** Comparison of the the coding sequence and non-coding sequence contents of several genomes

物种	基因组大小/Mb	编码区占比/%	非编码区占比/%	参考数据库
<i>Homo sapiens</i>	3107	2.8	97.2	UCSC Genome Browser (hg18)
<i>Drosophila melanogaster</i>	168.7	18.3	81.7	FlyBase
<i>Saccharomyces cerevisiae</i>	12.2	72.9	27.1	Saccharomyces Genome Database
<i>Escherichia coli</i> K12	4.6	88	12	EcoCyc

组研究的核心问题。1953年，DNA双螺旋结构模型的提出使DNA的研究进入了分子水平<sup>[3]</sup>。20世纪70年代以后，分子生物学飞速发展，Sanger测序<sup>[4]</sup>、重组DNA<sup>[5]</sup>、转基因<sup>[6]</sup>等技术方法逐步涌现。2001年，人类基因组测序计划的完成开启了基因组大规模测序的新时代<sup>[7-8]</sup>。2002年，脊髓灰质炎病毒基因组的合成首次证实了可以利用化学的方法从头再造具有生命活性的基因

组<sup>[9]</sup>。2012年，CRISPR/Cas技术的出现使得在不同的物种里面进行高效率的基因编辑成为可能<sup>[10-11]</sup>。相关技术的进步一直在不断拓展基因组研究的深度和广度。

通过测序读取基因组序列，通过基因组编辑研究基因组对应序列的功能以及通过化学合成从头再造基因组序列，基因组的“读”“改”和“写”技术正从不同的侧面逐步描绘生命这一复杂系统的序列和功能关系的蓝图（图1）。

### 1 基因组“读”技术

DNA测序技术在过去几十年间得到了快速的发展，从最初的Sanger测序发展到二代测序以及三代测序，人类读取基因组序列的能力得到了飞跃式的提升。不同的测序技术各自的优势使得它们在目前的测序领域都占有一席之地，其各自的不足之处也成为推动测序技术不断发展完善的动力源泉（表2）。



**图 1** 基因组“读-改-写”技术发展中的关键事件

**Fig. 1** The milestones in the development of the read, edit, and write techniques for genome research

**表 2** 不同测序技术比较

**Tab. 2** Comparison of different sequencing technologies

测序技术	应用	读长	优点	缺点
一代测序	早期简单基因组测序；日常PCR产物、质粒等的测序	约1000 bp	读长较长，准确度高(99.999%)	成本高，通量低
二代测序	目前大部分基因组、转录组的测序	200~500 bp	成本低，通量高，准确度大于99.94%	读长短
三代测序 (SMRT sequencing)	复杂基因组、全长转录组等的测序	10~100 kb	读长长，准确度高，可直接检测DNA或者RNA上的修饰	测序成本高，每个SMRT cell的数据产出有限(约10 Gb)，文库准备需要大量的起始材料，目前读长还比较有限(约80 kb)
三代测序 (nanopore sequencing)	复杂基因组的测序，结构变异的鉴定等	kb~Mb	超长读长，经济高效，可直接检测DNA或者RNA上的修饰	错误率高，对于多碱基重复存在系统误差，文库准备需要大量的起始材料

## 1.1 一代测序技术

1977年, Sanger测序法问世, 该方法后来又被称为一代测序技术<sup>[4]</sup>。Sanger测序法通过在4个独立的DNA合成反应体系中分别加入一定比例的ddNTP (ddATP, ddCTP, ddGTP, ddTTP), 利用ddNTP的3'位不含羟基从而无法形成磷酸二酯键的特性, 终止DNA合成反应。通过聚丙烯酰胺凝胶电泳和放射自显影, 根据4个反应体系中泳带的条带大小判定碱基序列。基于Sanger测序原理发明了最早的测序仪, 在人类基因组计划开展中起到了重要作用。

## 1.2 二代测序技术

随着人类基因组计划的开展, 人们对基因组测序的需求日益增加, 渴求开发成本更低、更快速方便的测序技术。21世纪初, 以Roche公司的454技术<sup>[12]</sup>、Illumina公司的HiSeq技术<sup>[13]</sup>和ABI公司的SOLiD技术<sup>[14]</sup>等为标志的二代测序技术诞生。二代测序又被称为高通量测序技术, 主要特征是可以开展平行测序, 一次对几百万条序列同时进行测定, 极大地节约了测序时间。不同的二代测序仪的原理有所不同, 以HiSeq测序技术为例, 它首先将待测序列变成单链DNA, 通过DNA合成过程中的碱基互补配对原则, 将带有不同荧光标记的四种dNTP添加到互补链上。每添加一个dNTP就会标记上特定的荧光并将上一步反应的荧光进行擦除, 从而可以通过每一步反应后该位点上荧光信号的颜色解析待测DNA的序列信息。

## 1.3 三代测序技术

二代测序技术虽然显著降低了测序的价格, 但较短的读长限制了其对基因组复杂序列和基因组结构变异的解析能力<sup>[15]</sup>。为此, 科研人员开发出了兼具单分子实时测序和长读长特征的三代测序技术。三代测序技术根据其原理目前主要可分为单分子实时测序 (single-molecule real time sequencing, SMRT sequencing)<sup>[16]</sup>和纳米孔测序 (nanopore sequencing)<sup>[17]</sup>两大类, 两者各有优势。

### 1.3.1 单分子实时测序技术

单分子实时测序技术采用的也是边合成边测序的思路<sup>[16]</sup>, 在每个纳米级的零模波导孔 (zero-mode waveguide, ZMW) 的底部固定有单个DNA聚合

酶, 通过实时检测插入碱基的荧光信号解析孔内的单个环状DNA模板 (SMRTbell) 的序列。由于ZMW的直径小于激发光的波长, 激发光在ZMW中的传播呈指数衰减模式, 只能激发孔底的荧光分子, 提高了检测的信噪比。模板中的修饰碱基 (如6mA) 影响了聚合酶插入碱基的速率, 因此SMRT技术可以通过监测DNA聚合反应的动力学变化, 实现对碱基修饰的直接检测<sup>[18]</sup>。在DNA聚合酶的活性周期内, 环状DNA模板的使用使得同一个位置的碱基可以被多次测序, 结合SMRT技术的随机错误模型, 采用多重比对可以获得高达99.999%的测序准确率<sup>[19]</sup>。因此, 在实际测序过程中需要平衡分子读长和测序精度, 目前Pacbio的读长大部分在10~100kb<sup>[20]</sup>。

### 1.3.2 纳米孔测序技术

纳米孔测序是根据不同的DNA序列穿过纳米孔时产生的电信号的差异区分不同的碱基<sup>[21]</sup>。具体来说, 将含有蛋白纳米孔的电阻膜浸在离子溶液中, 在膜的两侧施加恒定的电势, 进而在纳米孔中产生离子电流; DNA链在马达蛋白的牵引下, 解螺旋通过纳米孔, 通过高度灵敏的电流传感器检测不同的碱基所形成的特征性离子电流变化信号, 进而解析DNA的序列组成。与SMRT测序技术相比, 纳米孔测序的读长不受技术本身的限制, 而是受文库DNA分子自身长度的限制。近期, 通过改进相关方法, 研究者们获得了N50>100 kb (最长读长可达882 kb) 的人类基因组测序数据<sup>[22]</sup>, 成功组装了人类Y染色体着丝粒的高重复性DNA序列<sup>[23]</sup>。纳米孔测序技术的数据产出非常可观, 现在商用的PromethION平台单次运行可产出Tb级别的数据<sup>[20]</sup>。但是, 纳米孔测序技术的一大缺陷是其错误率非常高 (约15%)<sup>[24]</sup>, 现在改进版的“1D<sup>2</sup>”技术的错误率也还在3%左右<sup>[20]</sup>。

## 2 基因组“改”技术

测序技术的不断革新使得大量生物的基因组序列得到读取。然而读取基因组序列只是我们理解生命奥秘的第一步, 对野生型基因组序列进行改动, 进而研究突变体与野生型的功能差异则是解析基因组功能与调控机制的必要手段。

基因组序列在生物繁衍过程中发生的随机变化为生命体的进化提供了最初的原材料, 也是解析基因

组序列与功能关系的最早的研究对象。物理诱变剂、化学诱变剂和DNA分子的自发突变会引起多种形式的DNA损伤,对这些损伤的修复有可能导致基因组序列的变化<sup>[25]</sup>。诱变和自发突变所导致的基因组序列的改变是随机的,对这些修复机制的理解极大地促进了后续基因组定点编辑技术的开发。到目前为止,已经有多种各具特色的基因定点编辑技术得到开发。

## 2.1 前CRISPR时代的基因组定点编辑技术

20世纪80年代,包括Cre重组酶<sup>[26]</sup>和FLP倒位酶<sup>[27]</sup>在内的一系列位点特异性重组酶相继被发现。Cre重组酶能够特异性识别来源于噬菌体的一段反向重复序列loxP,通过控制两个loxP的位置和方向可以实现loxP位点间基因序列的敲除、易位和反转。FLP/FRT系统的工作原理同上。

近年来,包括锌指核酸酶(zinc finger nuclease, ZFN)、转录激活样效应因子核酸酶(transcription activator-like effector nuclease, TALEN)和Cas蛋白等在内的序列特异性识别蛋白相继被发现,基于这些蛋白发展而来的基因编辑技术得到了广泛的应用<sup>[28]</sup>。这些技术的工作原理具有类似之处,都需要通过这些蛋白对基因组特定序列的结合,进而实现对基因组特定位点的切割,激活细胞内的非同源末端连接或同源重组机制进行DNA的损伤修复。其中非同源末端连接修复方式直接将断裂的两个末端连接起来,这种修复方式易引起插入或缺失突变,是一种不精确的基因编辑方式。而同源重组修复时,如果细胞中有同源序列存在,同源序列能够与基因组序列进行重组,从而实现精确的基因编辑。

ZFN由锌指蛋白和FokI内切酶组成,前者特异性识别目标序列,后者对DNA序列进行切割,产生DNA双链断裂<sup>[29-30]</sup>。作为第一代基因编辑技术,ZFN技术实现了基因组的定点编辑,但一个锌指单元识别三个碱基的识别模式仍具有较大的不灵活性,需要64个具有高度特异性的锌指单元才有可能实现对所有序列的识别。TALEN技术的作用机制与ZFN类似,由特异性识别目标序列的TALE蛋白和介导切割的FokI内切酶组成<sup>[25]</sup>。相较于ZFN技术,TALEN技术的编辑蛋白设计相对简单,每个TALE单元识别单个碱基,且天然的TALE单元之间就以串联的形式

存在,不需要对连接序列进行优化选择<sup>[31-32]</sup>,也被称为第二代基因编辑技术。

## 2.2 CRISPR/Cas基因组编辑技术

### 2.2.1 CRISPR/Cas系统的多样性及应用场景

CRISPR-Cas系统来源于古细菌/细菌的免疫系统,是其切割外来核酸保护自身细胞免受侵害的一种保护性机制<sup>[10, 33]</sup>。来源于化脓性链球菌的Cas9蛋白能够在单个gRNA介导下实现高效的基因组定点编辑<sup>[10-11]</sup>。后续又从许多不同的物种里面找到了具有不同PAM识别序列、活性和特异性的Cas9蛋白,这些蛋白及其变体显著扩充了基因组的可编辑范围<sup>[34-38]</sup>(表3)。除了Cas9,目前已发现了多种CRISPR/Cas编辑系统,主要包括Cas12<sup>[39-41]</sup>、Cas13<sup>[42-43]</sup>、Cas14<sup>[44]</sup>等,它们在工作原理上各有优劣,应用场景也各不相同,将可编辑的对象从双链DNA(dsDNA)扩展到RNA以及单链DNA(ssDNA)(表4)。以CRISPR/Cas9为代表的第三代基因编辑技术具有构建简单、适用对象广泛和效率高等诸多优点,现已成为各实验室最常用的基因组编辑技术。

### 2.2.2 CRISPR/Cas系统的衍生应用

鉴于CRISPR/Cas系统的简易性,基于该系统衍生而来的基因组编辑工具得到快速的开发。通过突变Cas9蛋白的两个活性位点(D10A和H840A)得到了不具有切割能力的dCas9(dead Cas9)<sup>[45]</sup>。2016年,Komor等<sup>[46]</sup>首次将胞嘧啶核苷脱氨酶与dCas9融合,发展出了第一代碱基编辑器,实现了碱基C到碱基T的转换。目前已开发出胞嘧啶碱基编辑器(cytosine base editor, CBE)<sup>[46]</sup>和腺嘌呤碱基编辑器(adenine base editor, ABE)<sup>[47]</sup>,可分别实现碱基C到T以及A到G的转换。碱基编辑器的编辑过程不需要产生双链断裂,为其在治疗人类遗传疾病的应用方面提供了安全保障。然而,近期有系列的研究表明,这两种碱基编辑器在RNA或DNA水平上仍存在较严重的脱靶现象<sup>[48-51]</sup>,对其作用机制和应用仍需要更加深入的研究。2019年,另一种新型基因组编辑系统Prime editing也得到开发<sup>[52]</sup>,该系统包括Cas9-逆转录酶融合蛋白和pegRNA(prime editing

表3 Cas9蛋白的不同来源及相关改造

Tab. 3 Different sources of Cas9 and related engineerings

种类	大小(氨基酸数)	PAM	来源及特色	参考文献
<i>SpCas9</i>	1368	NGG	<i>Streptococcus pyogenes</i> , 应用广泛	[10-11]
<i>SaCas9</i>	1053	NNGRRT	<i>Staphylococcus aureus</i> 体积小, 单个 AAV 载体即可承载 <i>SaCas9</i> 和对应的 guide RNA	[34]
<i>NmeCas9</i>	1082	NNNGATT	<i>Neisseria meningitidis</i> 识别的 PAM 位点较长, 切割活性较 <i>SpCas9</i> 低, 但特异性较好	[35]
<i>CjCas9</i>	984	NNNNACAC 和 NNNNRYAC	<i>Campylobacter jejuni</i> 迄今为止发现的最小的 Cas9 蛋白, 单个 AAV 载体即可承载 <i>CjCas9</i> 和对应的 guide RNA	[36]
<i>SpCas9-NG</i>	1368	NGH	理性设计的 <i>SpCas9</i> 突变体, 具有与 <i>SpCas9</i> 类似的特异性	[37]
<i>xCas9</i>	1368	NG, GAA 和 GTA	用噬菌体辅助的连续进化技术对 <i>SpCas9</i> 改造而来, 相比 <i>SpCas9</i> 有更好的特异性	[38]

注: N=A, T, C or G; R=A or G; Y=T or C; H=A, T or C。

表4 不同 CRISPR/Cas 系统的比较

Tab. 4 Comparison of different CRISPR/Cas systems

Cas 蛋白	蛋白大小 (氨基酸数)	导向分子大小 (核苷酸数)	靶标核酸的 主要类型	靶标序列限制	活性控制	精确性	应用领域
Cas9	约 1000~1600	约 105 (sgRNA)	dsDNA	含有 PAM 序列 (对 <i>SpCas9</i> 而言, 为 NGG), PAM 位 点近端产生平末端	由 RuvC 和 HNH 结构域 负责靶 DNA 的切割, 结合 并切割靶向 DNA	可容忍 DNA 链 上的单个错配	基因组编辑、 基因表达调控、 成像等
Cas12	约 1300	约 42~44 (crRNA)	dsDNA	含有 PAM 序列 (对 Cas12a 而言, 为 TTTV), 在 PAM 位点远端产生 5' 突 出末端	由 RuvC 和 Nuc 结构域 负责靶 DNA 的切割, 结合 并切割靶向 DNA	可有效区分双 链 DNA 上一个碱 基的差异	基因组编辑 和基于双链 DNA 的分子诊断等
Cas13	约 1400	约 64 (crRNA)	ssRNA	RNA 结构影响 其活性, 不同来源 的 Cas13 蛋白对 PFS 序列的要求不 同, 但都相对较弱	由两个 HEPN 结构域 负责靶 RNA 的切割; 在 体外及细菌体内, Cas13a 识别并切割靶向序列后即 转入酶促“激活”状态, 结合 并切割其他的 ssRNA	很难区分一个 碱基的差异	RNA 敲低和 基于 RNA 的分 子诊断等
Cas14	约 400~700	约 140 (sgRNA)	ssDNA	对单链 DNA 没 有限制, 双链 DNA 必须含有 PAM 序 列(对 Cas14a 而言, 为 TTTR)	由 RuvC 结构域负责 靶 DNA 的切割, 识别并 切割靶向序列后即转入 酶促“激活”状态, 这时 它将结合并切割其他的 ssDNA	可有效区分单 链 DNA 上一个碱 基的差异	基于单链 DNA 的分子诊断等

guide RNA)。pegRNA 有两点作用: ①引导 Cas9-逆转录酶融合蛋白到达靶序列位点; ②自身含有目标序列的 RNA 序列, 作为逆转录模板。Cas9 蛋白在靶位点产生单链断裂, 而逆转录酶以 pegRNA 为模板合成目的 DNA 序列, 然后以目的 DNA 序列

为材料对切口进行修复, 实现基因组编辑。Prime editing 作用过程中不会产生双链断裂, 也不需要供体 DNA 模板, 不仅能实现嘌呤-嘌呤之间的转换, 也能实现嘌呤-嘧啶间的颠换。Prime editing 提供了一种单碱基替换和小片段删除、插入的新工具,

在基因治疗领域具有巨大潜力。2020年, 来自哈佛大学的Walton等在*Science*杂志上发表了基于spCas9蛋白的最新研究成果, 成功移除了对NGG这一PAM系列的依赖性<sup>[53]</sup>。通过基于三维结构的定向突变, 研究者们创建了PAM序列为NGN的SpG蛋白和PAM序列为NRN及NYN的SpRY蛋白, 显著扩大了可编辑的基因组序列范围。

### 3 基因组“写”技术

在基因组序列读取和编辑的基础上, 随着DNA合成成本的下降、组装和移植技术的不断改进, 人们开始逐步具备对全基因组进行从头设计与合成的能力。合成基因组学为人们提供了一种崭新的理解生命、改造生命的新方法。

#### 3.1 寡核苷酸的合成

寡核苷酸(oligonucleotide)的合成是合成基因组学的基本技术。目前Oligo的合成主要使用固相亚磷酸胺化学合成法, 通过去保护、偶联、加帽和氧化4个反应的循环往复进行寡核苷酸的合成, 每个循环添加一个碱基<sup>[54-55]</sup>。20世纪90年代初, 基于此原理的DNA合成仪得以开发, 寡核苷酸的合成实现了自动化和产业化。然而, 核苷酸的添加过程难免会发生错误, 从而导致合成的准确率随着合成序列的增长而逐渐降低, 目前寡核苷酸的合成长度一般不超过200个核苷酸<sup>[56]</sup>。此外, 柱法合成技术的通量较低, 微阵列介导的DNA合成技术的出现在一定程度上解决了这一问题, 并降低了寡核苷酸合成成本<sup>[57]</sup>。然而微阵列合成法在合成的长度、精确度以及所得寡核苷酸的产量上都不及前者, 还需进一步完善。

早在20世纪60年代, 就有人提出通过酶促法合成寡核苷酸<sup>[58-59]</sup>。酶促法具有多种优势, 合成过程不会用到有毒化合物, 理论上能够合成出长达数千个碱基的核苷酸链等。目前可用于酶促法合成的最具前景的DNA聚合酶是末端脱氧核苷酰转移酶(terminal deoxynucleotidyl transferase, TdT)。与大多数DNA聚合酶不同, TdT可以在没有模板的情况下在DNA分子的末端随机添加新的碱基。

如何精确地合成目的序列一直是尚待解决的问题。在酶促合成法提出约50年后, 研究者们才通过TdT酶与dNTP的共价结合技术为这一关键问题的解决提供了一个可行的方案<sup>[60]</sup>。目前, 酶促法合成寡核苷酸仍处于实验室技术研发阶段, 离应用尚有一段距离。

#### 3.2 体外拼接技术

通过DNA体外拼接技术, 将较短的寡核苷酸拼接成较长的基因片段, 是基因组合成过程的第二步, 较为常用的有以下几种方法。①PCA法(polymerase cycling assembly)是基于PCR的一种拼接方法, 寡核苷酸之间需要有15~25 bp的重叠区域, 重叠区域可互为引物在DNA聚合酶的作用下延伸形成较长的DNA片段。Smith等<sup>[61]</sup>用PCA法成功拼接了长达5386 bp的噬菌体φX174基因组。②Golden Gate拼接法依赖于II型限制性内切酶<sup>[62]</sup>。这种酶的切割位点位于识别位点以外, 因此可以通过对序列的设计, 使得识别位点在酶切后从目标序列被去除, 形成一个人工设计的黏性末端, 从而可以利用DNA连接酶实现片段间的无缝连接。③Gibson assembly体外组装技术只需要将具有同源序列的片段与包括DNA聚合酶、核酸外切酶和DNA连接酶在内的酶体系进行反应即可<sup>[63]</sup>。核酸外切酶降解DNA产生可以互补配对的3'突出同源末端, 通过DNA聚合酶和连接酶的作用实现多个片段的组装。

#### 3.3 全基因组合成技术

随着DNA合成成本的降低和体外组装技术的成熟, 人们开始逐步尝试全基因组的合成, 从简单的病毒基因组到支原体和大肠杆菌等原核基因组, 再到首个真核基因组(酿酒酵母基因组)即将完工, 人工合成的基因组越来越大, 也越来越复杂(表5)。

2002年Wimmer小组耗时3年合成了全长约7.7 kb的脊髓灰质炎病毒基因组, 在世界上首次证明人工化学再造的基因组可以产生有生命活力的病毒颗粒, 开创了合成基因组学的时代<sup>[9]</sup>。对于

表5 基因组合成对象复杂程度对比

物种	基因组大小	基因数目	合成时间
Poliiovirus	7.7 kb	10	2002年
PhiX174	5.4 kb	11	2003年
<i>Mycoplasma genitalium</i>	580 kb	525	2008年
<i>Mycoplasma mycoides</i>	1.2 Mb	985	2010年
<i>Escherichia coli</i>	3.97 Mb	3730	2019年
<i>Saccharomyces cerevisiae</i>	1.25 Mb	5770	2011年
<i>Homo sapiens</i>	3.3 Gb	约21 000	2016年

脊髓灰质炎病毒基因组和相关生物学特性的研究为这项工作的开展奠定了理论基础，而21世纪初DNA合成和组装技术（尤其是PCA法介导的DNA组装技术和限制性内切酶介导的酶联技术）的发展，为这项工作的实现提供了技术基础。

J. Craig Venter作为合成生物学领域的先驱，多年来一直致力于合成基因组学的研究。他所带领的研究组从2003年到2010年先后合成了长5386 bp的 $\phi$ X174噬菌体基因组<sup>[61]</sup>、583 kb的生殖支原体基因组<sup>[64]</sup>和1.08 Mb的蕈状支原体基因组（JCVI-syn1.0）<sup>[65]</sup>，并于2016年合成了蕈状支原体的最小基因组（JCVI-syn3.0）<sup>[66]</sup>。从支原体基因组的测序<sup>[67]</sup>，到利用转座子测定其必需基因<sup>[68]</sup>，再到Gibson assembly体外组装技术<sup>[65]</sup>、转化辅助的酵母体内组装技术（transformation-associated recombination, TAR）<sup>[69]</sup>以及原生质体融合介导的基因组移植技术<sup>[70]</sup>的开发，Venter研究组花费了将近20年的时间一步一个脚印地解决了细菌基因组“设计-构建-移植-复活”过程中的一系列技术和理论问题，构建了人类历史上第一个由全化学合成基因组支持存活的细菌<sup>[65]</sup>和第一个最小原核基因组<sup>[66]</sup>。在上述技术和CRISPR基因编辑技术的支持下，2019年，Jason W. Chin课题组报道了全合成的只有61个密码子的大肠杆菌基因组，首次实现了有义密码子的压缩<sup>[71]</sup>。

由Jef Boeke领衔，中、美、英、法、奥、新加坡等国家通力合作的首个真核生物基因组合成计划——酿酒酵母基因组合成计划（Sc 2.0）目前也已经完成了六条半染色体的合成工作<sup>[72-79]</sup>，剩余的染色体合成工作也将于近期完成。相较

于原核基因组，研究者在合成的酵母基因组中引入了更多的人为设计<sup>[75]</sup>，比如全面去除转座子、重复序列以及内含子，用以研究相关序列的生物学功能。在这个过程中，研究者们开发了BioStudio设计软件以及SwAP-in（switching auxotrophies progressively for integration）、Endoreduplication Backcross及MRA（meiotic recombination-mediated assembly）等DNA组装技术，实现了真核基因组的模块化设计和高效组装。到目前为止，可以说人们已经具备了合成真核细胞基因组的能力<sup>[80]</sup>，而国际基因组合成计划（Genome Project-Write, GP-Write）的目标已经瞄准了更加复杂的高等生物基因组合成<sup>[81]</sup>。

### 3.4 建物致知与建物致用

建物致知与建物致用是合成基因组学的两大愿景。在建物致知方面，最小原核基因组中仍然还有大量的未知功能的基因，对这些基因的研究将揭示细胞最核心的功能组成<sup>[66]</sup>；合成酵母基因组中引入了一系列的设计原则，以便为转座子及内含子的功能和存在的生物学意义提供直接的证据<sup>[75]</sup>；通过对核糖体编码序列的合成再造，我们揭示了这一基因组内拷贝数最高的序列对基因组高级结构的影响<sup>[82]</sup>。在建物致用方面，通过全基因组重编所释放的密码子可用于引入非天然氨基酸等，助力蛋白质相关的机制和应用研究<sup>[80]</sup>。通过在合成酵母基因组的每个非必需基因的终止密码子后第三个碱基处引入loxPsym位点，我们实现了Cre重组酶依赖的合成基因组序列的删除、倒换、重复和移位，产生了多样的基因型和表型，为合成菌株的进化和代谢工程应用铺平了道路<sup>[83-92]</sup>（表6）。

## 4 总结与展望

“读-改-写”的研究技术是解析基因组奥秘的有效手段，三者相互支撑。基因组序列的读取是后续修改和再造的基础；基因组序列的编辑是注释序列功能的有效手段，可为基因组的从头设计提供理论支撑；基因组的合成再造可对野生型序

表6 合成酵母 SCRaMbLE 系统的相关研究

Tab. 6 Studies on the synthetic yeast SCRaMbLE system

SCRaMbLE 对象	对后续合成菌株的应用所具有的价值	参考文献
synIXR(环形)	对 SCRaMbLE 系统的首次实验验证,证实了 SCRaMbLE 系统可以高效地产生删除、倒换和重复等基因组结构变异	[83]
synV	利用纳米孔测序技术解析重排基因组,证实了 SCRaMbLE 所造成的基因组重排可用于外源代谢途径特异的底盘细胞改良	[84]
含有 loxPsym 位点的质粒	建立光控的 Cre 重组酶活性控制系统(L-SCRaMbLE),实现对其活性更加精准的控制	[85]
synV	利用与门电路精准控制 Cre 重组酶的活性以及利用多次 SCRaMbLE 实现带有合成染色体的异源二倍体菌株中的外源代谢途径的产量提升	[86]
synII	通过对 loxPsym 等位点的设计改造建立 SCRaMbLE-in 技术,实现外源代谢途径的优化和底盘细胞的适配	[87]
synXII, synXIII 等	基于 loxP 与 loxPsym 位点的正交性建立 ReSCuES 基因组重排报告系统,实现重排菌株的高效筛选、耐受性进化和机制解析	[88]
synV, synX	表征了带有合成染色体的种内杂交菌株或种间杂交菌株的 SCRaMbLE 行为,为利用合成染色体优化工业菌株性状奠定基础	[89]
synV(环形)	环形染色体 SCRaMbLE 可以产生非合成染色体的数目变化,非整倍体菌株可用于提高 prodeoxyviolacein 的产量	[90]
$\beta$ -胡萝卜素途径	利用 SCRaMbLE 原理设计了基于结构变异的 DNA 文库体外构建方法,可用于外源代谢途径的体外优化	[91]
synV	将自动化样品制备、超快速 LC-MS 方法和条形码纳米孔测序相结合,解决了重排菌株代谢性能快速表征的技术障碍	[92]

列进行全局设计,是对基因组相关功能和调控机制的再验证和再利用。在未来的基因组研究中,以基因组的“读”和“改”为基因组的“写”提供更多的理论和技术支撑,以基因组的“写”验证基因组的“读”和“改”过程中发现的相关规律以及探索新的规律,实现“读-改-写”三位一体,将是推动基因组研究由浅到深、由点到面、由理论到应用的有效手段。

基因组学的发展对长读长的测序技术在成本和准确度等方面都提出了新的要求。对于 SMRT 测序技术而言,通过工程化改造 DNA 聚合酶以提高其持续聚合能力(processivity)和延长活性周期有望在保证准确率的前提下继续提升其读长,而优化相关试剂和仪器则有望提升其通量并降低测序成本。对于纳米孔测序技术而言,其错误率较高且这些错误更为系统化,通过增加测序的深度对降低错误率所起的作用有限。通过寻找新的纳米孔材料(如石墨烯纳米孔<sup>[93]</sup>和固态纳米孔<sup>[94]</sup>)、

控制 DNA 通过纳米孔的时间以及改进电信号检测仪器的灵敏度,有望提高单碱基分辨率和降低其错误率。值得一提的是,由于肽段通过纳米孔也会产生电流的变化,利用纳米孔技术实现单分子蛋白质测序将是值得研究的方向<sup>[95]</sup>。结合两者优势以及其他辅助组装技术,如 Hi-C 技术<sup>[96]</sup>和 BioNano Genomics 公司的纳米通道技术(nanochannel genome mapping)<sup>[97]</sup>,将为精准的基因组完成图的获取提供利器。

复杂的生命现象的解析需要更加精准、范围更广、通量更高的基因组编辑技术。除了能够对单个位点进行编辑以外,CRISPR/Cas 系统已经可以对基因组多个位点进行同时编辑。2019年,来自瑞士的科学家通过构建单质粒承载系统,成功利用 Cas12a 和 CRISPR array 实现了多达 25 个内源性靶点的编辑<sup>[98]</sup>。2011年,George Church 利用 MAGE (multiplex automated genome engineering) 和 CAGE (conjugal assembly genome engineering)

成功将大肠杆菌基因组中的全部314个TAG终止密码子替换为TAA终止密码子<sup>[99-100]</sup>。这也是目前最具有代表性的在全基因组范围内实现大规模编辑的一项工作。目前利用基因编辑技术实现基因组范围的大规模编辑仍然具有较大的困难,如何利用CRISPR/Cas技术实现高等生物中基因组范围的多位点编辑,比如替换人类基因组中的TAG终止密码子,将是未来的发展方向。

当代合成基因组学仍然处于发展的早期阶段,如何降低合成成本以及操作大型基因组片段(组装、移植和复活等),以满足高等生物复杂基因组的合成改造,还需要大量的研究。通过对酶和核苷酸底物的优化,DNA的酶法合成有望突破化学合成法在合成长度和准确度方面的限制,但还需要大量的突破性工作<sup>[101]</sup>;借助酵母自身的拼接系统,研究人员目前已经可以高效地进行10<sup>5</sup>bp级别的DNA组装,进一步提高多片段共转化效率以及酵母细胞内的同源重组效率有望实现Mb级别的DNA体内组装,而其他高效的外源DNA宿主(如枯草芽孢杆菌<sup>[102]</sup>)或者高效的同源重组宿主(如耐辐射球菌<sup>[103]</sup>)的开发则有望从另一个方向对DNA组装技术进行突破;在体内组装的大型DNA片段如何移植到目标宿主中进行功能测试是下一阶段基因组编写计划需要解决的重大技术问题,原生质体融合是目前所采用的方法之一<sup>[104]</sup>,但是成功率较低,需要进一步优化。除了在构建技术层面的发展,合成基因组的设计理论也必须跟上。如何深度设计基因组序列以探索特定的生物学问题或者实现应用价值的最大化是需要大力发展的方向。天然基因组的优化改造(比如合成最小的酵母基因组以探索真核生物的核心组成)以及非天然基因组的设计合成(如合成具有不同宿主特性的嵌合基因组以用作优良的代谢工程底盘)将是近期有可能突破的方向。

### 参 考 文 献

- [1] MCKUSICK V A. Genomics: structural and functional studies of genomes[J]. *Genomics*, 1997, 45(2):244-249.
- [2] ALEXANDER R P, FANG G, ROZOWSKY J, et al. Annotating non-coding regions of the genome[J]. *Nature Reviews Genetics*, 2010, 11(8):559-571.
- [3] CRICK F. The double helix: a personal view[J]. *Nature*, 1974, 248(5451):766-769.
- [4] SANGER F, NICKLEN S, COULSON A R. DNA sequencing with chain-terminating inhibitors[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1977, 74(12):5463-5467.
- [5] JACKSON D A, SYMONS R H, BERG P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1972, 69(10):2904-2909.
- [6] CHANG A C, COHEN S N. Genome construction between bacterial species *in vitro*: replication and expression of *Staphylococcus* plasmid genes in *Escherichia coli*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1974, 71(4):1030-1034.
- [7] VENTER J C, ADAMS M D, MYERS E W, et al. The sequence of the human genome[J]. *Science*, 2001, 291(5507):1304-1351.
- [8] International human genome sequencing C: finishing the euchromatic sequence of the human genome[J]. *Nature*, 2004, 431(7011):931-945.
- [9] CELLO J, PAUL A V, WIMMER E. Chemical synthesis of *Poliiovirus* cDNA: generation of infectious virus in the absence of natural template[J]. *Science*, 2002, 297(5583):1016-1018.
- [10] JINEK M, CHYLINSKI K, FONFARA I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity[J]. *Science*, 2012, 337(6096):816-821.
- [11] CONG L, RAN F A, COX D, et al. Multiplex genome engineering using CRISPR/Cas systems[J]. *Science*, 2013, 339(6121):819-823.
- [12] MARGULIES M, EGHOLM M, ALTMAN W E, et al. Genome sequencing in microfabricated high-density picolitre reactors[J]. *Nature*, 2005, 437(7057):376-380.
- [13] LEVY S E, MYERS R M. Advancements in next-generation sequencing[J]. *Annual Review of Genomics and Human Genetics*, 2016, 17:95-115.
- [14] VALOUEV A, ICHIKAWA J, TONTHAT T, et al. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning[J]. *Genome Research*, 2008, 18(7):1051-1063.
- [15] GOODWIN S, MCPHERSON J D, MCCOMBIE W R. Coming of age: ten years of next-generation sequencing technologies[J]. *Nature Reviews Genetics*, 2016, 17(6):333-351.
- [16] EID J, FEHR A, GRAY J, et al. Real-time DNA sequencing from single polymerase molecules[J]. *Science*, 2009, 323(5910):133-138.
- [17] ASHTON P M, NAIR S, DALLMAN T, et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island[J]. *Nature Biotechnology*, 2015, 33(3):296-300.
- [18] FLUSBERG B A, WEBSTER D R, LEE J H, et al. Direct detection of DNA methylation during single-molecule, real-time

- sequencing[J]. *Nature Methods*, 2010, 7(6):461-465.
- [19] TRAVERS K J, CHIN C S, RANK D R, et al. A flexible and efficient template format for circular consensus sequencing and SNP detection[J]. *Nucleic Acids Research*, 2010, 38(15):e159.
- [20] VAN DIJK E L, JASZCZYZYŃ Y, NAQUIN D, et al. The third revolution in sequencing technology[J]. *Trends in Genetics*, 2018, 34(9):666-681.
- [21] MANRAO E A, DERRINGTON I M, LASZLO A H, et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and *phi29* DNA polymerase[J]. *Nature Biotechnology*, 2012, 30(4):349-353.
- [22] JAIN M, KOREN S, MIGA K H, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads[J]. *Nature Biotechnology*, 2018, 36(4):338-345.
- [23] JAIN M, OLSEN H E, TURNER D J, et al. Linear assembly of a human centromere on the Y chromosome[J]. *Nature Biotechnology*, 2018, 36(4):321-323.
- [24] JAIN M, TYSON J R, LOOSE M, et al. MinION analysis and reference consortium: phase 2 data release and analysis of R9.0 chemistry[J]. *F1000Research*, 2017, 6:760.
- [25] FRIEDBERG E C. DNA damage and repair[J]. *Nature*, 2003, 421(6921):436-440.
- [26] ABREMSKI K, HOESS R. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein[J]. *Journal of Biological Chemistry*, 1984, 259(3):1509-1514.
- [27] SENECOFF J F, BRUCKNER R C, COX M M. The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1985, 82(21):7270-7274.
- [28] GAJ T, GERSBACH C A, BARBAS C F 3rd, ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering[J]. *Trends in Biotechnology*, 2013, 31(7):397-405.
- [29] KIM Y G, CHA J, CHANDRASEGARAN S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1996, 93(3):1156-1160.
- [30] URNOV F D, REBAR E J, HOLMES M C, et al. Genome editing with engineered zinc finger nucleases[J]. *Nature Reviews Genetics*, 2010, 11(9):636-646.
- [31] GUILLINGER J P, PATTANAYAK V, REYONN D, et al. Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity[J]. *Nature Methods*, 2014, 11(4):429-435.
- [32] CHANDRASEGARAN S, CARROLL D. Origins of programmable nucleases for genome engineering[J]. *Journal of Molecular Biology*, 2016, 428(5PtB):963-989.
- [33] BARRANGOU R, FREMAUX C, DEVEAU H, et al. CRISPR provides acquired resistance against viruses in prokaryotes[J]. *Science*, 2007, 315(5819):1709-1712.
- [34] RAN F A, CONG L, YAN W X, et al. *In vivo* genome editing using *Staphylococcus aureus* Cas9[J]. *Nature*, 2015, 520(7546):186-191.
- [35] LEE C M, CRADICK T J, BAO G. The *Neisseria meningitidis* CRISPR-Cas9 system enables specific genome editing in mammalian cells[J]. *Molecular Therapy : the Journal of the American Society of Gene Therapy*, 2016, 24(3):645-654.
- [36] KIM E, KOO T, PARK S W, et al. *In vivo* genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*[J]. *Nature Communications*, 2017, 8:14500.
- [37] HU J H, MILLER S M, GEURTSS M H, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity[J]. *Nature*, 2018, 556(7699):57-63.
- [38] NISHIMASU H, SHI X, ISHIGURO S, et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space[J]. *Science*, 2018, 361(6408):1259-1262.
- [39] ZETSCHKE B, GOOTENBERG J S, ABUDAYYEH O O, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system[J]. *Cell*, 2015, 163(3):759-771.
- [40] TENG F, CUI T, FENG G, et al. Repurposing CRISPR-Cas12b for mammalian genome engineering[J]. *Cell Discovery*, 2018, 4:1-15.
- [41] STRECKER J, JONES S, KOOPAL B, et al. Engineering of CRISPR-Cas12b for human genome editing[J]. *Nature Communications*, 2019, 10(1):212.
- [42] ADUDAYYEH O O, GOOTENBERG J S, KONERMAN S, et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector[J]. *Science*, 2016, 353(2699):aaf5573.
- [43] BARRANGOU R, GERSBACH C A. Expanding the CRISPR toolbox: targeting RNA with Cas13b[J]. *Molecular Cell*, 2017, 65:582-584.
- [44] HARRINGTON L B, BURSTEIN D, CHEN J S, et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes[J]. *Science*, 2018, 362(6416):839-842.
- [45] QI L S, LARSON M H, GILBERT L A, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression[J]. *Cell*, 2013, 152(5):1173-1183.
- [46] KOMER A C, KIM Y B, PACKER M S, et al. Programmable editing of a target base in genomic DNA without double-strand DNA cleavage[J]. *Nature*, 2016, 533(7603):420-424.
- [47] GAUDELLI N M, KOMER A C, REES H A, et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage[J]. *Nature*, 2017, 551(7681):464-471.
- [48] GRUNEWALD J, ZHOU R, GARCIA S P, et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors[J]. *Nature*, 2019, 569(7756):433-437.
- [49] JIN S, ZONG Y, GAO Q, et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice[J]. *Science*, 2019, 364(6437):292-295.
- [50] ZHOU C, SUN Y, YAN R, et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis [J]. *Nature*, 2019, 571(7764):275-278.
- [51] ZUO E, SUN Y, WEI W, et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos[J]. *Science*, 2019, 364(6437):289-292.

- [52] ANZALONE A V, RANDOLPH P B, DAVIS J R, et al. Search-and-replace genome editing without double-strand breaks or donor DNA[J]. *Nature*, 2019, 576(7758):149-157.
- [53] WALTON R T, CHRISTIE K A, WHITTAKER M N, et al. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants[J]. *Science*, 2020, 368(6488): 290-296.
- [54] BEAUCAGE S, CARUTHERS M. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis[J]. *Tetrahedron Letters*, 1981, 22(20): 1859-1862.
- [55] CARUTHERS M, BARONE A, BEAUCAGE S, et al. Chemical synthesis of deoxyoligonucleotides by the phosphoramidite method[J]. *Method Enzymol*, 1987, 154: 287-313.
- [56] LEPROUST E M, PECK B J, SPIRIN K, et al. Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process[J]. *Nucleic Acids Research*, 2010, 38(8):2522-2540.
- [57] KOSURI S, CHURCH G M. Large-scale *de novo* DNA synthesis: technologies and applications[J]. *Nature Methods*, 2014, 11(5):499-507.
- [58] BOLLUM F J. Oligodeoxyribonucleotide-primed reactions catalyzed by calf thymus polymerase[J]. *Journal of Biological Chemistry*, 1962, 237(6):1945-1949.
- [59] JENSEN M A, DAVIS R W. Template-independent Enzymatic Oligonucleotide Synthesis (TiEOS): its history, prospects, and challenges[J]. *Biochemistry*, 2018, 57(12):1821-1832.
- [60] PALLUK S, ARLOW D H, DE ROND T, et al. *De novo* DNA synthesis using polymerase-nucleotide conjugates[J]. *Nature Biotechnology*, 2018, 36(7):645.
- [61] SMITH H O, HUTCHISON C A 3rd, PFANNKOCH C, et al. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2003, 100(26):15440-15445.
- [62] ENGLER C, KANDZIA R, MARILLONNET S. A one pot, one step, precision cloning method with high throughput capability[J]. *PLoS One*, 2008, 3(11):e3647.
- [63] GIBSON D G, YOUNG L, CHUANG R Y, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases[J]. *Nature Methods*, 2009, 6(5):343-345.
- [64] GIBSON D G, BENDERS G A, ANDERWS-PFANNKOCH C, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome[J]. *Science*, 2008, 319(5867):1215-1220.
- [65] GIBSON D G, GLASS J I, LARTIGUE C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome [J]. *Science*, 2010, 329(5987):52-56.
- [66] HUTCHISON C A 3rd, CHUANG R Y, NOSKOV V N, et al. Design and synthesis of a minimal bacterial genome[J]. *Science*, 2016, 351(6280):aad6253.
- [67] FRASER C M, GOCAYNE J D, WHITE O, et al. The minimal gene complement of *Mycoplasma genitalium*[J]. *Science*, 1995, 270(5235):397-403.
- [68] HUTCHISON C A, PETERSON S N, GILL S R, et al. Global transposon mutagenesis and a minimal *Mycoplasma* genome [J]. *Science*, 1999, 286(5447):2165-2169.
- [69] GIBSON D G, BENDERS G A, AXELROD K C, et al. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome[J]. *Proceedings of the National Academy of Sciences*, 2008, 105(51):20404-20409.
- [70] LARTIGUE C, VASHEE S, ALGIRE M A, et al. Creating bacterial strains from genomes that have been cloned and engineered in yeast[J]. *Science*, 2009, 325(5948):1693-1696.
- [71] FREDENS J, WANG K, DE LA TORRE D, et al. Total synthesis of *Escherichia coli* with a recoded genome[J]. *Nature*, 2019, 569(7757):514-518.
- [72] DYMOND J S, RICHARDSON S M, COOMDES C E, et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design[J]. *Nature*, 2011, 477(7365): 471-476.
- [73] ANNALURU N, MULLER H, MITCHELL LA, et al. Total synthesis of a functional designer eukaryotic chromosome[J]. *Science*, 2014, 344(6179):55-58.
- [74] MITCHELL L A, WANG A, STRACQUADANIO G, et al. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond[J]. *Science*, 2017, 355(6322): aaf4831.
- [75] RICHARDSON S M, MITCHELL L A, STRACQUADANIO G, et al. Design of a synthetic yeast genome[J]. *Science*, 2017, 355(6329):1040-1044.
- [76] SHEN Y, WANG Y, CHEN T, et al. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome[J]. *Science*, 2017, 355(6322):aaf4791.
- [77] WU Y, LI B Z, ZHAO M, et al. Bug mapping and fitness testing of chemically synthesized chromosome X[J]. *Science*, 2017, 355(6322):aaf4706.
- [78] XIE Z X, LI B Z, MITCHELL L A, et al. "Perfect" designer chromosome V and behavior of a ring derivative[J]. *Science*, 2017, 355(6322):aaf4704.
- [79] ZHANG W, ZHAO G, LUO Z, et al. Engineering the ribosomal DNA in a megabase synthetic chromosome[J]. *Science*, 2017, 355(6322):aaf3981.
- [80] LUO Z, YANG Q, GENG B, et al. Whole genome engineering by synthesis[J]. *Science China Life Sciences*, 2018, 61(12):1515-1527.
- [81] BOEKE J D, CHURCH G, HESSEL A, et al. Genome Engineering. The Genome Project-Write[J]. *Science*, 2016, 353(6295):126-127.
- [82] MERCY G, MOZZICONACCI J, SCOLARI V F, et al. 3D organization of synthetic and scrambled chromosomes[J]. *Science*, 2017, 355(6322):aaf4597.
- [83] SHEN Y, STRACQUADANIO G, WANG Y, et al. SCRABLE generates designed combinatorial stochastic diversity in synthetic chromosomes[J]. *Genome Research*, 2016, 26(1): 36-49.

- [84] BLOUNT B A, GOWERS G F, HO J C H, et al. Rapid host strain improvement by *in vivo* rearrangement of a synthetic yeast chromosome[J]. *Nature Communications*, 2018, 9(1): 1932.
- [85] HOCHREIN L, MITCHELL L A, SCHULZ K, et al. L-SCRaMble as a tool for light-controlled Cre-mediated recombination in yeast[J]. *Nature Communications*, 2018, 9(1):1931.
- [86] JIA B, WU Y, LI B Z, et al. Precise control of SCRaMble in synthetic haploid and diploid yeast[J]. *Nature Communications*, 2018, 9(1):1933.
- [87] LIU W, LUO Z, WANG Y, et al. Rapid pathway prototyping and engineering using *in vitro* and *in vivo* synthetic genome SCRaMble-in methods[J]. *Nature Communications*, 2018, 9(1):1936.
- [88] LUO Z, WANG L, WANG Y, et al. Identifying and characterizing SCRaMbled synthetic yeast using ReSCuES[J]. *Nature Communications*, 2018, 9(1):1930.
- [89] SHEN M J, WU Y, YANG K, et al. Heterozygous diploid and interspecies SCRaMbleing[J]. *Nature Communications*, 2018, 9(1):1934.
- [90] WANG J, XIE Z X, MA Y, et al. Ring synthetic chromosome V SCRaMble[J]. *Nature Communications*, 2018, 9(1): 3783.
- [91] WU Y, ZHU R Y, MITCHELL L A, et al. *In vitro* DNA SCRaMble[J]. *Nature Communications*, 2018, 9(1):1935.
- [92] GOWERS G F, CHEE S M, BELL D, et al. Improved betulinic acid biosynthesis using synthetic yeast chromosome recombination and semi-automated rapid LC-MS screening[J]. *Nature Communications*, 2020, 11(1):868.
- [93] GARAJ S, HUBBARD W, REINA A, et al. Graphene as a sub-nanometre trans-electrode membrane[J]. *Nature*, 2010, 467: 190-193.
- [94] GOTO Y, AKAHORI R, YANAGI I, et al. Solid-state nanopores towards single-molecule DNA sequencing[J]. *Journal of Human Genetics*, 2020, 65(1):69-77.
- [95] KOLMOGOROV M, KENNEDY E, DONG Z X, et al. Single-molecule protein identification by sub-nanopore sensors[J]. *PLoS Computational Biology*, 2017, 13(5):e1005356.
- [96] MASCHER M, GUNDLACH H, HIMMELBACH A, et al. A chromosome conformation capture ordered sequence of the barley genome[J]. *Nature*, 2017, 544(7651):427-433.
- [97] SEO J S, RHIE A, KIM J, et al. *De novo* assembly and phasing of a Korean human genome[J]. *Nature*, 2016, 538(7624): 243-247.
- [98] CAMPA C C, WEISBACH N R, SANTINHA A J, et al. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts[J]. *Nature Methods*, 2019, 16(9):887-893.
- [99] ISAACS F J, CARR P A, WANG H H, et al. Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement[J]. *Science*, 2011, 333(6040):348-353.
- [100] LAJOIE M J, ROVNER A J, GOODMAN D B, et al. Genomically recoded organisms expand biological functions[J]. *Science*, 2013, 342(6156):357-360.
- [101] PERKEL J M. The race for enzymatic DNA synthesis heats up [J]. *Nature*, 2019, 566(7745):565.
- [102] JUHAS M, AJIOKA J W. High molecular weight DNA assembly *in vivo* for synthetic biology applications[J]. *Critical Reviews in Biotechnology*, 2017, 37(3):277-286.
- [103] OSTROV N, BEAL J, ELLIS T, et al. Technological challenges and milestones for writing genomes[J]. *Science*, 2019, 366(6463):310-312.
- [104] BROWN D M, CHAN Y A, DESAI P J, et al. Efficient size-independent chromosome delivery from yeast to cultured cell lines[J]. *Nucleic Acids Research*, 2017, 45(7):e50.



通讯作者：戴俊彪(1974—)，男，博士，研究员，主要研究方向为合成生物学。  
E-mail: junbiao.dai@siat.ac.cn



通讯作者：罗周卿(1990—)，男，博士，副研究员，主要研究方向为合成生物学。  
E-mail: zq.luo@siat.ac.cn



第一作者：王会(1993—)，女，硕士研究生，主要研究方向为合成生物学。  
E-mail: 15225377578@163.com