

## 特约评述

DOI: 10.12211/2096-8280.2022-055

## 面向微生物遗传操作的编辑序列设计工具的研究进展

杨毅<sup>1</sup>, 毛雨丰<sup>1</sup>, 杨春贺<sup>1,2</sup>, 王猛<sup>1</sup>, 廖小平<sup>1</sup>, 马红武<sup>1</sup>

(<sup>1</sup> 中国科学院天津工业生物技术研究所中国科学院系统微生物工程重点实验室, 天津 300308; <sup>2</sup> 天津科技大学生物工程学院, 天津 300457)

**摘要:** 以同源重组、CRISPR 等为代表的遗传操作技术是合成生物学的重要支撑技术。高效准确的遗传操作依赖于可精确定位编辑位点和实现序列改造的编辑序列 (如引物序列、同源臂序列、sgRNA 序列等)。由于遗传改造目标与遗传操作技术丰富多变, 加之近年来基于生物铸造厂 (BioFoundry) 的自动化遗传改造模式对高通量设计的需求日益增加, 使得通过计算机辅助设计工具实现精准快速的编辑序列设计愈发重要。本文针对不同阶段实现不同目标的微生物遗传操作技术和相应的编辑序列辅助设计工具的发展进行了综述。按照不同编辑序列类型和遗传操作应用场景, 将编辑序列设计工具划分为四种类型: 引物设计工具、DNA 组装的设计工具、sgRNA 设计工具、基因组编辑的全流程设计工具, 对各类编辑序列设计工具的应用及存在的问题进行了分析总结并对未来的研究方向进行了展望。编辑序列设计工具的发展将有助于实现合成生物学“设计—构建—测试—学习” (design-build-test-learn, DBTL) 工作循环中上游的基因型“设计”与下游“构建”两个关键环节之间的无缝衔接。

**关键词:** 编辑序列; DNA 组装; 基因组编辑; 全流程设计; 生物工厂

**中图分类号:** Q939.97 **文献标志码:** A

## Recent progress in computational tools for designing editing sequences used in microbial genetic manipulations

YANG Yi<sup>1</sup>, MAO Yufeng<sup>1</sup>, YANG Chunhe<sup>1,2</sup>, WANG Meng<sup>1</sup>, LIAO Xiaoping<sup>1</sup>, MA Hongwu<sup>1</sup>

(<sup>1</sup>Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China; <sup>2</sup>College of Biotechnology, Tianjin University of Science & Technology, Tianjin 300457, China)

**Abstract:** Genetic manipulations such as homologous recombination and CRISPR are basic technologies of synthetic biology aiming to design and construct artificial life. One key factor affecting the efficiency and accuracy of microbial genetic manipulations is the editing sequences (ES), namely the assisting sequences used for precisely locating and editing a target sequence in a genome, such as a primer, a homologous arm or a sgRNA sequence. For different genetic manipulation technologies, diverse manipulation types and multiple ESs are required for different stages of

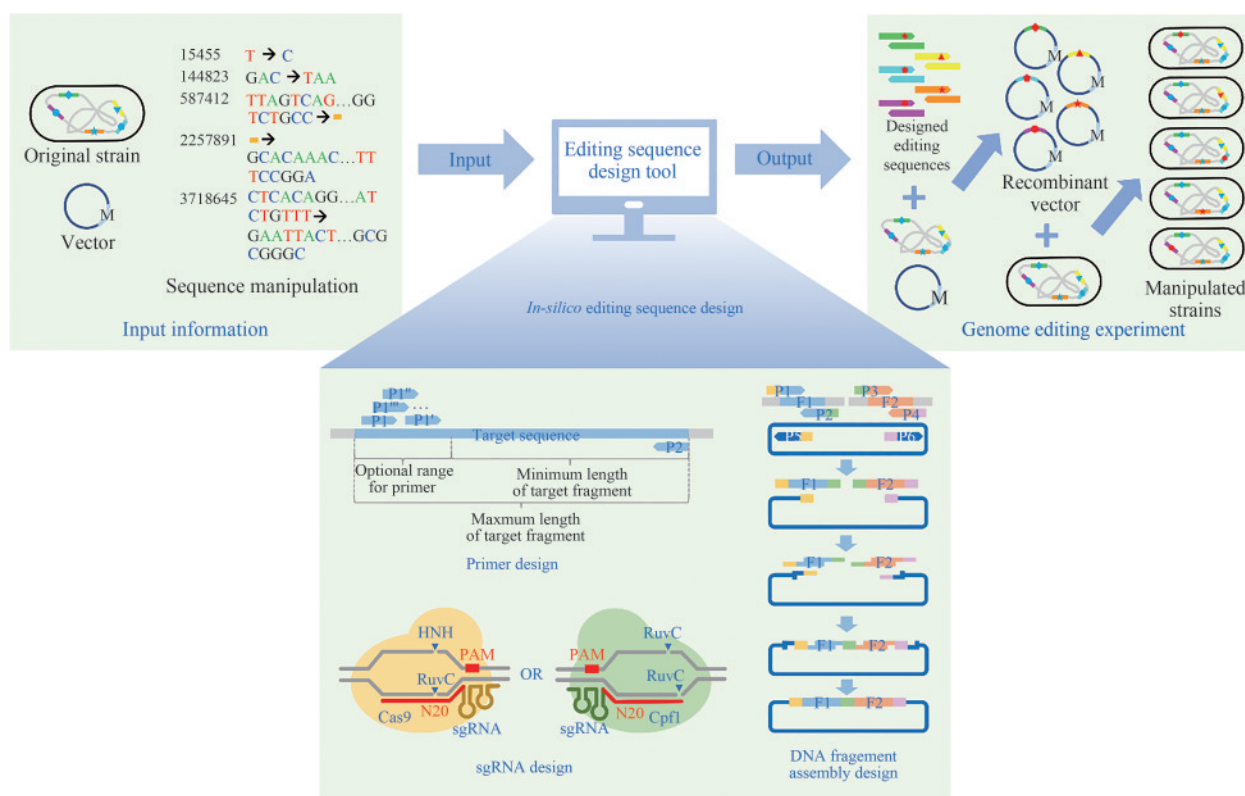
收稿日期: 2022-10-03 修回日期: 2022-11-23

基金项目: 国家重点研发计划“合成生物学”重点专项 (2018YFA0902900); 国家自然科学基金 (32101186); 中国科学院青年创新促进会; 天津市合成生物技术创新能力提升行动项目 (TSBICIP-PTJS-001, TSBICIP-KJGG-005)

引用本文: 杨毅, 毛雨丰, 杨春贺, 王猛, 廖小平, 马红武. 面向微生物遗传操作的编辑序列设计工具的研究进展[J]. 合成生物学, 2023, 4(1): 30-46

Citation: YANG Yi, MAO Yufeng, YANG Chunhe, WANG Meng, LIAO Xiaoping, MA Hongwu. Recent progress in computational tools for designing editing sequences used in microbial genetic manipulations[J]. Synthetic Biology Journal, 2023, 4(1): 30-46

manipulation processes, especially for the high-throughput design with recently developed biofoundries to enable automatic strain modifications. Therefore, it is becoming essential to use computational tools for precise, fast, high-throughput and whole workflow ES design. This article reviews tools for ES design at different stages using various microbial genetic manipulation technologies. The tools are classified into four categories based on the types of ESs and their application scenario: primer design, DNA assembly design, sgRNA design and whole workflow ES design. We first give a brief introduction to the tools used for basic primer design with an emphasis on the widely used open-source tool Primer3. Then we have an extensive discussion on the tools used in the design of ESs for DNA assembly using technologies like Gibson and Golden Gate assembly which are required for linking the editing sequences and/or the inserted sequences together as one big fragment to be transformed into target strains. Various tools for designing sgRNA used in the latest CRISPR technologies for genome and base editing are also evaluated and compared in detail. Moreover, we argue that one-stop ES design tools which integrate various design methods to cover the whole genetic manipulation workflow would be very important in addressing challenges for the high throughput design raised by automatic strain construction biofoundries to enable highly precise and efficient genome editing for different sequence manipulations at any location and in any organism. These ES design tools will seamlessly link the genotype “Design” step and the strain “Build” step in the “design-build-test-learn (DBTL)” working cycle of synthetic biology to facilitate the creation of artificial organisms.



**Keywords:** editing sequences; DNA assembly; genome editing; whole workflow design; biofoundry

合成生物学的一个重要目标是通过遗传操作技术创制具备理想新功能的人工生命，例如可以高效生产目标化合物的工程菌种。自1974年首个

基因工程微生物面世以来<sup>[1]</sup>，遗传操作技术与相应的菌种构建能力的升级迭代历经了多个发展阶段（图1）：①基于限制性内切酶的经典DNA组装

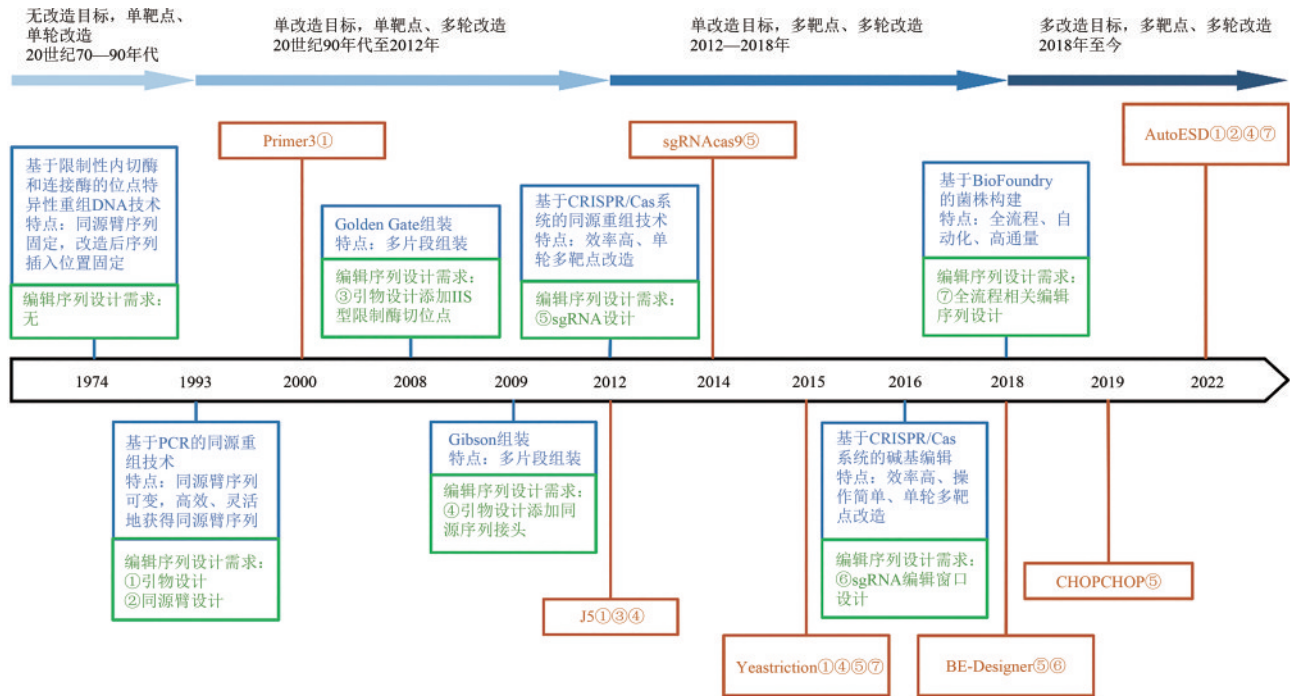


图1 遗传操作技术与编辑序列设计需求、工具的代表性进展

(蓝色框代表遗传操作技术; 绿色框代表编辑序列设计需求; 棕色框代表编辑序列设计工具; 棕色序号代表该编辑序列设计工具能够解决对应的编辑序列设计需求)

Fig. 1 Development of genetic manipulation technologies and corresponding editing sequences and tools for editing sequence design

(Blue rectangle: Genetic manipulation technologies; Green rectangle: Editing sequence design requirements; Brown rectangle: Editing sequence design tools; Brown serial numbers indicates that the editing sequence design tools can address the corresponding editing sequence design requirements)

技术, 构建质粒或者重组DNA片段, 对目标基因进行单基因克隆<sup>[1-2]</sup>、失活或者染色体插入<sup>[3]</sup>, 可以支持对单基因或者基因簇的扰动, 探究其对菌株生理性状或者产品生成的影响; ②凭借经典同源重组技术和新兴的DNA组装技术<sup>[4-5]</sup> (Gibson、Golden Gate等), 可迭代地编辑(插入、敲除与替换)基因组靶序列<sup>[6-7]</sup>, 可以支持以某一具体性状目标(快速生长或者产品高效生成)的多基因同时克隆和多轮迭代菌种改良; ③基于新兴的基因组编辑技术<sup>[8-10]</sup> (CRISPR/Cas同源重组与碱基编辑等), 进行与大片段组装, 实现精准高效的基因组多位点迭代编辑, 缩短了菌种构建改良的研发周期; ④基于自动化装备的合成生物铸造厂(biofoundry)<sup>[11-13]</sup>, 自动化高通量地进行基因克隆与基因组多位点迭代编辑以同步构建针对不同应用需求的大批量菌株, 有望变革传统的劳动密集型的研究范式, 创建出自动化与高通量的工程菌种铸造模式。

各种遗传操作技术中, 用于精准定位与编辑

目标DNA序列的相关序列可以被称为编辑序列<sup>[14]</sup>, 例如用于定位和扩增目标DNA序列的引物序列、用于定位和等位替换基因组上目标DNA序列的同源臂序列、用于定位和导引Cas酶的sgRNA序列等。编辑序列设计, 即按照合理的规则和逻辑设计功能优化的编辑序列, 以确保预期的编辑效率和精准度。例如, 设计位置和长度合适的同源臂序列, 以实现改造后序列在参考基因组上的精确定位和高效替换; 设计参数(熔解温度、二级结构生成倾向、二聚体生成倾向等)优化的引物序列, 以通过PCR高成功率地获得同源臂序列。早期的编辑序列设计主要是引物设计, 就已经包含很多涉及复杂算法或者大规模计算的步骤, 例如引物设计中基于多因素评价指标和复杂评价规则的评分和层级筛选过程, 往往需要借助诸如Primer3<sup>[15]</sup>等专业的计算机辅助设计工具才能得到优化的结果。随着遗传操作技术的迭代升级和新操作技术新变种的不断开发, 编辑序列的复杂度与范畴随之变化, 面向不同应用场景的编辑

序列设计工具也应需而生。例如，随着诸如 Gibson 等多片段 DNA 组装技术的应用推广，引物设计环节中需引入用于桥联不同片段的重叠序列，并需要考虑各个重叠序列的独特性，确保各个片段按照预期的顺序组装<sup>[4]</sup>。当多片段进行组装时，需要考量的因素呈指数增长，往往无法由人工完成设计，因此催生了 J5<sup>[16]</sup> 等专门用于 DNA 组装的设计工具。随着 CRISPR/Cas 等新一代基因组编辑技术的兴起，涌现出许多诸如 CHOPCHOP<sup>[17]</sup> 等专门用于设计效率高且脱靶风险低的 sgRNA 序列的工具。

得益于遗传操作技术的发展，微生物菌种构建能力大幅提升，但是由于细胞代谢的高度复杂性及人工细胞理性设计能力的缺失，目前仍然需要进行长期、反复的人工实验试错，才能逐步将工程菌种的化学品合成能力（产品产量、得率与生产速率）提升至有商业竞争力的水平。然而，海量的工程化试错实验超出了传统的劳动密集型研究范式的能力范畴，为此近年来国内外数十个研究机构已搭建或者正在搭建自动化合成生物铸造厂（Biofoundry）<sup>[18-20]</sup>，旨在提升合成生物实验对象、方法、技术和流程的标准化和模块化水平，实现海量工程试错的自动化闭环运行，从而实现具有高通量、低成本、多循环特性的“设计—构建—测试—学习”的合成生物系统创制技术能力。但由于单一的编辑序列设计工具往往只覆盖遗传操作的部分环节，例如 J5<sup>[16]</sup> 等 DNA 组装设计工具，只覆盖了质粒构建环节的设计，而面向基因组编辑还需要诸如 sgRNA、同源臂等功能序列的设计，必须联用诸如 CHOPCHOP<sup>[17]</sup> 等其他软件或工具。此外，由于不同软件间的数据交互一般是非标准化的，用户往往需要手动完成输入/输出信息的转换或加工，提高了用户的使用门槛。因此，包括笔者团队在内的研究者们开发了诸如 GeneTargeter<sup>[21]</sup> 和 AutoESD<sup>[14]</sup> 等全流程一站式的编辑序列设计自动化工具，以匹配现有硬件设施的操作通量，也为下一步与 Biofoundry 深度融合奠定了基础。

本文按照不同编辑序列类型和遗传操作应用场景，将编辑序列设计工具划分为引物设计工具、DNA 组装的设计工具、sgRNA 设计工具、基因组

编辑的全流程设计工具四种类型进行总结与分析。并对未来的发展方向进行了展望，为将来推进快速高效的“设计—构建—测试—学习”合成生物学工程循环的实现奠定基础。

## 1 引物设计工具

引物是聚合酶链反应（polymerase chain reaction, PCR）中用于定位待扩增序列并作为序列复制起点的寡核苷酸序列。在备选引物集合中筛选具有高可用性的引物的过程被称为引物设计。它是决定 PCR 效率的重要环节，也是制备与进一步准确组装同源臂等编辑序列片段的基础环节。早期的引物设计目标是确保 PCR 的效率，需要权衡多方面的参数，包括引物长度、退火温度（ $T_m$ ）、碱基对重复次数、鸟嘌呤/胞嘧啶（GC）含量、二级结构生成倾向性、同源/异源二聚体生成倾向性、定位特异性等<sup>[22-23]</sup>。鉴于众多的变量及相关的复杂搜索、比对和优化，相应的计算需要借助计算机软件实现。为此，一批商业化软件（Oligo、Primer Premier 等）涌现，旨在为生物学用户提供准确、高效且用户友好引物设计服务及操作界面。之前发表的论文<sup>[24-25]</sup>已对这些商业化软件的功能与使用方法进行总结，且商业化工具不易被整合，本文主要针对开源工具进行总结。2000 年，引物设计领域经典的开源软件 Primer3<sup>[26]</sup> 首次发布，提供多种可选参数来控制引物的位置、长度、GC 含量、熔解温度以及产物（目标序列）长度，以满足不同的应用需求。而在后续 Primer3 的升级中，2007 年修订版更新了计算熔解温度和盐浓度校正的公式，提高了对熔解温度预测的准确性<sup>[27]</sup>；2012 年的更新引入了更精确的热力学模型，进一步提高了对熔解温度预测的准确性，并且可以降低引物形成二级结构及二聚体的可能性。此外，该次升级也着重优化了编程接口的简洁性，提高了其与其他软件的可整合性<sup>[15]</sup>。鉴于 Primer3 引物设计的准确性、易用性以及开源性，后续的开源引物设计工具通常会基于 Primer3 进一步优化设计算法或者开发面向特定应用场景的新功能，例如批量引物设计<sup>[28]</sup>、基因组规模引物设计<sup>[29-31]</sup>、RT-PCR 引物设计<sup>[32-34]</sup>、用于检测不同生物体中的剪接变体<sup>[35-37]</sup> 或

单核苷酸多态性 (single nucleotide polymorphism, SNP)<sup>[38-39]</sup> 的引物设计等。

随着分子克隆技术的发展, PCR的目标和用途也不断扩展,不同的目标和用途产生了相应的特异性引物设计的需求,进而推动了相关引物设计工具的开发。表1总结了不同功能的主流引物设计工具。具体到遗传改造相关的PCR,其主要目标是扩增用于构建重组载体或重组片段(如插入、替换片段和用于同源重组的同源序列)的序列以及对改造后的目标序列的测序验证。相应的引物设计需要解决的特异性问题主要包括提高引物的特异性(即降低引物错误定位的风险)、引物的高通量设计以及重组载体或重组片段的组装,相关设计工具的引物特异性检验功能及设计通量在表1中做了总结,重组载体或重组片段组装相关的引物设计工具见“2 DNA组装的设计工具”一节。

尽管PCR的目标和用途呈多样化发展,但引物特异性检验(即降低引物错误定位的风险)始终是重要的质控环节。一方面引物的错误定位会减少与目标位点结合的引物数量,另一方面错误

定位的引物也有可能产生错误的扩增产物。在对微生物特别是基因组的重复性相对较高的真核微生物进行遗传操作时,需要视情况选择具有引物特异性分析和筛选功能的引物设计工具。引物的错误定位分析通常借助序列比对工具实现,例如,基于序列比对分析开发的Primer-BLAST<sup>[41]</sup>将序列比对工具BLAST和Needleman-Wunsch全局比对算法结合,可以更加精准地定位参考序列中的潜在错配。除了序列比对,引物与模板杂交的热力学也是影响引物特异性的主要因素之一<sup>[48]</sup>,部分工具尝试将二者以及更多因素结合进行引物特异性分析<sup>[49-50]</sup>。以MFEprimer-3.0为例,其使用k-mer索引算法来加速引物与潜在结合位点的搜索过程,并整合了快速识别引物二级结构和同源、异源引物二聚体的新算法来进行特异性评估<sup>[49]</sup>。

## 2 DNA组装的设计工具

面向基因组编辑的遗传操作需要将用于定位和编辑目标DNA序列的相关序列片段(如同源臂、

表1 不同功能的主流引物设计工具

Table 1 Major primer design tools with different functions

工具	基于 Primer3	特异性 检验	批量 设计	功能特色	可用性	URL
Primer3 <sup>[15, 26-27]</sup>	—	否	否	基础引物设计	在线/图形界面、 离线/命令行	<a href="https://bioinfo.ut.ee/primer3/">https://bioinfo.ut.ee/primer3/</a>
BatchPrimer3 <sup>[28]</sup>	是	否	是	批量引物设计	在线/图形界面	<a href="http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi">http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi</a>
ConservedPrimer2.0 <sup>[35]</sup>	是	是	是	Sanger测序引物设计	在线/图形界面	<a href="https://probes.pw.usda.gov/ConservedPrimers/index.html">https://probes.pw.usda.gov/ConservedPrimers/index.html</a>
PrimerDesign-M <sup>[40]</sup>	否	是	是	Sanger测序引物设计	在线/图形界面	<a href="http://www.hiv.lanl.gov/content/sequence/PRIMER_DESIGN/primer_design.html">www.hiv.lanl.gov/content/sequence/PRIMER_DESIGN/primer_design.html</a>
Primer-BLAST <sup>[41]</sup>	是	是	否	RT-PCR引物设计	在线/图形界面	<a href="https://www.ncbi.nlm.nih.gov/tools/primer-blast/">https://www.ncbi.nlm.nih.gov/tools/primer-blast/</a>
PCR designer <sup>[42]</sup>	否	否	否	SNP分析引物设计	在线/图形界面	<a href="http://primer1.soton.ac.uk/primer.html">http://primer1.soton.ac.uk/primer.html</a>
PRIMEGENS-v2 <sup>[37]</sup>	是	否	是	剪接变体引物设计	在线/图形界面、 离线/命令行	<a href="http://primegens.org/">http://primegens.org/</a>
PrimerSeq <sup>[43]</sup>	是	否	是	剪接变体引物设计	离线/图形界面	<a href="https://primerseq.sourceforge.net/">https://primerseq.sourceforge.net/</a>
MSP-HTPrimer <sup>[44]</sup>	是	否	是	DNA甲基化分析引物设计	离线/命令行	<a href="https://sourceforge.net/projects/msp-htprimer/">https://sourceforge.net/projects/msp-htprimer/</a>
MPD <sup>[45]</sup>	否	是	是	多重PCR引物设计	离线/命令行	<a href="https://wingolab-org.github.io/mpd-c/">https://wingolab-org.github.io/mpd-c/</a>
Mutation Maker <sup>[46]</sup>	是	是	是	多重PCR引物设计、氨基 酸突变引物设计	离线/命令行	<a href="https://github.com/ra100/Mutation_Maker">https://github.com/ra100/Mutation_Maker</a>
PerlPrimer <sup>[47]</sup>	否	否	是	Sanger测序引物设计、 RT-PCR引物设计、开放阅 读框(ORF)搜索及相关引 物设计	离线/图形界面	<a href="https://perlprimer.sourceforge.net/">https://perlprimer.sourceforge.net/</a>

待插入序列等) 组装成特定重组载体导入菌株中发挥功能来实现改造目标。因此, 精准高效的DNA组装(重组载体/片段)是遗传操作需解决的关键问题之一。通过在PCR两端的引物5'端引入特异的限制性核酸内切酶识别与切割位点, 依靠内切酶产生的不同黏性末端, 按照预定位置与方向连接不同的DNA片段, 是最经典的DNA组装方式<sup>[51]</sup>。以此为基础衍生出了多种技术: 基于同尾酶的BioBrick<sup>[52]</sup>和BglBrick<sup>[53]</sup>等技术, 该方法利用一对同尾酶(识别位点序列不同, 但酶切后产生的黏性末端相同的核酸内切酶)和两个非同尾酶将载体和DNA元件标准化, 形成可即插即用的元件库; 基于IIs型限制性内切酶的Golden Gate<sup>[54]</sup>技术, 该方法利用IIs型限制性内切酶识别位点与切割位点不同的特性, 借助相同的限制酶产生的不同黏性末端实现多个DNA片段的无痕组装, 该方法灵活性高, 可实现多达20个以上片段的组装。

基于限制性内切酶的组装技术前提是原始序列中不能包含相关内切酶识别位点, 这对此类技术的应用造成了一定限制。基于重叠延伸PCR(overlap extension PCR)原理<sup>[55]</sup>开发的组装方法则摆脱了对特异性限制性内切酶的依赖, 此类技术包括CPEC<sup>[56]</sup>、Gibson组装<sup>[4]</sup>、SLIC<sup>[57]</sup>和LCR<sup>[58]</sup>等。以Gibson组装为例(图2), 通过在引物中添加重叠序列, 借助PCR可以在待组装片段端部引入特异性重叠区域。然后利用T5核酸外切酶的5'外切酶活性, 酶切各片段产生同源序列单链接头, 经退火使各片段借接头连接, 再经DNA聚合酶与连接酶修补切口, 即可完成多片段DNA的无痕组装。该方法的3种酶最适温度接近, 可以在一个体系中一步实现片段组装。该方法可以实现十万碱基对长度的片段组装, 但可组装的片段数量较少, 通常不超过6个。

在进行质粒DNA组装的引物设计时, 需要在

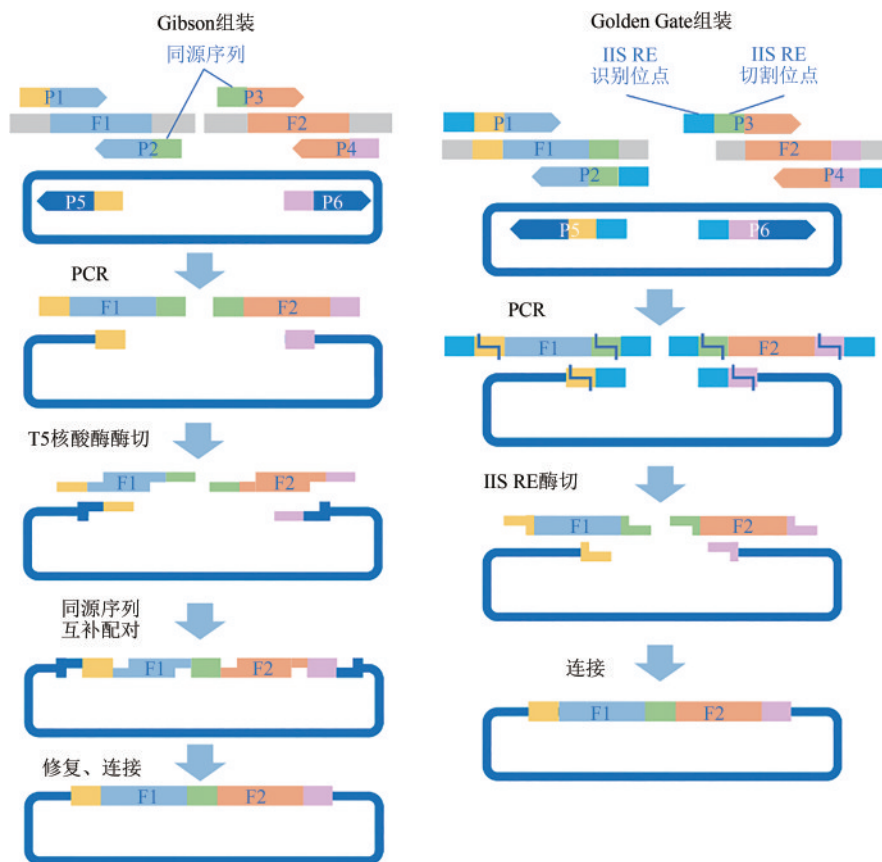


图2 Gibson 和 Golden Gate DNA 序列组装技术及相关编辑序列设计原理

F—片段; P—引物; RE—限制性内切酶

Fig. 2 Gibson and Golden Gate assembly and corresponding principles for editing sequence design

F—Fragment; P—Primer; RE—Restriction enzyme

引物的5'端添加接头序列(如Golden Gate技术所需的II s型限制酶识别位点与和切割位点序列、Gibson组装所需的重叠序列),借助PCR将接头序列引入待组装的DNA片段,实现多片段的正确组装(图2)。理论上可以通过人工手动在引物设计中引入接头序列实现组装,但这不仅需要研究者具有较高的DNA组装技术经验,而且时间和劳动力成本较高。此外,由于认知偏好,不同研究者的人工设计结果也会有差异,这与合成生物学日益完善的标准化理念冲突。为此,研究人员开发了多种面向不同DNA片段组装技术的CAD工具。表2总结了代表性工具的特点和适用范围。2012年Jay D. Keasling团队开发了面向质粒DNA组装设计的网站工具J5<sup>[16]</sup>。J5使用Primer3作为底层引物设计算法,集成了Golden Gate、Gibson、CPEC和SLIC等多种组装技术,支持高通量设计和质粒文库构建。凭借实用的实验成本预估功能、完善的可视化功能和优良的用户友好性,该工具在代谢工程及合成生物学领域获得了广泛应用。为了优化Gibson组装过程,后续的研究者又开发了Raven<sup>[59]</sup>,该工具借助核心算法产生优化的片段组装计划并支持基于实验结果以交互方式重新设计组装计划。为了支持包括Lambda Red重组等更多的重组技术和追踪整个过程的数据,2016年研究人员开发了MCDS<sup>[60]</sup>。

随着研究者对质粒DNA组装技术的不断深入研究,相关技术不断升级和更新,相关设计工具也在不断发展。以Golden Gate组装技术为例,早期的设计工具诸如J5<sup>[16]</sup>、MCDS<sup>[60]</sup>在进行切割位点选取的时候只考虑了切割位点序列之间的相似性以确保组装效率。近年来,随着针对不同黏性

末端序列组合对组装效率和特异性的影响的研究数据不断丰富<sup>[62-63]</sup>,更多切割位点选取新规则被纳入新的设计工具的考量范围<sup>[61-62]</sup>。比如2020年,研究人员首先通过一种基于三代测序的方法检查Golden Gate组装的切割位点序列的相互结合能力,并基于此开发了NEBbridge Golden Gate组装优化设计工具<sup>[61]</sup>。通过对切割位点序列的优化设计,该工具能够实现多达35个DNA片段的同时组装。此外,相比于基因组编辑而言,质粒DNA组装因其相对简单,更易进行模块化、标准化处理。因此相关的生物实验操作流程已经成功地与自动化设施整合,可以支持高通量自动化的质粒DNA组装<sup>[11, 13, 64]</sup>。其中,美国伊利诺伊大学Zhao团队<sup>[13]</sup>基于iBioFAB自动化铸造平台,开发了在线的编辑序列设计工具PlasmidMaker,可以允许用户从网站前端界面提交质粒构建需求,由系统自动化、高通量地完成相关引物设计,并根据生成的任务清单触发机器人系统集成的后端开展自动化实验构建。

### 3 sgRNA设计工具

借助基因组编辑技术(基于筛选标记的经典同源重组技术、CRISPR/Cas介导的同源重组技术以及碱基编辑技术等)直接对基因组DNA序列进行精准的靶向改造(敲除、插入和替换),是构建稳定、安全的工程菌种,实现绿色、高效的生物制造的常用手段。近年来,CRISPR/Cas系统因其易于构建、编辑效率高等优点逐渐成为应用最为广泛的基因组编辑工具,其借助单链向导RNA

表2 面向DNA片段组装的代表性设计工具

Table 2 Representative design tools for DNA assembly

工具	基于Primer3	支持的DNA片段组装技术	批量设计	可用性	URL
J5 <sup>[16]</sup>	是	Golden Gate/Gibson/ CPEC/SLIC	是	在线/图形界面	<a href="https://j5.jbei.org/">https://j5.jbei.org/</a>
Geneious	否	Gibson	是	在线/图形界面	<a href="https://www.geneious.com/">https://www.geneious.com/</a>
Raven <sup>[59]</sup>	否	Gibson	是	在线/图形界面	<a href="http://ravencad.org/">http://ravencad.org/</a>
MCDS <sup>[60]</sup>	否	GoldenGate/Gibson/ SLIC/Gateway	否	离线/命令行	<a href="https://github.com/errisy/MCDS">https://github.com/errisy/MCDS</a>
NEBbridge Golden Gate <sup>[61]</sup>	未知	Golden Gate	是	在线/图形界面	<a href="https://goldengate.neb.com/">https://goldengate.neb.com/</a>
iBioCAD GGA <sup>[62]</sup>	未知	Golden Gate	是	在线/图形界面	<a href="https://ibiocad.igb.illinois.edu/">https://ibiocad.igb.illinois.edu/</a>
PlasmidMaker <sup>[13]</sup>	是	Golden Gate/Gibson	是	在线/图形界面	<a href="https://biofoundry.web.illinois.edu/">https://biofoundry.web.illinois.edu/</a>

(single-guide RNA, sgRNA) 识别前间区序列邻近基序 (protospacer adjacent motif, PAM) 并引导 Cas 核酸酶结合和切割目标 DNA 产生 DNA 双链断裂 (double stranded break, DSB), 后续通过非同源末端连接 (non-homologous end-joining, NHEJ) 修复或同源重组修复 (homology-directed repair, HDR) 等方式进行修复, 达到基因组编辑的目的 (图3)<sup>[65]</sup>。CRISPR/Cas 系统的作用效果取决于其组分 (sgRNA 和 Cas 酶) 的特性, 而在特定的 Cas 核酸酶 (Cas9、Cpf1 等) 下, sgRNA 的编辑效率和特异性 (on-target and off-target) 是决定 CRISPR 系统效率的主要因素。因此, sgRNA 设计也是应用 CRISPR/Cas 系统所需的编辑序列设计的核心任务, 需要针对不断新发现的 Cas 系统开发对应的设计工具。

sgRNA 序列通常由两部分组成<sup>[8, 66]</sup>: pre-crRNA (与目标 DNA 序列反向互补的序列, 经 RNA 酶修饰后形成 crRNA, 负责引导 sgRNA-Cas 蛋白复合体结合靶序列, 通常为 sgRNA 的 5' 端前 20 bp 且靶序列下游必须含有 Cas 蛋白特异的 PAM 序列) 和 tracrRNA (与 Cas 蛋白特异性结合的序列, 一方面负责启动 RNA 酶将 pre-crRNA 加工为成熟的 crRNA, 另一方面 sgRNA-Cas 蛋白复合体结合靶序列后负责激活 Cas 蛋白切割目标 DNA 形成 DSB)。因此, 在特定的 Cas 核酸酶 (Cas9、Cpf1 等) 系统中, sgRNA 设计的核心是针对 pre-crRNA 序列的设计。理论上, 如果 sgRNA 中的 5' 端 20 bp 序列与目标 DNA 序列完全互补, 则 sgRNA-Cas9 复合物应该与该位点结合并进行切割, 但大量的研究报道显示不同 sgRNA 序列的效

率差异很大<sup>[67-69]</sup>。sgRNA 的编辑效率不仅取决于 PAM 位置以及 sgRNA 和靶序列之间的特异性碱基配对, 还受到 DNA 序列特征<sup>[70]</sup>、染色质特征<sup>[71-72]</sup>、sgRNA 特定位置的碱基偏好性<sup>[72-74]</sup>、靶序列下游的 DNA 序列<sup>[72, 75]</sup> 和某些表观遗传因素<sup>[76]</sup> 等的影响。而在脱靶效应 (特异性) 方面, 除了目标序列在参考基因组中的特异性<sup>[77-78]</sup> 这一主要因素外, 还受到染色质状态<sup>[79-80]</sup>、Cas 酶特征<sup>[81-82]</sup> 等因素的影响。例如, *SpCas9* (*Streptococcus pyogenes* originated Cas9) 识别的最佳 PAM 是 NGG, 但也能以低频率与 NAG 或 NGA 结合<sup>[83-84]</sup>; sgRNA 的 5' 端 20 bp 序列和靶序列之间对碱基错配有一定的容忍度<sup>[85-86]</sup> 等。鉴于如此多的影响因素, sgRNA 的设计通常需要依靠计算机软件实现。

sgRNA 的编辑效率和特异性预测的策略基本相同。首先通过序列搜索或比对获取候选 sgRNA 及其潜在脱靶位点, 再基于一定的特征和评价标准借助特定的算法或模型预测编辑效率和特异性, 进而优选 sgRNA。候选 sgRNA 一般通过搜索目标序列上的 PAM 序列 (如 *SpCas9* 的 5'-NGG-3') 获得, 无需借助专门的计算工具。获取 sgRNA 潜在脱靶位点需要借助一定的算法在全基因组范围内进行搜索和比对, 计算量较大, 需要借助专门的计算工具。多种经典的序列比对工具 (如 Bowtie<sup>[87]</sup>、BWA<sup>[88]</sup>、SeqMap<sup>[89]</sup> 等) 和为 CRISPR 技术定制的比对工具 (如 Cas-OFFinder<sup>[90]</sup>、CasOT<sup>[91]</sup>、CRISPRitz<sup>[92]</sup> 等) 被改造和开发用于搜索 sgRNA 潜在脱靶位点, 并被整合到多种 sgRNA 设计工具中<sup>[17, 93-97]</sup>。编辑效率和特异性的预测是 sgRNA 设计的核心, 相关的设计工具按照基本的预测原理

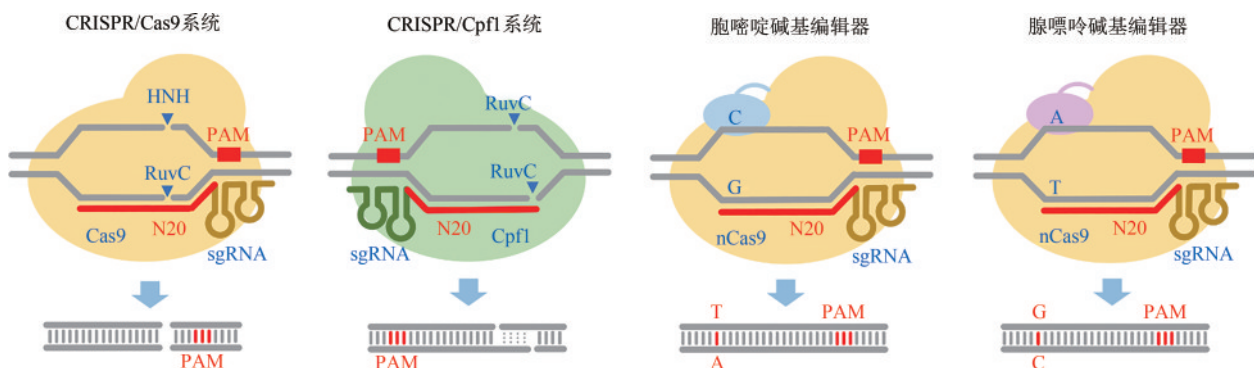


图3 代表性的CRISPR/Cas系统介导的基因组编辑原理

Fig. 3 Principles of genome editing mediated by representative CRISPR/Cas systems

可分为两类：基于假设先验的工具和基于学习（机器学习或深度学习）的工具。本节主要对这两类sgRNA设计工具进行概述，相关代表性工具在表3中做了总结。

基于假设先验的工具基于已有的知识和经验综合考虑多种因素对sgRNA编辑效率或特异性的影响，按设定的规则对候选sgRNA或sgRNA潜在脱靶位点进行评估和打分。基于假设先验的工具开发较早，一般都包含编辑效率和特异性预测（即 on-target 和 off-target 评分），支持多种 Cas 酶（PAM 类型）以及多种改造类型，普适度较高且一般会开发面向实验生物学用户的网站服务<sup>[17, 94, 98, 101-103]</sup>。基于学习的方法利用已知的sgRNA编辑效率或特异性数据训练模型，再用训练好的模型预测未知sgRNA的编辑效率或特异性。但由于受限于数据的规模和来源（数据来源通常基于研究最深入、应用最广泛的 *SpCas9*，PAM 位点为 NGG），通常只支持一种 Cas 酶类型<sup>[107-109]</sup>。此外，基于一个物种的数据训练的sgRNA编辑效率或特异性预测模型一般无法应用到另外的物种，甚至在同一个物种不同的细胞系中的预测效果都不甚理想<sup>[110]</sup>，所以目前包含基于学习的sgRNA评估打分模块的设计工具大多只适用于面向人类基因组改造的sgRNA设计，而面向更广泛物种的sgRNA设计工具通常采用基于假设先验的策略。以 CHOPCHOP 为例，经过 3 个版本的迭代更新<sup>[17, 102-103]</sup>，通过方便的网站服务，已经实现对超

过 100 个物种和多种编辑技术（CRISPR/Cas9，CRISPR/Cas9 nickase，CRISPR/Cpf1，CRISPR/Cas13，TALEN）的支持，并且提供了不同编辑效率和特异性打分的选择，极大地方便了实验生物学家。但我们也注意到在一些第三方的评测中，基于学习特别是深度学习的工具（比如 DeepHF<sup>[111]</sup>）在人类细胞系等相关数据集上展现出相比于传统基于假设先验工具更好的性能<sup>[110]</sup>。未来，随着更多不同条件的sgRNA编辑效率和特异性实验数据的获得，整合不同来源的基因组编辑数据，可能确定无偏见的sgRNA设计规则和影响sgRNA的编辑效率和脱靶效应的关键特征，这也是sgRNA设计的重要未来方向。

此外，除了常用的 CRISPR/Cas 同源重组系统，基于 CRISPR 的碱基编辑技术在微生物里也有广泛的应用（图3）。面向碱基编辑sgRNA设计，需要考虑碱基编辑器类型、编辑窗口、目标碱基位置等，有时还需要根据编辑后改变的翻译产物选择编辑位点和sgRNA，因此常规的sgRNA设计工具不能直接用于碱基编辑sgRNA设计。2018年，研究人员开发了面向碱基编辑的在线工具 BE-Designer<sup>[112]</sup>，可以针对不同的参考基因组，根据用户需求选择不同 PAM、碱基编辑器类型、编辑窗口进行sgRNA设计，输出结果为包含可编辑序列、相对目标位置、GC 含量和潜在的脱靶位点的候选sgRNA列表。2019年，研究人员开发了 beditor<sup>[113]</sup>，可以根据用户提供的目标核苷酸或氨

表3 基于假设先验方法与基于学习方法的代表性sgRNA设计工具

Table 3 Representative hypothesis-driven and learning-based sgRNA design tools

工具	编辑效率预测方法	特异性预测方法	PAM	支持的改造类型	URL
E-CRISP <sup>[98]</sup>	假设先验	假设先验	NGG	敲除、激活、抑制	<a href="http://www.e-crisp.org/E-CRISP/index.html">http://www.e-crisp.org/E-CRISP/index.html</a>
CRISPR-ERA <sup>[99]</sup>	假设先验	假设先验	NGG	敲除、激活、抑制	<a href="http://CRISPR-ERA.stanford.edu">http://CRISPR-ERA.stanford.edu</a>
EuPaGDT <sup>[100]</sup>	假设先验	假设先验	NGG, TTTN, NGA 等	敲除	<a href="http://grna.ctegd.uga.edu/">http://grna.ctegd.uga.edu/</a>
CRISPOR <sup>[94, 101]</sup>	假设先验/机器学习	假设先验	NGG, TTTN, NGA 等	敲除	<a href="http://crispor.tefor.net/">http://crispor.tefor.net/</a>
CHOPCHOP <sup>[17, 102-103]</sup>	假设先验/机器学习	假设先验	NGG, TTTN, NGA 等	敲除、敲入、激活、抑制	<a href="http://chopchop.cbu.uib.no/">http://chopchop.cbu.uib.no/</a>
CRISPETA <sup>[104]</sup>	假设先验	数据库搜索	NGG	敲除	<a href="http://crispeta.org.eu">http://crispeta.org.eu</a>
CRISPR-P <sup>[105]</sup>	假设先验	假设先验	NGG, NRG, TTN 等	敲除	<a href="http://crispr.hzau.edu.cn/CRISPR2/">http://crispr.hzau.edu.cn/CRISPR2/</a>
CRISPRscan <sup>[106]</sup>	机器学习	假设先验	NGG, TTTN, NGN 等	敲除	<a href="https://www.crisprscan.org">https://www.crisprscan.org</a>
WU-CRISPR <sup>[107]</sup>	机器学习	假设先验	NGG	敲除	<a href="http://crisprdb.org/wu-crispr/">http://crisprdb.org/wu-crispr/</a>
CRISTA <sup>[108]</sup>	机器学习	机器学习	NGG	敲除	<a href="https://crista.tau.ac.il/">https://crista.tau.ac.il/</a>
Elevation <sup>[109]</sup>	机器学习	机器学习	NGG	敲除	<a href="https://crispr.ml/">https://crispr.ml/</a>

基酸突变类型、PAM、碱基编辑器类型、编辑窗口灵活地设计用于碱基编辑的 sgRNA，并且实现了对非模式生物的基因组的支持。此外，beditor 引入了针对 sgRNA 的脱靶风险的打分函数，考虑错配与 PAM 的相对位置，能够更准确地预估 sgRNA 的潜在脱靶风险。此外，还有面向碱基编辑的特定应用开发的 sgRNA 设计工具，例如将编码区序列中的非终止密码子编辑为终止密码子以实现基因失活是胞嘧啶碱基编辑器（cytosine base editor, CBE）最常规的应用。与常规设计相比，相应的 sgRNA 设计需要额外考虑 sgRNA 特定位置的编辑窗口中是否包含编辑后转变为终止密码子的密码子。为此，笔者所在团队前期开发的在线工具 gBIG<sup>[114]</sup>，实现了面向 CBE 和腺嘌呤碱基编辑器（adenine base editor, ABE）介导的基于引入终止密码子的基因失活的碱基编辑设计。用户通过在线选择 PAM 序列、碱基编辑器类型和参考基因组，并上传包含一个或多个目标基因序列的 FASTA 格式文本文件进行 sgRNA 设计。

#### 4 基因组编辑的全流程设计工具

一个完整的基因组编辑流程一般包含多个编辑序列设计任务。以 CRISPR/Cas 系统介导的同源重组技术实现靶序列敲除为例（其原理如图 4）。首先 CRISPR/Cas 系统对靶序列的定位和高效切割需要在靶序列中搜索具有高靶向切割效率和低脱靶风险的 sgRNA [图 4(d)]。改造后序列（本例中为空序列）对靶序列的替换由依赖于上下游同源臂序列的同源重组双交换实现。通过在靶序列上下游适当区域中选择具有优化的评价指标的引物，借助高成功率的 PCR 可以获得同源臂序列 [图 4(a)和(b)]。包含 crRNA、tracrRNA 以及相关启动子和终止子序列的 sgRNA 片段，也需要通过设计特定引物进行 PCR 制备 [图 4(c)]。随后这些重组片段需要借助特定的 DNA 组装技术（如 Gibson 或 Golden Gate 技术）组装成重组质粒，因此在设计片段扩增引物时，需要提前考虑引入接头序列 [图 4(a)~(d)]。重组质粒通过转化进入宿主细胞，经过转录翻译产生 CRISPR/Cas 系统发挥效应，在靶序列特定位置产生 DSB。DSB 借助重组

质粒同源臂参与的同源重组得以修复，同时实现靶序列的删除。最后需要在靶序列同源臂序列上下游适当区域中设计测序验证引物以实现基因组编辑后序列的测序验证 [图 4(e)]。

可见，完成一个基因组编辑包含的编辑序列设计任务需要多种类型的设计工具配合使用。例如，使用 J5<sup>[16]</sup> 进行重组载体片段组装设计的同时可以实现片段的 PCR 扩增引物设计，使用 CHOPCHOP<sup>[103]</sup> 可以实现 CRISPR/Cas 系统的 sgRNA 设计、使用 Primer3<sup>[15]</sup> 可以实现改造后针对目标序列的测序验证引物的设计。但这种多工具联用的模式也存在着诸多问题，例如，各设计步骤间可能涉及不同格式的输入/输出文件的判断、筛选、转换等人工操作，这不但耗用户的时间和劳动力，而且可能导致无法预见的错误；需要用户熟练掌握多种生信工具，学习成本较高等。此外，自动化菌种铸造的高速发展使高通量的基因组编辑成为可能，人工或者多软件联用的“半自动化”的编辑序列设计模式已经无法匹配现有硬件设施的操作通量。为此，研究者们着手于开发可以覆盖基因组编辑全流程的设计工具（表 4）。

2015 年，荷兰代尔夫特理工大学的研究团队开发了首个针对基因组编辑全流程设计的网站工具 Yeastription v0.1<sup>[115]</sup>。该工具支持基于 CRISPR/Cas 系统的面向基因敲除（单基因或双基因）的引物及 sgRNA 等编辑序列的一站式设计。但支持的 CRISPR/Cas 系统限定在基于 *SpCas9*（PAM: NGG）和 pMEL10-pMEL17 或者 pROS10-pROS17 系列的 sgRNA 表达工具，且仅支持 33 种不同的酵母菌株基因组。2021 年，笔者所在团队针对谷氨酸棒杆菌的基因组点突变构建的编辑序列设计场景，开发了在线工具 GEDpm-cg<sup>[116]</sup>，可以在 5 min 内完成 10<sup>4</sup> 级别的单碱基插入、敲除与替换的编辑序列设计一站式自动化设计，并借助实验验证了设计结果的可靠性。2022 年，美国麻省理工学院的学者开发了面向镰状疟原虫 3D7 (*Plasmodium falciparum* 3D7) CRISPR/Cas 系统基因组编辑的在线工具 GeneTargeter<sup>[21]</sup>。该工具可以实现多目标基因的敲除及条件抑制所需编辑序列的全流程一站式设计。然而上述工具通常是专门针对特定物种

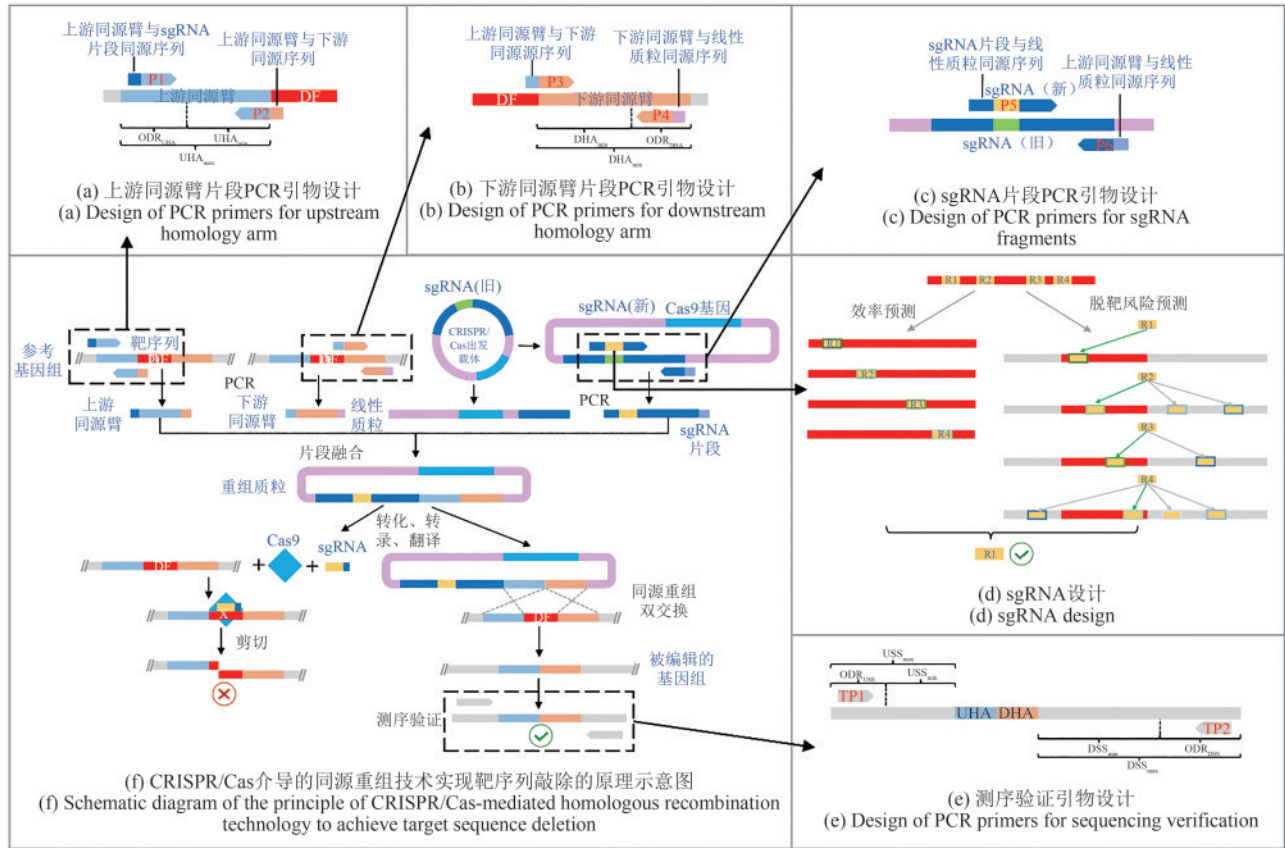


图4 CRISPR/Cas系统介导的同源重组技术实现靶序列敲除相关编辑序列设计

DF—敲除的片段；P—引物； $UHA_{min}$ —上游同源臂长度最小值； $UHA_{max}$ —上游同源臂长度最大值； $ODR_{UHA}$ —上游同源臂上游引物的可选设计区域； $DHA_{min}$ —下游同源臂长度最小值； $DHA_{max}$ —下游同源臂长度最大值； $ODR_{DHA}$ —下游同源臂下游引物的可选设计区域；R—gRNA（sgRNA边框线的绿色和蓝色的深度分别代表 on-target 效率和 off-target 风险）；TP—证引物；UHA—上游同源臂；DHA—下游同源臂； $USS_{min}$ —游间隔序列（上游验证引物的3'末端到上游同源臂5'末端的序列）的长度最小值； $USS_{max}$ —游间隔序列的长度最大值； $ODR_{USS}$ —游验证引物可选设计区域； $DSS_{min}$ —游间隔序列（下游验证引物的3'末端到下游同源臂3'-末端的序列）的长度最小值； $DSS_{max}$ —游间隔序列的长度最大值； $ODR_{DSS}$ —游验证引物可选设计区域

Fig. 4 Schematic diagram for CRISPR/Cas-based homologous recombination and its editing sequences

DF—Deleted fragment; P—Primer;  $UHA_{min}$ —Minimum length of upstream homology arm;  $UHA_{max}$ —Maximum length of upstream homology arm;  $ODR_{UHA}$ —Optional design region of the upstream primer of the upstream homologous arm;  $DHA_{min}$ —Minimum length of downstream homology arm;  $DHA_{max}$ —Maximum length of downstream homology arm;  $ODR_{DHA}$ —Optional design region of the downstream primer of downstream homologous arm; R—sgRNA (dark green and blue of the sgRNA border lines represent on-target efficiency and off-target risk, respectively); TP—Test primer; UHA—Upstream homologous arm; DHA—Downstream homologous arm;  $USS_{min}$ —Minimum length of upstream spacer sequence (sequence from the 3'-end of the upstream verification primer to the 5'-end of the upstream homology arm);  $USS_{max}$ —Maximum length of upstream spacer sequence;  $ODR_{USS}$ —Optional design region of the upstream verification primer;  $DSS_{min}$ —Minimum length of downstream spacer sequence (sequence from the 3'-end of the downstream verification primer to the 3'-end of the downstream homology arm);  $DSS_{max}$ —Maximum length of downstream spacer sequence;  $ODR_{DSS}$ —Optional design region of the downstream verification primer

以及有限的遗传操作类型的传统人工操作场景开发的，且靶序列往往局限于编码序列，忽略了对非编码序列（启动子、RBS等）的改造需求。这一方面限制了工具的应用推广，另一方面难以适配自动化高通量的Biofoundry的发展需求。

为了实现跨物种、任意基因组位点、任意操作类型的高通量基因组编辑设计，笔者所在团队

开发了面向微生物遗传操作的全流程自动化编辑序列设计在线工具AutoESD<sup>[14]</sup>。通过对经典同源重组技术的模块化解构和标准化处理，AutoESD实现了全流程、自动化的编辑序列设计，支持绝大多数经典的基于筛选标记的同源重组技术，理论上支持所有基因组序列已知的微生物的编辑序列设计，并可以在单批次任务中处理针对不同目

表4 面向基因组编辑全流程的编辑序列设计工具

Table 4 Editing sequence design tools for the whole workflow of genome editing

工具	编辑序列设计	遗传改造类型	是否高通量	可支持物种	可编辑位点	web应用
Yeastription <sup>[115]</sup>	引物设计 sgRNA设计	删除	否	<i>Saccharomyces cerevisiae</i> (33个菌株)	已被注释的基因	<a href="http://yeastription.tnw.tudelft.nl/#/">http://yeastription.tnw.tudelft.nl/#/</a>
GeneTargeter <sup>[21]</sup>	引物设计 sgRNA设计	删除、表达弱化	是	<i>Plasmodium falciparum</i>	已被注释的基因	<a href="http://genetargeter.mit.edu/">http://genetargeter.mit.edu/</a>
GEDpm-cg <sup>[116]</sup>	引物设计、同源臂设计	单点突变	是	<i>Corynebacterium glutamicum</i>	基因组任意区域	<a href="https://gedpm-cg.biodesign.ac.cn/">https://gedpm-cg.biodesign.ac.cn/</a>
AutoESD <sup>[14]</sup>	引物设计、同源臂设计	插入、删除、替换、单点突变	是	多种原核、真核微生物	基因组任意区域	<a href="https://autoesd.biodesign.ac.cn/">https://autoesd.biodesign.ac.cn/</a>

标序列（基因组上任意 CDS 或基因间区域）的多种类型的遗传操作（序列敲除、插入和替换）。用户在网站界面无需登录即可选择参考基因组和遗传操作技术，上传目标操作序列与用户自己的载体和筛选标记序列，即可自动实现编辑序列设计供用户下载，并可通过网站进行可视化分析。同时，AutoESD 还提供失败任务原因分析、同源序列脱靶风险评估等功能，用户可以根据这些结果，更改默认的参数，进行优化再设计。此外，AutoESD 的开发采用了基于云端的无服务器架构，确保了高可靠性、稳健性和可扩展性，能够在几分钟内并行处理包含上千个编辑序列设计目标的数百个设计任务，大大提高了当前编辑序列设计的通量，可以为世界范围内的合成生物铸造厂提供免费、灵活、用户友好的编辑序列设计自动化支持。

## 5 总结和展望

编辑序列设计是影响遗传操作效率和准确性的关键步骤，也是合成生物学 DBTL 工作循环中负责衔接上游基因型、生物元件“设计”模块与下游菌种遗传操作“构建”模块的重要环节。在可预见的合成生物学发展趋势中，以传统实验室为研究单元的劳动密集型的研究范式必然会被高度标准化、自动化且高通量的基于 Biofoundry 的合成生物铸造模式所颠覆。为此，已有多个可以覆盖特定遗传操作技术全流程的编辑序列设计工具<sup>[14, 21, 115-116]</sup>被开发出来，旨在提供标准化、自动化的编辑序列设计解决方案。但在工具的适用范

围、各种类型编辑序列设计的精准度、对接实际生物学家改造需求、与自动化设施硬件的融合等方面仍然有诸多局限，亟待进一步优化研究。

首先，全流程设计工具的适用范围主要受到可支持技术种类与变种数目的限制，例如 AutoESD 虽然通过整合多种经典的同源重组技术，从而理论上可以支持跨物种的设计，但由于尚未整合更加流行的 CRISPR/Cas 系统，很大程度限制了它的应用场景。因此，整合更多的基因组编辑技术（例如 CRISPR 系统介导的同源重组技术<sup>[70]</sup>、碱基编辑技术<sup>[114]</sup>、先导编辑技术<sup>[117]</sup>等）和质粒 DNA 组装技术（Golden Gate<sup>[63]</sup>等）是下一阶段编辑序列设计工具研究开发的一个重要方向。

其次，实验技术的升级必然会催生出编辑序列设计的新规则、新算法和新需求，编辑序列设计工具需要紧跟实验技术的发展方向并不断升级，否则就会因用户的流失而被淘汰。例如，对于引物设计的接头序列添加问题，传统设计工具的策略都是先设计引物，再将接头序列直接添加至引物 5' 端，这导致最终引物存在潜在的引物二聚体和发卡结构等风险。为此，Hiraga 等在开发蛋白表达质粒的设计工具 Mutation Maker<sup>[46]</sup> 时提供了两种新算法：①暴力算法，先计算所有突变位点周围的所有候选引物，然后使用各种约束系统的过滤候选引物得到满足约束的引物；②快速逼近算法，动态扩展候选诱变引物集合直到其满足设计标准。上文提到的针对 Golden Gate 组装的设计工具的改进<sup>[61-62]</sup>以及针对 sgRNA 设计工具 CHOPCHOP 的迭代升级<sup>[17, 102-103]</sup>也是很好的例子。

再次，在工程菌种构建的具体应用场景中，实验生物学用户的绝大多数改造需求是针对目标

基因的强化、弱化与失活。这些常可以通过不同的编辑手段如启动子、RBS替换,全基因删除或插入终止密码子等实现。目前大多数的编辑序列设计工具都需要用户首先确定具体如何改造,如用一个特定强启动子序列替换某一基因上游的启动子序列,然后将相应的序列作为编辑序列设计工具的输入进行设计。而对大多数实验生物学用户来说,选择强度合适的启动子和RBS以实现预期的改造目标也是一个很有挑战的计算设计问题,也需要借助相关的软件工具设计以达到相应改造目标。当一开始的改造手段选择不当时(例如设计得到的强启动子在特定生物特定基因中表达强度并不高),无论编辑序列设计如何精准都无法达到期望的菌种改造目标。在实际使用诸如AutoESD等设计工具时,仍然需要耗费一定精力进行改造策略和生物元件的设计与选取。因此,将上游设计工具(包括代谢合成途径与基因型设计工具<sup>[118-120]</sup>、生物元件设计工具<sup>[121-122]</sup>)和编辑序列设计工具进行整合是未来实现菌种改造生物设计全流程自动化的关键步骤。

最后,虽然面向质粒DNA组装的设计工具PlasmidMaker<sup>[13]</sup>已经可以实现与iBioFAB自动化铸造平台的信息交互与任务触发,但面向更为复杂的基因组编辑的设计工具仍然缺乏与自动化设施的对接功能,如对包含引物、DNA模板、试剂等信息的自动化移液工作站操作任务的执行和管理功能。因此,进一步开发集成更多的遗传操作技术、更精准的编辑序列设计算法的编辑序列设计工具,并将其与上下游模块进行整合,将有望推动工程菌种从设计到构建的全流程自动化铸造模式的早日实现。

## 参 考 文 献

- [1] COHEN S N, CHANG A C, BOYER H W, et al. Construction of biologically functional bacterial plasmids *in vitro*[J]. Proceedings of the National Academy of Sciences of the United States of America, 1973, 70(11): 3240-3244.
- [2] 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.
- [3] KLECKNER N, ROTH J, BOTSTEIN D. Genetic engineering *in vivo* using translocatable drug-resistance elements: new methods in bacterial genetics[J]. Journal of Molecular Biology, 1977, 116(1): 125-159.
- [4] 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.
- [5] ENGLER C, GRUETZNER R, KANDZIA R, et al. Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes[J]. PLoS One, 2009, 4(5): e5553.
- [6] SCHERER S, DAVIS R W. Replacement of chromosome segments with altered DNA sequences constructed *in vitro*[J]. Proceedings of the National Academy of Sciences of the United States of America, 1979, 76(10): 4951-4955.
- [7] MURPHY K C. Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*[J]. Journal of Bacteriology, 1998, 180(8): 2063-2071.
- [8] 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.
- [9] KOMOR A C, KIM Y B, PACKER M S, et al. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage[J]. Nature, 2016, 533(7603): 420-424.
- [10] 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(7785): 149-157.
- [11] STORCH M, HAINES M C, BALDWIN G S. DNA-BOT: a low-cost, automated DNA assembly platform for synthetic biology[J]. Synthetic Biology, 2020, 5(1): ysaa010.
- [12] SI T, CHAO R, MIN Y H, et al. Automated multiplex genome-scale engineering in yeast[J]. Nature Communications, 2017, 8: 15187.
- [13] ENGHAD B, XUE P, SINGH N, et al. PlasmidMaker is a versatile, automated, and high throughput end-to-end platform for plasmid construction[J]. Nature Communications, 2022, 13: 2697.
- [14] YANG Y, MAO Y F, WANG R Y, et al. AutoESD: a web tool for automatic editing sequence design for genetic manipulation of microorganisms[J]. Nucleic Acids Research, 2022, 50(W1): W75-W82.
- [15] UNTERGASSER A, CUTCUTACHE I, KORESSAAR T, et al. Primer3—new capabilities and interfaces[J]. Nucleic Acids Research, 2012, 40(15): e115.
- [16] HILLSON N J, ROSENGARTEN R D, KEASLING J D. J5 DNA assembly design automation software[J]. ACS Synthetic Biology, 2012, 1(1): 14-21.
- [17] MONTAGUE T G, CRUZ J M, GAGNON J A, et al. CHOP-CHOP: a CRISPR/Cas9 and TALEN web tool for genome editing[J]. Nucleic Acids Research, 2014, 42(W1): W401-W407.
- [18] 唐婷,付立豪,郭二鹏,等.自动化合成生物技术与工程化设施平台[J].科学通报,2021,66(3): 300-309.
- TANG T, FU L H, GUO E P, et al. Automation in synthetic bi-

- ology using biological foundries[J]. Chinese Science Bulletin, 2021, 66(3): 300-309.
- [19] ARNOLD C. Cloud labs: Where robots do the research[J]. Nature, 2022, 606(7914): 612-613.
- [20] WANG Y, LIU Y, LIU J, et al. MACBETH: Multiplex automated *Corynebacterium glutamicum* base editing method[J]. Metabolic Engineering, 2018, 47: 200-210.
- [21] CÁRDENAS P, ESHERICK L Y, CHAMBONNIER G, et al. GeneTargeter: automated *in silico* design for genome editing in the malaria parasite, *Plasmodium falciparum*[J]. The CRISPR Journal, 2022, 5(1): 155-164.
- [22] KÄMPKE T, KIENINGER M, MECKLENBURG M. Efficient primer design algorithms[J]. Bioinformatics, 2001, 17(3): 214-225.
- [23] LINZ U, DELLING U, RÜBSAMEN-WAIGMANN H. Systematic studies on parameters influencing the performance of the polymerase chain reaction[J]. Journal of Clinical Chemistry and Clinical Biochemistry (Zeitschrift Fur Klinische Chemie Und Klinische Biochemie), 1990, 28(1): 5-13.
- [24] ABD-ELSALAM K A. Bioinformatic tools and guideline for PCR primer design[J]. African Journal of Biotechnology, 2003, 2(5): 91-95.
- [25] 张新宇, 高燕宁. PCR引物设计及软件使用技巧[J]. 生物信息学, 2004, 2(4): 15-18, 46.  
ZHANG X Y, GAO Y N. To design PCR primers with oligo 6 and primer premier 5[J]. Bioinformatics, 2004, 2(4): 15-18, 46.
- [26] ROZEN S, SKALETSKY H. Primer3 on the WWW for general users and for biologist programmers[J]. Methods in Molecular Biology, 2000, 132: 365-386.
- [27] KORESSAAR T, REMM M. Enhancements and modifications of primer design program Primer3[J]. Bioinformatics, 2007, 23(10): 1289-1291.
- [28] YOU F M, HUO N X, GU Y Q, et al. BatchPrimer3: a high throughput web application for PCR and sequencing primer design[J]. BMC Bioinformatics, 2008, 9: 253.
- [29] ANDRESON R, REPO E, KAPLINSKI L, et al. GENOME-MASKER package for designing unique genomic PCR primers[J]. BMC Bioinformatics, 2006, 7: 172.
- [30] BOUTROS R, STOKES N, BEKAERT M, et al. UniPrime2: a web service providing easier Universal Primer design[J]. Nucleic Acids Research, 2009, 37(suppl\_2): W209-W213.
- [31] SHEN Z Y, QU W B, WANG W, et al. MPprimer: a program for reliable multiplex PCR primer design[J]. BMC Bioinformatics, 2010, 11: 143.
- [32] ZIESEL A C, CHRENEK M A, WONG P W. MultiPriDe: automated batch development of quantitative real-time PCR primers[J]. Nucleic Acids Research, 2008, 36(9): 3095-3100.
- [33] VIJAYA SATYA R, KUMAR K, ZAVALJEVSKI N, et al. A high-throughput pipeline for the design of real-time PCR signatures[J]. BMC Bioinformatics, 2010, 11: 340.
- [34] ARVIDSSON S, KWASNIEWSKI M, RIAÑO-PACHÓN D M, et al. QuantPrime—a flexible tool for reliable high-throughput primer design for quantitative PCR[J]. BMC Bioinformatics, 2008, 9: 465.
- [35] YOU F M, HUO N X, GU Y Q, et al. ConservedPrimers 2.0: a high-throughput pipeline for comparative genome referenced intron-flanking PCR primer design and its application in wheat SNP discovery[J]. BMC Bioinformatics, 2009, 10: 331.
- [36] YOU F M, WANJUGI H, HUO N X, et al. RJPrimers: unique transposable element insertion junction discovery and PCR primer design for marker development[J]. Nucleic Acids Research, 2010, 38(suppl\_2): W313-W320.
- [37] SRIVASTAVA G P, HANUMAPPA M, KUSHWAHA G, et al. Homolog-specific PCR primer design for profiling splice variants[J]. Nucleic Acids Research, 2011, 39(10): e69.
- [38] TSAI M F, LIN Y J, CHENG Y C, et al. PrimerZ: streamlined primer design for promoters, exons and human SNPs[J]. Nucleic Acids Research, 2007, 35(suppl\_2): W63-W65.
- [39] PIRIYAPONGSA J, NGAMPHIW C, ASSAWAMAKIN A, et al. REXPrimer: an integrated primer designing tool increases PCR effectiveness by avoiding 3' SNP-in-primer and mis-priming from structural variation[J]. BMC Genomics, 2009, 10(Suppl 3): S4.
- [40] YOON H, LEITNER T. PrimerDesign-M: a multiple-alignment based multiple-primer design tool for walking across variable genomes[J]. Bioinformatics, 2015, 31(9): 1472-1474.
- [41] YE J, COULOURIS G, ZARETSKAYA I, et al. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction[J]. BMC Bioinformatics, 2012, 13: 134.
- [42] KE X Y, COLLINS A, YE S. PCR designer for restriction analysis of various types of sequence mutation[J]. Bioinformatics, 2002, 18(12): 1688-1689.
- [43] TOKHEIM C, PARK J W, XING Y. PrimerSeq: design and visualization of RT-PCR primers for alternative splicing using RNA-seq data[J]. Genomics, Proteomics & Bioinformatics, 2014, 12(2): 105-109.
- [44] PANDEY R V, PULVERER W, KALLMEYER R, et al. MSP-HTPrimer: a high-throughput primer design tool to improve assay design for DNA methylation analysis in epigenetics[J]. Clinical Epigenetics, 2016, 8:101.
- [45] WINGO T S, KOTLAR A, CUTLER D J. MPD: multiplex primer design for next-generation targeted sequencing[J]. BMC Bioinformatics, 2017, 18(1):14.
- [46] HIRAGA K, MEJZLIK P, MARCISIN M, et al. Mutation maker, an open source oligo design platform for protein engineering[J]. ACS Synthetic Biology, 2021, 10(2): 357-370.
- [47] MARSHALL O J. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR[J]. Bioinformatics, 2004, 20(15): 2471-2472.
- [48] SANTALUCIA J. Physical principles and visual-OMP software for optimal PCR design[J]. Methods in Molecular Biology, 2007, 402: 3-34.

- [49] WANG K, LI H W, XU Y, et al. MFEprimer-3.0: Quality control for PCR primers[J]. *Nucleic Acids Research*, 2019, 47(W1): W610-W613.
- [50] KECHIN A, BOROBOVA V, BOYARSKIKH U, et al. NGS-PrimerPlex: High-throughput primer design for multiplex polymerase chain reactions[J]. *PLoS Computational Biology*, 2020, 16(12): e1008468.
- [51] RAIBAUD O, MOCK M, SCHWARTZ M. A technique for integrating any DNA fragment into the chromosome of *Escherichia coli*[J]. *Gene*, 1984, 29(1/2): 231-241.
- [52] KNIGHT T. Idempotent vector design for standard assembly of biobricks[R]. MIT Artificial Intelligence Laboratory, MIT Synthetic Biology Working Group, 2003.
- [53] ANDERSON J C, DUEBER J E, LEGUIA M, et al. BglBricks: a flexible standard for biological part assembly[J]. *Frontiers in Medicine*, 2010, 4(1): 1.
- [54] 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.
- [55] HORTON R M, HUNT H D, HO S N, et al. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension[J]. *Gene*, 1989, 77(1): 61-68.
- [56] QUAN J Y, TIAN J D. Circular polymerase extension cloning of complex gene libraries and pathways[J]. *PLoS One*, 2009, 4(7): e6441.
- [57] LI M Z, ELLEDGE S J. Harnessing homologous recombination *in vitro* to generate recombinant DNA *via* SLIC[J]. *Nature Methods*, 2007, 4(3): 251-256.
- [58] DE KOK S, STANTON L H, SLABY T, et al. Rapid and reliable DNA assembly *via* ligase cycling reaction[J]. *ACS Synthetic Biology*, 2014, 3(2): 97-106.
- [59] APPLETON E, TAO J, HADDOCK T, et al. Interactive assembly algorithms for molecular cloning[J]. *Nature Methods*, 2014, 11(6): 657-662.
- [60] SHI Z, VICKERS C E. Molecular Cloning Designer Simulator (MCDS): all-in-one molecular cloning and genetic engineering design, simulation and management software for complex synthetic biology and metabolic engineering projects[J]. *Metabolic Engineering Communications*, 2016, 3: 173-186.
- [61] PRYOR J M, POTAPOV V, KUCERA R B, et al. Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design[J]. *PLoS One*, 2020, 15(9): e0238592.
- [62] HAMEDIRAD M, WEISBERG S, CHAO R, et al. Highly efficient single-pot scarless golden gate assembly[J]. *ACS Synthetic Biology*, 2019, 8(5): 1047-1054.
- [63] POTAPOV V, ONG J L, KUCERA R B, et al. Comprehensive profiling of four base overhang ligation fidelity by T4 DNA ligase and application to DNA assembly[J]. *ACS Synthetic Biology*, 2018, 7(11): 2665-2674.
- [64] CASINI A, CHANG F Y, ELUERE R, et al. A pressure test to make 10 molecules in 90 days: external evaluation of methods to engineer biology[J]. *Journal of the American Chemical Society*, 2018, 140(12): 4302-4316.
- [65] 李洋, 申晓林, 孙新晓, 等. CRISPR 基因编辑技术在微生物合成生物学领域的研究进展[J]. *合成生物学*, 2021, 2(1): 106-120.
- LI Y, SHEN X L, SUN X X, et al. Advances of CRISPR gene editing in microbial synthetic biology[J]. *Synthetic Biology Journal*, 2021, 2(1): 106-120.
- [66] DELTCHEVA E, CHYLINSKI K, SHARMA C M, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III[J]. *Nature*, 2011, 471(7340): 602-607.
- [67] GRAHAM D B, ROOT D E. Resources for the design of CRISPR gene editing experiments[J]. *Genome Biology*, 2015, 16: 260.
- [68] CHUAI G H, WANG Q-L, LIU Q. *In silico* meets *in vivo*: towards computational CRISPR-based sgRNA design[J]. *Trends in Biotechnology*, 2017, 35(1): 12-21.
- [69] 王远立, 啜国晖, 闫继芳, 等. 计算机辅助 CRISPR 向导 RNA 设计[J]. *生物工程学报*, 2017, 33(10): 1744-1756.
- WANG Y L, CHUAI G H, YAN J F, et al. *In silico* CRISPR-based sgRNA design[J]. *Chinese Journal of Biotechnology*, 2017, 33(10): 1744-1756.
- [70] DOUDNA J A, CHARPENTIER E. The new frontier of genome engineering with CRISPR-Cas9[J]. *Science*, 2014, 346(6213): 1258096.
- [71] DOENCH J G, FUSI N, SULLENDER M, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9[J]. *Nature Biotechnology*, 2016, 34(2): 184-191.
- [72] XU H, XIAO T F, CHEN C H, et al. Sequence determinants of improved CRISPR sgRNA design[J]. *Genome Research*, 2015, 25(8): 1147-1157.
- [73] WANG T, WEI J J, SABATINI D M, et al. Genetic screens in human cells using the CRISPR-Cas9 system[J]. *Science*, 2014, 343(6166): 80-84.
- [74] WU X B, SCOTT D A, KRIZ A J, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells[J]. *Nature Biotechnology*, 2014, 32(7): 670-676.
- [75] DOENCH J G, HARTENIAN E, GRAHAM D B, et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation[J]. *Nature Biotechnology*, 2014, 32(12): 1262-1267.
- [76] CHARI R, MALI P, MOOSBURNER M, et al. Unraveling CRISPR-Cas9 genome engineering parameters *via* a library-on-library approach[J]. *Nature Methods*, 2015, 12(9): 823-826.
- [77] HSU P D, SCOTT D A, WEINSTEIN J A, et al. DNA targeting specificity of RNA-guided Cas9 nucleases[J]. *Nature Biotechnology*, 2013, 31(9): 827-832.
- [78] KIM D, KIM S, KIM S, et al. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-

- seq[J]. *Genome Research*, 2016, 26(3): 406-415.
- [79] SINGH R, KUSCU C, QUINLAN A, et al. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction[J]. *Nucleic Acids Research*, 2015, 43(18): e118.
- [80] KIM D, KIM J S. DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA[J]. *Genome Research*, 2018, 28(12): 1894-1900.
- [81] MORGENS D W, WAINBERG M, BOYLE E A, et al. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens[J]. *Nature Communications*, 2017, 8: 15178.
- [82] AGUIRRE A J, MEYERS R M, WEIR B A, et al. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting[J]. *Cancer Discovery*, 2016, 6(8): 914-929.
- [83] JIANG W Y, BIKARD D, COX D, et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems[J]. *Nature Biotechnology*, 2013, 31(3): 233-239.
- [84] ZHANG Y L, GE X L, YANG F Y, et al. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells[J]. *Scientific Reports*, 2014, 4: 5405.
- [85] PATTANAYAK V, LIN S, GUILINGER J P, et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity[J]. *Nature Biotechnology*, 2013, 31(9): 839-843.
- [86] CRADICK T J, FINE E J, ANTICO C J, et al. CRISPR/Cas9 systems targeting  $\beta$ -globin and CCR5 genes have substantial off-target activity[J]. *Nucleic Acids Research*, 2013, 41(20): 9584-9592.
- [87] LANGMEAD B, TRAPNELL C, POP M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome[J]. *Genome Biology*, 2009, 10(3): R25.
- [88] LI H, DURBIN R. Fast and accurate short read alignment with Burrows-Wheeler transform[J]. *Bioinformatics*, 2009, 25(14): 1754-1760.
- [89] JIANG H, WONG W H. SeqMap: mapping massive amount of oligonucleotides to the genome[J]. *Bioinformatics*, 2008, 24(20): 2395-2396.
- [90] BAE S S, PARK J, KIM J S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases[J]. *Bioinformatics*, 2014, 30(10): 1473-1475.
- [91] XIAO A, CHENG Z C, KONG L, et al. CasOT: a genome-wide Cas9/gRNA off-target searching tool[J]. *Bioinformatics*, 2014, 30(8): 1180-1182.
- [92] CANCELLIERI S, CANVER M C, BOMBIERI N, et al. CRISPRitz: rapid, high-throughput and variant-aware *in silico* off-target site identification for CRISPR genome editing[J]. *Bioinformatics*, 2020, 36(7): 2001-2008.
- [93] STEMMER M, THUMBERGER T, DEL SOL KEYER M, et al. CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool[J]. *PLoS One*, 2017, 12(4): e0124633.
- [94] HAEUSSLER M, SCHÖNIG K, ECKERT H, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR[J]. *Genome Biology*, 2016, 17(1): 148.
- [95] XIE S S, SHEN B, ZHANG C B, et al. sgRNAs9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites[J]. *PLoS One*, 2014, 9(6): e100448.
- [96] JACQUIN A L S, ODOM D T, LUKK M. Crisflash: open-source software to generate CRISPR guide RNAs against genomes annotated with individual variation[J]. *Bioinformatics*, 2019, 35(17): 3146-3147.
- [97] MCKENNA A, SHENDURE J. FlashFry: a fast and flexible tool for large-scale CRISPR target design[J]. *BMC Biology*, 2018, 16(1): 74.
- [98] HEIGWER F, KERR G, BOUTROS M. E-CRISP: fast CRISPR target site identification[J]. *Nature Methods*, 2014, 11(2): 122-123.
- [99] LIU H L, WEI Z, DOMINGUEZ A, et al. CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation[J]. *Bioinformatics*, 2015, 31(22): 3676-3678.
- [100] PENG D, TARLETON R. EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens[J]. *Microbial Genomics*, 2015, 1(4): e000033.
- [101] CONCORDET J P, HAEUSSLER M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens[J]. *Nucleic Acids Research*, 2018, 46(W1): W242-W245.
- [102] LABUN K, MONTAGUE T G, GAGNON J A, et al. CHOP-CHOP v2: a web tool for the next generation of CRISPR genome engineering[J]. *Nucleic Acids Research*, 2016, 44(W1): W272-W276.
- [103] LABUN K, MONTAGUE T G, KRAUSE M, et al. CHOP-CHOP v3: expanding the CRISPR web toolbox beyond genome editing[J]. *Nucleic Acids Research*, 2019, 47(W1): W171-W174.
- [104] PULIDO-QUETGLAS C, APARICIO-PRAT E, ARNAN C, et al. Scalable design of paired CRISPR guide RNAs for genomic deletion[J]. *PLoS Computational Biology*, 2017, 13(3): e1005341.
- [105] LEI Y, LU L, LIU H Y, et al. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants[J]. *Molecular Plant*, 2014, 7(9): 1494-1496.
- [106] MORENO-MATEOS M A, VEJNAR C E, BEAUDOIN J D, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting *in vivo*[J]. *Nature Methods*, 2015, 12(10): 982-988.
- [107] WONG N, LIU W J, WANG X W. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system[J]. *Genome Biology*, 2015, 16: 218.
- [108] ABADI S, YAN W X, AMAR D, et al. A machine learning ap-

- proach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action[J]. *PLoS Computational Biology*, 2017, 13(10): e1005807.
- [109] LISTGARTEN J, WEINSTEIN M, KLEINSTIVER B P, et al. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs[J]. *Nature Biomedical Engineering*, 2018, 2(1): 38-47.
- [110] KONSTANTAKOS V, NENTIDIS A, KRITHARA A, et al. CRISPR-Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning[J]. *Nucleic Acids Research*, 2022, 50(7): 3616-3637.
- [111] WANG D Q, ZHANG C D, WANG B, et al. Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning[J]. *Nature Communications*, 2019, 10: 4284.
- [112] HWANG G H, PARK J, LIM K, et al. Web-based design and analysis tools for CRISPR base editing[J]. *BMC Bioinformatics*, 2018, 19(1): 542.
- [113] DANDAGE R, DESPRÉS P C, YACHIE N, et al. Beditor: a computational workflow for designing libraries of guide RNAs for CRISPR-mediated base editing[J]. *Genetics*, 2019, 212(2): 377-385.
- [114] WANG Y, LIU Y, LI J W, et al. Expanding targeting scope, editing window, and base transition capability of base editing in *Corynebacterium glutamicum*[J]. *Biotechnology and Bioengineering*, 2019, 116(11): 3016-3029.
- [115] MANS R, VAN ROSSUM H M, WIJSMAN M, et al. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*[J]. *FEMS Yeast Research*, 2015, 15(2): fov004.
- [116] YANG Y, MAO Y F, LIU Y, et al. GEDpm-cg: genome editing automated design platform for point mutation construction in *Corynebacterium glutamicum*[J]. *Frontiers in Bioengineering and Biotechnology*, 2021, 9: 768289.
- [117] TONG Y J, JØRGENSEN T S, WHITFORD C M, et al. A versatile genetic engineering toolkit for *E. coli* based on CRISPR-prime editing[J]. *Nature Communications*, 2021, 12: 5206.
- [118] YANG X, YUAN Q Q, LUO H, et al. Systematic design and *in vitro* validation of novel one-carbon assimilation pathways[J]. *Metabolic Engineering*, 2019, 56: 142-153.
- [119] LIAO X P, MA H W, TANG Y J. Artificial intelligence: a solution to involution of design-build-test-learn cycle[J]. *Current Opinion in Biotechnology*, 2022, 75: 102712.
- [120] SABZEVARI M, SZEDMAK S, PENTTILÄ M, et al. Strain design optimization using reinforcement learning[J]. *PLoS Computational Biology*, 2022, 18(6): e1010177.
- [121] WANG Y, WANG H C, WEI L, et al. Synthetic promoter design in *Escherichia coli* based on a deep generative network[J]. *Nucleic Acids Research*, 2020, 48(12): 6403-6412.
- [122] SALIS H M, MIRSKY E A, VOIGT C A. Automated design of synthetic ribosome binding sites to control protein expression[J]. *Nature Biotechnology*, 2009, 27(10): 946-950.



**通讯作者:** 马红武(1970—),男,博士,研究员,博士生导师。研究方向为计算生物学、系统生物学等。  
E-mail: ma\_hw@tib.cas.cn



**第一作者:** 杨毅(1986—),男,博士研究生。研究方向为生物信息学、合成生物学、代谢工程。  
E-mail: yangyi@tib.cas.cn