

一种基于 T7 核酸内切酶 I 纠错的病毒基因组高效组装方法

张绪伟¹, 文斌¹, 王飞², 王学军², 刘丽艳², 王淑美^{1*}, 王升启^{2*}

1 广东药科大学 药学院, 广东 广州 510006

2 军事医学研究院 生物信息中心, 北京 100850

张绪伟, 文斌, 王飞, 王学军, 刘丽艳, 王淑美, 王升启. 一种基于 T7 核酸内切酶 I 纠错的病毒基因组高效组装方法[J]. 生物工程学报, 2025, 41(1): 385-396.

ZHANG Xuwei, WEN Bin, WANG Fei, WANG Xuejun, LIU Liyan, WANG Shumei, WANG Shengqi. An efficient assembly method for a viral genome based on T7 endonuclease I-mediated error correction[J]. Chinese Journal of Biotechnology, 2025, 41(1): 385-396.

摘要: 基因合成技术是支撑合成生物学发展的使能技术。现有的基因从头合成技术存在操作步骤多、效率低、错误率高且长度有限等问题,难以支撑合成生物学日益拓展的庞大需求。其中 DNA 片段的组装和纠错是基因合成的关键环节。本研究首先通过平衡序列设计软件能力、PCR 扩增能力以及组装酶组装能力等参数,将约 10 kb 病毒基因组序列进行合理拆分后设计寡核苷酸序列;然后使用高保真聚合酶进行两步 PCR 反应,完成 3.0 kb DNA 片段的从头合成,并使用 T7 核酸内切酶 I 分别对不同阶段的 PCR 产物进行纠错反应;最后将从头合成并经过纠错的 3.0 kb DNA 片段进行约 10 kb 片段的组装并测序验证。结果表明,本方法可成功获得约 10 kb DNA 片段,有效降低组装过程中的大片段突变概率,组装错误率最低可至 0.36 errors/kb。综上,本研究开发了一种高效从头合成约 10 kb 病毒基因组方法,辅以 T7 核酸内切酶 I 纠错,可 1 d 内获得约 10 kb 病毒基因组片段,5 d 内获得病毒基因组的正确质粒。该方法优化了基因从头合成流程,降低了错误率,简化了合成与组装步骤,并进一步降低了病毒基因组组装成本。

关键词: 基因组组装; 从头合成; 纠错; T7 核酸内切酶 I

资助项目: 国家重点研发计划(2018YFA0902300)

This work was supported by the National Key Research and Development Program of China (2018YFA0902300).

*Corresponding authors. E-mail: WANG Shumei, gdupuwsm@126.com; WANG Shengqi, sqwang@bmi.ac.cn

Received: 2024-04-07; Accepted: 2024-06-14; Published online: 2024-06-17

An efficient assembly method for a viral genome based on T7 endonuclease I-mediated error correction

ZHANG Xuwei¹, WEN Bin¹, WANG Fei², WANG Xuejun², LIU Liyan², WANG Shumei^{1*}, WANG Shengqi^{2*}

1 School of Pharmacy, Guangdong Pharmaceutical University, Guangzhou 510006, Guangdong, China

2 Bioinformatics Center of Academy of Military Medical Sciences, Beijing 100850, China

Abstract: Gene synthesis is an enabling technology that supports the development of synthetic biology. The existing approaches for *de novo* gene synthesis generally have tedious operation, low efficiency, high error rates, and limited product lengths, being difficult to support the huge demand of synthetic biology. The assembly and error correction are the keys in gene synthesis. This study first designed the oligonucleotide sequences by reasonably splitting the virus genome of approximately 10 kb by balancing the parameters of sequence design software ability, PCR amplification ability, and assembly enzyme assembly ability. Then, two-step PCR was performed with high-fidelity polymerase to complete the *de novo* synthesis of 3.0 kb DNA fragments, and error correction reactions were performed with T7 endonuclease I for the products from different stages of PCR. Finally, the virus genome was assembled by 3.0 kb DNA fragments from *de novo* synthesis and error correction and then sequenced. The experimental results showed that the proposed method successfully produced the DNA fragment of about 10 kb and reduced the probability of large fragment mutations during the assembly process, with the lowest error rate reaching 0.36 errors/kb. In summary, this study developed an efficient *de novo* method for synthesizing a viral genome of about 10 kb with T7 endonuclease I-mediated error correction. This method enabled the synthesis of a 10 kb viral genome in one day and the correct plasmid of the viral genome in five days. This study optimized the *de novo* gene synthesis process, reduced the error rate, simplified the synthesis and assembly steps, and reduced the cost of viral genome assembly.

Keywords: gene assembly; *de novo* synthesis; error correction; T7 endonuclease I

基因的从头合成是合成生物学的基石，随着人类对生物基因组的认知逐渐深入，使得任何需要的DNA都可能被合成出来，这对于探索难以获取天然序列的基因信息至关重要。基因的从头合成不仅在检测试剂研发、药物与疫苗研制以及信息存储等应用研究发挥重要作用，也是基因功能甚至生命功能探索不可或缺的技术手段。

近30年来合成基因组学蓬勃发展，2002年

脊髓灰质炎病毒基因组被成功合成，标志着基因组合成时代的正式开始^[1]；2016年Hutchison等^[2]设计并制造了最简单的人工合成生命体，证明我们已经跨入了创造生命的新阶段；2017年*Science*杂志的7篇文章报道了5条酵母染色体的从头设计与合成^[3-10]，2023年研究人员完成了酵母全部16条染色体合成，最新成果于11月份发表在*Cell Genomics*^[11]，这是合成生物学的一个里程碑。DNA从头合成技术主要涉及“设

计-寡核苷酸(oligos)合成-DNA 短片段合成-DNA 片段纠错-DNA 片段拼接组装”这 5 个方面。

基因组设计需根据序列特性并辅以热力学原理、启发式算法等使基因合成的初始元件在一定范围内具有相对统一的 T_m 值、较少的二级结构和错配等,最终将其拆分为几十个甚至上百条 oligos^[12]。目前 oligos 合成主要使用固相亚磷酸胺化学法通过自动化合成仪进行^[13-14],通过不断循环“脱保护-偶联-加帽-氧化”的步骤延长 oligos 的长度,得到粗品后根据需要选择脱盐、寡核苷酸纯化柱(oligonucleotide purification cartridge, OPC)、聚丙烯酰胺凝胶电泳(polyacrylamide gel electrophoresis, PAGE)及高效液相色谱(high performance liquid chromatography, HPLC)等纯化方式纯化 oligos,其产量和纯度也随着 oligos 长度的增加而不断下降。

长片段 DNA 分子组装较难且在体外操作时易断裂,因此通常先进行短片段 DNA 的组装。短片段 DNA 组装的主要方式为胞外组装,常用组装方法包括连接酶链式组装反应(ligase chain reaction, LCR)和聚合酶链式组装反应(polymerase chain assembly, PCA)^[15-16]。其中 LCR 法所使用的 oligos 之间高度重叠无间隙,并需要对其 5'端进行磷酸化处理,通过耐热连接酶将 oligos 连接为双链 DNA 短片段。PCA 法使用具有 15-25 nt 互补末端的 oligos,通过 PCR 扩增的原理获得较小的 DNA 片段,然后以此为模板进一步 PCR 富集目的片段^[17]。该方法无需额外对 oligos 进行磷酸化处理且所需数量减少,降低了操作难度和成本,但由于没有杂交处理过滤错误使其错误率略高。现有的 DNA 拼接技术主要包括依赖 IIS 型限制性内切酶的 Golden gate 拼接技术、依赖同源序列的 Gibson 拼接技术和不依赖序列和连接的克隆技术(sequence and ligation independent cloning,

SLIC)等^[18-19]。

目前,基因合成的错误率为 1-10 errors/kb^[20]。合成错误可能由起始 oligos 纯度、PCR 扩增或者连接过程中的引物错配、片段性的插入或缺失、单碱基的突变等因素造成。为了实现基因的高效率合成,纠错就显得尤为重要。现有的纠错酶主要包括具有错配切割酶功能的 T4 核酸内切酶 VII、T7 核酸内切酶 I、核酸内切酶 V 等,可以识别碱基错配、异源双链 loop 环和分支 DNA,并对其进行切割;商品化纠错酶 CorrectASE 和 Surveyor,也可识别错配双链并进行切割;来源于生物体内的一对 MutS 和 MutL 蛋白,可以通过共同作用对错配碱基进行识别和修复^[21-22]。

常规的基因合成与组装方法往往采取逐级拼装的方式:首先合成 1.0 kb 以内的短片段,对合成的质粒不断重复“酶切-拼接-质粒构建”的操作,从而实现长片段的合成与组装,该方法费时费力且多为重复性劳动。因此,本研究开发了一种高效的从头合成与组装技术,结合 T7 核酸内切酶 I 纠错,以期缩短 DNA 长片段的组装时间并降低长片段组装的错误率。

1 材料与方法

1.1 试剂和仪器

Q5 热启动高保真 Mix、T7 核酸内切酶 I、NEB Builder 高效 DNA 组装 Mix (纽英伦生物技术有限公司);DNA 产物磁珠纯化试剂盒(上海碧云天生物技术有限公司);质粒提取试剂盒(北京全式金生物技术股份有限公司);氨苄青霉素(北京索莱宝科技有限公司);BigDye Terminator v3.1 循环测序试剂盒(赛默飞世尔科技有限公司);测序产物磁珠纯化试剂盒(北京擎科生物科技股份有限公司)。

PCR 仪(伯乐生命医学产品有限公司);凝

胶成像系统(上海天能生命科学有限公司); 标准振荡培养箱(Rigns Biotechnology Co. Ltd.); 细菌恒温培养箱(上海一恒科学仪器有限公司); 离心机(艾本德国际贸易有限公司); 涡旋仪(Scilogex llc 公司); DNA 合成仪(江苏领坤生物科技有限公司); DNA 分析仪(赛默飞世尔科技有限公司)。

1.2 菌株、载体及培养基

使用 pACYC 载体(本实验室保存)构建病毒基因组质粒; 使用 NEB10-beta 感受态细胞(北京博迈德基因技术有限公司)对连接产物进行转化; 使用 LB 培养基对菌株进行常规培养, 其中 LB 液体培养基由 10 g/L 胰蛋白胨、5 g/L 酵母浸出粉和 10 g/L NaCl 配制而成, LB 固体培养基为 LB 液体培养基加入 15 g/L 琼脂粉, LB 抗性培养基要额外添加浓度为 100 $\mu\text{g}/\text{mL}$ 的氨苄抗生素。

1.3 病毒基因组、引物设计及合成

以增加 CMV 启动子、重组序列等其他辅助序列后的登革热病毒 2 型(dengue virus type 2, DENV2)基因组(10 963 bp)和人肠道病毒 71 型(human enterovirus type 71, EV71)基因组(9 251 bp)为组装模板, 分别在基因组的 5' 端和 3' 端添加 80 bp pACYC 载体同源序列, 基于现有的 oligos 设计能力、PCR 聚合酶扩增能力和组装酶组装能力间的平衡, 将 DENV2 基因组序列大致等分为 4 段约 3.0 kb 的 DNA 片段, EV71 基因组序列大致等分为 3 段约 3.0 kb 的 DNA 片段, 每个片段之间含有 80 bp 的同源序列。

使用本地化 DNAWorks 软件设计组装引物, 每条 oligo 设置为 60 nt, 退火温度设置为 65 $^{\circ}\text{C}$, 运行程序设计 oligos, 对 T_m 值超过 3 $^{\circ}\text{C}$ 的引物再次调试 oligos 长度和退火温度直至 T_m 值 ≤ 3 $^{\circ}\text{C}$ 。引物设计完成后由本实验室合成。

1.4 DNA 片段合成、纠错与组装

1.4.1 3.0 kb DNA 片段合成

将 3.0 kb 拆分的 oligos 溶解后分别等体系混合为 oligos mix, 按表 1 所示配置 PCR 反应体系, 并按说明书进行 PCR 反应。

表 1 第一步 PCR 体系

Table 1 The first step of PCR system

Component	Volume (μL)
Q5 High-Fidelity 2 \times Master Mix	10.0
Oligos Mix (10.0 $\mu\text{mol}/\text{L}$)	4.0
Nuclease-free water	6.0
Total volume	20.0

第一轮 PCR 结束后, 从产物中取出 1.0 μL 作为第二轮 PCR 反应的模板, 使用每个 3.0 kb 片段 oligos 的第一条与最后一条作为第二轮扩增的引物。第二轮反应体系和程序均参照说明书进行。

两轮 PCR 反应结束后, 使用 Qubit 4.0 检测 PCR 产物浓度和纯度, 使用琼脂糖凝胶电泳检测目的条带正确性。

1.4.2 3.0 kb DNA 片段纠错

为验证和提高 T7 核酸内切酶 I 的纠错效果, 分别于第一轮 PCR 结束后、第二轮 PCR 结束后及两轮 PCR 结束后均进行纠错反应。纠错反应流程参照 T7 核酸内切酶 I 说明书并进行些许改动。按照表 2 所示配置纠错反应体系。

体系配置完成后, 95 $^{\circ}\text{C}$ 孵育 5 min 进行变性。随后从 95 $^{\circ}\text{C}$ 以 2 $^{\circ}\text{C}/\text{s}$ 的速度进行梯度退火至 85 $^{\circ}\text{C}$, 再以 0.1 $^{\circ}\text{C}/\text{s}$ 的速度进行梯度退火至 25 $^{\circ}\text{C}$ 。退火完成后加入 T7 核酸内切酶 I 1.0 μL , 37 $^{\circ}\text{C}$ 孵育 15 min。

表 2 纠错体系

Table 2 Error correction system

Component	Volume/Quantity
DNA	200 ng
10 \times NEBuffer 2	2 μL
Nuclease-free water	To 19 μL

第一轮 PCR 反应结束后的纠错反应于第一轮 PCR 反应结束后进行, 取 200 ng PCR 产物进行纠错, 纠错结束后取 80 ng 纠错产物进行第二轮 PCR 反应, 随后再进行病毒全基因组组装; 第二轮 PCR 反应结束后的纠错反应于第二轮 PCR 反应结束后进行, 取 400 ng PCR 产物进行纠错, 纠错结束后使用产物进行组装; 两轮 PCR 均纠错的反应于第一轮 PCR 反应结束后, 取 200 ng PCR 产物进行纠错, 纠错结束后取 80 ng 产物进行第二轮 PCR 反应, 反应结束后取 400 ng PCR 产物进行纠错, 纠错结束后使用产物进行组装。

1.4.3 病毒全基因组组装

使用 NEB Builder 无缝克隆组装试剂盒构建病毒基因组质粒, 每个 3.0 kb 的 DNA 片段与 pACYC 载体按照等摩尔数进行配比, 体系配置完成后 50 °C 孵育 1 h。使用 NEB10-beta 感受态细胞转化连接产物, 菌液复壮后涂布于含有氨苄抗性的 LB 固体培养基中过夜培养。

挑取单菌落进行菌液 PCR 鉴定, 使用跨越 3.0 kb DNA 片段的引物(表 3)进行 PCR 扩增, 均能扩增出目的条带的菌落为阳性菌落。将阳性菌落扩繁后进行质粒提取。质粒提取使用质粒提取试剂盒并参照说明书进行。质粒提取完

成后, 使用测序引物(表 4)对病毒基因组进行一代 Sanger 测序, 测序结果使用 SnapGene 软件进行比对。

1.4.4 病毒全基因组序列拼接

根据质粒测序结果挑选 2-3 个质粒并从中选择正确序列的部分, 从全基因组合成的引物中挑选合适引物(相邻片段的 oligo 之间含有 overlap 序列)扩增正确序列, PCR 反应体系和程序均参照说明书进行。将 PCR 产物使用磁珠纯化试剂盒进行纯化后与载体进行组装, 随后进行转化、质粒提取(挑选 3 个菌落)并测序, 以期获得全基因组序列正确的质粒。

2 结果与分析

本研究建立了一种病毒基因组从头合成及组装技术, 通过合理设计 oligos、两轮 PCR 扩增、T7 核酸内切酶 I 纠错及 Gibson 组装等系列操作可高效组装约 10 kb 病毒基因组, 其具体组装流程如图 1 所示。

2.1 3.0 kb DNA 片段合成及纠错

使用 DNAWorks 软件针对 DENV2 基因组设计了 316 条 oligos, EV71 基因组设计了 243 条 oligos。通过第一轮 PCR 反应合成 3.0 kb 片段, 通过第二轮 PCR 反应富集目的片段。使用 T7

表 3 阳性菌落鉴定引物

Table 3 Primers for positive colony identification

Virus type	Primer name	Sequence (5'→3')
DENV2	81-P1F	GTGAAGTTAACTATTATGACAGGAGACATCAAAGGAATCATGCAGGCAGGAAAACGATCT
	81-P1R	TGTCATTGAGTGCACCTTTCTATCCAATAACCCATATCGGCATGGACGGCTCTGTTGTCTT
	81-P2F	GCATATGTGAGTGCTATAGCCCAGACTGAAAAAGCATTGAAGACAACCCAGAGATCGAA
	81-P2R	GTGGGTCACCTGGCATAGGTCCTGCCAGAATCACCCGCTCTTACCATCTGTTAGTATGAC
	81-P3F	TCTTCATCCCGCCAGAAAAGTGTGACACATTATTGTGTGACATAGGGGAGTCATCACCAA
	81-P3R	TGGTCTTGGTCATAGTGCCATGATGTTTCATGCTCTTGTCTTTATTTTTCTATTCTTTTC
EV71	71-P1F	AATGGGAGTTTGTGTTTGGCACC
	71-P1R	TGACCCAGTAAGGTGGATTGCCATGGTCCATTTTCGCCCAGGATCGGCTCTAAACACCGCA
	71-P2F	TCCACCGTAGACTTCATTCCACCAATGGCTTCTCTCGAGGAGAAGGGAGTTTCCTTACC
	71-P2R	TCTATCTCTAAGTTGGATAGCTTTCCACACACTAATGGGCTGCAACGTTTGAATTGTC

表 4 病毒基因组测序引物

Table 4 Sequencing primers for virus genomes

Virus type	Primer name	Sequence (5'→3')
DENV2	CMV-F	CGCAAATGGGCGGTAGGCGTG
	Seq-81-1	ACGTACAAGTGTCCCCTTCT
	Seq-81-2	ATGGGGAAATGGATGTGGAC
	Seq-81-3	ATGGAACAATAGTTATCAGA
	Seq-81-4	TATCCAGAAAGCCCATGAAG
	Seq-81-5	CATCTAGGTAAGCTTGAGAT
	Seq-81-6	TGGAAAAGTATCAATTGGCA
	Seq-81-7	GGCTGAACTGGAAGATGGAG
	Seq-81-8	TTCCAATAAGATACCAGACC
	Seq-81-9	AGGAAGAAATCCAAAAAATG
	Seq-81-10	GATCTTTTTATTCTTGATGA
	Seq-81-11	ATCATGAAAAACCCAAGTGT
	Seq-81-12	TGTCAACATATGGGTGGAAT
	Seq-81-13	ACAGATGGCAATGACAGACA
	Seq-81-14	AACTAGCCGAGGCCATTTTC
Seq-81-15	TTGGGTTCCAACAAGTCGAA	
SV40-R	GAAATTTGTGATGCTATTGC	
EV71	CMV-F	CGCAAATGGGCGGTAGGCGTG
	Seq-71-1	TGCTTATGGTGACAATTA
	Seq-71-2	ATCATCAGTACCTTTAAGTG
	Seq-71-3	ATCACTGTAACAGGGACCCT
	Seq-71-4	CTACCGATCAGGGTTCTGCA
	Seq-71-5	AATCCACCTTACTGGGTCAG
	Seq-71-6	GCTGAGACCACTCTTGATAG
	Seq-71-7	TAAGCCAAGTGGTGCCAGTC
	Seq-71-8	TGATCAGAAAGTGATTACGAC
	Seq-71-9	ACAGTGATGGATGATTTGTG
	Seq-71-10	ATCCATCGCCACAGTAGTGG
	Seq-71-11	CGCAGGCCTCAAAGGAGTT
	Seq-71-12	TGGGACGATAACTGGATCAG
Seq-71-13	GTGCCCTCCTAGCATGGCATA	
SV40-R	GAAATTTGTGATGCTATTGC	

核酸内切酶 I 对 PCR 反应的不同阶段辅以产物纠错。未经纠错的两轮 PCR 后的 3.0 kb 目的片段如图 2 (1A 和 2A)所示, 显示出特异性目的条带, 无拖尾、非特异和引物二聚体; 只进行第一轮 PCR 产物纠错后的凝胶电泳结果如图 2 (1B 和 2B)所示, 无明显目的条带且出现大量弥散, 表明第一轮 PCR 产物产量极低且错误序列被

T7 核酸内切酶 I 进行了切割; 只进行第二轮 PCR 产物纠错后的凝胶电泳结果如图 2 (1C 和 2C)所示, 出现目的条带但有拖尾现象, 表明第二轮 PCR 可有效富集目的片段, 且错误序列被大量切割; 两轮 PCR 后均进行纠错的第一次纠错后的凝胶电泳结果如图 2 (1D 和 2D)所示, 同样无明显目的条带且出现大量弥散, 与图 2 (1B 和 2B)结果相同, 第二次纠错后的凝胶电泳如图 2 (1E 和 2E)所示, 目的条带明显且较图 2 (1C 和 2C)减少了弥散情况, 表明纠错两次效果更优。

2.2 病毒全基因组菌落鉴定

分别将未纠错和各轮纠错后的 3.0 kb DNA 片段与载体连接构建质粒, 对获得的单克隆进行阳性鉴定。各类质粒的鉴定结果如表 5 所示, 其中未纠错的产物和只针对第一轮 PCR 反应后纠错的产物所构建的单克隆均具有较高的阳性率, 约为 75%–90%, 而只针对第二轮 PCR 反应后纠错和两轮 PCR 反应后都进行纠错的产物构建的单克隆阳性率较低, 只有 10%左右。

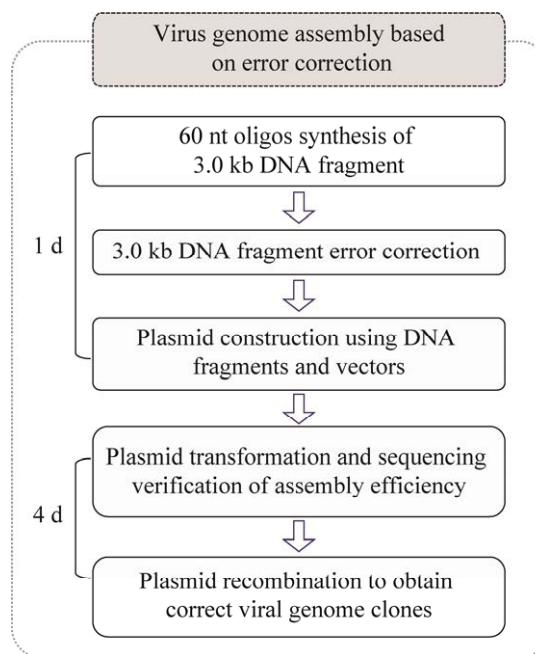


图 1 病毒基因组组装流程图

Figure 1 Flow chart of viral genome assembly.

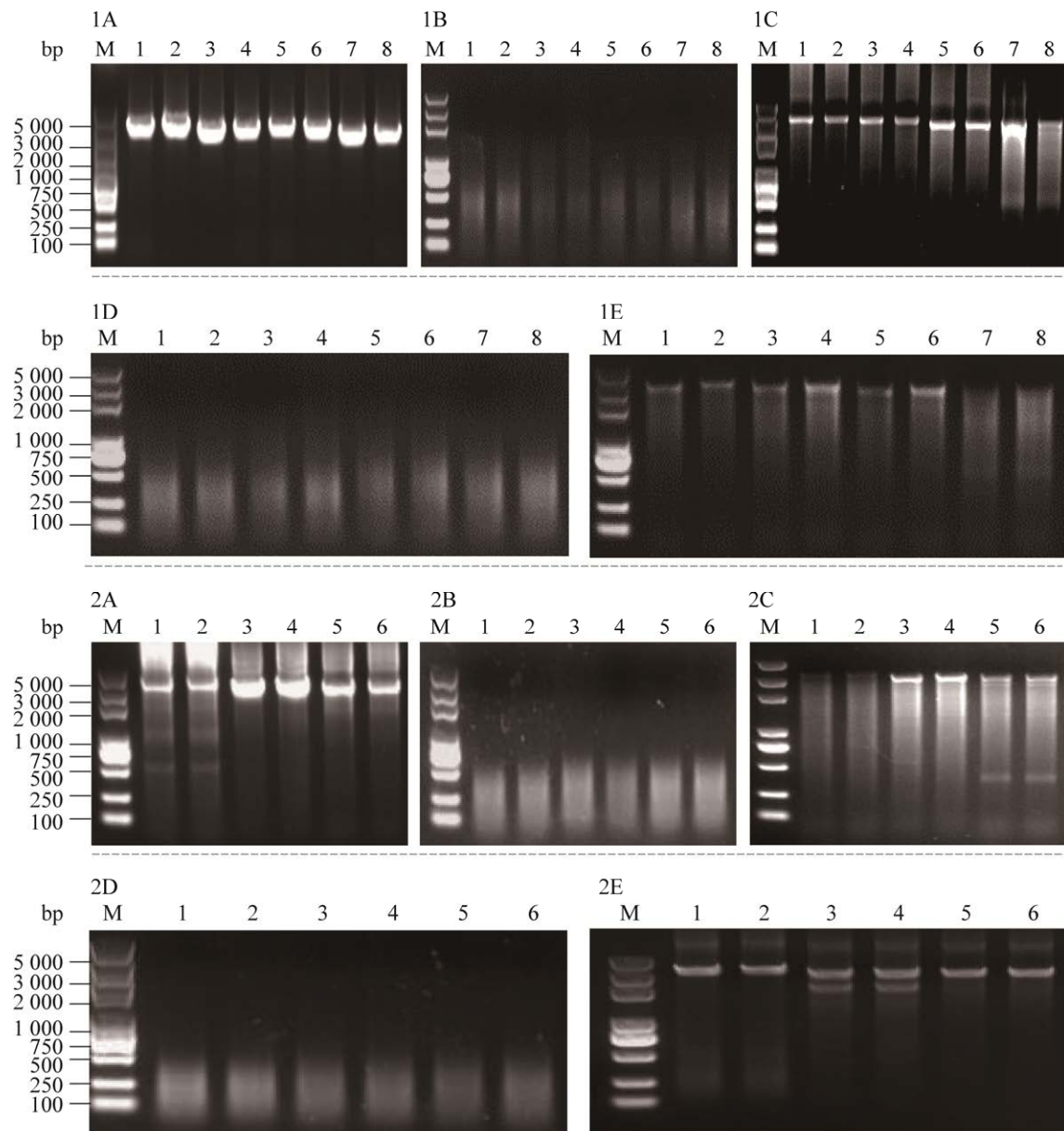


图 2 3.0 kb DNA 片段合成及纠错后琼脂糖凝胶电泳图

Figure 2 Agarose gel electrophoresis of 3.0 kb DNA fragment synthesis and error correction. 1 and 2 represent DENV2 and EV71, respectively; 1A and 2A are 3.0 kb DNA fragments after 2 rounds of PCR that have not been corrected; 1B and 2B are 3.0 kb DNA fragments corrected after the first round of PCR; 1C and 2C are 3.0 kb DNA fragments corrected after the second round of PCR; 1D and 2D are the 3.0 kb DNA fragments corrected after the first round of PCR in two rounds of error correction; 1E and 2E are 3.0 kb DNA fragments corrected after the second round of PCR in two rounds of error correction. Lane 1 and 2 in 1A–1E represent the first 3.0 kb DNA fragment of DENV 2; Lane 3 and 4 represent the second 3.0 kb DNA fragment; Lane 5 and 6 represent the third 3.0 kb DNA fragment; Lane 7 and 8 represent the fourth 3.0 kb DNA fragment. Lane 1 and 2 in 2A–2E represent the first 3.0 kb DNA fragment of EV71; Lane 3 and 4 represent the second 3.0 kb DNA fragment; Lane 5 and 6 represent the third 3.0 kb DNA fragment. M: D5000 marker.

2.3 基于纠错反应的 10 kb 病毒全基因组质粒错误率

使用 Sanger 一代测序法对 DENV2 和 EV71 的全基因组质粒进行测序, 并使用 SnapGene 软件对测序结果进行比对。分别统计了两个病毒质粒的错误率, 结果如表 6 所示。其中 DENV2 在 3' UTR 区域大约 100 bp 的范围极容易发生约 84–88 bp 的片段缺失或插入, 而使用 T7 核酸内切酶 I 对其进行两轮纠错可有效改善这一现象, 在测序的 16 个质粒中未出现片段的缺失和插入现象。在统计 DENV2 基因组组装错误率时, 为了研究 T7 核酸内切酶 I 对单碱基突变的纠错效率, 将 84–88bp 的大片段缺失或插入部分单独统计, 不列入错误率统计范围。在组装过程中对 PCR 产物只进行一轮纠错可降低错误率 11.19% (第一轮 PCR 后纠错) 和 22.38% (第二轮 PCR 后纠错), 而对 PCR 产物进行两轮纠错可降低错误率 35.50%, 且获得 1 个质粒只发生了 4 个碱基突变, 错误率仅为 0.365 errors/kb。EV71 所有单克隆未出现片段性的插入和缺失, 在组装过程中对 PCR 产物只进行一轮纠错可降低错误率 43.46% (第一轮 PCR 后纠错) 和 29.53% (第二轮 PCR 后纠错), 而对 PCR 产物进行两轮纠错可降低错误率 55.07% (表 6)。

分别统计测序质粒中不同错误类型的错误

率, 结果如图 3 所示。缺失是本研究中主要的错误类型。在缺失错误中, 更容易出现 A 碱基的缺失; 插入错误未表现出明显的偏好性; 而在突变错误中, 在所有的质粒中均未统计到 T 突变为 A 的情况, 并且 A 突变为 T, G 突变为 C 和 C 突变为 G 的频率也极低。T7 核酸内切酶 I 对所有错误类型没有表现出明显的偏好性。

2.4 病毒全基因组序列合成

以错误率最低的质粒为拼接主体, 从中首先挑选出正确序列的部分, 再从其他测序质粒中挑选 1–2 个质粒补充主体质粒错误部分的正确序列, 跨片段的引物选择具有 overlap 的部分, 分别通过 PCR 扩增的方法获得正确的片段, 随后将正确的片段与载体连接后构建质粒。

表 5 病毒全基因组阳性菌落鉴定表

Table 5 Statistical table for identification of positive colonies in the whole genome of viruses

Virus type	Error correction stage	Colony positivity rate (%)
DENV2	No correcting errors	85
	After the first round of PCR	90
	After the second round of PCR	9
	After two rounds of PCR	10
EV71	No correcting errors	79
	After the first round of PCR	75
	After the second round of PCR	13
	After two rounds of PCR	9

表 6 病毒基因组组装错误率统计表

Table 6 Statistical table for error rate of virus genome assembly

Virus type	Error correction stage	Sequencing plasmids numbers	Large mutant plasmids numbers	Error rate (errors/kb)	Error rate range
DENV2	Not correcting errors	18	7	1.448	1.095–2.280
	After the first round of PCR	19	6	1.286	0.639–2.280
	After the second round of PCR	15	6	1.124	0.730–2.463
	After two rounds of PCR	16	0	0.934	0.365–1.553
EV71	Not correcting errors	10	0	3.129	2.180–5.777
	After the first round of PCR	10	0	1.769	1.199–2.507
	After the second round of PCR	10	0	2.205	1.744–3.379
	After two rounds of PCR	10	0	1.406	0.763–2.180

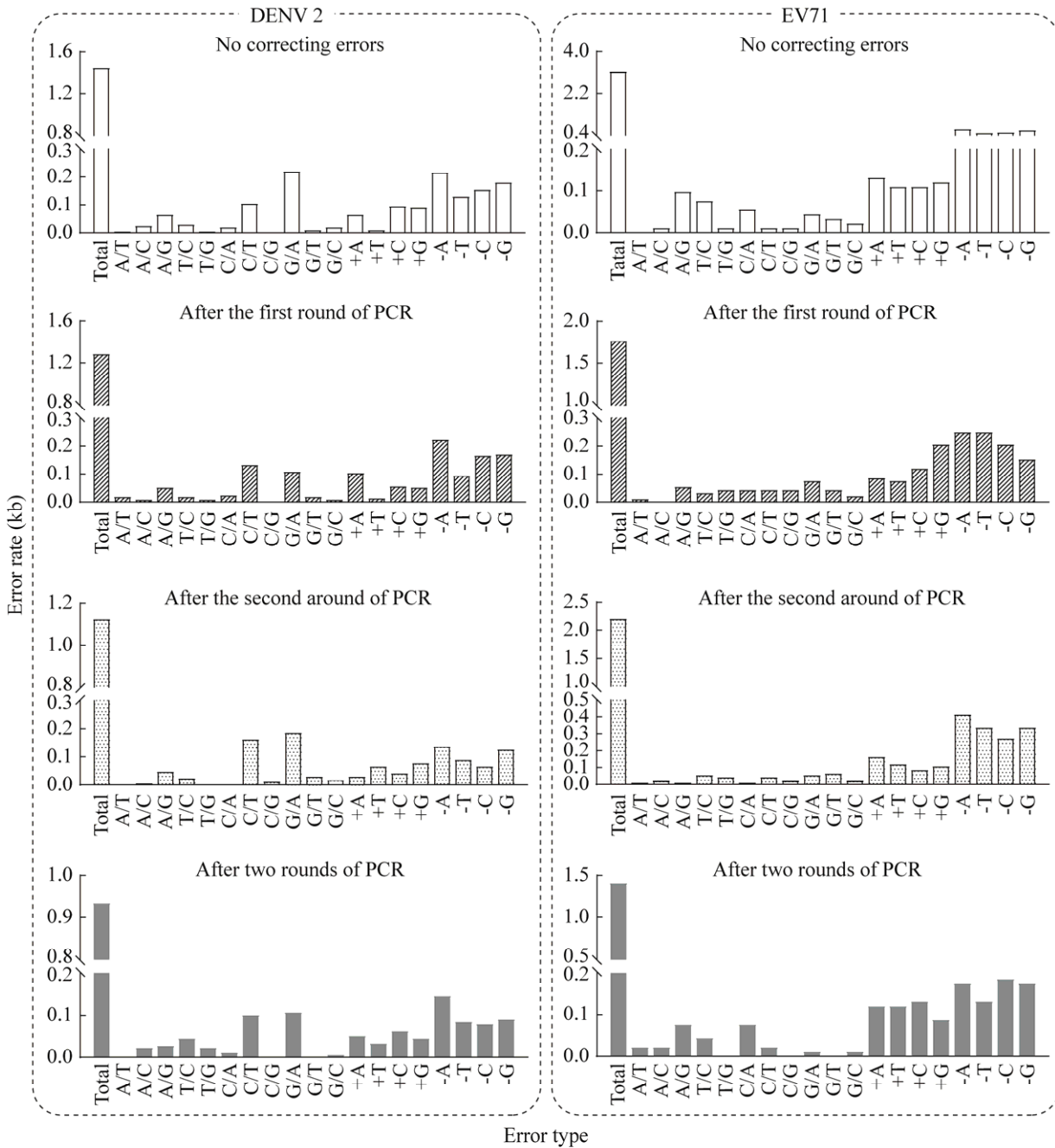


图 3 T7 核酸内切酶 I 去除错误的效率分析

Figure 3 Analysis of T7 endonuclease I removal errors.

DENV2 和 EV71 均通过拼接方法在 3 个质粒中获得了正确质粒。

3 讨论与结论

基因的从头合成是人类深度认知生命的有

效工具。本研究建立了一种高效的 10 kb 左右病毒基因组的组装方法, 通过从头合成 DENV2 进行方法开发, 并通过 EV71 从头合成验证了该方法的通用性。

序列分析设计是基因从头合成的关键步

骤。现有的序列设计软件,无法一步设计出一个约 10 kb 病毒全基因组的 oligos,且无法通过 PCR 将 10 kb 的病毒全基因组一次性从头合成,因此序列的拆分就尤为重要。通过平衡序列设计软件的能力、PCR 扩增的能力以及组装酶的组装能力,研究将序列拆分为约 3.0 kb 的 DNA 片段。该方式既可以在较短时间内设计出 oligos,又可一次性在 PCR 扩增时获得足够产量的目的序列,还可通过组装酶一次性将所有 DNA 片段组装成功。同样 oligos 的设计也非常关键,首先需确定 oligos 的长度,根据现有 DNA 合成仪的合成能力,片段越长产量越低,纯度也越低。短链 oligos 虽然能获得高纯度产物,但增加了错配风险且成本较高。最终,通过平衡纯度、DNA 片段合成准确率及成本,本研究将 oligos 长度设定在 60 nt。同样为了保证 PCR 聚合酶的扩增效率及特异性,将 T_m 值设定为 65 °C。

T7 核酸内切酶 I 能识别并切割不完全配对 DNA、Holliday 结构以及异源双链 DNA,利用其特点,本研究选择两轮 PCR 后均进行纠错的操作流程,结果发现该纠错酶除对个别突变类型没有明显改善外,对大部分的错误类型都可完成有效切割,广谱性较高。另外实验结果发现使用纠错后产物进行组装时,会产生单克隆阳性率降低现象,这可能是因为 T7 核酸内切酶 I 可以识别并切割错误片段,而较长错误片段的切割会导致基因片段缺失,缺失后的基因片段也可能构建成质粒并转化进大肠杆菌感受态细胞中。因为在对单菌落进行 PCR 鉴定以及后续的测序质控时,发现大部分假阳性单菌落不是空载体,而是含有部分病毒基因组的片段。

Sequeira 等^[23]于 2016 年分别表达了 T7 核酸内切酶 I 及融合麦芽糖结合蛋白基因的 T7 核酸内切酶 I,以测试其在体外基因合成过程中切

割错配纠正错误的的能力。在该研究中,Sequeira 等^[23]以包含有 lacZ 启动子的绿色荧光蛋白基因作为基因合成模板,全长共计 967 nt,通过两轮 PCR 的方法扩增该基因片段,随后进行变性-退火步骤形成含有错配的异源双链,再分别使用 T7 核酸内切酶 I 和融合 T7 核酸内切酶 I 进行一次切割,最后进行第三轮 PCR 以增加正确序列的比例。通过该方法 T7 核酸内切酶 I 将错误率从 3.45 errors/kb 降低为 0.84 errors/kb,融合 T7 核酸内切酶 I 将错误率降低为 0.43 errors/kb。该研究^[23]证明了 T7 核酸内切酶 I 在体外基因合成过程中纠正错误序列的有效性,但其合成片段较短(967 nt)、操作流程复杂。本研究从头合成了 3.0 kb DNA 片段,并用 T7 核酸内切酶 I 进行纠错,进而组装成 10 kb DNA 片段,错误率最低可至 0.36 errors/kb,且有效修正了 DNA 长片段合成与组装过程中出现的片段缺失问题。本研究在更长的 DNA 片段合成中测试了 T7 核酸内切酶 I 的纠错效果,结果表明,该酶是建立 DNA 长片段精准快速合成方法的重要支撑。

病毒基因组序列特殊复杂,且与工程菌株相互影响,不同基因序列合成难易程度存在差异^[20]。本研究虽然使用 T7 核酸内切酶 I 有效提高了基因组合成的正确率,但是在合成与组装 EV71 基因组时错误率仍然颇高。为获得完全正确的序列,质粒的再次拼接就显得尤为重要。本研究首先以 EV71 错误率最低的质粒为“拼接主体”,扩增其正确片段的部分;再从其他质粒中寻找“拼接主体”有突变而该质粒正确的部分进行扩增,扩增引物只需要选择基因合成时的 oligos,仅通过一轮 PCR 扩增后进行亚克隆就获得了正确的 EV71 基因组单克隆。

常规的基因组拼装方法从 1.0 kb 以内的 DNA 短片段开始,通过不断将短片段构建到质

粒上测序, 再将质粒酶切后重新拼装成更大的 DNA 片段的循环步骤来逐级拼接, 通常一个 10 kb 的基因组序列可能需要重复 3-4 次才能完成。这期间需要不断重复质粒酶切、质粒构建、质粒提取及质粒测序的步骤, 耗费大量人力物力。本研究建立的 10 kb 病毒基因组组装方法, 通过合理地拆分序列, 选取高保真聚合酶进行合成, 辅以 T7 核酸内切酶 I 的两次纠错, 通过一次性 Gibson 组装, 在 1 d 内即可获得 10 kb 的基因组序列, 随后通过质粒测序和一次拼接, 大约 5 d 内可得到正确的 10 kb 病毒基因组质粒。本研究开发的方法节省了时间和成本, 还可根据该方案合理开发自动化的基因组组装方法, 为基因组合成的标准化流程和自动化基因组组装提供了方法支持和基础。

REFERENCES

- [1] CELLO J, PAUL AV, WIMMER E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template[J]. *Nature*, 2002, 297(5583):1016-1018.
- [2] HUTCHISON CA 3rd, CHUANG RY, NOSKOV VN, ASSAD-GARCIA N, DEERINCK TJ, ELLISMAN MH, GILL J, KANNAN K, KARAS BJ, MA L, PELLETIER JF, QI ZQ, RICHTER RA, STRYCHALSKI EA, SUN LJ, SUZUKI Y, TSVETANOVA B, WISE KS, SMITH HO, GLASS JI, MERRYMAN C, GIBSON DG, VENTER JC. Design and synthesis of a minimal bacterial genome[J]. *Science*, 2016, 351(6280): aad6253.
- [3] KANNAN K, GIBSON DG. Yeast genome, by design[J]. *Science*, 2017, 355(6329): 1024-1025.
- [4] ZHANG WM, ZHAO GH, LUO ZQ, LIN YC, WANG LH, GUO YK, WANG A, JIANG SY, JIANG QW, GONG JH, WANG Y, HOU S, HUANG J, LI TY, QIN YR, DONG JK, QIN Q, ZHANG JY, ZOU XZ, HE X, et al. Engineering the ribosomal DNA in a megabase synthetic chromosome[J]. *Science*, 2017, 355(6329): eaaf3981.
- [5] MERCY G, MOZZICONACCI J, SCOLARI VF, YANG K, ZHAO GH, THIERRY A, LUO YS, MITCHELL LA, SHEN M, SHEN Y, WALKER R, ZHANG WM, WU Y, XIE ZX, LUO ZQ, CAI YZ, DAI JB, YANG HM, YUAN YJ, BOEKE JD, et al. 3D organization of synthetic and scrambled chromosomes[J]. *Science*, 2017, 355(6329): eaaf4597.
- [6] XIE ZX, LI BZ, MITCHELL LA, WU Y, QI X, JIN Z, JIA B, WANG X, ZENG BX, LIU HM, WU XL, FENG Q, ZHANG WZ, LIU W, DING MZ, LI X, ZHAO GR, QIAO JJ, CHENG JS, ZHAO M, KUANG Z, et al. "Perfect" designer chromosome V and behavior of a ring derivative[J]. *Science*, 2017, 355(6329): eaaf4704.
- [7] WU Y, LI BZ, ZHAO M, MITCHELL LA, XIE ZX, LIN QH, WANG X, XIAO WH, WANG Y, ZHOU X, LIU H, LI X, DING MZ, LIU D, ZHANG L, LIU BL, WU XL, LI FF, DONG XT, JIA B, et al. Bug mapping and fitness testing of chemically synthesized chromosome X[J]. *Science*, 2017, 355(6329): eaaf4706.
- [8] SHEN Y, WANG Y, CHEN T, GAO F, GONG JH, ABRAMCZYK D, WALKER R, ZHAO HC, CHEN SH, LIU W, LUO YS, MÜLLER CA, PAUL-DUBOIS-TAINE A, ALVER B, STRACQUADANIO G, MITCHELL LA, LUO ZQ, FAN YQ, ZHOU BJ, WEN B, et al. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome[J]. *Science*, 2017, 355(6329): eaaf4791.
- [9] MITCHELL LA, WANG A, STRACQUADANIO G, KUANG Z, WANG XY, YANG K, RICHARDSON S, MARTIN JA, ZHAO Y, WALKER R, LUO YS, DAI HJ, DONG K, TANG ZJ, YANG YL, CAI YZ, HEGUY A, UEBERHEIDE B, FENYŐ D, DAI JB, et al. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond[J]. *Science*, 2017, 355(6329): eaaf4831.
- [10] RICHARDSON SM, MITCHELL LA, STRACQUADANIO G, YANG K, DYMOND JS, DiCARLO JE, LEE D, HUANG CLV, CHANDRASEGARAN S, CAI YZ, BOEKE JD, BADER JS. Design of a synthetic yeast genome[J]. *Science*, 2017, 355(6329): 1040-1044.
- [11] SHEN Y, GAO F, WANG Y, WANG YR, ZHENG J, GONG JH, ZHANG JT, LUO ZQ, SCHINDLER D, DENG Y, DING WC, LIN T, SWIDAH R, ZHAO HC, JIANG SY, ZENG C, CHEN SH, CHEN T, WANG Y, LUO YS, et al. Dissecting aneuploidy phenotypes by constructing Sc2.0 chromosome VII and SCRaMbLEing synthetic disomic yeast[J]. *Cell Genomics*, 2023, 3(11): 100364.

- [12] HOOVER DM, LUBKOWSKI J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis[J]. *Nucleic Acids Research*, 2002, 30(10): e43.
- [13] CARUTHERS MH, BEAUCAGE SL, BECKER C, EFCAVITCH JW, FISHER EF, GALLUPPI G, GOLDMAN R, DEHASETH P, MATTEUCCI M, McBRIDE L. Deoxyoligonucleotide synthesis via the phosphoramidite method[J]. *Gene Amplification and Analysis*, 1983, 3: 1-26.
- [14] HAO M, QIAO JJ, QI H. Current and emerging methods for the synthesis of single-stranded DNA[J]. *Genes*, 2020, 11(2): 116.
- [15] de KOK S, STANTON LH, SLABY T, DUROT M, HOLMES VF, PATEL KG, PLATT D, SHAPLAND EB, SERBER Z, DEAN J, NEWMAN JD, CHANDRAN SS. Rapid and reliable DNA assembly via ligase cycling reaction[J]. *ACS Synthetic Biology*, 2014, 3(2): 97-106.
- [16] MA SY, TANG N, TIAN JD. DNA synthesis, assembly and applications in synthetic biology[J]. *Current Opinion in Chemical Biology*, 2012, 16(3/4): 260-267.
- [17] 常汉臣, 王琛, 王培霞, 周见庭, 李炳志. DNA 组装技术[J]. *生物工程学报*, 2019, 35(12): 2215-2226.
CHANG HC, WANG C, WANG PX, ZHOU JT, LI BZ. DNA assembly technologies: a review[J]. *Chinese Journal of Biotechnology*, 2019, 35(12): 2215-2226 (in Chinese).
- [18] 李雷, 芦银华, 姜卫红. DNA 组装新方法的研究进展[J]. *生物工程学报*, 2013, 29(8): 1113-1122.
LI L, LU YH, JIANG WH. Perspective on the novel methods for DNA assembly[J]. *Chinese Journal of Biotechnology*, 2013, 29(8): 1113-1122 (in Chinese).
- [19] 罗周卿, 戴俊彪. 合成基因组学: 设计与合成的艺术[J]. *生物工程学报*, 2017, 33(3): 331-342.
LUO ZQ, DAI JB. Synthetic genomics: the art of design and synthesis[J]. *Chinese Journal of Biotechnology*, 2017, 33(3): 331-342 (in Chinese).
- [20] MA SY, SAAEM I, TIAN JD. Error correction in gene synthesis technology[J]. *Trends in Biotechnology*, 2012, 30(3): 147-154.
- [21] QIU RY, SAKATO M, SACHO EJ, WILKINS H, ZHANG XD, MODRICH P, HINGORANI MM, ERIE DA, WENINGER KR. MutL traps MutS at a DNA mismatch[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(35): 10914-10919.
- [22] YANG XW, HAN XP, HAN C, LONDON J, FISHEL R, LIU JQ. MutS functions as a clamp loader by positioning MutL on the DNA during mismatch repair[J]. *Nature Communications*, 2022, 13: 5808.
- [23] SEQUEIRA AF, GUERREIRO CIPD, VINCENTELLI R, FONTES CMGA. T7 endonuclease I mediates error correction in artificial gene synthesis[J]. *Molecular Biotechnology*, 2016, 58(8): 573-584.