

文章编号: 1004-0374(2013)10-0983-10

· 技术与应用 ·

## 合成生物学中的DNA组装技术

赵 鹏, 王 霞, 李炳志\*, 程景胜, 元英进

(天津大学化工学院系统生物工程教育部重点实验室, 天津 300072)

**摘 要:** 合成生物学旨在应用工程学的研究思路及手段去设计或改造生物系统, 是一个综合了科学与工程学的拥有发展潜力的新兴学科, 在生物医药、农业、能源、环保等方面发挥着巨大作用。DNA 组装技术是合成生物学中的关键技术, 也是合成生物学快速发展的限制性技术。综述了众多 DNA 组装技术的发展及其在合成生物学研究中的意义和应用。

**关键词:** 合成生物学; DNA 组装; 基因合成; 生物途径构建; 基因组合成

**中图分类号:** O621.3; Q523; TB383 **文献标志码:** A

### DNA assembly in synthetic biology

ZHAO Juan, WANG Xia, LI Bing-Zhi\*, CHENG Jing-Sheng, YUAN Ying-Jin

(Key Laboratory of System Biological Engineering of the Ministry of Education,  
School of Chemical Engineering, Tianjin University, Tianjin 300072, China)

**Abstract:** Synthetic biology aims to apply the engineering principles to the design of biological system. It is an emerging field that combines science and engineering, and has great potentials in promoting the development in biomedical research, agriculture, bio-energy and bioremediation. DNA assembly is a pivotal technique implementing the vision of synthetic biology, and is also a limiting technology in advancing synthetic biology. This paper will review the development and current status of DNA assembly, and discuss its applications and significance in synthetic biology.

**Key words:** synthetic biology; DNA assembly; gene synthesis; pathway construction; genome synthesis

合成生物学是一个综合了科学与工程的新兴研究领域, 在生物、医药、农业、能源及环保等方面具有巨大的应用潜力。合成生物学旨在应用工程学的研究思路及手段去设计改造基因、生物元件、生物途径, 甚至整个基因组; 而 DNA 组装技术则是实现合成生物学各种目的的关键技术。由于基因合成的成本降低, 基因组的全合成增加了对人工设计合成细胞的期望。目前, 设计并合成基因组的能力还远远落后于已掌握的合理设计生物元件的能力。合成生物学的大部分工作还处在利用模块化的生物组件组合构建各种网络和代谢途径。很明显, DNA 组装是合成生物学快速发展的限制性技术。因此, 合理而可靠的组装技术是发展合成生物学所必需的。

#### 1 从寡核苷酸到基因的组装

在阐明基因功能、分析蛋白质与核酸相互作用、

基因异源表达等研究中, DNA 片段的化学合成是一个高效的技术手段。在多数情况下, 如模板 DNA 不可得或是密码子需要优化, 化学合成是唯一的选择。20 世纪 60 年代、70 年代初, 一群有机化学家, 尤其是 Khorana 开发了寡核苷酸的合成技术, 为基因的化学合成奠定了基础。1970 年, Khorana 等<sup>[1]</sup>合成了编码酵母丙氨酸 tRNA 的基因, 这是第一条人工合成基因。随后, 一系列蛋白质的编码基因被合成并在大肠杆菌中表达<sup>[2-4]</sup>。这一系列研究充分证明了化学合成基因的可行性。20 世纪 80 年代中期, 大于 100 bp 的寡核苷酸的合成得以

收稿日期: 2013-03-14; 修回日期: 2013-04-07

基金项目: 国家高技术研究发展计划项目(“863”计划)(2012AA02A708)

\*通信作者: E-mail: bzli@tju.edu.cn

实现<sup>[5]</sup>,大大地推动了基因合成技术的发展。通过连接酶介导法和 *FokI* 法等,寡核苷酸被组装成具有功能的基因<sup>[6-7]</sup>。然而,早期的合成技术通常只能合成小于 1 kb 的 DNA 片段。

自 20 世纪 90 年代,PCR(polymerase chain reaction) 技术被用于基因合成,并在长期的研究发展中日渐成熟。连接酶链式反应法(ligase chain reaction, LCR)与 PCR 法的结合使用,在基因合成研究中发挥了重要作用<sup>[8]</sup>。该法通过热循环反应,采用嗜热性 DNA 连接酶将首尾相连、重叠杂交的寡核苷酸片段连接起来,再以连接产物为模板、首尾寡核苷酸片段为引物进行 PCR 扩增,从而合成基因片段(图 1a)。LCR 法还被用于 microRNAs 的检测<sup>[9]</sup>以及单核苷酸多态性的检测<sup>[10]</sup>等。1995 年,Stemmer 等<sup>[11]</sup>描述了一种只依赖于 DNA 聚合酶的基因合成法,将 56 个 40 bp 的寡核苷酸片段通过 PCR 法组装成 1.1 kb 的 TEM-1  $\beta$  内酰胺酶基因。采用同样的方法,他们还将 134 个 40 bp 的寡核苷酸片段合成一个 2.7 kb 的质粒<sup>[11]</sup>。随后,Withers 等<sup>[12]</sup>对该法进行优化,合成了 2.1 kb 的恶性疟原虫基因 *pfsub-1*。

近年来,以 PCR 为基础的基因合成方法得到了进一步的演变与发展。2003 年,Smith 等<sup>[13]</sup>将连接酶拼接法(LCR)与重叠延伸 PCR 法(overlap extension PCR, OE-PCR)相结合(图 2a),合成了噬菌体  $\phi$ X174 基因组全长序列(5 386 bp)。2004 年,Young 和 Dong<sup>[14]</sup>将双重不对称 PCR(dual asymmetrical PCR, DA-PCR)与 OE-PCR 相结合,在 DA-PCR 阶段,每四个相邻的带有重叠区的寡核苷酸互补结合成小片段,这些小片段再经过 OE-PCR 合成全长基因(图 2b)。

Kodumal 等<sup>[15]</sup>、Xiong 等<sup>[16-17]</sup>、Reisinger 等<sup>[18]</sup>采用以 PCR 为基础的两步法 DNA 合成方法,将大量带有重叠区的寡核苷酸合成长基因。在这种方法中,设计合成的寡核苷酸涵盖基因的两条链,所有寡核苷酸在一个 PCR 反应体系中经过重叠延伸,再以首尾寡核苷酸片段为引物扩增合成基因全长(图 1b)。PTDS (PCR-based two-step DNA synthesis) 是 Xiong 等<sup>[16]</sup>描述的另一类以 PCR 为基础的两步法 DNA 合成方法(图 3),它包含两个关键步骤:首先由寡核苷酸合成几个独立的 DNA 小片段;其次,通过 OE-PCR 将这些 DNA 小片段组装成全长基因。

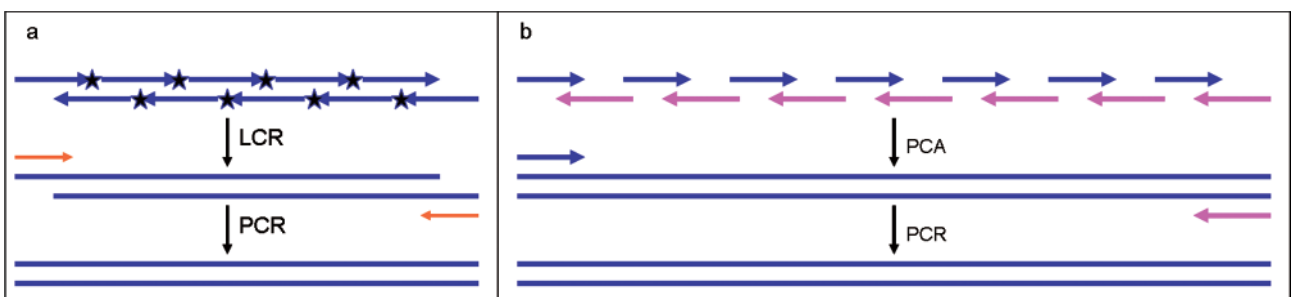


图1 LCR-PCR法合成基因(a)及PCA-PCR法合成基因(b)的示意图

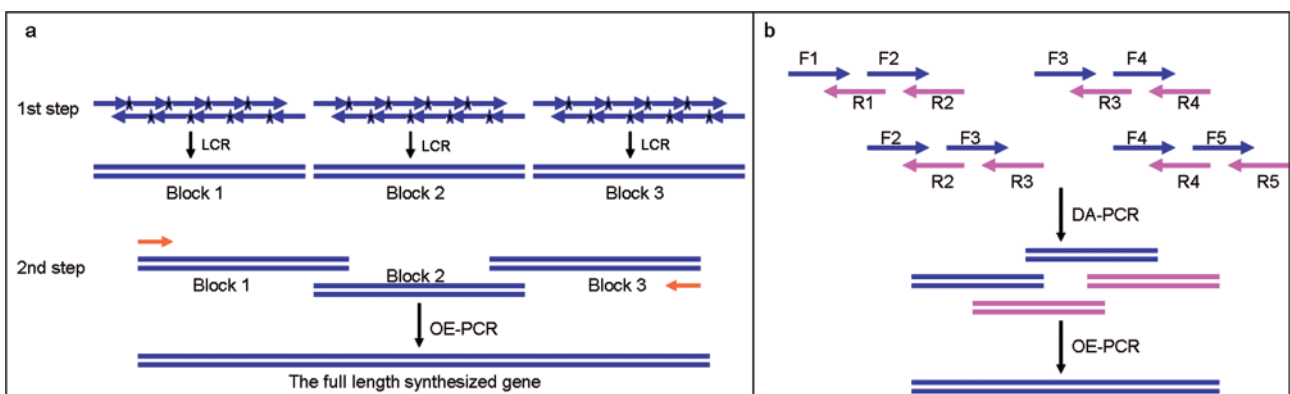


图2 LCR+OE-PCR法合成基因(a)及DA-PCR+OE-PCR法合成基因(b)的示意图

通过这种方法, 他们合成了 657 bp 的水稻转录因子 *OsDREB1B*<sup>[19]</sup>、1 230 bp 的 *Peniophoralycii* 植酸酶基因<sup>[17]</sup>、1 245 bp 的 HBV 表面抗原基因 *PRS-SIS2S*<sup>[20]</sup>、2 382 bp 的 *vip3aI* 基因及 5 367 bp 的 *CrtEBWY* 基因<sup>[16]</sup> 等。这种方法对于 GC 含量高、重复序列多或二级结构复杂的基因合成有一定的适用性。

另一种以 PCR 为基础的两步法基因合成法是由 Gao 等<sup>[21]</sup> 建立的热力学平衡由内而外合成法 (thermodynamically balanced inside-out, TBIO)(图 4),

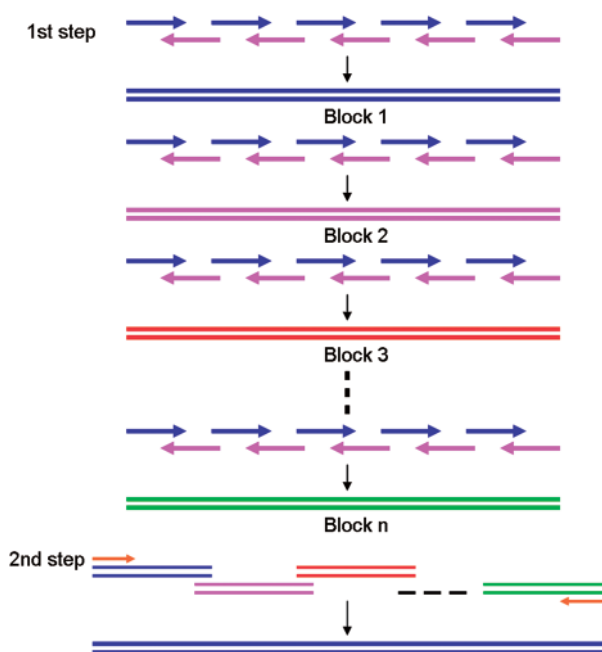


图3 PTDS法合成基因的示意图

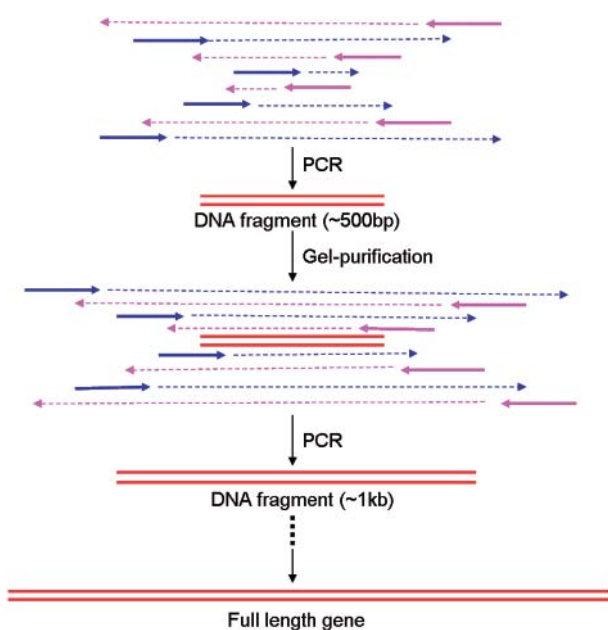


图4 TBIO法合成基因的示意图

该法包含几个关键的步骤: (1) 引物的设计, 正向引物涵盖基因 N 端的一半序列, 反向引物涵盖基因 C 端的一半序列; (2) 将位于中间的 4~6 对引物混合进行内外双向延伸 PCR 反应, 合成一个小片段; (3) 将合成的小片段进行纯化, 并作为模板, 与另外几对稍微外侧的引物进行又一次的内外双向延伸 PCR 反应, 合成一个较长的片段; (4) 片段纯化及 PCR 反应继续进行, 直到获得全长基因。这种合成方法的突变率 (0~0.3%) 低于许多其他方法 (0.1%~1%)。用于基因合成的另一策略是连续延伸 PCR (successive extension PCR)<sup>[22-24]</sup>, 它的引物设计与 TBIO 相同, 然后将所有引物混合进行延伸反应, 获得全长基因。Peng 等<sup>[25]</sup> 应用该法将 26 个寡核苷酸合成 *BtcryIA(c)* 基因; 但对于较长基因 (>1 kb) 的合成, 采用这种方法通常得不到目的片段, 且非特异条带非常明显。因此, Xiong 等<sup>[16]</sup> 优化了该法, 采用不等量引物混合及分段合成法进行较长基因的合成, 取得成功。

为了提高基因合成通量并降低成本, 研究者们开始探索基因合成的微型化和自动化技术。2004 年, Tian 等<sup>[26]</sup> 和 Zhou 等<sup>[27]</sup> 研制出用 DNA 芯片合成大量寡核苷酸片段进行基因组装的方法。随后的研究在提高基因组装效率和选择性扩增寡核苷酸库等方面取得了一定进展, 但却没有解决微型化和自动化的问题。2011 年, Tian 实验室研发出一种新的芯片基因合成技术, 将寡核苷酸合成、扩增和连接组装等步骤集成到一块塑料芯片上, 从而初步解决了这一难题<sup>[28]</sup>。基因芯片在生物、医学、农业等领域得到广泛应用<sup>[29-30]</sup>。

除了体外合成法, 在生物体内也可以进行基因的合成。2009 年, Gibson<sup>[31]</sup> 将 38 个寡核苷酸片段和一个线性化的质粒在酿酒酵母体内合成了一个携带 1 170 bp 基因序列的环形质粒。相邻寡核苷酸片段之间具有 30 bp 的重叠区, 首尾寡核苷酸片段与线性化质粒具有 20 bp 的重叠区; 利用酵母细胞的同源重组功能, 将同时转化进酿酒酵母体内的寡核苷酸片段和线性化质粒自动修复整合成为一个携带基因的环状质粒。

## 2 从元件到功能基因的组装

一个有功能的基因包含带有核糖体结合位点 (RBS) 的启动子、开放阅读框 (ORF) 及终止子。当然也存在其他元件, 诸如上下游调节区、内含子、RNA 可折叠区等, 但这些元件并不是必需的。将

这些独立元件组装成一个有功能的基因需要按顺序进行,即启动子在 ORF 之前,终止子在 ORF 之后, Knight<sup>[32]</sup> 建立的 BioBrick<sup>™</sup> 组装法则是一个有效的方法。在各个独立元件的序列两端加入标准的侧翼序列,如此则可以通过一个低成本的、简单的、标准化的酶切连接法将各个元件有序地组装起来(图 5a)。采用 BioBrick<sup>™</sup> 法可以构建并储存一系列的生物模块供大家分享。在这种方法兴起的十年内,合成生物学中生产了一大批有价值的生物组件和途径<sup>[33-35]</sup>。此方法还可与 PCR 结合使用来合成 DNA 重组片段<sup>[36]</sup>。BioBrick<sup>™</sup> 最主要的缺点是在元件相互连接处存在 8 bp 序列痕迹,而这段痕迹在某些情况下是不可接受的,如某些基因的 RBS 必须紧挨 ORF。又由于这 8 bp 的序列含有一个终止密码子,因此用于融合蛋白的构建。为了解决这一问题,研究者们对传统的 BioBrick<sup>™</sup> 组装法进行改进,建立了两种用于构建融合蛋白的方法(图 5b、c)<sup>[37-38]</sup>。2010 年,Anderson 等<sup>[39]</sup> 创建了一种新的标准化组装法 BglBricks,这种方法只在连接处留下 6 bp 的

序列痕迹,可用于融合蛋白的构建。此外,该法采用常见的高效的限制性内切酶,且这些酶的识别序列不被最常见的 DNA 甲基化酶所阻断。

然而,不管是 BioBrick<sup>™</sup> 及其改进方案还是 BglBricks,都不能实现无痕组装,且受酶切位点序列限制。要实现不受序列限制的无痕组装,重叠延伸 PCR (OE-PCR) 是一个不错的选择。OE-PCR 由 Horton 等<sup>[40]</sup> 建立的,它采用具有互补末端的引物,使 PCR 产物产生重叠区,从而在随后的扩增反应中通过重叠区的互补结合,将不同来源的扩增片段拼接起来(图 6a)。该法操作简单、周期短,适用于 0.5~5 kb 的 DNA 片段的组装。OE-PCR 由于其技术优势,广泛应用于基因克隆<sup>[41]</sup>、定点突变<sup>[42]</sup>、基因剪接<sup>[43]</sup> 等研究中。研究表明,OE-PCR 不仅能用于基因的组装,也具有组装整个质粒的能力。2009 年,Quan 和 Tian<sup>[44]</sup> 描述的聚合酶环形成克隆 (circular polymerase extension cloning, CPEC) 即是根据 OE-PCR 原理,通过无引物 PCR 将片段连接进入线性化质粒,形成的环形质粒可直接转化大

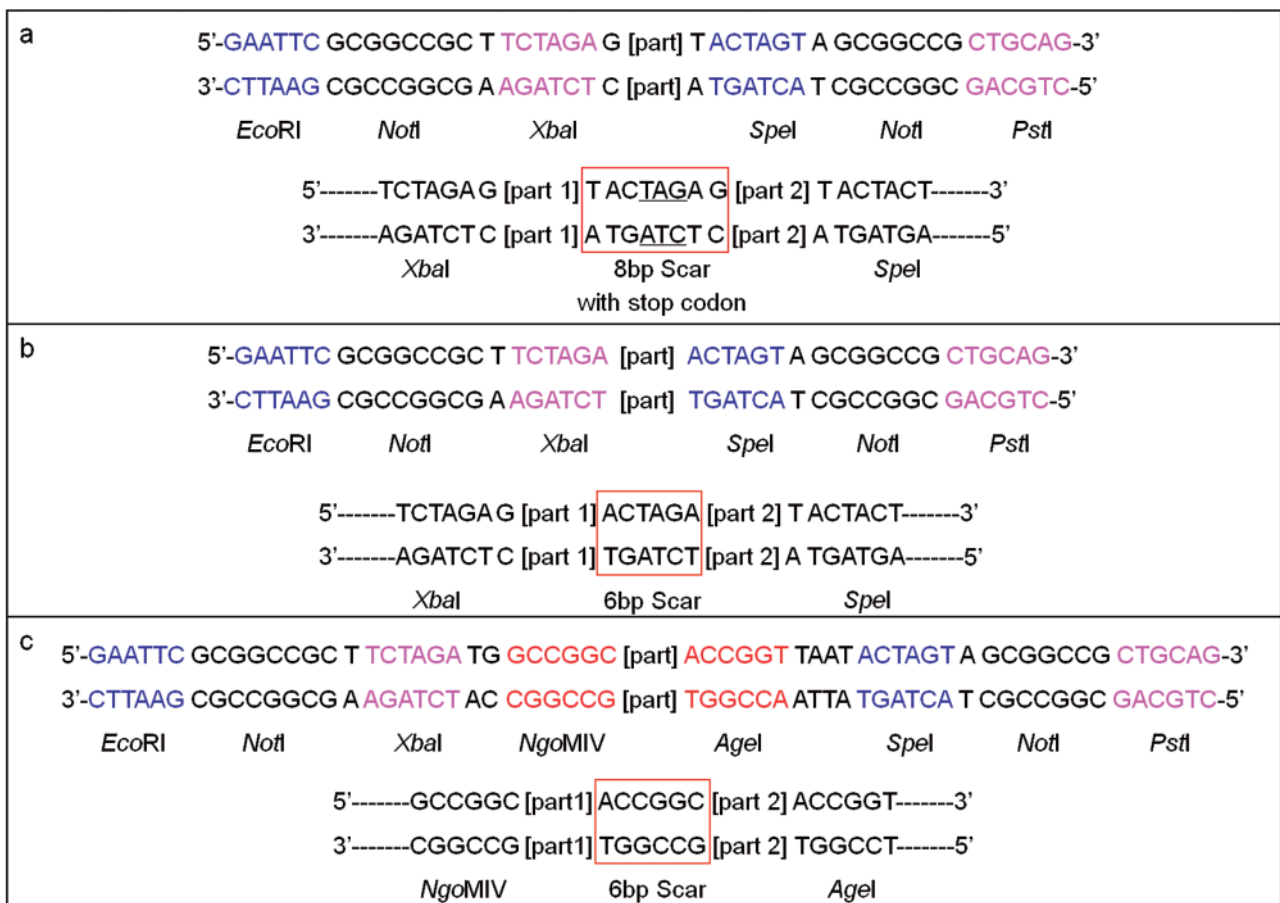


图5 BioBrick<sup>™</sup> 组装法(a)及其改进策略(b、c)



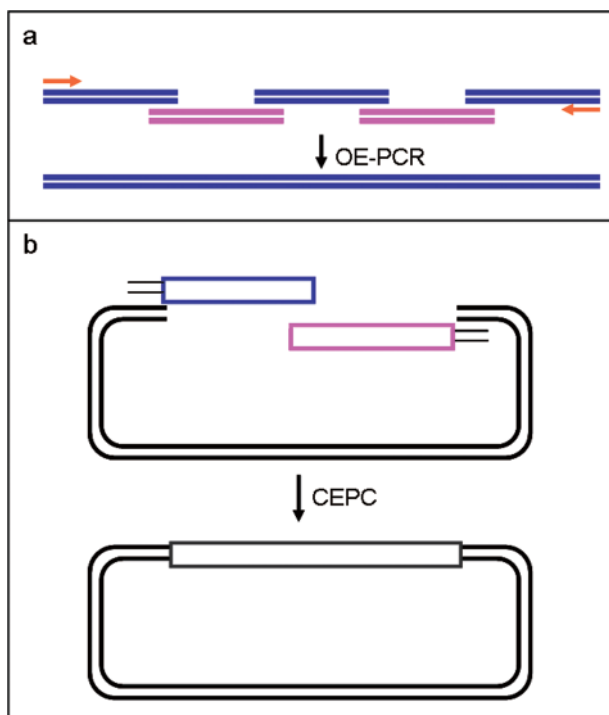


图6 OE-PCR(a)和CEPC(b)的原理示意图

肠杆菌感受态细胞 (图 6b)。这种方法依赖于片段与载体之间的同源序列, 一般为 20~25 bp。该法简便高效, 既可用于单片段的克隆, 也可用于多片段与载体的组装。2010 年, 一个基于 CPEC 的改良技术<sup>[45]</sup>显示, 用于克隆的载体不需要进行线性化, 只需在反应完成后采用限制性内切酶 *DpnI* 对产物进行处理。该酶对甲基化敏感, 可以将原始的环形载体消化而只留下经 PCR 组装而成的新的质粒。这两种技术为 OE-PCR 提供了一个新的强大的视角, 可以快速地完成了从元件到基因到整个质粒的组装而不需要使用连接酶。但是, 当序列高度重复或 GC 含量高时, 所有的 PCR 法都不能确保组装的精确性。

2013 年, 赵国屏研究组建立了一种新的无序列依赖性的 DNA 无痕组装法, 称之为 MASTER 连接法 (methylation-assisted tailorable ends rational ligation method)<sup>[46]</sup>。该法所用的一个关键酶是同时具有 IIm 型和 IIs 特性的限制性内切酶 *MspII*, 该酶只识别甲基化的 4 bp 的位点, <sup>m</sup>CNNR (R=A or G), 然后在识别位点的外侧进行切割, 留下黏性末端供连接使用。运用此法, 他们组装成了 29 kb 的天蓝色链霉菌放线紫红素的生物合成基因簇, 但是, *MspII* 是一个比较昂贵的酶, 在进行大量组装时应考虑其经济成本。

### 3 从基因到生物途径的组装

在合成生物学领域, 将原件组装成功能基因只是一个最基本的步骤, 大多数研究都致力于将多个基因重组为具有特定功能的生物途径, 如 iGEM 项目构建的具有独创性的调节网络<sup>[47]</sup>、通过代谢工程实现的生物燃油的工业合成<sup>[48]</sup>等。在生物途径的构建中, 基因之间的序列痕迹对其功能的发挥没有明显的影响。因此, BioBrick<sup>TM</sup>、BglBricks 等方法也适用于将基因组装成生物途径; 但由于这些方法受序列限制, 且耗时长, 因而只适合构建小型的途径。OE-PCR, 特别是 CPEC 也适用于小型途径的构建, 这些方法自身的特点及缺陷限制了它们进行更大规模的组装。

IIs 型限制性内切酶的使用让更大规模的组装得以实现。这些酶在其识别序列的外侧进行切割<sup>[49]</sup>, 留下 4 bp 的黏性末端, 因而只需在与其进行连接的另一片段上合理地设计 4 bp 的互补序列, 就可以进行无痕组装 (图 7)。Kodumal 等<sup>[50]</sup>利用 IIs 型限制性内切酶法组装了一个 32 kb 的基因簇。Golden Gate 组装法就是依据这种酶的特性建立的, 它可以通过一步酶切连接构建得到重组质粒, 正确率几乎 100%<sup>[51]</sup>。这种方法可以有序、并行的进行无痕组装, 在合成生物学领域倍受推崇。从理论上讲, 4 bp 的序列可以组成 256 种不同的互补区, 因而该法可用于多片段的拼接。但研究表明, 相似的互补序列容易形成错配而造成错误的组装, 所以该法更适合于元件到功能基因的组装以及小型生物途径的构建。

2010 年, Blake 等<sup>[52]</sup>描述了一种比较复杂的 IIs 型限制性内切酶克隆法, 专用于将 DNA 大片段

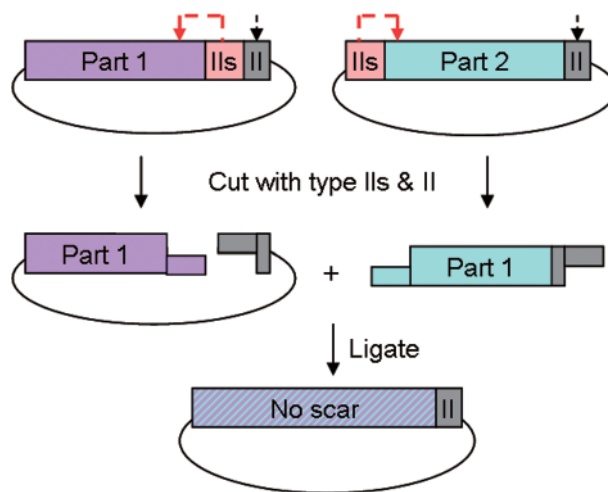


图7 IIs型限制性内切酶进行无痕组装的示意图

组装成较大的组件或途径。这种方法的主体是两端带有 65 bp 的标准化标签的片段, 以及两种载体, 每种载体携带两个不同的无启动子的抗生素抗性标记。IIs 型限制性内切酶的识别序列位于标签内, 片段经 IIs 型内切酶处理后, 连接到一种载体的特定位置; 在此位置, 标签序列恰好可充当抗生素抗性标记的启动子而使其发挥作用, 便于筛选。两个质粒的插入片段再次经过 IIs 型内切酶消化, 并暴露出 4 bp 的互补区; 然后这两个片段被连接进入带有另两种抗性标记的载体中, 组装成一个更大的片段。重复上述过程, 他们将多个 1~2 kb 的片段经过六次循环组装成了 91 kb 的大片段。

此外, 有一系列被称作 overlap 法的组装技术被用于生物途径的构建中。这些方法不需要使用限制性内切酶, 只要求片段之间具有同一个约 20 bp 的序列。这段同源序列经过酶处理从而使片段拼接在一起。著名的商业化的质粒构建试剂盒 Gateway™ 就是采用这种方法, 带有同源序列的片段经过一种特殊的重组酶处理而组装在一起<sup>[53]</sup>。由于每个连接处都会留下相同的约 20 bp 的序列, 因而这种方法并不适合于大规模片段的平行组装。相对而言, In-Fusion™ 则是一种更受欢迎的商业化克隆试剂盒<sup>[54]</sup>, 它利用专有的酶混合物可以将任何具有 15 bp 同源序列的片段组装在一起, 不遗留连接痕迹, 特别适用于多个 DNA 片段的平行连接(图 8)。

另一种 overlap 法被称作 USER 克隆<sup>[55]</sup>, 首先采用含有至少一个尿嘧啶的引物扩增片段, 然后利

用尿嘧啶 DNA 糖基酶将尿嘧啶切除, 余下一个无碱基位点。这个位点经 AP 裂解酶处理, 则留下一个与原始引物序列互补的 3' 悬挂区。依赖于这种悬挂区将片段连接起来, 而不需要使用连接酶。这种组装方法在片段之间不留下痕迹, 可用于多片段的平行组装, 但该法也存在一个缺点, 它要求在片段末端附近至少存在一个胸腺嘧啶, 因而并不是完全无序列依赖性。

2007 年, Li 和 Elledge<sup>[56]</sup> 描述的 SLIC 法则是一种既不需要连接酶也无序列依赖性的组装方法。该法依赖于 T4 DNA 聚合酶的 3'-5' 核酸外切酶活性, 片段经 30 min 的孵育处理, 产生约 30 bp 的序列悬挂区; 相邻片段的悬挂区是互补序列, 经过退火处理, 结合在一起; 加入 *RecA* 使结合稳固, 模拟体内 DNA 修复机制使片段连接在一起。SLIC 法可进行多片段的平行组装, 是构建生物途径的一个得力技术。

#### 4 基因组合成

DNA 组装技术的发展使科学家们能够制造完整的基因, 最终合成微生物的基因组。目前, 对人工合成微生物基因组的研究空前活跃, 并已经取得了一些成果。2002 年, Wimmer 小组合成了脊髓灰质炎病毒的全基因组, 全长约 7.7 kb<sup>[57]</sup>。经实验证明, 这些合成的病毒基因组可以合成与天然病毒完全相同的蛋白质, 而且具有侵染宿主细胞的活力。2003 年, Venter 研究组采用 LCR-PCR 法在两周内从头合成了噬菌体 φX174 基因组, 全长约 5.3 kb<sup>[13]</sup>。2009 年, Venter 研究组又创建了一种恒温的一步法组装技术, 被称作“Gibson 恒温组装法”(图 9), 即利用 5' 核酸外切酶、DNA 聚合酶及连接酶的协调作用在体外将多个带有重叠区的 DNA 片段组装起来<sup>[58]</sup>。采用这种方法, Gibson 等<sup>[58]</sup> 成功地将 4 个超过 100 kb 的片段组装成了约 580 kb 的生殖道支原体基因组。通过实验证明, 该法对于较小 DNA 片段之间的组装也是有效的, 因而同样可用于生物途径的构建。同时, 经过调节几种酶的比例, Gibson 等<sup>[59]</sup> 又将 600 个带有 20 bp 互补序列的 60 bp 的寡核苷酸组装成 16.3 kb 的小鼠线粒体基因。这种技术具有强大的并行性, 可以通过一步反应组装环状质粒或细菌人工染色体。

在利用“Gibson 恒温组装法”合成生殖道支原体基因组的同时, Venter 研究组同时采用改进的 TAR (transformation assisted recombination) 法合成

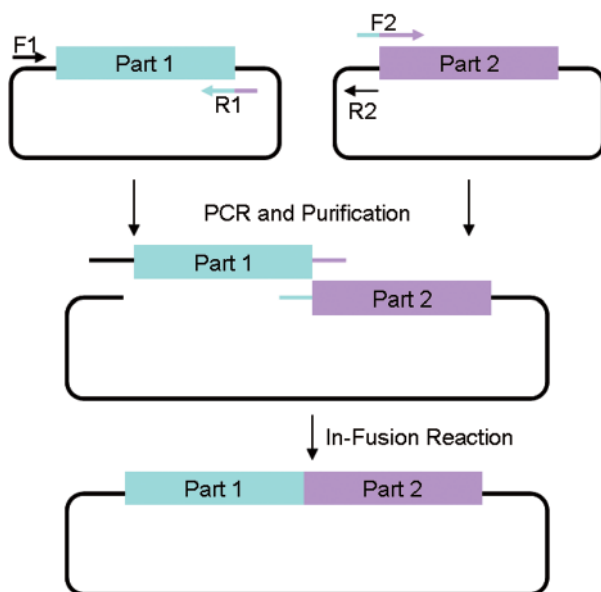


图8 In-Fusion的原理示意图

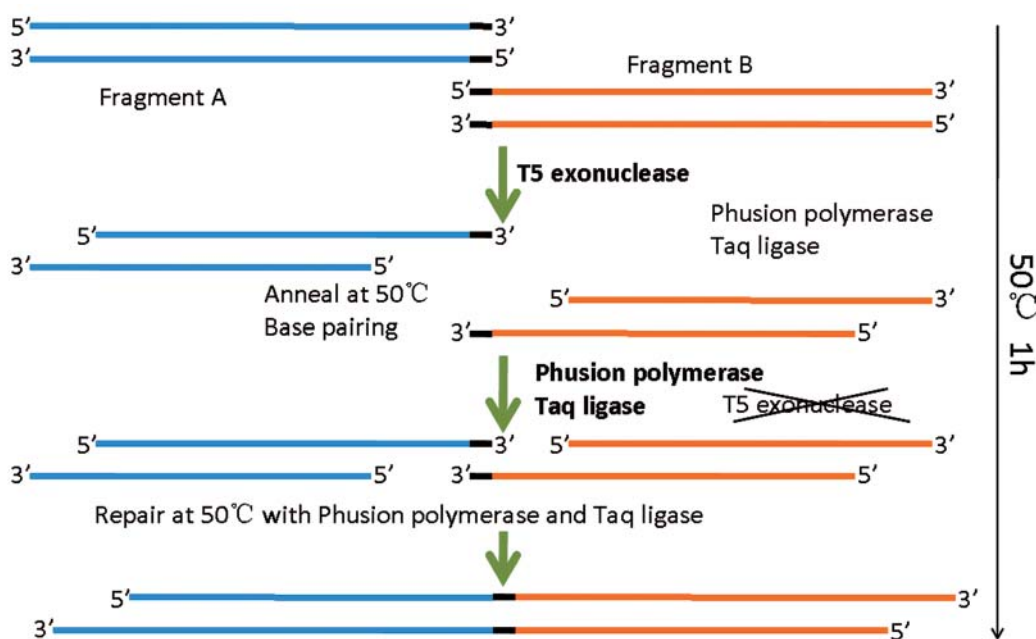


图9 Gibson恒温组装法的原理示意图

该基因组 (图 10)。TAR 是利用酵母体内同源重组来操纵 DNA 的普遍手段, 主要用于将功能基因或生物途径整合到酵母染色体的特定位点。Venter 研究组对其进行了改进, 将 25 个约 24 kb 的片段以及一个线性化的载体组装成一个携带 580 kb 基因组的环形质粒<sup>[58]</sup>。2010 年, Gibson 等<sup>[60]</sup>采用该技术成功合成了 1.08 Mb 的丝状支原体基因组。将该人工合成基因组移植入去除基因组的受体细胞内, 一段时间后, 该细胞完全由合成的基因组控制, 被称之为“合成细胞”。Shao 等<sup>[61]</sup>采用类似的“DNA Assembler”法将两个生物合成途径共 8 个基因构建到一个载体上, 并成功地检测到了代谢产物。酵母体内天然的同源重组酶具有超高的保真性, 因而酵母体内组装具有高度的精确性。

约翰·霍普金斯大学的 Jef Boeke 教授发起了酵母基因组人工合成计划 (Sc. 2.0 计划), 旨在通过人工设计、改造并合成酿酒酵母 16 条染色体序列,

且具有功能性。2011 年, Boeke 教授实验室在 *Nature* 上发表了一篇具有里程碑意义的论文<sup>[62]</sup>。该论文指出, 他们成功地人工合成了世界上首个多功能专门设计的酵母染色体臂, 并且展示了其功能性。自此, 许多国际实验室都加入了这项人工合成整个酵母基因组的研究。本实验室承担第 V 号染色体的合成, 利用 PCA-PCR 法将寡核苷酸合成 942 个约 750 bp 的小片段, 再采用 OE-PCR 法、Gibson 恒温组装法、酵母体内组装法成功将小片段组装成 266 个覆盖 V 号染色体全序列 2.5~3.5 kb 的 DNA 大片段。

在合成基因组的研究中, 酵母并不是唯一可用的底盘细胞。2005 年, Itaya 等<sup>[63]</sup>采用“蠕动延伸法”(IWe) 将 3.5 Mb 的集胞藻 PCC6803 的基因组克隆到 4.2 Mb 的枯草芽孢杆菌基因组中, 形成一个 7.7 Mb 的复合基因组。在该方法的基础上, 2007 年, Itaya 等<sup>[64]</sup>又建立了一种新的利用同源重组的“多米诺法”, 用于合成 16.3 kb 的小鼠线粒体基因组。该方法采用的是一种特殊的枯草芽孢杆菌, 该菌的基因组中整合了 pBR322 序列, 被称为芽孢杆菌基因组宏基因克隆 (BGM)。这种组装是顺序进行的, 需要将待合成的基因组分成带有同源序列的约 5 kb 的片段, 这些片段被交替克隆到带有不同抗性标记的两个质粒上。首先, 将第一个质粒转化到枯草芽孢杆菌中, 第一个片段以及抗性标记则被重

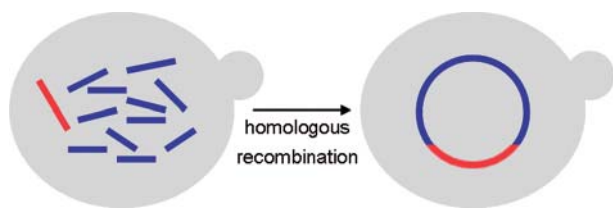


图10 酵母体内DNA组装



组到基因组的 BGM 载体位点；然后进行第二个质粒的转化，第二个片段及抗性标记通过同源重组替换第一个抗性标记；重复上述步骤，直至合成整个基因组。随后，通过在此法中引入更多的抗性标记，134.5 kb 的水稻叶绿体基因组被成功合成。值得一提的是，该基因组具有高重复性，采用其他方法几乎无法合成。然而，无论是“蠕动延伸法”还是“多米诺法”都不能进行快速的平行组装。

## 5 展望

综上所述，DNA 组装技术正处于一个快速发展的阶段；但是，目前并没有一个理想的组装方法同时满足无序列依赖性、无痕或只留痕于不受痕迹序列影响的位置、可按预设顺序进行快速平行组装，并同时适用于基因、途径及基因组的合成。不同的组装技术有不同的适用性，通常的策略是几项技术串联使用。

毫无疑问，未来的 DNA 组装将大大受助于计算机工具。各种计算机软件的发展促进了寡核苷酸的设计和合成，设计、改写、合成基因将更高效、更精确、更廉价。目前，对自动化组装的研究相当活跃，主要包括对策略的理论分析以及计算机辅助设计软件的开发<sup>[65-66]</sup>。不少软件工具已经被应用于实验研究中，而新的组装方法在必要时则需要受助于更多更先进的软件工具。美国生命技术公司 (Life Technologies) 就提供在线的 Gibson 恒温组装法及酵母组装法的设计软件。下一阶段的设计软件应该针对于相邻片段之间顺序依赖性的模块化理解，并应适合于多组装法的混合使用。联合生物能源研究院 (The Joint BioEnergy Institute) 就开发了一个专门针对此目的的软件包 J5，能够结合软件利用机器人进行自动化组装。DNA 组装的自动化将大大提高生产效率，大规模的基因、生物途径、生物网络及基因组的合成让我们更加容易地阐明广阔而深邃的生物机制。

另一个值得考虑的发展方向是体内法，利用微生物体内的甲基化酶和限制性内切酶对 DNA 序列进行编程。目前已有技术被用于对细胞体内基因组进行直接编辑，如 MAGE (multiplex-automated genomic engineering)<sup>[67]</sup>。随着全基因组测序成本的迅速下降，验证细胞内 DNA 的多重变化是可以承受的。

相信在可预见的将来，DNA 合成将会更加快速、高效而廉价。大规模的基因、基因组的合成将成为现实。DNA 组装技术的应用将越来越广泛，

很有可能取代现有的分子克隆技术而被广泛应用于生物、医药、农业、能源、材料及环保等各研究领域。

## [参 考 文 献]

- [1] Agarwal KL, Büchi H, Caruthers MH, et al. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature*, 1970, 227(5253): 27-34
- [2] Itakura K, Hirose T, Crea R, et al. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science*, 1977, 198(4321): 1056-63
- [3] Goeddel DV, Kleid DG, Bolivar F, et al. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc Natl Acad Sci USA*, 1979, 76(1): 106-10
- [4] Ohsuye K, Nomura M, Tanaka S, et al. Expression of chemically synthesized  $\alpha$ -neo-endorphin gene fused to *E. coli* alkaline phosphatase. *Nucleic Acids Res*, 1983, 11(5): 1283-94
- [5] Caruthers MH. Gene synthesis machines: DNA chemistry and its uses. *Science*, 1985, 230(4723): 281-5
- [6] Smith J, Cook E, Fortheringham I, et al. Chemical synthesis and cloning of a gene for human  $\beta$ -urogastrone. *Nucleic Acids Res*, 1982, 10(15): 4467-82
- [7] Mandecki W, Bolling TJ. *FokI* method of gene synthesis. *Gene*, 1988, 68(1): 101-7
- [8] Jayaraman K, Fingar SA, Shah J, et al. Polymerase chain reaction-mediated gene synthesis: synthesis of a gene coding for isozyme c of horseradish peroxidase. *Proc Natl Acad Sci USA*, 1991, 88(10): 4084-8
- [9] Yan J, Li Z, Liu C, et al. Simple and sensitive detection of microRNAs with ligase chain reaction. *Chem Commun*, 2010, 46(14): 2432-4
- [10] Cheng Y, Du Q, Wang L, et al. Fluorescently cationic conjugated polymer as an indicator of ligase chain reaction for sensitive and homogeneous detection of single nucleotide polymorphism. *Anal Chem*, 2012, 84(8): 3739-44
- [11] Stemmer WP, Cramer A, Ha KD, et al. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, 1995, 164(1): 49-53
- [12] Withers-Martinez C, Carpenter EP, Hackett F, et al. PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome. *Protein Eng*, 1999, 12(12): 1113-20
- [13] Smith HO, Hutchison III CA, Pfannkoch C, et al. Generating a synthetic genome by whole genome assembly:  $\phi$ X174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci USA*, 2003, 100(26): 15440-5
- [14] Young L, Dong QH. Two-step total gene synthesis method. *Nucleic Acids Res*, 2004, 32(7): e59
- [15] Kodumal SJ, Patel KG, Reid R, et al. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc Natl Acad Sci USA*, 2004, 101(44): 15573-8
- [16] Xiong AS, Yao QH, Peng RH, et al. A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA



- synthesis method for long gene sequences. *Nucleic Acids Res*, 2004, 32(12): e98
- [17] Xiong AS, Yao QH, Peng RH, et al. PCR-based accurate synthesis of long DNA sequences. *Nat Protoc*, 2006, 1(2): 791-7
- [18] Reisinger SJ, Patel KG, Santi DV. Total synthesis of multi-kilobase DNA sequences from oligonucleotides. *Nat Protoc*, 2007, 1(6): 2596-603
- [19] Qin Q, Liu J, Zhang Z, et al. Isolation, optimization, and functional analysis of the cDNA encoding transcription factor OsDREB1B in *Oryza Sativa* L. *Mol Breed*, 2007, 19(4): 329-40
- [20] Lou XM, Yao QH, Zhang Z, et al. Expression of the human hepatitis B virus large surface antigen gene in transgenic tomato plants. *Clin Vaccine Immunol*, 2007, 14(4): 464-9
- [21] Gao X, Yo P, Keith A, et al. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res*, 2003, 31(22): e143
- [22] Xiong AS, Peng RH, Li X, et al. Influence of signal peptide sequences on the expression of heterogeneous proteins in *Pichia pastoris*. *Acta Biochim Biophys Sin*, 2003, 35(2): 154-60
- [23] Xiong AS, Yao QH, Peng RH, et al. Isolation, characterization, and molecular cloning of the cDNA encoding a novel phytase from *Aspergillus niger* 113 and high expression in *Pichia pastoris*. *J Biochem Mol Biol*, 2004, 37(3): 282-91
- [24] Xiong AS, Yao QH, Peng RH, et al. High level expression of a recombinant acid phytase gene in *Pichia pastoris*. *J Appl Microbiol*, 2005, 98(2): 418-28
- [25] Peng RH, Xiong AS, Li X, et al. PCR-aided synthesis and stable expression in *E. coli* of the cryIA (c) Bt gene. *Acta Biochim Biophys Sin*, 2001, 33(2): 219-24
- [26] Tian J, Gong H, Sheng N, et al. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*, 2004, 432(7020): 1050-4
- [27] Zhou X, Cai S, Hong A, et al. Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Nucleic Acids Res*, 2004, 32(18): 5409-17
- [28] Quan J, Saaem I, Tang N, et al. Parallel on-chip gene synthesis and application to optimization of protein expression. *Nat Biotechnol*, 2011, 29(5): 449-52
- [29] 郭育奇, 赵春燕. 电子基因微芯片在医学研究中的应用. *中国医药生物技术*, 2012, 7(4): 298-301
- [30] 肖翔, 官春云, 尹明智, 等. 基因芯片技术在农业中应用的研究进展. *中国农学通报*, 2012, 28(33): 187-93
- [31] Gibson DG. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res*, 2009, 37(20): 6984-90
- [32] Knight T. Idempotent vector design for standard assembly of BioBricks[EB/OL]. DSpace, 2003, <http://dspace.mit.edu/handle/1721.1/21168>
- [33] Danino T, Mondragón-Palomino O, Tsimring L, et al. A synchronized quorum of genetic clocks. *Nature*, 2010, 463(7279): 326-30
- [34] Friedland AE, Lu TK, Wang X, et al. Synthetic gene networks that count. *Sci Signal*, 2009, 324(5931): 1199-202
- [35] Basu S, Gerchman Y, Collins CH, et al. A synthetic multicellular system for programmed pattern formation. *Nature*, 2005, 434(7037): 1130-4
- [36] He TPY, Campbell A, Zhou SY. PCR-ligation assembly standard for BioBrick parts. DSpace, 2011, <http://dspace.mit.edu/handle/1721.1/67700>[EB/OL]
- [37] Phillips I, Silver P. BioBricks Foundation RFC 23: A new biobrick assembly strategy designed for facile protein engineering. DSpace, 2006, <http://dspace.mit.edu/handle/1721.1/32535>[EB/OL]
- [38] Grünberg R, Arndt K, Müller K. BioBricks Foundation RFC 25: Fusion Protein (Freiburg) Biobrick assembly standard. DSpace, 2009, <http://dspace.mit.edu/handle/1721.1/45140>[EB/OL]
- [39] Anderson JC, Dueber JE, Leguia M, et al. BglBricks: a flexible standard for biological part assembly. *J Biol Eng*, 2010, 4(1): 1-12
- [40] Horton RM, Hunt HD, Ho SN, et al. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, 1989, 77(1): 61-8
- [41] 翟玲, 董娜, 张雅慧, 等. 不需酶切位点的重叠延伸 PCR 在克隆表达中的应用. *南昌大学学报:理科版*, 2012, 36(3): 273-6
- [42] Luo LN, Wang S, Wang Y. Site-directed mutagenesis based on overlap extension PCR. *Agricul Sci Technol: Hunan*, 2012, 13(4): 719-22
- [43] Wei H, Hu J, Wang L, et al. Rapid gene splicing and multi-sited mutagenesis by one-step overlap extension PCR. *Anal Biochem*, 2012, 429(1): 76-8
- [44] Quan J, Tian J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One*, 2009, 4(7): e6441
- [45] Bryksin AV, Matsumura I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques*, 2010, 48(6): 463-5
- [46] Chen WH, Qin ZJ, Wang J, et al. The MASTER (methylation-assisted tailorable ends rational) ligation method for seamless DNA assembly. *Nucleic Acids Res*, 2013, 41(8): e93
- [47] Lou C, Liu X, Ni M, et al. Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch. *Mol Syst Biol*, 2010, 6(1): 350
- [48] Lee SK, Chou H, Ham TS, et al. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr Opin Biotechnol*, 2008, 19(6): 556-63
- [49] Szybalski W, Kim SC, Hasan N, et al. Class-IIS restriction enzymes—a review. *Gene*, 1991, 100: 13-26
- [50] Kodumal SJ, Patel KG, Reid R, et al. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc Natl Acad Sci USA*, 2004, 101(44): 15573-8
- [51] Engler C, Kandzia R, Marillonnet S. A one pot, one step,

- precision cloning method with high throughput capability. PLoS One, 2008, 3(11): e3647
- [52] Blake WJ, Chapman BA, Zindal A, et al. Pairwise selection assembly for sequence-independent construction of long-length DNA. Nucleic Acids Res, 2010, 38(8): 2594-602
- [53] Sasaki Y, Sone T, Yoshida S, et al. Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the Multisite Gateway system. J Biotechnol, 2004, 107(3): 233-43
- [54] Sleight SC, Bartley BA, Lieviant J A, et al. In-Fusion BioBrick assembly and re-engineering. Nucleic Acids Res, 2010, 38(8): 2624-6
- [55] Nour-Eldin HH, Geu-Flores F, Halkier BA. USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories. Methods Mol Biol, 2010, 643: 185-200
- [56] Li MZ, Elledge SJ. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. Nat Methods, 2007, 4(3): 251-6
- [57] Cello J, Paul AV, Wimmer E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science, 2002, 297(5583): 1016-8
- [58] Gibson DG, Young L, Chuang RY, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods, 2009, 6(5): 343-5
- [59] Gibson DG, Smith HO, Hutchison III CA, et al. Chemical synthesis of the mouse mitochondrial genome. Nat Methods, 2010, 7(11): 901-3
- [60] Gibson DG, Glass JI, Lartigue C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science, 2010, 329(5987): 52-6
- [61] Shao Z, Zhao H. DNA assembler: a synthetic biology tool for characterizing and engineering natural product gene clusters. Methods Enzymol, 2011, 517: 203-24
- [62] Dymond JS, Richardson SM, Coombes CE, et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature, 2011, 477(7365): 471-6
- [63] Itaya M, Fujita K, Kuroki A, et al. Bottom-up genome assembly using the *Bacillus subtilis* genome vector. Nat Methods, 2007, 5(1): 41-3
- [64] Itaya M, Tsuge K, Koizumi M, et al. Combining two genomes in one cell: stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. Proc Natl Acad Sci USA, 2005, 102(44): 15971-6
- [65] Cai Y, Hartnett B, Gustafsson C, et al. A syntactic model to design and verify synthetic genetic constructs derived from standard biological parts. Bioinformatics, 2007, 23(20): 2760-7
- [66] Densmore D, Hsiao THC, Kittleson JT, et al. Algorithms for automated DNA assembly. Nucleic Acids Res, 2010, 38(8): 2607-16
- [67] Wang HH, Isaacs FJ, Carr PA, et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature, 2009, 460(7257): 894-8