

Research
Synthetic Biology—Article

DNA 组装效率的无偏差快速检验——基于 qPCR 而不依赖于转化的方法

马晓焉^{a,#}, 梁昕鑫^{a,#}, 霍毅欣^{a,b,*}^a Key Laboratory of Molecular Medicine and Biotherapy, School of Life Science, Beijing Institute of Technology, Beijing 100081, China^b UCLA Institute for Technology Advancement (Suzhou), Suzhou 215123, China

ARTICLE INFO

Article history:

Received 1 February 2018

Revised 29 July 2018

Accepted 19 March 2019

Available online 26 June 2019

关键词

组装效率

DNA 组装

qPCR

二级结构

转化

摘要

合成生物学正朝着大规模、复杂化的方向发展,这在很大程度上依赖于基因模块的高效组装。传统评估DNA组装效率(AE)的方法需要进行转化,整个过程耗时长达10 h,而且容易受到各种因素的干扰。为快速、可靠地测定组装效率,本研究建立了一种基于qPCR的不依赖于转化的测定方法,用连接上的片段占初始加入片段的比例表征组装效率,3 h即可完成测定。利用该方法测定了酶切连接、Golden Gate组装以及Gibson组装法的双片段或多片段组装效率,所得结果与菌落计数法表征的组装效率呈显著正相关。该方法消除了转化过程的随机性,降低了测定偏差,优于传统的菌落计数法。随后,用此方法研究了DNA片段末端的二级结构对组装效率的影响。结果显示,所有依赖于末端序列互补的组装技术,其组装效率主要受重叠序列整体性质的影响,而发夹结构可显著降低组装效率。这种基于qPCR的测定方法将促进DNA组装技术的发展,并有助于对组装效率影响因素的评估。

© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

合成生物学旨在通过改变生物行为或创造新的生命形式以实现全新的生物功能[1]。这门学科以DNA和蛋白质等生物分子为基础,遵循从生物反应、代谢途径、个体细胞到最终种群的层级发展规律[1,2]。然而,向更复杂的生物系统前行时面临两大挑战:首先是人工生物组件的多样性问题[3],目前建库时可用于自由组装的基因片段数目($10^3\sim 10^5$ 个)低于理想水平[4,5];第二个挑战是多个遗传模块的整合[6],目前仍难以对10个以上的生物模块进行组装[7]。DNA组装技术为应对这些挑战提供了更加简单高效的途径[8]。

组装效率(AE)反映了DNA片段发生组装的概率,

是评估组装方法的关键参数,通常由转入连接产物后平板上形成的菌落数表征[9,10]。组装效率的测定开始于转化,该步骤需要1~1.5 h,然后过夜培养至少8 h以形成菌落。最后,还需烦琐的计数及验证过程。除去感受态细胞的准备时间,以上过程总计消耗约10 h。因此,组装效率的快速测定方法将大大加速组装技术的检验与优化过程。

DNA分子的组装技术已从依赖于特定位点的双片段组装发展到不依赖于位点特异性的多片段组装[11]。现有的组装方法如体内组装(IVA)[12]、双引物组装(TPA)[13]、不依赖基因序列的连接反应克隆(SLIC)[10]以及环状聚合酶延伸克隆(CPEC)[9]都具有较高的效率,每单位DNA可形成400~56 000个菌落。然而,这些数据

* Corresponding author.

E-mail address: huoyixin@bit.edu.cn (Y.-X. Huo).

These authors contributed equally to this work.

DNA片段的末端，PCR结束后用DpnI消除模板质粒。在50 μL 的反应体系中，加入1 U的两种内切酶，切割1 μg 插入片段及1 μg 载体片段。此处1 U的定义为，在37 $^{\circ}\text{C}$ 50 μL 反应体系中，消化1 μg λDNA 所需的限制性内切酶的量。反应体系在37 $^{\circ}\text{C}$ 温浴2 h，然后用胶回收试剂盒（Thermo Fisher Scientific Inc.）对片段进行纯化。在20 μL 反应体系中加入400 CEU（黏性末端单位）的T4 DNA连接酶及60~350 ng的DNA片段（插入片段与载体片段的物质的量之比为1:1或5:1），22 $^{\circ}\text{C}$ 下反应1 h。所用酶均来自NEB。

2.2. Golden Gate 组装

本实验一共测定了8组Golden Gate样品的组装效率。通过PCR扩增将BbsI限制性酶切位点添加到待组装DNA片段末端。在总体积为20 μL 的反应体系中加入2 μL 10 \times T4连接酶缓冲液、200 CEU T4 DNA连接酶、5 U BbsI酶以及200~500 ng插入片段与载体片段（物质的量之比分别为1:1、3:1或5:1）。反应进行30个循环，每个循环的程序为37 $^{\circ}\text{C}$ 5 min及22 $^{\circ}\text{C}$ 5 min。循环结束后55 $^{\circ}\text{C}$ 15 min失活限制酶，并在85 $^{\circ}\text{C}$ 反应15 min失活T4连接酶。所用酶和缓冲液均来自NEB（New England Biolabs, Inc.）。

2.3. Gibson 组装

用PCR扩增将引物上的重叠序列添加至待组装片段的末端。一般两个DNA片段的组装需引入20个碱基对（bp）的重叠序列，多片段组装需将重叠区域扩展到40 bp。将各50 fmol的载体片段和插入片段加入到实验室自制的11.25 μL Gibson混合体系，其中，含0.06 U T5外切核酸酶、0.375 U Phusion[®]高保真DNA聚合酶、60 CEU Taq DNA连接酶、13.3 $\text{mmol}\cdot\text{L}^{-1}$ MgCl_2 、13.3 $\text{mmol}\cdot\text{L}^{-1}$ 1,4-二硫苏糖醇（DTT）和5 \times 等温反应缓冲液。每1 \times 等温缓冲液含pH为7.5的0.1 $\text{mol}\cdot\text{L}^{-1}$ 三羟甲基氨基甲烷-盐酸盐（Tris-HCl）、0.2 $\text{mmol}\cdot\text{L}^{-1}$ 脱氧核糖核苷三磷酸（dNTP）、1 $\text{mmol}\cdot\text{L}^{-1}$ 烟酰胺腺嘌呤二核苷酸（NAD）、50 $\text{g}\cdot\text{L}^{-1}$ 聚乙二醇（PEG）8000。总体积为15 μL 的反应体系在50 $^{\circ}\text{C}$ 温育1 h（双片段组装）或2.5 h（多片段组装）。所用酶均来自NEB。本次实验共测定了6组样品。

2.4. 转化

转化所用的感受态细胞是实验室制备的转化效率为

每微克pUC19 DNA 7×10^6 CFU（ $\text{CFU}\cdot\mu\text{g}^{-1}$ ）的大肠杆菌菌株XL10-Gold（Integrated Science & Technology, Inc.）。将1~6 μL 组装产物加入50 μL 感受态细胞中，在950 μL SOC培养液中复苏培养1 h后涂板。如果插入片段为紫色蛋白（iGEM号：BBa_K1033906）编码基因，则平板上紫色菌落为阳性菌。对于没有指示基因的组装产物，随机挑选10个单菌落，用菌落PCR进行验证。组装效率由每微克DNA所得的阳性CFU表征。

2.5. 定量 PCR

在定量PCR之前，用0.5 μL （5 U）T5外切核酸酶消化组装体系中残余的线性片段，步骤如下：37 $^{\circ}\text{C}$ 消化15 min，然后85 $^{\circ}\text{C}$ 温育15 min失活该酶。定量反应的总体积为20 μL ，其中含1.5 μL 组装产物（Supplementary data, Table S1）。所用定量试剂来自于TaKaRa（SYBR[®] Premix DimerEraser[™]）。将等量DNA片段作为定量反应的参比，该参比DNA无需T5外切核酸酶处理。将T5消化后的线性载体片段与插入片段的混合物用作空白对照试验的模板。用具有SYBR Green I检测通道的Roche LightCycler[®] 96系统进行定量。反应步骤依次为：95 $^{\circ}\text{C}$ 预变性30 s，然后循环以下三步：95 $^{\circ}\text{C}$ 5 s、56 $^{\circ}\text{C}$ 30 s、72 $^{\circ}\text{C}$ 30 s，循环40次。以0.1 $^{\circ}\text{C}\cdot\text{s}^{-1}$ 的升温速率从60 $^{\circ}\text{C}$ 升至95 $^{\circ}\text{C}$ ，每升高1 $^{\circ}\text{C}$ 进行5次信号采集，用于溶解曲线分析。每个样品设置两个技术重复。组装效率的计算参考 ΔC_t （目标基因与参比基因循环阈值的差值）定量法[22]。对于目标DNA和参比DNA循环阈值的详细描述如下：

$$X_t = X_0 \times E_X^{C_{t,X}} \quad (1)$$

$$R_t = R_0 \times E_R^{C_{t,R}} \quad (2)$$

式中， X_0 与 R_0 分别代表目标DNA与参比DNA； E_X 代表目标DNA的扩增效率， E_R 代表参比DNA的扩增效率； $C_{t,X}$ 代表目标DNA的循环阈值， $C_{t,R}$ 代表参比DNA的循环阈值。相应的， X_0 与 R_0 的比值表示如下：

$$\frac{X_0}{R_0} = \frac{X_t}{R_t} \times \frac{E_R^{C_{t,R}}}{E_X^{C_{t,X}}} = R \quad (3)$$

本方法同时对组装产物及线性片段中的目标序列进行测定。因此，可以认为 X_t 与 R_t 的比值是一个接近于1的常数。此外， ΔC_t 计算方式假定目标DNA和参比DNA扩增效率相同。而本方法所靶向的目标DNA和参比DNA

一致，故本假设更加可信。理想状况下，PCR每循环一次产物加倍，即扩增效率为2。因此，式(3)可以转变为：

$$R = 2^{-\Delta C_t} \quad (4)$$

式中， R 代表已组装的片段与初始片段含量的比值； ΔC_t 即目标DNA与参比DNA循环阈值之差 ($C_{t,X} - C_{t,R}$)，即本实验中已组装片段与初始线性片段之间 C_t 的差值。

2.6. 数据分析

采用皮尔逊双尾相关系数来研究基于qPCR与CFU方法所测得的组装效率间的相关性。组装效率之间的差异用独立样本 t 检验分析。用增强树法分析重叠区DNA性质与Gibson组装效率之间的关系，该分析采用R (2.7.2) 与gbmplus (1.5-17) 软件包完成[23]。

3. 实验结果

3.1. 不同方法测定的酶切连接效率

基于CFU的测定结果显示，7组样品的组装效率从 $0.5 \times 10^3 \text{ CFU} \cdot \mu\text{g}^{-1}$ 到 $1.6 \times 10^4 \text{ CFU} \cdot \mu\text{g}^{-1}$ 不等[图2 (a)]。第6组和第7组连接产物的大小为2.1 kb，平均组装效率比

其他组（连接产物的大小在5.5~8.7 kb之间）高11倍。该结果与用qPCR测定的结果非常类似，基于qPCR得到的比例在0.021~0.209之间[图2 (a)]。但是，组装效率最高的组（RL7）和最低的组（RL1）之间的差异从基于CFU数所得的35倍下降到由qPCR计算的10倍。CFU数之间的显著差异可能是由转化效率导致的，从RL7至RL1，组装产物的大小增加了4倍，这会降低转化效率。因此，对于RL1等大分子组装，由CFU数所表征的组装效率可能会低于由qPCR比率所表征的实际值。整体而言，两种测定方法所得出的组装效率呈正相关[图2 (d)]。

3.2. 不同方法测定的 Golden Gate 组装效率

两种方法测得的8组样品的组装效率从GG1到GG8逐渐增大[图2 (b)]。用CFU法测得的效率从 $4 \times 10^3 \text{ CFU} \cdot \mu\text{g}^{-1}$ 增加到 $2.6 \times 10^4 \text{ CFU} \cdot \mu\text{g}^{-1}$ ，增幅为6.5倍，而用qPCR得到的比率从0.013增加到0.072，增幅为5.5倍。基于qPCR的测定结果显示，GG3~GG5的连接效率相比于GG1或GG2提高了4倍，这种显著增加的趋势持续到GG6，而GG7和GG8则趋于稳定。相比之下，用CFU方法测得的组装效率从GG7到GG8显著增加，而从GG1

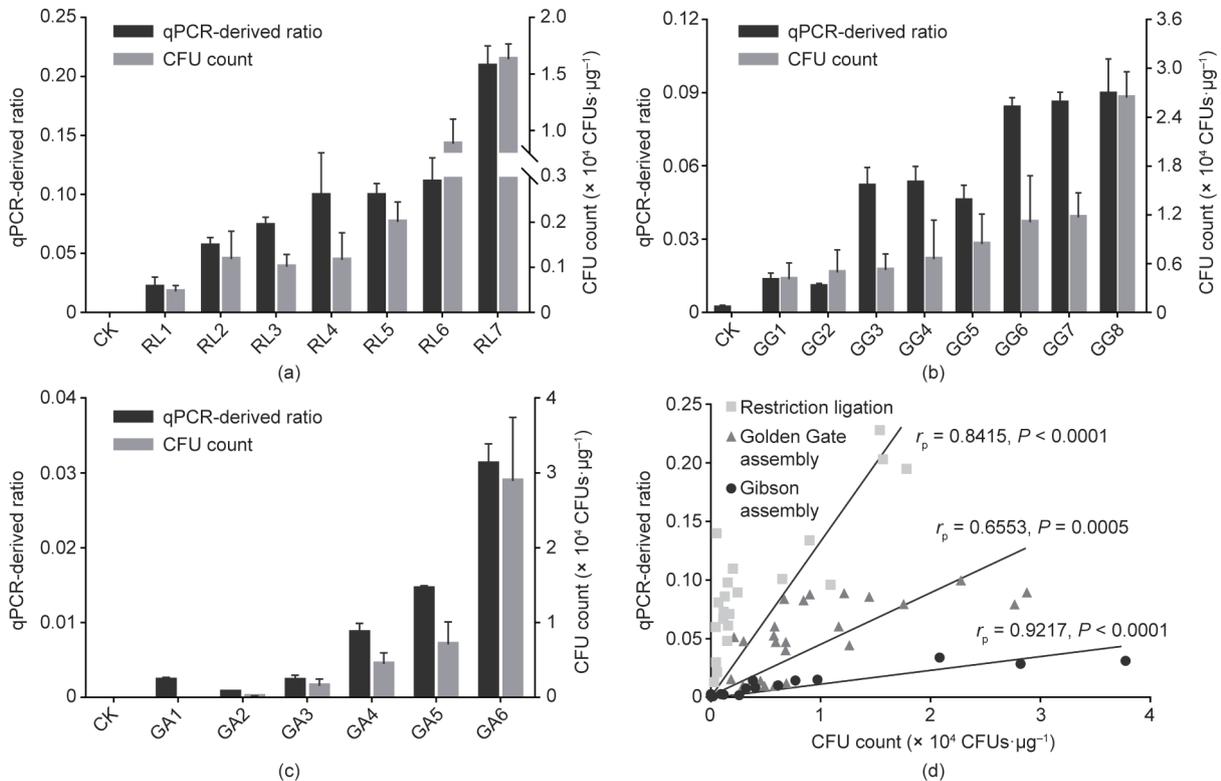


图2. 用CFU数及qPCR比率测定各组装技术的组装效率。(a) 酶切连接；(b) Golden Gate组装；(c) Gibson组装；(d) 两种测定方法之间的相关性。CK表示经T5外切酶消化后残留的线性片段。基于qPCR法及CFU法得到的测定结果用各自的平均值表示，误差线代表标准偏差 ($n=3$)。 r_p : 皮尔逊相关系数。

到GG6则增幅不明显。后者可能是由于转化过程中存在较大偏差。即使严格控制DNA分子的大小及转化条件,转化的误差也不可避免。尽管存在以上差异,基于qPCR的测定结果与CFU方法所得的结果也呈正相关[图2(d)]。

3.3. 不同方法测定的 Gibson 组装效率

根据CFU测定的结果可知,6组DNA片段(GA1~GA6)的组装效率从0到 2.9×10^4 CFU $\cdot\mu\text{g}^{-1}$ 不等[图2(c)]。组装产物的片段越大,组装效率越低。6组样品的组装效率用qPCR所得比率与CFU测得的结果排序一致,两组数据呈正相关[图2(d)]。值得注意的是,在以上几组样品中组装产物的基因片段最大的GA1没有形成单克隆,但是用qPCR得到的值为0.002,远高于用T5消化的空白对照。以上结果可能是由于DNA片段太大对转化不利而导致的。

此方法也被用于测定多个DNA片段的组装效率。当组装片段的数目从两个(M2)增加到三个(M3)时,基于qPCR得到的值从0.028大幅下降到0.015,下降了48%(图3)。当组装四个片段时,用qPCR比值表征的组装效率持续下降到0.007,下降了55%。CFU数量的变化与基于qPCR所得比值的变化趋势一致。相较于两个片段的组装,三个片段组装的CFU数下降了54%。而当组装四个片段时,CFU数则继续减少84%($1.6 \times 10^4 \sim 3 \times 10^3$ CFU $\cdot\mu\text{g}^{-1}$)。总之,以上结果均证实了当组装多个片段时,组装效率会下降。

3.4. 重叠区域二级结构对 Gibson 组装效率的影响

为了研究重叠区二级结构对Gibson组装效率的影响,本研究设计了11组会生成二级结构的20 bp短序列。

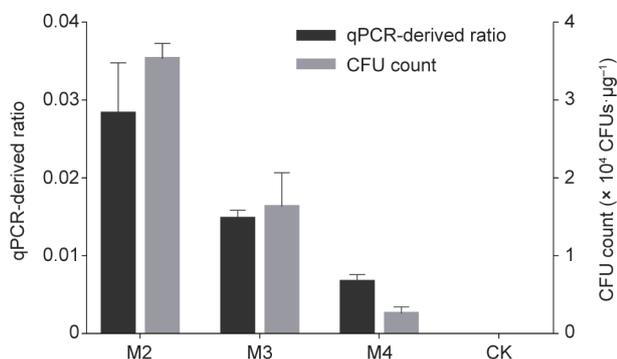


图3. 用CFU数和qPCR比率测定的双片段和多片段的组装效率。以上两种方法对每个样品的测定结果用各自的平均值表示,误差线代表标准偏差($n=3$)。

二级结构的稳定性由自由能(ΔG)表征[24,25], ΔG 越低代表二级结构稳定性更高[26]。11组序列中,会形成发夹结构的序列的 ΔG 从 -1.2 kcal $\cdot\text{mol}^{-1}$ 到 -9.2 kcal $\cdot\text{mol}^{-1}$ (1 kcal=4186 J)不等,而相同序列间生成二聚体的 ΔG 从 -3.3 kcal $\cdot\text{mol}^{-1}$ 到 -11.6 kcal $\cdot\text{mol}^{-1}$ 。此外,本实验还设计了两组没有二级结构($\Delta G=0$)的序列,并将这些短序列引入线性pUC19载体及紫色蛋白编码基因的插入片段(长约1 kb)末端(Supplementary data, Table S2),如图1(d)所示,总共有12对基因片段。其中,一对片段(OL1,作为对照)在重叠区域不含二级结构,9对(OL2~OL10)只在一个重叠区域有二级结构,其余两对(OL11和OL12)在两个重叠区域都有二级结构。同一管组装产物同时用于两种方法的测定。

结果显示,从OL1到OL6,组装效率基本呈增加趋势,而从OL7到OL12,则急剧下降(图4)。CFU数从OL1(没有二级结构)的 4×10^3 CFU $\cdot\mu\text{g}^{-1}$ 增加到OL6(发夹结构的 ΔG 为 -5.3 kcal $\cdot\text{mol}^{-1}$)的 1.1×10^4 CFU $\cdot\mu\text{g}^{-1}$ 。当发夹结构的 ΔG 下降至 -6.9 kcal $\cdot\text{mol}^{-1}$ (OL7)时,CFU数显著下降了89%。样品(OL10)在重叠区域的发夹结构的 ΔG 最低,CFU数也最少,为 2×10^2 CFU $\cdot\mu\text{g}^{-1}$,只有OL1的5%。而基于qPCR的数据显示,OL6比OL1有更高的AE,OL3最高(比率为0.156)。从OL6到OL7因二级结构更加稳定,其比率从0.136下降到0.052,降幅达到62%,到OL10下降至最低水平。虽然从 ΔG (从 -4.5 kcal $\cdot\text{mol}^{-1}$ 到 -6.4 kcal $\cdot\text{mol}^{-1}$)来看,两个重叠区域都有二级结构的两组样品(OL11和OL12)在所有样品中的结构稳定性处于中等水平,但相比于OL1(0.032),组装效率显著下降。

由 ΔG 表征的末端二级结构并不是影响Gibson组装的唯一因素。重叠序列的整体性质比二级结构的稳定性

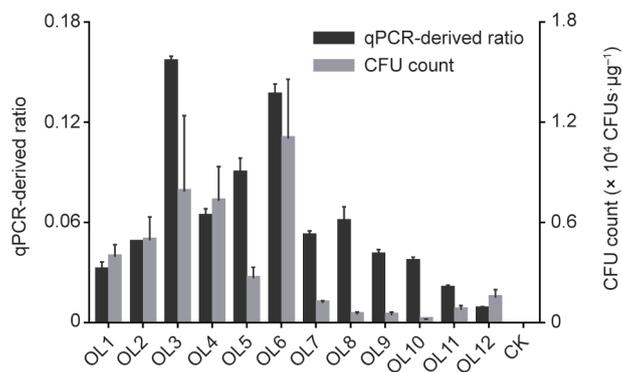


图4. 用CFU数及qPCR比率测定重叠区含有二级结构的DNA片段的组装效率。以上两种方法对每个样品的测定结果用各自的平均值表示,误差线代表标准偏差($n=3$)。

对组装效率有更大的影响(图5)。从12组Gibson组装产物的组装效率数据生成的模型可得,重叠区域中退火温度(T_m)和鸟嘌呤与胞嘧啶(GC)的含量是影响组装效率的主要因素(占总差异的56%)。组装效率随GC含量(50%~70%)的增加而升高[图5(c)],且当 T_m 值高于60 °C时,组装效率将显著提高[图5(b)]。相比于二聚体结构,重叠区域内部形成的发夹结构对组装效率的影响更大。详细来讲,序列自身形成发夹结构的 T_m 、 ΔG 及GC含量对组装效率的影响占31%,而序列之间形成二聚体的以上参数对组装效率的影响仅占13%。一旦发夹结构的 ΔG 降至 $-4 \text{ kcal}\cdot\text{mol}^{-1}$ 时,组装效率会大幅度下降[图5(d)]。当发夹结构的 T_m 为22~30 °C时,组装效率与 T_m 之间呈负相关[图5(e)]。当发夹结构的 T_m 为22~30 °C时,组装效率与 T_m 之间呈负相关[图5(e)]。

4. 讨论

本研究采用qPCR方法,建立了一种不依赖于转化的快速测定组装效率的方法。用该方法测定了酶切连接、Golden Gate组装以及Gibson组装的组装效率,与传统的基于CFU数得到的结果具有可比性,有助于探究组装效率的决定因素。

使用基于qPCR方法进行组装效率测定可以排除转化过程中各种因素的干扰,从而比传统的基于CFU的方法更加可靠。在这项研究中,由qPCR比率所表征的组装效率的平均标准偏差如下:酶切连接为14%,Golden Gate组装为13%,Gibson组装为17%(图2)。相比之下,即使采用同一批次的感受态细胞,且严格控制转

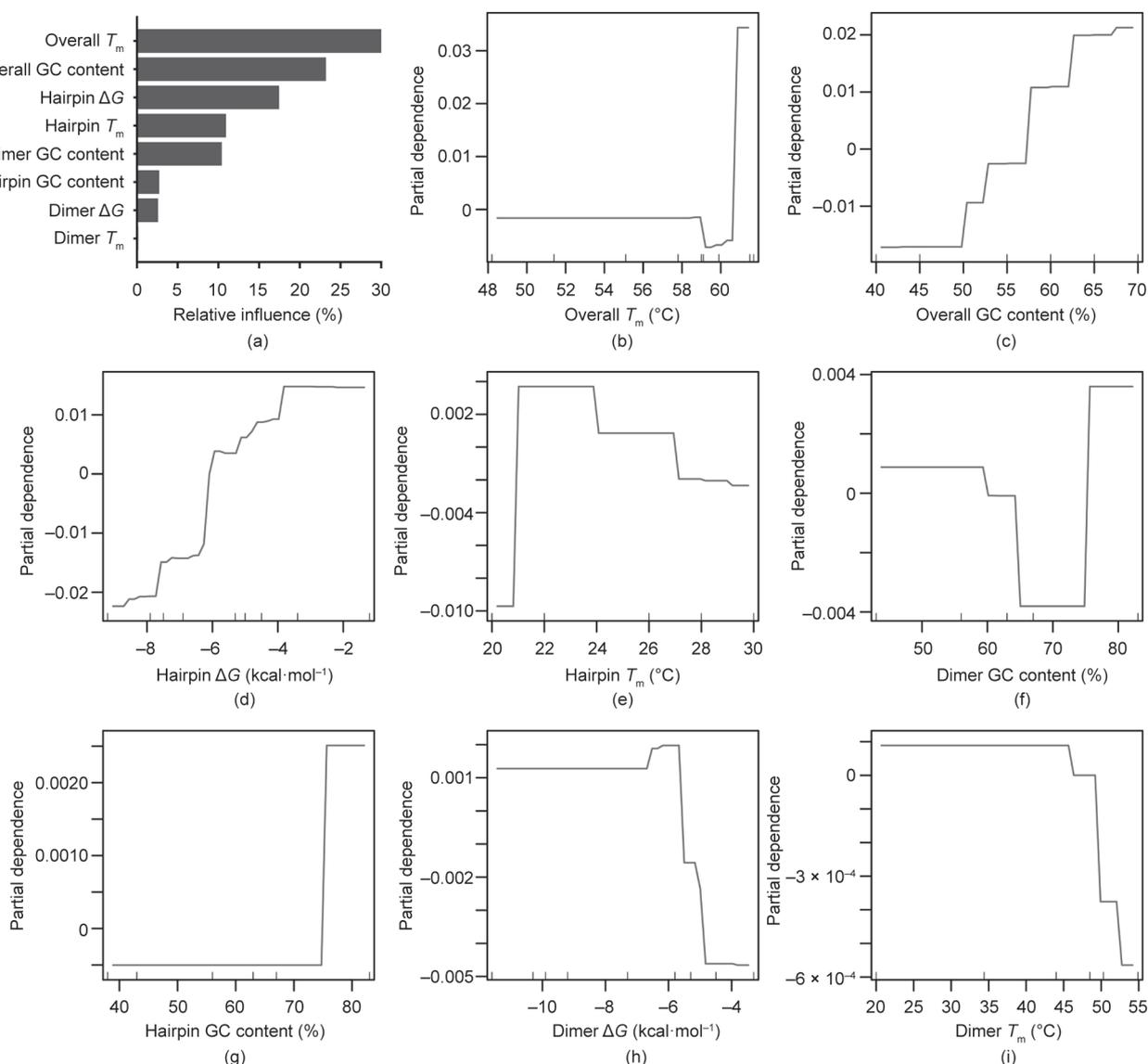


图5. (a) 末端二级结构及重叠区域的整体和局部性状对组装效率的相对影响; (b) 整体 T_m 的影响; (c) 整体GC含量的影响; (d) 发夹结构 ΔG 的影响; (e) 发夹结构 T_m 的影响; (f) 二聚体GC含量的影响; (g) 发夹结构GC含量的影响; (h) 二聚体 ΔG 的影响; (i) 二聚体 T_m 的影响。

化条件，由CFU数得到的组装效率的平均偏差仍高达31%~48%，比对应的qPCR比率的测量偏差超出2.2~3.7倍。转化法重复性较差，这可能掩盖了GG6和GG5等样品之间的差异，而用qPCR检测的结果显示GG6的组装效率明显高于GG5。此外，qPCR方法也可以避免片段大小带来的测定偏差。已有一些研究报道，转入片段越大，转化的成功率越低[27,28]，从而会低估组装效率。这为两种方法测得GA1的组装效率不一致提供了可能的解释[图2(c)]。由于GA1得到的10 kb组装产物无法成功转化，组装效率只能由其他不依赖于转化的方式检测。对于RL7和RL1，基于CFU得到的组装效率之间的差异相较于qPCR方法更显著，这种偏差可能也是由片段大小导致的。因此，qPCR的方法比传统基于CFU的方法具有更高的灵敏度和可靠性。

在比较不同样品间的组装效率时，应尽可能降低随机误差。用依赖于转化的CFU方法耗时耗力。首先，为排除在DNA富集和稳定性方面的差异，应该用同一基因型的感受态细胞；其次，转化效率必须保持稳定；为了确保这一点，感受态细胞应该来源于同一批次，并且转化必须遵循相同的步骤，例如，保持感受态细胞解冻时间一致等，这些步骤的偏差可能会导致转化效率高达10000倍的差异[29]。即使严格控制以上条件，组装效率仍然会受胞内干扰因素的影响，例如，含有重复序列DNA分子的重组[30,31]。相比之下，基于qPCR的检测方法可以通过遵循qPCR的基本规则，从而更容易地控制检测条件。而且对于目标DNA和参比DNA，定量相同的序列可进一步简化工作流程。因此，定量检测只需要一对引物，从而降低了目标DNA和参比DNA之间扩增效率的差异。需要注意的是，组装产物及参比基因的浓度应保持一致。此外，组装产物混合体系中剩余的线性片段必须彻底消化，外切核酸酶需彻底灭活，以避免其继续消化定量引物。遵循以上规则，即可实现快速测定组装效率。通常需要用30 min消化残留片段，15 min失活外切核酸酶，2 h准备和运行qPCR。仅需不足3 h即可快速测定组装效率，占CFU方法所需时长的四分之一。

本研究建立的基于qPCR的测定方法并不是为了取代转化，而是提供一种可靠、快速地测定组装效率的方法。因此，这种方法尤其适用于DNA组装技术的开发。当对测试通量和数据可靠性要求较高时，可以评估低效组装的真正原因。通过一次3 h的qPCR即可同时检测如重叠区域长度、片段之间的比例、DNA分子的数量和组装时间等因素对组装效率的影响，从而确定转化的最佳

条件。体外测试只针对组装效率，而非转化步骤和组装过程的综合效率，有助于解析转化子数量偏低的多种原因。例如，如果用qPCR得到的组装效率的值高于CFU测得的结果，很可能由低效的转化效率导致的。如果提升转化效率仍不能得到更多的转化子，可能表明插入的基因片段表达的蛋白是有毒的[32,33]，此问题可以通过换用一个严紧型表达载体以防止毒性基因的泄漏表达得以解决。此外，DNA组装的阳性率可以通过定量法来确定。qPCR的定量引物靶向组装产物的接合处，而非单个待组装片段，因此，只有当片段按照正确的顺序进行组装时才能被检测到。基于这种假设，用已组装片段与待组装游离片段之间的比值来表征组装产物的阳性率[Fig. S1(a)]。本实验用以上方法成功表征了五组Gibson组装阳性率的差异[Supplementary data, Fig. S1(b)]。

基于qPCR的测定方法揭示了末端二级结构对Gibson组装效率的影响。一般而言，在重叠区域内部形成稳定的发夹结构时，组装效率会显著降低。但是，应该注意到，重叠区域的整体性质，如碱基组成和热稳定性，可能比局部的二级结构对DNA组装有更大的影响。具体而言， T_m 较高或GC丰富的重叠序列可能更有助于同源末端结合，组装效率更高。在本研究中，重叠区域的 T_m 值范围为48~62 °C，总GC含量为40%~70%，基本覆盖了通常DNA组装所涉及的范围[9,12,13,15]。因此，为提高Gibson组装的成功率，除了确保GC含量和 T_m 在正常的范围，还应尽可能引入更多的鸟嘌呤和胞嘧啶。重叠区域内部形成的二级结构对于DNA组装影响较弱，但应避免高GC回文序列等可以形成稳定发夹结构的序列。这些规则也适用于SLIC、TPA、IVA，以及其他依赖于同源末端互补的DNA组装技术。

5. 结论

综上所述，本研究建立了一种基于qPCR测定DNA片段组装效率的方法。该结果与常用的CFU法测得的组装效率具有可比性。qPCR法降低了转化法测量过程中出现的系统和随机误差，因而优于转化法。这种快速检测的流程可用于DNA组装技术的开发以及对组装效率影响因素的快速评估。

致谢

本研究得到了国家重点研发计划

(2017YFD0201400)、国家自然科学基金 (21676026), 以及中央高校基本科研业务费的支持。

Compliance with ethics guidelines

Xiaoyan Ma, Xinxin Liang, and Yi-Xin Huo declare that they have no conflicts of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eng.2019.06.002>.

References

- [1] Andrianantoandro E, Basu S, Karig DK, Weiss R. Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol* 2006;2(1):2006.0028.
- [2] Choffnes ER, Relman DA, Pray L. The science and applications of synthetic and systems biology: workshop summary. Report. Washington, DC: National Academies Press; 2011.
- [3] Ramon A, Smith HO. Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. *Biotechnol Lett* 2011;33(3):549–55.
- [4] Wingler LM, Cornish VW. Reiterative recombination for the in vivo assembly of libraries of multigene pathways. *Proc Natl Acad Sci USA* 2011;108(37):15135–40.
- [5] Dietrich JA, McKee AE, Keasling JD. High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu Rev Biochem* 2010;79(1):563–90.
- [6] Purnick PEM, Weiss R. The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol* 2009;10(6):410–22.
- [7] Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, et al. Genetic circuit design automation. *Science* 2016;352(6281):aac7341.
- [8] Chao R, Yuan Y, Zhao H. Recent advances in DNA assembly technologies. *FEMS Yeast Res* 2015;15(1):1–9.
- [9] Quan J, Tian J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One* 2009;4(7):e6441.
- [10] Li MZ, Elledge SJ. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods* 2007;4(3):251–6.
- [11] Cobb RE, Ning JC, Zhao H. DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J Ind Microbiol Biotechnol* 2014;41(2):469–77.
- [12] García-Nafria J, Watson JF, Greger IH. IVA cloning: a single-tube universal cloning system exploiting bacterial in vivo assembly. *Sci Rep* 2016;6(1):27459.
- [13] Liang J, Liu Z, Low XZ, Ang EL, Zhao H. Twin-primer non-enzymatic DNA assembly: an efficient and accurate multi-part DNA assembly method. *Nucleic Acids Res* 2017;45(11):e94.
- [14] Jin P, Ding W, Du G, Chen J, Kang Z. DATEL: a scarless and sequence-independent DNA assembly method using thermostable exonucleases and ligase. *ACS Synth Biol* 2016;5(9):1028–32.
- [15] Zhang Y, Werling U, Edlmann W. SLICE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res* 2012;40(8):e55.
- [16] Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* 2008;3(11):e3647.
- [17] Fu C, Donovan WP, Shikapwashya-Hasser O, Ye X, Cole RH. Hot fusion: an efficient method to clone multiple DNA fragments as well as inverted repeats without ligase. *PLoS One* 2014;9(12):e115318.
- [18] Yoshida N, Sato M. Plasmid uptake by bacteria: a comparison of methods and efficiencies. *Appl Microbiol Biotechnol* 2009;83(5):791–8.
- [19] Aune TEV, Aachmann FL. Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. *Appl Microbiol Biotechnol* 2010;85(5):1301–13.
- [20] Sleight SC, Bartley BA, Lieviant JA, Sauro HM. In-Fusion BioBrick assembly and re-engineering. *Nucleic Acids Res* 2010;38(8):2624–36.
- [21] Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 2009;6(5):343–5.
- [22] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using realtime quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 2001;25(4):402–8.
- [23] De'ath G. Boosted trees for ecological modeling and prediction. *Ecology* 2007;88(1):243–51.
- [24] SantaLucia J Jr, Hicks D. The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct* 2004;33(1):415–40.
- [25] Lomzov AA, Vorobjev YN, Pyshnyi DV. Evaluation of the Gibbs free energy changes and melting temperatures of DNA/DNA duplexes using hybridization enthalpy calculated by molecular dynamics simulation. *J Phys Chem B* 2015;119(49):15221–34.
- [26] Gao Y, Wolf LK, Georgiadis RM. Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. *Nucleic Acids Res* 2006;34(11):3370–7.
- [27] Lu YP, Zhang C, Lv FX, Bie XM, Lu ZX. Study on the electro-transformation conditions of improving transformation efficiency for *Bacillus subtilis*. *Lett Appl Microbiol* 2012;55(1):9–14.
- [28] Tee KL, Grinham J, Othustise AM, González-Villanueva M, Johnson AO, Wong TS. An efficient transformation method for the bioplastic-producing “Knallgas” bacterium *Ralstonia eutropha* H16. *Biotechnol J* 2017;12(11):1700081.
- [29] Russo R, Panangala VS, Wood RR, Klesius PH. Chemical and electroporated transformation of *Edwardsiella ictaluri* using three different plasmids. *FEMS Microbiol Lett* 2009;298(1):105–10.
- [30] Bzymek M, Lovett ST. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc Natl Acad Sci USA* 2001;98(15):8319–25.
- [31] Lovett ST. Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol Microbiol* 2004;52(5):1243–53.
- [32] Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnol J* 2012;7(5):620–34.
- [33] Pelicic V, Reytrat JM, Gicquel B. Expression of the *Bacillus subtilis* sacB gene confers sucrose sensitivity on mycobacteria. *J Bacteriol* 1996;178(4):1197–9.