

细菌最小基因组研究进展

李金玉^{1,2,3}, 杨珊^{2,3}, 崔玉军^{2,3}, 王涛¹, 滕越^{2,3}

1. 天津大学生命科学学院, 天津 300110
2. 军事科学院军事医学研究院微生物流行病学研究所, 北京 100071
3. 病原微生物生物安全国家重点实验室, 北京 100071

摘要: 具有最小基因组的细菌只包含维持自我生命复制所必需的基因, 其作为一种潜在的工业生产平台具有诸多优势。由于高通量 DNA 测序和合成技术的发展, 目前已经构建了多种缩减基因组的菌株。本文首先介绍了最小基因组的概念, 其次总结了细菌必需基因的相关研究进展, 然后梳理了人工缩减与合成微生物基因组的相关工作, 最后探讨了在设计和组装基因组的过程中遇到的技术障碍和限制, 以期为人工合成基因组的实验与应用提供理论参考。

关键词: 细菌; 最小基因组; 必需基因; 合成生物学; 人工设计

Research progress of bacterial minimal genome

Jinyu Li^{1,2,3}, Shan Yang^{2,3}, Yujun Cui^{2,3}, Tao Wang¹, Yue Teng^{2,3}

1. School of Life Sciences, Tianjin University, Tianjin 300110, China
2. Beijing Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences, Beijing 100071, China
3. State Key Laboratory of Pathogen and Biosecurity, Beijing 100071, China

Abstract: Bacteria with the smallest genome contain genes necessary for self-sustaining replication only, giving the organisms advantages to serve as a potential industrial production platform. Many strains with reduced genomes have been constructed, owing to the development of high-throughput DNA sequencing and synthesis technology. This review first describes the concept of minimal genomes, summarizes the relevant research progress of bacterial essential genes, then systematically lists the work related to artificial reduction and synthesis of bacterial genomes, finally discusses the technical obstacles and limitations encountered in the process of designing and constructing reduced genomes, hoping to provide a theoretical basis for the experiment and application of artificially synthesized genomes.

Keywords: bacteria; minimal genome; essential genes; synthetic biology; artificial design

收稿日期: 2020-11-23; 修回日期: 2021-01-04

基金项目: 国家科技重大专项(编号: 2018ZX10101003-002-011)和“艾滋病和病毒性肝炎等重大传染病防治”科技重大专项(编号: 2018ZX10712-001-003)资助[Supported by the National Science and Technology Major Project (No. 2018ZX10101003-002-011), and the Major Infectious Diseases Such as AIDS and Viral Hepatitis Prevention and Control Technology Major Projects (No. 2018ZX10712001-003)]

作者简介: 李金玉, 在读硕士研究生, 专业方向: 生物学。E-mail: jyli1996@163.com

通讯作者: 滕越, 博士, 副研究员, 研究方向: 合成生物学。E-mail: yueteng@sklbp.org

DOI: 10.16288/j.ycz.20-301

网络出版时间: 2021/2/1 15:12:50

URL: <https://kns.cnki.net/kcms/detail/11.1913.R.20210201.1343.010.html>

人们很早就认识到支原体属(*Mycoplasma*)等物种的基因组比大多数细菌要小得多^[1~3]。虽然最初认为这些小基因组生物代表了生物进化的原始祖先,但到20世纪80年代,基于16S核糖体RNA序列的系统发育研究证实:*Mycoplasma*^[4]以及其他具有小基因组的胞内细菌,如立克次氏体科(*Rickettsiaceae*)中的一些物种^[5],皆来源于具有更大基因组的细菌。20世纪初,基因组测序结果证实了这些发现,并在不同的细菌类群中发现了独立的通过缩减基因组进化的案例,如柔膜菌纲中的*Mycoplasma genitalium*^[6]、 α -变形菌纲中的普氏立克次体(*Rickettsia prowazekii*)^[7]、 γ -变形菌纲中的蚜虫内共生菌(*Buchnera aphidicola*)^[8]和螺旋体纲中的伯氏疏螺旋体(*Borrelia burgdorferi*)^[9]。这些微生物的基因组保留了满足其自身在宿主中生存的关键基因,证明不同细菌类群在特定情况下具有丢失多种基因的“能力”,这表明控制基因组缩减的机制和动力可能存在于多数细菌中。

随着全基因组测序技术的发展,通过对几个胞内共生细菌的基因组测序发现存在更小的基因组,这突破了早期测序达到的约500 kb(即500个基因左右)的下限;其中一些基因组还包括其他极端特征,如极快的基因进化、密码子重排和核苷酸组成的极端偏向性等^[10,11]。这些变化的主要驱动因素包括较小的种群规模以及无性繁殖背景下的选择、突变和遗传漂变^[11~13]。这类生物不仅编码很少的蛋白质,而且其高度修饰的蛋白质很容易发生错误折叠,需要大量分子伴侣以保证蛋白质具有正常功能^[10]。尽管与宿主相互作用的关键基因得以保留,但其基因组通过持续删除特定基因而不断被缩减。共生生物和宿主的共适应可能促进一些基因丢失,但与从线粒体和胞质体向细胞核的基因转移不同,细菌基因向宿主基因组的转移似乎与细菌基因组缩减无关。

20世纪初,伴随重组DNA技术的出现,合成生物学可以设计并构建基于合成遗传回路和通路的新颖高潜力生物系统。在合成生物学的“设计-构建-测试-学习”循环中,系统生物学为其设计步骤提供了关于复杂生物过程的完整信息。因此,借助系统生物学和合成生物学,可最小化细菌基因组,使其仅包含细胞复制和产品生产所必需的基因,并以此设计和构建可预测、高效和精于生产的细胞(图1)。本文首先介绍了细菌最小基因组概念,其次总结了

必需基因的研究进展,然后回顾了在建构人工缩减与合成基因组方面的研究工作,最后探讨了在设计在建构缩减基因组的过程中遇到的技术障碍和限制,以期为人工合成基因组的实验与应用提供理论参考。

1 最小基因组

1.1 最小基因组的概念

最小基因组被定义为在无外界压力(营养丰富和无应激)条件下足以维持生命活动的基因集。许多实验研究进一步将其定义为在丰富培养基中支持纯种培养的基因集。因为大多数生物在大自然的生态环境中需要额外的基因,所需的基因集会随着环境条件的变化而有所不同。最小基因组概念出现和发展的时间与细菌全基因组测序时代相同,都始于20世纪90年代中期^[14]。生物的基因组大小从几十万个碱基对到一千多亿个碱基对不等。其中,布赫纳氏菌(*Buchnera* spp.)是一种胞内共生菌,与大肠杆菌(*Escherichia coli*)具有共同祖先;与其祖先相比,基因组缩减达75%,仅为250 kb^[15,16],是自然界中基因组大规模缩减的一个明显示例。最小基因组不仅存在于布赫纳氏菌,也存在许多其他共生细菌中。*Nasuia deltocephalinicola*是一种昆虫共生细菌,具有最小的自我复制基因组(112 kb)^[17]。此外,本课题组还聚焦于最近发现的为食液昆虫提供营养的共生细菌中的极端微小基因组,包括来自4个远缘细菌类群的5个独立的极端基因组缩减实例,其基因组大小都不到生殖支原体基因组的一半(甚至不到1/4)(表1)。这些例子显示出源于生态位适应的基因组进化趋势。共生生物不需要环境条件响应相关的基因,因为其宿主提供了稳定的营养供应,并保护其免受恶劣环境变化的影响,因此,在长期进化过程中,这些不必要的基因已经从它们的基因组中移除。

许多研究表明,物种种群规模较小、无性繁殖^[18~20]以及细菌所固有的基因删除偏倚^[19,21],是基因组缩减的主要原因。由于特定宿主的限制和不同宿主中的菌株之间缺乏重组,导致了高度遗传漂变,使有益但非必需的基因失活和缺失。这种菌群结构的另一个影响是最初发现于*B. aphidicola*和*M. genitalium*中的微小基因组基因序列快速进化,及其

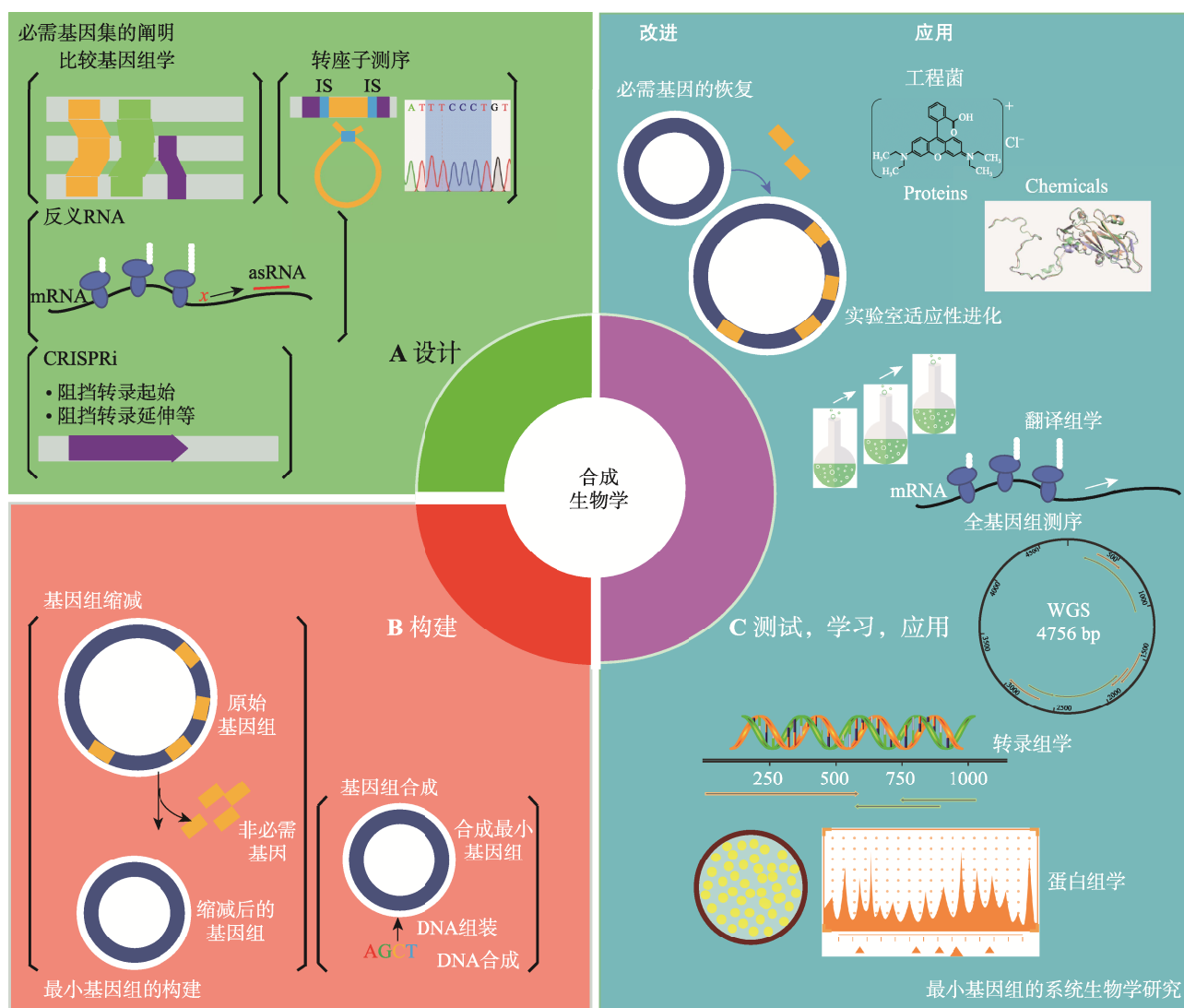


图 1 设计、构建和测试最小基因组的研究

Fig. 1 Design, build and test the minimal genome research

以合成生物学为基础的最小基因组的构建, 包括设计、构建、测试、学习、应用的循环周期。A: 通过不同方法确定物种的必需基因集, 包括比较基因组学、转座子测序、反义 RNA、CRISPRi 等。B: 构建最小基因组的过程, 主要遵循自上而下, 自下而上两种原则。C: 对构建的最小基因组进行系统生物学研究, 以测序技术、组学技术等为手段进行测试、学习、应用。

对蛋白质二级结构稳定性的影响^[22,23]。基因组减少的早期阶段以一些较新进化的共生生物为代表, 其特征是移动元件的增加, 假基因的形成, 多个基因组重排和染色体片段的缺失(图 2)^[24-26]。然而, 在更早期进化的共生体中, 如 *B. aphidicola*, 移动元件和大多数假基因已被移除。

通过对不同生物基因库的计算分析^[14,27-32], 以及纯种培养生物的基因诱变破坏实验^[33-36], 可对最小基因组的基因完整性进行预测。这种诱变实验只限于可纯种培养的生物, 如 *M. genitalium*, 它的基

因组是迄今为止纯种培养生物中最小的。研究表明, 通用基因集很小, 且不包括许多已被实验确定为特定生物所必需的基因。因此, 不同生物可能以完全不同的方式完成基本生命活动。例如, 通过运输而不是从头合成获得所需的化合物, 或者使用不相关的基因和不同的途径使特定 tRNA 装载正确的氨基酸^[37]。某些通用或接近通用的基因缺失并不致死^[32]。不同的研究预测了不同的最小基因集, 虽然这些基因的分布较广泛, 但都包括参与细胞基本功能的基因。有些新陈代谢相关基因分布较广泛, 在自然环

表 1 具有代表性的细菌、共生生物、病毒和细胞器基因组的比较

Table 1 Comparison of representative genomes of bacteria, symbiotic organisms, viruses and organelles

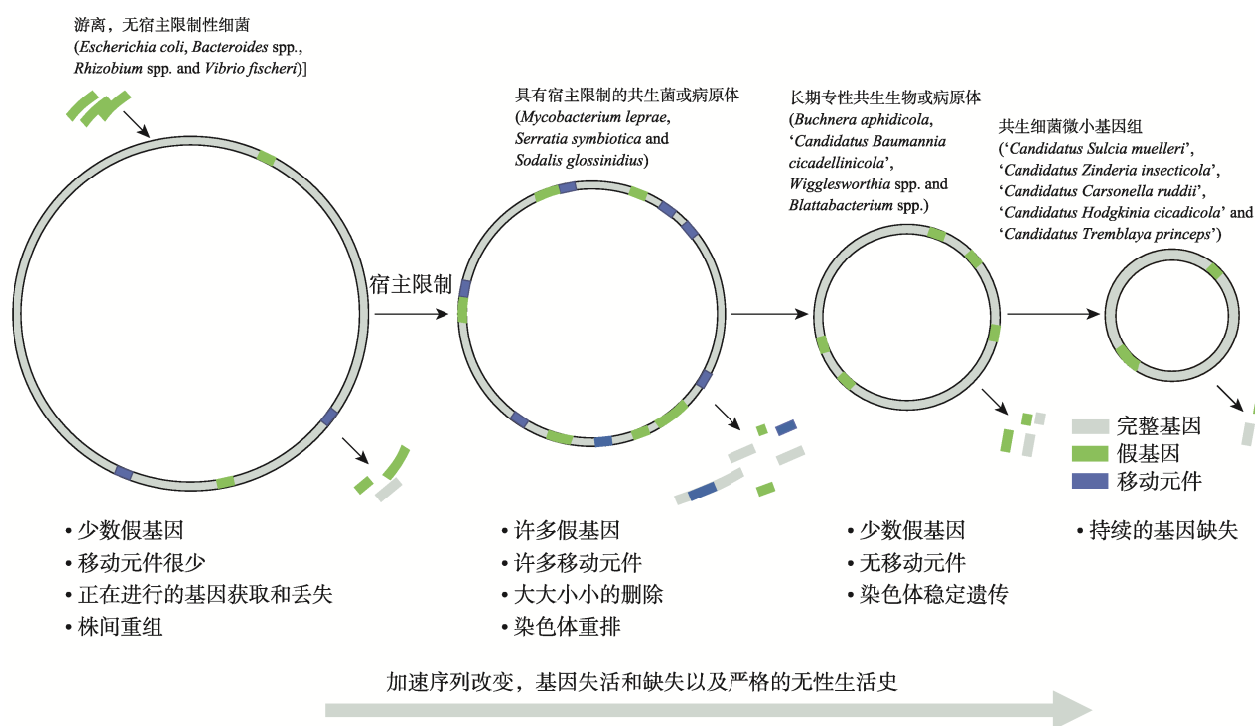
微生物	分类 ^a	基因组 大小(bp)	GC 含 量(%)	蛋白质 编码 序列数	位于 COGs M 簇和 I 簇 的基因数目 ^b	细胞形态	杆状决定基因			
							<i>ftsZ</i>	<i>ispA</i>	<i>mreB</i>	<i>rodA</i>
游离细菌										
<i>Escherichia coli</i>	γ-变形菌纲	4,639,675	50.8	4145	323	杆状 ^[46]	+	+	+	+
<i>Bacillus subtilis</i>	柔膜菌纲	4,215,606	43.5	4176	307	杆状 ^[46]	+	+	+	+
<i>Rickettsia prowazekii</i>	α-变形菌纲	1,111,523	29.0	835	100	杆状 ^[47]	+	–	+	+
<i>Mycobacterium genitalium</i>	Mollicutes	580,076	31.7	475	17	'烧瓶形' ^[48]	+	–	–	–
缩减基因组共生生物										
<i>Candidatus Blochmannia floridanus</i>	γ-变形菌纲	705,557	27.4	583	77	杆状 ^[49]	+	–	+	+
<i>Wigglesworthia glossinidia</i>	γ-变形菌纲	697,724	22.5	611	85	杆状 ^[50]	+	+	+	+
<i>Candidatus Baumannia cicadellinicola</i>	γ-变形菌纲	686,194	33.2	595	62	球状 ^[51]	+	+	+	+
<i>Buchnera aphidicola</i> str. APS	γ-变形菌纲	640,681	26.3	564	41	球状 ^[52]	+	+	–	–
<i>Candidatus Moranella endobia'</i>	γ-变形菌纲	538,294	43.5	406	36	球状 ^[53]	+	–	+	+
<i>Buchnera aphidicola</i> str. Cc	γ-变形菌纲	416,380	20.2	357	10	球状 ^[54]	+	–	–	–
极端缩减基因组共生生物										
<i>Candidatus Sulcia muelleri</i> str. GWSS	黄杆菌纲	245,530	22.4	227	5	多形性管状 ^[55]	–	–	–	–
<i>Candidatus Zinderia insecticola</i>	β-变形菌纲	208,564	13.5	202	1	多形性斑点 ^[56]	–	–	–	–
<i>Candidatus Carsonella ruddii</i>	γ-变形菌纲	159,662	16.6	182	1	多形性管状 ^[57]	–	–	–	–
<i>Candidatus Hodgkinia cicadicola</i>	α-变形菌纲	143,795	58.4	169	0	多形性管状 ^[58]	–	–	–	–
<i>Candidatus Tremblaya princeps</i>	β-变形菌纲	138,927	58.8	121	0	多形性斑点 ^[53]	–	–	–	–
具有较大基因组的细胞器										
<i>Cucurbita pepo</i>	线粒体	982,833	42.8	38	NA	NA	–	–	–	–
<i>Floydiella terrestris</i>	叶绿体	521,168	34.5	74	NA	NA	–	–	–	–
<i>Porphyra purpurea</i>	叶绿体	191,028	33.0	209	NA	NA	–	–	–	–
<i>Reclinomonas americana</i>	线粒体	69,034	26.1	67	NA	NA	–	–	–	–
巨大病毒										
<i>Acanthamoeba polyphaga mimivirus</i>	拟菌病毒科	1,181,404	28.0	1262	NA	NA	NA	NA	NA	NA
<i>Cafeteria roenbergensis virus</i> (CroV)	拟菌病毒科	617,453	23.4	544	NA	NA	NA	NA	NA	NA
<i>Coccolithovirus</i>	藻类 DNA 病毒科	407,339	40.2	472	NA	NA	NA	NA	NA	NA

^a: 指个体所属的纲, 细胞器或病毒家族类型; ^b: 蛋白质直系同源基因簇, 即指位于蛋白质直系同源基因簇中 M 簇(细胞膜生物发生, 外膜)和 I 簇(脂质代谢)的基因数; NA: 不适用。

境下对特定生物至关重要, 但并非必不可少。

细菌基因组含有数百万个碱基对。例如, 研究最广泛的模式生物大肠杆菌的基因组>5 Mb, 包含 4000 多个基因, 其中 1000 多个为未知功能的基因^[38,39]。

大肠杆菌可在多种环境下繁殖, 如好氧和厌氧, 以及不同营养物质、pH 值和温度等。大肠杆菌的基因组中有许多基因负责处理环境压力和利用各种营养物质; 然而, 在环境确定的实验室条件下, 不再需



宿主限制性细菌的基因组减少阶段;

种群规模较小以及无性生活史导致突变固定。

请注意, 某些可以在外部环境中持续存在并且重新感染宿主的共生菌和病原体(例如根瘤菌和费氏弧菌)通常不会经历这些基因组变化。

图 2 细菌基因组的减少阶段及其进化特征

Fig. 2 Genome reduction stage of bacteria, and its genome evolution characteristics

要某些应激反应基因; 因此许多基因可以被删除, 而不会对细胞生长产生负面影响^[40]。同时, 大肠杆菌基因组还编码了许多实验室培养和工业发酵不需要的基因, 这些基因会导致能量和生物物质前体的浪费, 不必要的基因组片段的复制, 以及功能冗余或无用的转录物、蛋白质和代谢物的合成。因此, 删除这些不必要基因的生物体或可成为人工条件下产品生产的全新平台^[41]。

1.2 线粒体基因组简化

线粒体比较基因组学使人们了解线粒体祖先的基因组是什么样的, 以及它包含的基因是什么。从它所包含的基因组的角度看, 线粒体具有毫无疑问的细菌起源, 起源于 α -变形杆菌(*Alpha proteobacteria*)^[42]。因此, 内共生假说认为真核细胞中的线粒体从细菌祖细胞通过共生进化而来。基于系统发育的分析表明, 线粒体基因组进化的特征是某些谱系大量扩增, 另一些谱系则极度减少和压缩, 通过内共生基因转移将线粒体基因组中的许多初始遗传

信息重新定位到细胞核^[43]。 α -变形杆菌基因组之间的比较表明, 线粒体的细菌祖先包含约 3000~5000 个基因, 线粒体基因组中直系同源基因的祖先簇的上限约为 1700 个, 这些估计表明, 从细菌共生体到原始细胞器的过渡过程中丢失了大约 3000 个基因^[44]。可以预见, 基因组学和蛋白质组学数据的不断增长将提供关于线粒体结构和功能的新的数据和见识, 并将不断重塑和完善目前的线粒体进化理论。

1.3 构建和设计最小基因组

随着基因组学革命以及系统生物学的不断进步, 合成生物学在过去 10 年中飞速发展, 尤其是在基于系统和合成生物学的微生物底盘工程和改造领域。在系统生物学中, 理想的底盘可代指具有简化的基因组且可以实现全部功能的生物, 以及能够更有效地合成所需产物的代谢网络; 在合成生物学中, 底盘是指通过提供允许其运转的资源来容纳和支持遗传成分的生物。

构建合成微生物底盘包括两种方法: 自上向下

和自下向上。自上而下的方法是通过去除不必要的细胞基因来了解基因组架构并改善其特性的减少基因组大小的策略。基于大规模 DNA 分析的出现,通过对来自不同生物体的基因组进行比较分析,通常可以揭示对于细胞生命和相似和/或截然不同的代谢途径必不可少的基因。接下来,再通过不同的实验方法来实现缺失。工程设计和修改合成微生物底盘揭示生命的基本原理,是加强其在卫生医药和食品工业中的应用的最好方法之一。有关这个主题的研究计划已在多个国家获得资助。“最小基因组工厂”(minimum genome factory, MGF)项目于 2001 年在日本启动,其目标是构建具有较小基因组的微生物以用于工业用途。基于比较大肠杆菌和布赫纳氏菌获得的非必需基因信息,Hiroshi 等^[45]通过删除大肠杆菌 W3110 中的大肠杆菌特异性基因构建了大肠杆菌 MGF-01,其构建方法与其他缺失方法一致,即连续缺失结合 P1 转导。简言之,用含有 *sacB* 和 *cat* 基因的 DNA 组件同源替换删除靶 DNA 序列。然后,用另一个 DNA 组件替换选择标记,用蔗糖进行负筛选获得无标记克隆。最后,在 28 个 P1 转导循环中积累了 53 个缺失,从而使总缺失长度达 1.03 Mb。MGF-01 的最终生物量是野生型亲本的 1.5 倍。据称,MGF-01 相比亲本菌株葡萄糖利用效率更高,这是因为显著减少了过量的醋酸盐积累(从大肠杆菌 W3110 中的 1.37 g/L 降至 MGF-01 中的 0.50 g/L)。此外,与野生型相比,MGF-01 的苏氨酸浓度和产量分别提高了 2.44 倍和 1.69 倍,醋酸盐副产物则显著降低。MGF-01 的这些意想不到的有益特性与其他具有与野生型祖先相似表型的缩减基因组大肠杆菌不同,但需要进一步研究 MGF-01 的差异性。

2 细菌必需基因

构建最小基因组,首先要阐明维持生命所必需的基因。20 世纪 90 年代,只测得少数微小细菌的全基因组序列。Mushegian 和 Koonin 等^[59]对生殖支原体和流感嗜血杆菌(*Haemophilus influenza*)进行比较,表明这两种细菌都有相对较小的基因组,但表现出完全不同的进化轨迹。根据比较基因组学原则,在多种生物中保守的基因很可能有必需的功能。在

生殖支原体和流感嗜血杆菌中,共发现 240 个直系同源基因;但此分析遗漏了几个编码细胞必需功能的基因,如磷酸甘油酸变位酶和核苷二磷酸激酶基因。这两个物种有不同且不相关的磷酸甘油酸变位酶,因此非同源基因偶尔可以取代古老的基因并破坏其保守性。最终,研究结果证明包括非直系同源基因取代在内,预测共有 262 个基因构成了两个物种的核心生物功能。尽管这两种细菌和它们的共同祖先之间经历了 15 亿年的进化,仍有约 50% 的基因被保存下来。

自 2000 年以来,已获得数万基因组序列。科学家比较了数百个基因组,以确定物种间普遍存在的基因,然而这样的基因并不常见。Harris 等^[60]报告了来自直系同源簇(COG)数据库的 80 个核心基因 Brown 等^[61]比较了 45 个基因组,发现只有 23 个保守基因;Koonin 等^[62]报道了普遍存在于 100 个基因组的 63 个基因;Charlebois 等^[63]比较了 14 个门中的 147 个原核生物基因组,仅发现 34 个通用基因。虽然报告的保守基因数目不同,但都这些研究同时证明只有少数保守基因绝对不足以维持生命。因此,比较基因组学对必需基因的理论预测仅限于功能未知、非同源取代和具有数十亿年进化历史的基因。

除了计算预测,研究者还使用实验方法来探究必需基因。最简单的方法是从基因组中删除一个特定的基因位点,观察其致死效应。在迄今为止已知的任何自由生活的生物体中,生殖支原体具有最小的基因组(580 kb,编码 517 个基因),其中许多基因可能通过转座子作用而失活^[64]。在枯草芽孢杆菌(*Bacillus subtilis*)中,通过对其基因组的 79 个区域进行随机突变^[65],发现只有 6 个位点是必不可少的,大小约 318~526 kb,与生殖支原体基因组大小相似。因此,人们设计了各种方法来确定细菌中的必需基因。如通过重组、转座子插入和反义 RNA (asRNA) 直接使单个基因失活。通过插入非复制质粒失活枯草芽孢杆菌中的单个基因,Kobayashi 等^[66]发现 4101 个基因中只有 271 个是必需基因。在大肠杆菌中,4288 个开放阅读框(open reading frame, ORF)中的 3985 个可被敲除,表明剩余的 303 个 ORFs 对生命至关重要^[67]。尽管靶向基因敲除研究提供了基因必需性的直接证据,但该方法耗时长,需要进行数千

次删除实验。

为了克服这些限制,可采用基于转座子诱变失活的高通量方法。转座子是一种基因元件,可在基因组内随机移动并插入,破坏基因的正常功能。如果基因组中的必需基因被插入转座子,则此突变体不能存活。利用此特征,可以通过鉴定存活突变体中的转座子插入位点来区分必需和非必需基因。Glass 等^[68]和 Hutchison 等^[69]分别利用 1300 和 3000 个突变体,在生殖支原体中构建了全基因组转座子插入图谱,发现 265~382 个必需编码序列(coding sequences, CDSs)。由 3000 个唯一插入位点组成的转座子插入图谱,其平均分辨率约为每 200 bp 一个插入位点。此分辨率无法检测功能性 RNAs (如 tRNAs 和 ncRNAs)等小遗传元件的必需性。此外,一些必需基因在 3'端对转座子插入检测具有抗性,因为短截尾或延伸并不影响其功能。为了规避这一限制,研究者发明了一种通过 asRNAs 使基因失活的方法^[70]。与转座子不可逆的基因破坏不同,利用 asRNAs 可以特异性敲低某个基因。因此,一旦构建出 asRNA 文库,研究者可以在多种环境条件和迭代过程中评估基因必需性或适合性,而无需重复构建敲除菌株或转座子突变体。

转座子插入位点的鉴定依赖于单个克隆的分离和现有的高通量测序技术的发展;通过与高通量测序技术相结合,可以并行识别多个插入位点。因此,转座子诱变与下一代测序技术(Tn-Seq)相结合,即利用转座子诱变以更高的分辨率检测必需基因。对 2×10^5 大肠杆菌转座子突变体文库的统计分析,共鉴定出 620 个必需基因^[71]。Tn-Seq 和单个基因敲除研究之间的差异可能源于细胞增殖。一个重要基因,即使不是严格意义上的必需基因,其失活也会导致严重的生长缺陷;在细胞繁殖过程中,这种缺陷可能未被充分表达或逐渐消失。此外,不同的统计临界值和实验条件也可能导致这种差异。

最后,基于成簇规律间隔短回文重复序列(clustered regularly interspaced short palindromic repeats, CRISPR)引领的技术革命,催化失活 Cas9 (dead Cas9, dCas9)可以在转录水平抑制靶基因的表达^[72]。由于 CRISPR 系统的特异性只取决于嵌合单导向 RNA (sgRNA)中 20 nt 的原间隔序列,因此易于通过 DNA 合成构建大规模全基因组 sgRNA 文库。

利用全基因组 CRISPR 干扰(CRISPR interference, CRISPRi),通过约 59 000 个 sgRNAs 抑制所有基因,在大肠杆菌中鉴定出 379 个必需基因^[73]。利用 CRISPRi 技术估计的必需基因数量略大于其他方法估计的数量。在细菌中,许多基因是多顺反子转录的,故前导多顺反子的破坏会使操纵子中包含的下游基因失活。因此,即使前导基因是非必需的,其转录抑制也会导致下游必需基因的致死效应,即多顺反子结构导致必需基因的高估。利用目前的高通量 DNA 合成技术,可以高效合成庞大的 sgRNA 文库,并将其用于鉴定具有更大基因组的生物(如人类)的必需基因^[74,75]。

总之,人们已经采用多种方法来阐明各种生物体中的必需基因。尽管不同方法得到的必需基因的确切数目不同,但 500 个基因被认为足以维持生命。利用单基因敲除研究、Tn-Seq 和 CRISPRi 直接检查单个基因是否必需无法规避各种技术固有的局限性。因为这些方法依赖于单个基因的移除或失活,而同时失活两个以上的基因从未被测试过。例如,假设两个基因在大肠杆菌中具有相同的基本功能,这两个基因可以单独删除,因为另一个基因可以恢复其功能;但是,两个基因同时失活则产生致死效应。即使最简单的细菌生殖支原体中,构建双基因敲除库也需要 25 万个菌株。考虑到构建约 4 000 个大肠杆菌单基因敲除菌株所需的巨大时间和费用成本^[40],需要采用新的技术来探索基因高阶组合的必需性。

3 人工缩减基因组微生物

大肠杆菌是研究最广泛的模式生物,已为其构建了多种缩减基因组,缺失大小从 300 kb 到 1.38 Mb 不等(表 2)。大肠杆菌 K-12 菌株的基因组约为 4.64 Mb,缺失大小约为原始基因组的 6.8%~29.7%。下面简介了一些目前构建的缩减基因组大肠杆菌,还回顾了其他物种中基因组缩减案例。

3.1 大肠杆菌 CDΔ3456

2002 年有两例关于缩减基因组大肠杆菌的报道。其中之一是 Yu 等^[76]报道的 CDΔ3456,缺失片段大小超过 300 kb。该菌株构建过程中,位于 Cre 位点

表 2 缩减基因组大肠杆菌及其特征

Table 2 Reduced genomes *E. coli* and its characteristics

原始菌株	缩减基因组菌株名称	原始基因组大小(Mb)	缩减量(%)	特征
<i>Escherichia coli</i> str. K-12 substr. MG1655	CDA3456	4.64	313 kb (6.8)	生长速度正常
<i>Escherichia coli</i> str. K-12 substr. MG1655	MDS12	4.64	376 kb (8.1)	生长速度正常, 细菌生长密度提高 10%
<i>Escherichia coli</i> str. K-12 substr. MG1655	MDS42	4.64	708 kb (15.3)	生长速度正常, 更高的转化效率
<i>Escherichia coli</i> str. K-12 substr. MG1655	MS56	4.64	1068 kb (23.0)	生长速度正常, 遗传稳定性提高
<i>Escherichia coli</i> str. K-12 substr. MG1655	Δ16	4.64	1377 kb (29.7)	较低的生长速度, 异常的拟核结构
<i>Escherichia coli</i> str. K-12 substr. W3110	MGF-01	4.65	1030 kb (22.2)	细菌生长密度提高 50%, 更高的苏氨酸产量
<i>Bacillus subtilis</i> str. 168	Δ6	4.22	320 kb (7.7)	生长速度正常
<i>Bacillus subtilis</i> str. 168	PG10	4.22	1456 kb (34.5)	较低的生长速度
<i>Bacillus subtilis</i> str. 168	PG38	4.22	1535 kb (36.4)	较低的生长速度
<i>Bacillus subtilis</i> str. 168	MGB469	4.22	469 kb (11.1)	生长速度正常
<i>Bacillus subtilis</i> str. 168	MG1M	4.22	991 kb (23.5)	生长速度正常
<i>Bacillus subtilis</i> str. 168	MBG874	4.22	874 kb (20.7)	较低的生长速度, 更高的蛋白产量
<i>Streptomyces avermitilis</i>	SUKA17	9.03	1674 kb (18.5)	更高的抗生素产量
<i>Schizosaccharomyces pombe</i>	A8	12.6	657 kb (5.2)	更高的蛋白产量

特异性重组酶识别的两个 *loxP* 位点之间的一个较大的基因组区域被删除。使用转座子将 *loxP* 位点预先插入到基因组中的任意位置。具体方法是, 用两种抗生素抗性基因盒构建两个转座子突变库, 并定位所有转座子插入位点。筛选转座子分别定位在目标缺失区域一端的两个突变体, 其中包含两个不同的标记基因, 并通过 P1 转导将其基因组融合。使用两种不同的抗生素抗性基因标记来筛选成功转导的细胞。然后, 融合基因组中的两个 *loxP* 位点通过 Cre 重组酶重组, 产生较大的靶向缺失。由此构建了 6 个缺失菌株, 并进一步利用 P1 转导将多个缺失累积到一个基因组中。在此过程中, 某些基因组区域可以单独从基因组中删除, 但是特定的基因组合不能同时删除, 这种关系被称为合成致死, 其原因可能是存在基因重复或直系同源物^[77]。当两个基因的一个拷贝被删除时, 另一个拷贝可以维持生物体正常功能, 而双突变体则不能。为了避免合成致死组合, 4 个区域被整合在 CDA3456 中, 该菌株缺少 287 个 ORFs, 其中包含 179 个未知基因, 以及与组氨酸生物合成、菌毛和数个转运蛋白相关的基因, 最终获得克隆的生长速率与亲本大肠杆菌相当。

3.2 大肠杆菌 MS56 及其近缘菌株

2002 年报道了缩减基因组大肠杆菌 MDS12, 其缺少大肠杆菌 K-12 菌株的 12 个 K-岛^[78]。K-岛是 K-12 通过水平基因转移获得的基因组区域。K-岛含有非必需基因, 如原噬菌体和转座子。利用 *I-SceI* 巨核酶和双链断裂修复系统, 用无痕删除法将 12 个 K-岛依次删除。具体方法是, 将含有氯霉素抗性基因的 DNA 组件引入大肠杆菌 MG1655, 以取代同源靶区。虽然靶标被删除, 但下一轮删除需要去除抗性基因, 因此采用 *I-SceI* 将该基因删除。

之后, 双链断裂被 *RecA* 修复, 形成无痕缺失株。经过 12 次迭代删除和 P1 转导, 删除长度合计为 376 kb, 包含 409 个 ORFs。由于被删除的基因没有必需功能, MDS12 的生长速度和 DNA 转化效率与原始菌株无差异。MDS12 的最终细胞密度比野生型大肠杆菌高约 10%。在 MDS12 中, 节省的能量和物质可以转化为生物量, 显示出缩减基因组的优势。

4 年后, Pósfai 等^[79]报告了大肠杆菌 MDS41、42 和 43, 它们是 MDS12 的后代, 含有额外的缺失(与野生型相比, 总缺失为 663~708 kb)。MDS42 不含

任何可转位和插入序列(insertion sequence, IS)元件。据报道,在大肠杆菌中,20%~25%的突变与 IS 元件有关^[79]。MDS42 未显示 IS 介导的基因失活。利用该特性, MDS42 可稳定复制携带毒性 ORFs 的质粒 DNA, 且基因组构成稳定。大肠杆菌有一个沉默的 *bgl* 操作子, 能够利用水杨苷; 因此, 通常条件下, 大肠杆菌不利用水杨苷作为其唯一碳源。当使用水杨苷作为唯一碳源培养时, MDS41 的 *bgl* 操纵子激活率低于 MG1655 (8%)。此外, MDS42 产生的苏氨酸比野生型大肠杆菌多 80%以上^[80]。

成功构建 MDS 菌株之后, Park 等^[81]在 MDS42 中又引入了 14 个缺失, 以此构建大肠杆菌 MS56。除 MDS42 的缺失外, 还删除了非必需基因和遗传元件, 如氢化酶、菌毛样黏附素和厌氧呼吸酶, 基因组缩减总量达 1.07 Mb。编码抑制大肠杆菌生长的人蛋白异源基因可在 MS56 中成功表达, 不会有任何 IS 介导的失活, 而正常含 IS 的大肠杆菌中会迅速发生失活。

3.3 大肠杆菌 $\Delta 16$

Hashimoto 等^[82]构建了大肠杆菌 $\Delta 16$, 其基因组缩减量为 1.377 Mb (占其原始基因组的 29.7%), 是报道的最大缩减量。与其他缩减基因组一样, $\Delta 16$ 的构建方法也是同源重组连续缺失结合 P1 转导。Hashimoto 等^[82]采用正筛选和负筛选两种方法进行基因删除。首先, 构建靶区序列末端同源 DNA 组件。该盒还含有三个基因: 氯霉素乙酰转移酶(*cat*)、*rpsL* 和 *sacB*。将此 DNA 组件导入后, 用氯霉素选择目标基因组区域被此组件取代的克隆。然后, 用新的同源 DNA 组件取代该序列, 接着用 *rpsL* 和 *sacB* 基因进行负筛选, 使细胞对链霉素和蔗糖敏感。通过 P1 转导将 16 个无痕缺失进行组合, 获得最终菌株 $\Delta 16$, 累计缺失 1.377 Mb。在连续缺失过程中, 倍增时间从 MG1655 的 26.2 min 依次增加到 $\Delta 16$ 的 45.4 min, 细胞形状变长变宽, 拟核呈不规则分布, 其复制与细胞分裂不同步。

3.4 枯草芽孢杆菌 PG38 及其近缘菌株

枯草芽孢杆菌是研究最广泛的革兰氏阳性菌之一, 因其具有蛋白分泌系统而成为各种蛋白的优良生产宿主。Westers 等^[83]通过去除 6 个基因组位点,

包括聚酮、蛋白质抗生素生物合成、原噬菌体和原噬菌体样元件相关基因, 构建了基因组缩减枯草芽孢杆菌 $\Delta 6$ 。这些位点共包含 332 个基因(320 kb)。通过使用整合质粒 pG + host4 整合并切除选择标记实现基因删除^[84]。基因组缩减枯草芽孢杆菌 $\Delta 6$ 在细胞生理上没有明显变化。与亲本菌株相比, 该菌株的生长速度、葡萄糖/醋酸盐代谢通量、异源蛋白分泌和生物量均相同。出乎意料的是, 尽管没有删除与细胞运动相关的基因, $\Delta 6$ 在琼脂糖平板上显示出细胞运动性增加。基因组缩减菌株偶尔会出现此类意外表型, 说明对即使是微小细菌的基因组理解也是有限的^[85]。

枯草芽孢杆菌 $\Delta 6$ 基因组被进一步缩减, 构建形成 PG10 和 PG38。Reu 等^[86]构建了两个独立的基因组缩减菌株 PG10 和 PG38, 分别包含 88 和 94 个迭代缺失, 缺失部分包含产孢、运动、抗生素合成和次生代谢相关的非必需基因。在连续缺失过程中, 中间菌株逐渐丧失 DNA 整合所需的感受态。因此, 在基因组中引入了额外的感受态蛋白(ComK 和 ComS), 以有效导入缺失所需的 DNA。两种衍生菌株的生长速度较慢(倍增时间延长约 40%), 细胞形态呈长丝状。虽然这两个菌株的生长速度都有所降低, 与之前认为细胞存活所需的基因组相比, 它们的基因组缩减比例是迄今为止最大的(1.46 和 1.54 Mb; 超过了它们原始基因组的 1/3)。

3.5 枯草芽孢杆菌 MGB874 及其近缘菌株

Ara 等^[87]基于枯草芽孢杆菌 168 构建了一株缩减基因组枯草芽孢杆菌 MGB469。缩减菌株缺少 9 个与原噬菌体和原噬菌体样元件相关的基因组位点和 2 个抗生素合成基因(巴斯他汀和聚酮)。首先从枯草芽孢杆菌的基因组中逐一删除这些位点, 以确认其不包含必需基因。然后, 依次合并删除, 使删除区域达 469 kb, 由此得到菌株 MGB469, 其生长速度和蛋白质产率与原始菌株无差别, 从而达到进一步缩减基因组的目的。该文作者发现一个可以增加蛋白质产率的基因组缩减位点, 最终构建出比原始菌株具有更高生产率的基因组缩减菌株。从 MGB469 中进一步删除单独缺失时增加纤维素酶产率的 6 个基因组位点, 构建了总基因组缩减 991 kb 的菌株 MG1M。然而, MG1M 中纤维素酶和蛋白酶的生产

没有发生明显的变化。

由于枯草芽孢杆菌 MGB469 和 MG1M 的表型与野生型相当, 并且在进一步基因组缩减后没有产生有益表型, Morimoto 等^[88]重新利用中间菌株 MGB469 构建具有优势特征的新型基因组缩减枯草芽孢杆菌。该文作者测试了包括原噬菌体和次级代谢基因在内的 74 个基因组区域的缺失。在 63 个可发生缺失的区域中, 11 个连续缺失被引入 MGB469, 从而得到缺失长度为 874 kb 的枯草芽孢杆菌 MGB874。虽然 MGB874 具有正常的细胞形态和拟核结构, 但其生长速度降低至野生型枯草芽孢杆菌的 70%。与枯草芽孢杆菌的其他缩减基因组不同, MGB874 产生的纤维素酶和蛋白酶比枯草芽孢杆菌 168 分别提高 1.7 倍和 2.5 倍。转录组研究发现了许多转录组水平的基因表达变化, 如产孢、降解酶分泌和 σ 因子, 推测是其生产率提高的原因。MGB874 证明具有缩减基因组的菌株可作为工业蛋白生产的优良宿主。

3.6 阿维链霉菌(*Streptomyces avermitilis*) SUKA17

放线菌门中的链霉菌属是一类工业和临床应用中重要的放线菌, 已知其可产生抗生素等大多数具有生物活性的次级代谢产物。链霉菌属基因组在细菌中相对较大, 其染色体 DNA 呈线性而非环状。基于应用最广的工业菌种之一阿维链霉菌, Komatsu 等^[89]构建了菌株 SUKA17。阿维链霉菌中含有 20 多个次级代谢物生物合成基因簇, 主要位于基因组的末端, 称为亚端粒区。对灰色链霉菌(*Streptomyces griseus*)、天蓝色链霉菌(*Streptomyces coelicolor*)和阿维链霉菌进行比较基因组学研究, 揭示了位于基因组中心的必需核心基因。使用同源重组和 Cre 介导的重组, 从阿维链霉菌中删除亚端粒区主要次级代谢相关基因(包括萜烯代谢)。与大肠杆菌和枯草芽孢杆菌不同, 阿维链霉菌不易发生重组。因此, 两个 *loxP* 位点首先通过同源重组从环状 DNA 模板导入到基因组中。然后, 通过 Cre 介导重组删除基因组中大片段目标位点。该缺失包括 1272 个 ORFs 的基因组区域, 总长度为 1.67 Mb。由于其基因组较简单, 无需承担不必要的次级代谢物合成, 含有异源基因簇的 SUKA17 比原生灰色链球菌和棒状链霉菌(*Streptomyces clavuligerus*)产生更多的链霉素和头霉素 C。

4 合成基因组学

4.1 基于合成基因组学的微生物设计与改造

上面讨论的缩减基因组都是使用自上而下的方式对基因组进行缩减而构建的。现今, 研究者已实现从头合成病毒、噬菌体和细菌基因组。借助基因组编辑和合成工具, 可以进行全基因组水平上进行工程设计。研究者已实现设计并构建基于合成遗传回路和通路的新高潜力生物系统。借助系统生物学和合成生物学而发展的合成基因组学, 可最小化细菌基因组, 从头开始合成基因组可以对基因组结构和功能进行前所未有的修饰, 获得对生命基本原理的新见解并致力于设计有价值的应用, 例如设计并构建可预测的、高效的或是精于生产的细胞(图 3)。

自 1995 年关于生殖支原体基因组大小为 525 个基因的第一份报道以来, 该细菌和该属的其他成员已成为探索赋予生命最少数量的细胞成分的最有用的试验台。事实上, 到目前为止, 某些支原体菌株已成为我们最了解的生物系统, 其所有组成部分及其相互作用的各个方面都经过了非常详细的探索和量化。里程碑式的工作包括通过转座子诱变检查每个基因的必要性, 实现支原体基因组的完全化学合成等。

支原体是小基因组细菌, 其将成为探索生命所需的最低限度的基因组大小的理想平台^[90]。Gibson 等^[91]构建的支原体 JCVI-syn1.0 含有一个完全化学合成的基因组, 即略加修饰的 1.08 Mb 蕈状支原体(*Mycoplasma mycoides*)基因组的复制体。通过转座子诱变阐明蕈状支原体的必需基因后, 合成了最小基因组蕈状支原体 JCVI-syn3.0^[92](包含 473 个基因, 总长度为 531 kb)。进一步的实验验证表明, 基于分子生物学的系统知识以及有限的转座子诱变数据相结合的初始设计未能产生活细胞。随后, 改良的转座子诱变方法揭示了稳健生长所需的一类准必需基因, 这解释了其最初设计的失败。最后, 研究者通过设计, 合成和测试的 3 个循环(保留了基本必需的基因)产生了 JCVI-syn3.0, 相对于其亲本菌株而言, 基因组的净基因组大小减少了约 50% (从约 1079 kb 减少到 531 kb), 并将基本生物学信息精简到了 473 个基因, 其基因组比自然界中任何自主复制的细胞

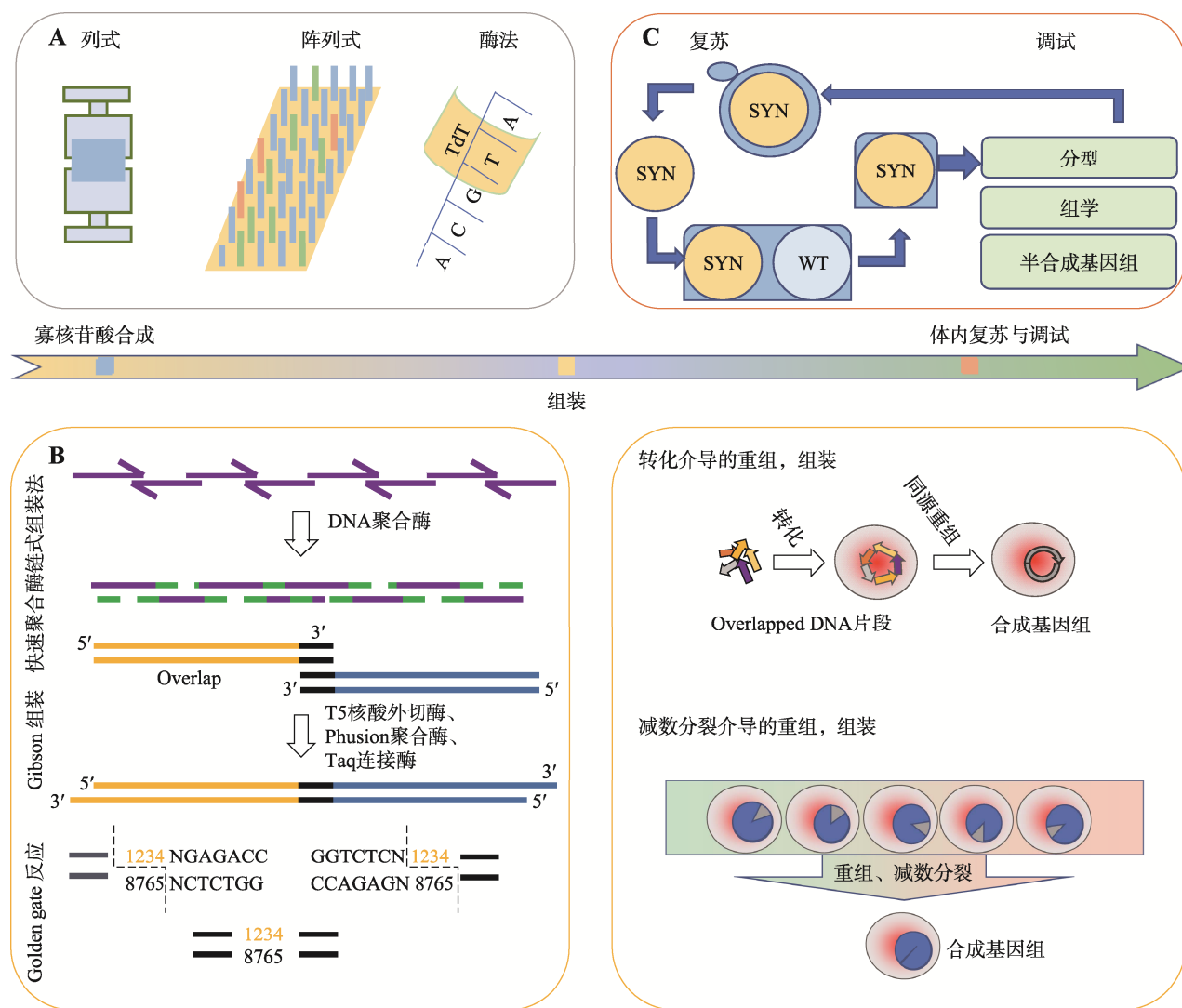


图 3 合成基因组的过程：从设计序列到体外组装到体内复苏

Fig. 3 The process of synthesizing genome research: from design sequence to *in vitro* assembly to *in vivo* recovery

A: 在体外进行寡核苷酸的合成。常用方法包括列式合成、阵列式合成、酶法合成等。B: 通过 PCA、Gibson 组装等进行体外组装，或是导入细胞内进行由转化以及减数分裂介导的重组、组装。C: 最后进行分型、调试。

都要小。虽然 JCVI-syn3.0 的倍增时间是 JCVI-syn1.0 的 3 倍 (~180 min)，但远小于具有天然最小基因组的生殖支原体 (~16 h)。尽管全基因组合成需要漫长的时间和巨大的费用^[93]，但无疑是基因组缩减的一种极有吸引力的替代方法。

JCVI-syn3.0 是一个多功能平台，可用于研究生活的核心功能并探索全基因组设计。最小细胞的概念乍看之下似乎很简单，但基于实验的实践验证却变得更加复杂。除了必需和非必需基因外，还有许多准必需基因，它们对生存力不是绝对关键，但对于强劲的生长却是必需的。因此，在最小化基因组

的过程中，需要在基因组大小和生长速率之间进行权衡。JCVI-syn3.0 近似为最小细胞基因组，为最小基因组与实验生物可行生长速率间的折中，它保留了几几乎所有与大分子合成和加工有关的基因。出乎意料的是，它还包含 149 个生物学功能未知的基因，表明存在着生命必需的未发现功能。

4.2 大肠杆菌基因组密码子的简化

大自然使用 64 个密码子编码来自基因组的蛋白质合成，并从多达 6 个同义词中选择 1 个有义密码子编码每个氨基酸。同义密码子选择具有多种重

要作用, 且许多同义替换是有害的。合成基因组不仅模仿模板 DNA, 而且允许遗传密码重编程。Lajoie 等^[94]将 80 个大肠杆菌菌株中 42 个高度表达的必需基因中 13 个稀有密码子的所有实例与编码相同氨基酸的密码子交换, 显示了在活细胞中以全基因组规模进行编码的可行性。

生命科学的技术进步加快了我们处理基因组编码信息的能力, 包括接合组装基因组改造 (conjugative assembly genome engineering, CAGE) 技术等。通过定义的同义密码子对目标密码子进行全基因组取代, 可以减少用于编码规范氨基酸的密码子数量, 最后, 通过重新编码 18 214 个密码子, 创建了一个具有 61 个密码子, 基因组大小为 4 Mb 的大肠杆菌变体, 其利用 59 个密码子编码 20 个氨基酸, 并能够删除以前必不可少的转移 RNA^[95]。使用减少数量的有义密码子(59)编码 20 个必需氨基酸产生的生物体表明, 生命可以在减少数量的同义有义密码子下运转。研究者开发的将设计的基因组分为多个片段、部分, 通过 Rexer、定向缀合等方法进行融合, 最后进行无缝和强大的集成来实现设计的策略, 为将来的基因组合成提供了一个蓝图。

4.3 DNA 组装技术

作为合成生物学的基石, DNA 组装过程允许使用被定义的、标准化的、特征明确的组件构建新颖的生物学系统。实际上是将多个 DNA 片段首尾相连的一种物理方式, 从而创建目标高阶装配, 然后将其连接到载体。采用限制酶切消化和逐个元素克隆的传统技术既耗时又低廉成本。因此, 人们正在付出巨大的努力来开发更好的克隆策略和 DNA 组装技术, 以使多基因系统能够被更快, 更有效地构建。这样, 构建具有复杂遗传功能的菌株将变得更加容易。这些新技术还有助于提高重组菌株的活力、可转化性或是稳定性。

DNA 组装技术的加速使用使得研究人员能够进行更复杂的合成项目, 常见的组装方法除了上面提到的 Gibson 组装, 还有 Golden Gate 组装(GG)。Golden Gate 模块化克隆系统利用 II 型限制酶建立了标准化且可互换的 DNA 部分文库, 随后将其一步一步组装到一个预先设计的 4nt 悬臂支架。目前, Golden Gate 组装已成为许多基因组编辑试剂盒的核

心, 许多实验室已经具备在单个反应中按定义的顺序装配许多不同零件的能力。GG 的优点之一是它允许组合组装, 可用于高效地生成库。此外, 使用 GG 的启动子改组策略用于筛选表达的每个转录单位的最佳启动子-基因对, 然后使用最佳启动子组合工程改造脂质过量生产菌株。

5 缩减基因组的障碍与挑战

解析突变和选择对细菌 GC 含量的不同影响一直是个难题。1962 年, 研究者提出一个模型, 将基因组中 GC 平均含量描述为由(G 或 C)→(A 或 T)和(A 或 T)→(G 或 C)碱基替换率差异所驱动的严格中性突变的过程函数。之后, 人们将细菌类群之间基因组 GC 含量的变异归因于对各种全基因组特性的选择等^[96]。最近发表的两篇论文指出存在一种固有且普遍的(G 或 C)→(A 或 T)突变偏倚, 并表明有利于高 GC 含量的选择过程是决定细菌基因组碱基组成的主要因素^[97,98]。

缩减的细菌基因组往往具有更高的 AT 含量, 且有时 AT 含量会急剧增加。具有最高 AT 偏倚的细胞基因组来自共生菌 '*Candidatus Zinderia insecticola*' (β-变形菌, 基因组为 209 kb, GC 含量为 13.5%^[99]) 和 '*Candidatus. Carsonella ruddii*' (γ-变形菌, GC 含量为 16.6%^[100])。已提出几种假设来解释严格内共生菌的 AT 偏倚, 包括选择或群体遗传结合突变模式变化^[101~103]。微小基因组倾向于删除许多参与 DNA 修复的基因, 这可能导致更多的 A 或 T 突变, 因为脱氧核糖核酸损伤, 如胞嘧啶脱氨和鸟嘌呤氧化, 往往会导致(G 或 C)→(A 或 T)^[104]。穆勒棘轮效应 (Muller's ratchet) 和纯化选择压力减轻将产生更多轻微有害突变, 此类突变偏向 A 或 T 突变, 并在种群中得到固定。这几种因素的共同作用使基因组 GC 平均含量移向 AT 含量较高的新平衡点。

值得注意的是, 两个最小的细菌基因组, *Ca. Tremblaya princeps*^[105] (基因组 GC 含量为 58.8%, 四重简并密码子第三位的 GC 含量为 66.6%) 和 *Ca. Hodgkinia cicadicola*^[106] (α-变形菌, 基因组大小为 144 kb, 基因组 GC 含量为 58.4%, 四重简并密码子第三位的 GC 含量为 62.5%), 打破了极端缩减基因组和低 GC 含量之间原本普遍的联系。在 *Ca.*

Hodgkinia cicadicola 中, 缩减基因组表现为偏向 GC 的突变压力, 因为在蛋白质序列水平上, 四重简并密码子第三位几乎没有选择压力^[106]。由此提出一种假设, 即在基因组缩减的过程中, *Ca. Hodgkinia cicadicola* 以某种方式维持了自由生活 α -变形菌中的高 GC 含量突变偏倚。通过后续研究 *Ca. Hodgkinia cicadicola* 和 *Ca. Tremblaya princeps* 的突变方向, 可以确定它们的 GC 突变偏倚是否具有普遍性, 或它们是否像大多数(或所有)其他细菌一样, 存在 AT 突变偏倚^[97,98]。因此, 其独特基因组组成的一个最明显的解释是全基因组范围内偏向 GC 的选择压力。

在细胞和细胞器基因组中, 多种遗传密码独立于“通用”密码^[107], UGA 从终止密码子重新分配到色氨酸密码子是最常见的密码变化之一。在细菌中, 这种密码子改变发现于柔膜菌纲的一个谱系中^[108], 这也是直到最近唯一报道的细菌编码重组事件。对来自昆虫共生细菌的微小基因组进行测序, 发现了这种终止子-色氨酸重编码的两个新案例: *Ca. Hodgkinia cicadicola*^[106]和 *Ca. Zinderia insecticola*^[99]。这是柔膜菌纲谱系以外的细菌中唯一发现的密码子重新分配的两个案例, 而 *Ca. Hodgkinia cicadicola* 是唯一已知的在高 GC 含量的基因组中发生 UGA 密码子重新分配的案例。这一发现, 以及基于几种线粒体基因组对密码子重新分配机制的分析^[109], 似乎与广泛引用的观点不符, 即低 GC 含量是密码子重新分配的先决条件(“密码子捕获”假说)^[110-112]。McCutcheon 等^[106]设想了一种基于“模糊翻译”的机制, 其中 tRNA^{Trp} 突变允许色氨酸(UGG)和终止(UGA)密码子的混杂解码, 从而允许通过持续的基因组缩减移除识别 UGA 密码子的肽链释放因子 2 (RF-2; 由 *prfB* 编码)。在此假设中, UGA 密码子重新分配是对关键基因 *prfB* 缺失的一种遗传共适应, 而不是为了使基因组(或翻译)更加精简的适应性事件^[113], 也不是通过 GC 含量偏倚导致的密码子丢失和重新分配的完全中性事件。

大部分缩减基因组菌株的生长速度与其原始菌株相当。其中一些菌株的生长速率和最终生物量甚至高于其亲本菌株。此外, 缩减基因组还具有诸如更高的转化效率^[79]和生物化学法生产^[80]等有利特性。少数情况下, 尽管与生长、细胞周期和形态相关的

基因未发生改变, 但缩减基因组表现出意想不到的表型, 如生长迟缓和异常细胞形态。最近一项研究提出了一种实验室适应进化(ALE)技术, 以改善缩减基因组大肠杆菌的生长表型^[85]。对进化菌株的多组学分析表明, 不平衡代谢通过 ALE 重组代谢扰动, 诱导生长延迟和转录组及翻译组重构。上述研究说明我们对基因功能、代谢和基因组的认识还不全面, 亟需对细菌基因组进行更全面的研究以填补我们知识空白, 解答基因组缩减生物体所呈现的表型特征。

6 结语与展望

回顾全文, 首先介绍了基于各种方法阐明必需基因集的研究, 并依此设计最小(或缩减)基因组。通过比较多个基因组对必需基因进行计算研究, 表明只有不到 300 个基因是细菌生存复制必需的。然而, 这种比较低估了必需基因的数量, 因为基因的功能可以被其他基因取代。利用转座子诱变、敲除、asRNA 和 CRISPRi 技术进行实验评估, 可揭示必需基因的直接和精确信息。但是, 由于上述实验方法使基因破坏或失活, 所以无法研究基因失活组合(如合成基因致死), 需要开发更精细的方法以阐明基因之间相互关联的上位相互作用。

然后, 介绍了以前构建的缩减基因组的特点及其构建方法。科学家们试图通过缩减已有基因组, 使之成为最小基因组。现有的研究结果证明, 缩减基因组是一把“双刃剑”, 一方面, 基因组缩减菌株可保留原始菌株优势特性, 甚至还具有诸如更高的转化效率和生物化学法生产等有利特性; 另一方面, 仍会出现意想不到的表型, 如生长迟缓和异常细胞形态等。这说明我们对基因组、基因功能和代谢的认识还不全面, 还需要对细菌基因组进行更全面的研究以填补知识空白。

现今, 高速发展的生物技术使设计基因网络, 生物合成途径乃至整个基因组的构建成为可能, 将遗传学和基因组学领域从描述性应用转移到了合成应用。在合成小型病毒基因组之后, DNA 组装和重写技术的进步使细菌基因组(例如生殖支原体)的层次化合成成为可能^[114], 并且通过将密码子的数量从 64 个减少到 57 个来重新编码大肠杆菌基因组。目前, 合成基因组学已经发展到合成整个真核基因组

的地步。合成酵母基因组计划(Sc2.0)正在进行中,旨在重写所有 16 个酿酒酵母染色体;到 2018 年,已经设计并合成了 6.5 条染色体。使用自下而上的装配并应用全基因组的改变将增进对基因组结构和功能的理解。这种方法不仅将为真核染色体的系统研究提供平台,还将产生可能适用于医学和工业应用的各种“简化”菌株。Sc2.0 的目的是设计和完全化学合成 16 条染色体,这些染色体包含来自啤酒酵母的 1250 万个碱基和一个带有所有 tRNA 基因的“新染色体”(http://syntheticyeast.org/)^[115]。该项目不仅将为真核染色体的系统研究提供一个平台,还将通过其“从构建到理解”的过程扩展生物学知识的范围。

从小型病毒基因组的合成到酵母染色体 Sc2.0,从罕见的密码子替代到全基因组重新编码,合成基因组学的可行性和能力已得到一次又一次的证明。Sc2.0 着重介绍了用于生产非天然药物和工业化合物(如青蒿素)的酵母“底盘”的开发。基于合成染色体,可以将多种异源途径直接整合到酵母基因组中。合成基因组学还使改组基因组以快速生成新的基因组并筛选所需的特性成为可能。

最后,合成最小基因组是合成基因组学领域的一个里程碑,高通量基因合成技术正在持续发展过程中,凭借其已被证实的优势,使得最小基因组生物成为科学、工业和许多其他应用的理想生物平台,相信随着系统和合成生物学前沿技术的不断发展,终将解决目前存在的许多障碍和挑战。

参考文献(References):

- [1] Bak AL, Black FT, Christiansen C, Freundt EA. Genome size of mycoplasmal DNA. *Nature*, 1969, 224(5225): 1209–1210.
- [2] Maniloff J, Morowitz HJ. Cell biology of the mycoplasmas. *Bacteriol Rev*, 1972, 36(3): 263–290.
- [3] Wallace DC, Morowitz HJ. Genome size and evolution. *Chromosoma*, 1973, 40(2): 121–126.
- [4] Woese CR, Maniloff J, Zablen LB. Phylogenetic analysis of the mycoplasmas. *Proc Natl Acad Sci*, 1980, 77(1): 494–498.
- [5] Weisburg WG, Woese CR, Dobson ME, Weiss E. A common origin of rickettsiae and certain plant pathogens. *Science*, 1985, 230(4725): 556–558.
- [6] Fraser CM. The minimal gene complement of *Mycoplasma genitalium*. *Science*, 1995, 270, 397–403.
- [7] Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UC, Podowski RM, Näslund AK, Eriksson AS, Winkler HH, Kurland CG. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, 1998, 396(6707): 133–140.
- [8] Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, 2000, 407(6800): 81–86.
- [9] Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, White O, Ketchum KA, Dodson R, Hickey EK, Gwinn M, Dougherty B, Tomb JF, Fleischmann RD, Richardson D, Peterson J, Kerlavage AR, Quackenbush J, Salzberg S, Hanson M, Vugt RV, Palmer N, Adams MD, Gocayne J, Weidman J, Utterback T, Watthey L, McDonald L, Artiach P, Bowman C, Garland S, Fuji C, Cotton MD, Horst K, Roberts K, Hatch B, Smith HO, Venter JC. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature*, 1997, 390(6660): 580–586.
- [10] McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*, 2011, 10(1): 13–26.
- [11] Moran NA. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA*, 1996, 93(7): 2873–2878.
- [12] Mira A, Ochman H, Moran NA. Deletional bias and the evolution of bacterial genomes. *Trends Genet*, 2001, 17(10): 589–596.
- [13] Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JCD, Andersson DI. Bacterial genome size reduction by experimental evolution. *Proc Natl Acad Sci USA*, 2005, 102(34): 12112–12116.
- [14] Mushegian AR, Koonin EV. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA*, 1996, 93(19): 10268–10273.
- [15] Mccutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*, 2011, 10(1): 13–26.
- [16] Moran NA, Mira A. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol*, 2001, 2(12): h51–h54.
- [17] Bennett GM, Moran NA. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biol Evol*, 2013, 5(9): 1545–1554.

- 1675–1688.
- [18] Moran NA. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA*, 1996, 93(7): 2873–2878.
 - [19] Mira A, Ochman H, Moran NA. Deletional bias and the evolution of bacterial genomes. *Trends Genet*, 2001, 17(10): 589–596.
 - [20] Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JCD, Andersson DI. Bacterial genome size reduction by experimental evolution. *Proc Natl Acad Sci USA*, 2005, 102(34): 12112–12116.
 - [21] Kuo CH, Ochman H. Deletional bias across the three domains of life. *Genome Biol. Evol*, 2009, 1: 145–152.
 - [22] Fares MA., Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E. Endosymbiotic bacteria: groEL buffers against deleterious mutations. *Nature*, 2002, 417(6887): 398.
 - [23] Fernandez A, Lynch M. Non-adaptive origins of interactome complexity. *Nature*, 2011, 474(7352): 502–505.
 - [24] Hidehiro T, Weiss BL, Perkin SAH, Yamashita A, Oshima K, Hattori M, Aksoy S. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res*, 2016, 16(2): 149–156.
 - [25] Ochman H, Davalos LM. The nature and dynamics of bacterial genomes. *Science*, 2006, 311(5768): 1730–1733.
 - [26] Burke GR, Moran NA. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol*, 2011, 3: 195–208.
 - [27] Itaya M. An estimation of minimal genome size required for life. *FEBS Lett*, 1995, 362(3): 257–260.
 - [28] Mushegian A. The minimal genome concept. *Curr Opin Genet Dev*, 1999, 9(6): 709–714.
 - [29] Koonin EV. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature Rev Microbiol*, 2003, 1(2): 127–136.
 - [30] Harris JK, Kelley ST, Spiegelman GB, Pace NR. The genetic core of the universal ancestor. *Genome Res*, 2003, 13(3): 407–412.
 - [31] Charlebois RL, Doolittle WF. Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res*, 2004, 14(12): 2469–2477.
 - [32] Koonin EV. How many genes can make a cell: The minimal-gene-set concept. *Annu Rev Genomics Hum Genet*, 2000, 1: 99–116.
 - [33] Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science*, 1999, 286(5447): 2165–2169.
 - [34] Akerley BJ, Rubin EJ, Novick VL, Amaya K, Judson N, Mekalanos JJ. A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc Natl Acad Sci USA*, 2002, 99(2): 966–971.
 - [35] Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Débarbouille M, Dervyn E, Deuerling E, Devine K, Devine SK, Dreesen O, Errington J, Fillinger S, Foster SJ, Fujita Y, Galizzi A, Gardan R, Eschevins C, Fukushima T, Haga K, Harwood CR, Hecker M, Hosoya D, Hullo MF, Kakeshita H, Karamata D, Kasahara Y, Kawamura F, Koga K, Koski P, Kuwana R, Imamura D, Ishimaru M, Ishikawa S, Ishio I, Coq DL, Masson A, Mauël C, Meima R, Mellado RP, Moir A, S Moriya, E Nagakawa, Nanamiya H, Nakai S, Nygaard P, Ogura M, Ohanan T, O'Reilly M, O'Rourke M, Pragai Z, Pooley HM, Rapoport, Rawlins JP, Rivas LA, Rivolta C, Sadaie A, Sadaie Y, Sarvas M, Sato, Saxild HH, Scanlan E, Schumann W, Seegers JFML, Sekiguchi J, Sekowska A, Séror SJ, Simon M, Stragier P, Studer R, Takamatsu H, Tanaka T, Takeuchi M, Thomaides HB, Vagner V, Diji JMV, Watabe K, Wipat A, Yamamoto, Yamamoto M, Yamamoto Y, Yamane K, Yata K, Yoshida K, Yoshikawa H, Zuber U, Ogasawara N. Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci USA*, 2003, 100(8): 4678–4683.
 - [36] Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, Smith HO, Venter JC. Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA*, 2006, 103(2): 425–430.
 - [37] Curnow AW, Hong KW, Yuan R, Kim SI, Martins O, Winkler W, Henkin TM, Söll D. Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc Natl Acad Sci USA*, 1997, 94(22), 11819–11826.
 - [38] Ghatak S, King ZA, Sastry A, Palsson BO. The y-ome defines the 35% of *Escherichia coli* genes that lack experimental evidence of function. *Nucleic Acids Res*, 2019, 47(5): 2446–2454.
 - [39] Riley M, Abe T, Arnaud MB, Berlyn MKB, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, Plunkett G, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D,

- Wanner BL. *Escherichia coli* K-12: a cooperatively developed annotation snapshot--2005. *Nucleic Acids Res*, 2006, 34(1): 1–9.
- [40] Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, 2006, 2: 2006.0008.
- [41] Moya A, Gil R, Latorre A, Peretó J, Garcillán-Barcia MP, Cruz FDL. Toward minimal bacterial cells: evolution vs. design. *FEMS Microbiol Rev*, 2009, 33(1): 225–235.
- [42] Fan L, Wu DF, Goremykin V, Xiao J, Xu YB, Garg S, Zhang CL, Martin WF, Zhu RX. Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. *Nat Ecol Evol*, 2020, 4(9): 1213–1219.
- [43] Gray MW. Mitochondrial evolution. *Cold Spring Harb Perspect Biol*, 2012, 4(9): a11403.
- [44] Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SGE. Computational inference of scenarios for α -proteobacterial genome evolution. *Proc Natl Acad Sci USA*, 2004, 101(26): 9722–9727.
- [45] Mizoguchi H, Sawano Y, Kato JI, Mori H. Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome. *DNA Res*, 2008, 15(5): 277–284.
- [46] Stewart GC. Taking shape: control of bacterial cell wall biosynthesis. *Mol Microbiol*, 2005, 57(5): 1177–1181.
- [47] Silverman DJ, Wisseman Jr CL, Waddell A. *In vitro* studies of Rickettsia-host cell interactions: ultrastructural study of Rickettsia prowazekii-infected chicken embryo fibroblasts. *Infect Immun*, 1980, 29(2): 778–790.
- [48] Tully JG, Taylor-Robinson D, Cole RM, Rose DL. A newly discovered mycoplasma in the human urogenital tract. *Lancet*, 1981, 1(8233): 1288–1291.
- [49] Schröder D, Deppisch H, Obermayer M, Krohne G, Stackebrandt E, Hölldobler B, Goebel W, Gross R. Intracellular endosymbiotic bacteria of Camponotus species (carpenter ants): systematics, evolution and ultrastructural characterization. *Mol Microbiol*, 1996, 21(3): 479–489.
- [50] Aksoy S. Wigglesworthia gen. nov. and Wigglesworthia glossinidia sp. nov., Taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *Int J Syst Bacteriol*, 1995, 45(4): 848–851.
- [51] Moran NA, Dale C, Dunbar H, Smith WA, Ochman H. Intracellular symbionts of sharpshooters (insecta: hemiptera: cicadellinae) form a distinct clade with a small genome. *Environ Microbiol*, 2003, 5(2): 116–126.
- [52] Griffiths GW, Beck SD. Effects of antibiotics on intracellular symbionts in the pea aphid, Acyrthosiphon pisum. *Cell Tissue Res*, 1974, 148(3): 287–300.
- [53] von Dohlen CD, Kohler S, Alsop ST, McManus WR. Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature*, 2001, 412(6845): 433–436.
- [54] Gómez-Valero L, Soriano-Navarro M, Pérez-Brocail V, Heddi A, Moya A, García-Verdugo JM, Latorre A. Coexistence of Wolbachia with Buchnera aphidicola and a secondary symbiont in the aphid Cinara cedri. *J. Bacteriol*, 2004, 186(19): 6626–6633.
- [55] Moran NA, Tran P, Gerardo NM. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the Bacterial phylum Bacteroidetes. *Appl Environ Microbiol*, 2005, 71(12): 8802–10.
- [56] McCutcheon JP, Moran NA. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol. Evol*, 2010, 2: 708–718.
- [57] Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M. The 160-Kilobase genome of the bacterial endosymbiont carsonella. *Science*, 2006, 314(5797): 267.
- [58] McCutcheon JP, McDonald BR, Moran NA. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet*, 2009, 5(7): e1000565.
- [59] Mushegian AR, Koonin EV. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA*, 1996, 93(19): 10268–10273.
- [60] Harris JK, Kelley ST, Spiegelman GB, Pace NR. The genetic core of the universal ancestor. *Genome Res*, 2003, 13(3): 407–412.
- [61] Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ. Universal trees based on large combined protein sequence data sets. *Nat Genet*, 2001, 28(3): 281–285.
- [62] Koonin EV. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol*, 2003, 1(2): 127–136.
- [63] Charlebois RL, Doolittle WF. Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res*, 2004, 14(12): 2469–2477.
- [64] Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC. Global transposon mutagenesis and a minimal Mycoplasma genome. *Science*, 1999, 286(5447): 2165–2169.

- [65] Itaya M. An estimation of minimal genome size required for life. *FEBS Lett*, 1995, 362(3): 257–260.
- [66] Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Débarbouille M, Dervyn E, Deuerling E, Devine K, Devine SK, Dreesen O, Errington J, Fillinger S, Foster SJ, Fujita Y, Galizzi A, Gardan R, Eschevins C, Fukushima T, Haga K, Harwood CR, Hecker M, Hosoya D, Hullo MF, Kakeshita H, Karamata D, Kasahara Y, Kawamura F, Koga K, Koski P, Kuwana R, Imamura D, Ishimaru M, Ishikawa S, Ishio I, Le Coq D, Masson A, Mauël C, Meima R, Mellado RP, Moir A, Moriya S, Nagakawa E, Nanamiya H, Nakai S, Nygaard P, Ogura M, Ohanan T, O'Reilly M, O'Rourke M, Pragai Z, Pooley HM, Rapoport G, Rawlins JP, Rivas LA, Rivolta C, Sadaie A, Sadaie Y, Sarvas M, Sato T, Saxild HH, Scanlan E, Schumann W, Seegers JFML, Sekiguchi J, Sekowska A, Séror SJ, Simon M, Stragier P, Studer R, Takamatsu H, Tanaka T, Takeuchi M, Thomaidis HB, Vagner V, van Dijl JM, Watabe K, Wipat A, Yamamoto H, Yamamoto M, Yamamoto Y, Yamane K, Yata K, Yoshida K, Yoshikawa H, Zuber U, Ogasawara N. Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci USA*, 2003, 100(8): 4678–4683.
- [67] Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, 2006, 2: 2006.0008.
- [68] Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, Smith HO, Venter JC. Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA*, 2006, 103(2): 425–430.
- [69] Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science*, 1999, 286(5447): 2165–2169.
- [70] Ji Y, Zhang B, Van SF, Warren HP, Woodnutt G, Burnham MK, Rosenberg M. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science*, 2001, 293(5538): 2266–2269.
- [71] Gerdes SY, Scholle MD, Campbell JW, Balázsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, Bhattacharya A, Kapatral V, D'Souza M, Baev MV, Grechkin Y, Mseeh F, Fonstein MY, Overbeek R, Barabási AL, Oltvai ZN, Osterman AL. Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol*, 2003, 185(19): 5673–5684.
- [72] Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 2013, 152(5): 1173–1183.
- [73] Rousset F, Cui L, Siouvé E, Becavin C, Depardieu F, Bikard D, Blokesch M. Genome-wide CRISPR-dCas9 screens in *E. coli* identify essential genes and phage host factors. *PLoS Genet*, 2018, 14(11): e1007749.
- [74] Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen T, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, 2014, 343(6166): 84–87.
- [75] Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, 2014, 343(6166): 80–84.
- [76] Yu BJ, Kim C. Minimization of the *Escherichia coli* genome using the Tn5-targeted Cre/loxP excision system. *Methods Mol Biol*, 2008, 416: 261–277.
- [77] Nijman SMB. Synthetic lethality: General principles, utility and detection using genetic screens in human cells. *FEBS Lett*, 2011, 585(1): 1–6.
- [78] Kolisnychenko V, Plunkett G, Herring CD, Fehér T, Pósfai J, Blattner FR, Pósfai G. Engineering a reduced *Escherichia coli* genome. *Genome Res*, 2002, 12(4): 640–647.
- [79] Pósfai G, Plunkett G, Fehér T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR. Emergent properties of reduced-genome *Escherichia coli*. *Science*, 2006, 312(5776): 1044–1046.
- [80] Lee JH, Sung BH, Kim MS, Blattner FR, Yoon BH, Kim JH, Kim SC. Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. *Microb Cell Fact*, 2009, 8: 2.
- [81] Park MK, Lee SH, Yang KS, Jung SC, Lee JH, Kim SC. Enhancing recombinant protein production with an *Escherichia coli* host strain lacking insertion sequences. *Appl Microbiol Biotechnol*, 2014, 98(15): 6701–6713.
- [82] Hashimoto M, Ichimura T, Mizoguchi H, Tanaka K, Fujimitsu K, Keyamura K, Ote T, Yamakawa T, Yamazaki Y, Mori H, Katayama T, Kato JI. Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol Microbiol*, 2005, 55(1): 137–149.
- [83] Westers H, Dorenbos R, van Dijl JM, Kabel J, Flanagan T, Devine KM, Jude F, Seror SJ, Beekman AC, Darmon

- E, Eschevins C, de Jong A, Bron S, Kuipers OP, Albertini AM, Antelmann H, Hecker M, Zamboni N, Sauer U, Bruand C, Ehrlich DS, Alonso JC, Salas M, Quax WJ. Genome engineering reveals large dispensable regions in *Bacillus subtilis*. *Mol Biol Evol*, 2003, 20(12): 2076–2090.
- [84] Biswas I, Gruss A, Ehrlich SD, Maguin E. High-efficiency gene inactivation and replacement system for gram-positive bacteria. *J Bacteriol*, 1993, 175(11): 3628–3635.
- [85] Choe D, Lee JH, Yoo M, Hwang S, Sung BH, Cho S, Palsson B, Kim SC, Cho BK. Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nat Commun*, 2019, 10(1): 935.
- [86] Reuß DR, Altenbuchner J, Mäder U, Rath H, Ischebeck T, Sappa PK, Thürmer A, Guérin C, Nicolas P, Steil L, Zhu BY, Feussner I, Klumpp S, Daniel R, Commichau FM, Völker U, Stülke J. Large-scale reduction of the *Bacillus subtilis* genome: consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res*, 2017, 27(2): 289–299.
- [87] Ara K, Ozaki K, Nakamura K, Yamane K, Sekiguchi J, Ogasawara N. Bacillus minimum genome factory: effective utilization of microbial genome information. *Biotechnol Appl Biochem*, 2007, 46(Pt 3): 169–178.
- [88] Morimoto T, Kadota R, Endo K, Tohata M, Sawada K, Liu SG, Ozawa T, Kodama T, Kakeshita H, Kageyama Y, Manabe K, Kanaya S, Ara K, Ozaki K, Ogasawara N. Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. *DNA Res*, 2008, 15(2): 73–81.
- [89] Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc Natl Acad Sci USA*, 2010, 107(6): 2646–2651.
- [90] Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science*, 1999, 286(5447): 2165–2169.
- [91] Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, Smith HO, Venter JC. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 2010, 329(5987): 52–56.
- [92] Hutchison CA, 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. *Science*, 2016, 351(6280): aad6253.
- [93] Sleator RD. The story of *Mycoplasma mycoides* JCVI-syn1.0: the forty million dollars microbe. *Bioeng Bugs*, 2010, 1(4): 229–230.
- [94] Lajoie MJ, Kosuri S, Mosberg JA, Gregg CJ, Zhang D, Church GM. Probing the limits of genetic recoding in essential genes. *Science*, 2013, 342(6156): 361–363.
- [95] Fredens J, Wang KH, de la Torre D, Funke LFH, Robertson WE, Christova Y, Chia T, Schmied WH, Dunkelmann DL, Beránek V, Uttamapinant C, Llamazares AG, Elliott TS, Chin JW. Total synthesis of *Escherichia coli* with a recoded genome. *Nature*, 2019, 569(7757): 514–518.
- [96] Rocha EP, Feil EJ. Mutational patterns cannot explain genome composition: Are there any neutral sites in the genomes of bacteria? *PLoS Genet*, 2010, 6(9): e1001104.
- [97] Hildebrand F, Meyer A, Eyre-Walker A. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet*, 2010, 6(9): e1001107.
- [98] Hershberg R, Petrov DA. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet*, 2011, 6(9): e1001115.
- [99] McCutcheon JP, Moran NA. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol Evol*, 2010, 2: 708–718.
- [100] Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science*, 2006, 314(5797): 267.
- [101] Moran NA. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA*, 1996, 93(7): 2873–2878.
- [102] Rocha EP, Danchin A. Base composition bias might result from competition for metabolic resources. *Trends Genet*, 2002, 18(6): 291–294.
- [103] Bentley SD, Parkhill J. Comparative genomic structure of prokaryotes. *Annu Rev Genet*, 2004, 38: 771–792.
- [104] Lind PA, Andersson DI. Whole-genome mutational biases in bacteria. *Proc Natl Acad Sci USA*, 2008, 105(46): 17878–17883.
- [105] McCutcheon JP, von Dohlen CD. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol*, 2011, 21(16): 1366–1372.
- [106] McCutcheon JP, McDonald BR, Moran NA. Origin of

- an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet*, 2009, 5(7): e1000565.
- [107] Knight RD, Freeland SJ, Landweber LF. Rewiring the keyboard: evolvability of the genetic code. *Nature Rev. Genet*, 2001, 2(1): 49–58.
- [108] Maniloff J. Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol Rev*, 1998, 62(4): 1094–156.
- [109] McCutcheon JP, McDonald BR, Moran NA. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet*, 2009, 5(7): e1000565.
- [110] Osawa S, Jukes TH. Evolution of the genetic code as affected by anticodon content. *Trends Genet*, 1988 4(7), 191–198.
- [111] Osawa S, Jukes TH, Watanabe K, Muto A. Recent evidence for evolution of the genetic code. *Microbiol. Rev*, 1992, 56(1): 229–264.
- [112] Santos MA, Moura G, Massey SE, Tuite MF. Driving change: the evolution of alternative genetic codes. *Trends Genet*, 2004, 20(2): 95–102.
- [113] Andersson SG, Kurland CG. Genomic evolution drives the evolution of the translation system. *Biochem. Cell Biol*, 1995, 73(11–12): 775–787.
- [114] Liang QF, Wang Q, Qi QS. Synthetic biology and rearrangements of microbial genetic material. *Hereditas (Beijing)*, 2011, 33(10): 1102–1112.
梁泉峰, 王倩, 祁庆生. 合成生物学与微生物遗传物质的重构. *遗传*, 2011, 33(10): 1102–1112.
- [115] Xu HM, Xie ZX, Liu D, Wu Y, Li BZ, Yuan YJ. Design and synthesis of yeast chromosomes. *Hereditas(Beijing)*, 2017, 39(10): 865–876.
徐赫鸣, 谢泽雄, 刘夺, 吴毅, 李炳志, 元英进. 酿酒酵母染色体设计与合成研究进展. *遗传*, 2017, 39(10): 865–876.

(责任编辑: 谢建平)