

## 特约评述

DOI: 10.12211/2096-8208.2020-050

## 微生物细胞工厂的设计构建：从诱变育种到全基因组定制化创制

袁姚梦<sup>1</sup>, 邢新会<sup>1,2</sup>, 张翀<sup>1</sup>

(<sup>1</sup> 工业生物催化教育部重点实验室, 清华大学化工系生物化工研究所, 清华大学合成与系统生物学研究中心, 北京 100084; <sup>2</sup> 清华大学深圳国际研究生院生物医药与健康工程研究院, 广东 深圳 518055)

**摘要:** 微生物细胞工厂 (microbial cell factories, MCFs) 被广泛用于生产丰富多样的化学品、食品、药品和能源, 是绿色生物制造的核心环节。早期主要通过天然微生物的筛选和诱变育种的方式获得高产菌种, 然而作为一种“以时间(人力)换水平”的非理性策略, 其创制效率极低。随着分子生物学和基因工程研究方法的不断发展, 对微生物系统认知和改造能力的进步促使代谢工程学科诞生。基于生物学知识的理性/半理性代谢工程设计和构建策略, 目前已发展了从分子、途径到基因组层次不同的MCFs设计和工程化构建策略。本文结合实际案例对MCFs的设计及构建策略进行综述, 首先回顾传统诱变育种和代谢工程指导的理性/半理性设计策略, 探讨如何突破代谢工程经典框架的限制, 实现全基因组水平定制化MCFs的快速构建, 最后对这一新的构建范式的未来进行展望。

**关键词:** 生物制造; 微生物细胞工厂; 代谢工程; 基因型-表型关联; 数据驱动

**中图分类号:** Q81 **文献标志码:** A

## Progress and prospective of engineering microbial cell factories: from random mutagenesis to customized design in genome scale

YUAN Yaomeng<sup>1</sup>, XING Xinhui<sup>1,2</sup>, ZHANG Chong<sup>1</sup>

(<sup>1</sup>MOE Key Laboratory for Industrial Biocatalysis, Institute of Biochemical Engineering, Department of Chemical Engineering, Center for Synthetic & Systems Biology, Tsinghua University, Beijing 100084, China; <sup>2</sup>Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, Shenzhen 518055, Guangdong, China)

**Abstract:** As the core of green bio-manufacturing and bioeconomy, microbial cell factories (MCFs) are widely used to produce a variety of chemicals, foods, medicines and fuels. In the early years, isolation and random mutagenesis of natural microbes were time-consuming but widely used for developing well-performed MCFs. With the development of molecular biology and genetic engineering, the advancements in understanding of microbial systems prompted the establishment of metabolic engineering. Nowadays, different MCF-construction strategies in terms of protein, pathway and genome-wide engineering have been well developed based on rational or semi-rational metabolic engineering strategies. However, due to limited biological knowledge, these strategies mainly rely on the iterative cycle of ‘Design-Build-Test-Learning (DBTL)’,

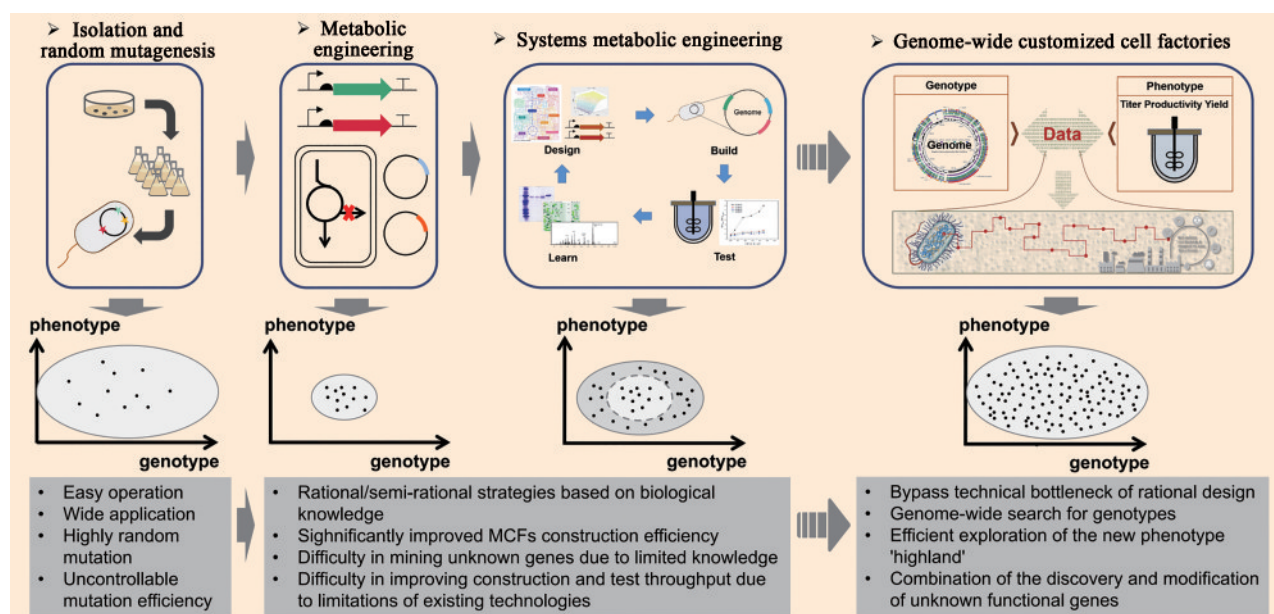
收稿日期: 2020-04-16 修回日期: 2020-09-26

基金项目: 国家自然科学基金重点项目 (2193000018); 国家重点研发计划 (2019YFA0904802)

引用本文: 袁姚梦, 邢新会, 张翀. 微生物细胞工厂的设计构建: 从诱变育种到全基因组定制化创制[J]. 合成生物学, 2020, 1(6): 656-673

Citation: YUAN Yaomeng, XING Xinhui, ZHANG Chong. Progress and prospective of engineering microbial cell factories: from random mutagenesis to customized design in genome scale[J]. Synthetic Biology Journal, 2020, 1(6): 656-673

usually taking 50–300 person-years and hundreds of millions of dollars to develop a MCF that can meet industrial demands. Combining high-throughput genome editing and phenotype screening and selection technologies, the genome-wide customized engineering allows one to obtain large-scale genotype-phenotype association (GPA) data sets quickly. Based on these results, data science technologies can be further applied to mine a large number of unknown genes or loci associated with the specific phenotypes. This strategy has a wider search scope for genotype (genome-wide) and does not rely on existing biological knowledge (data-driven), thus making it possible to explore phenotypes that could not be achieved by the above mentioned rational/semi-rational strategies and develop MCFs with superior performance more efficiently. This paper reviews general strategies and application cases for the design and construction of MCFs. We will firstly summarize the overview of random mutagenesis strategies, and the history and latest progress in metabolic engineering for the construction of MCFs. Then we discuss the potential of the newly emerging MCF-design and construction paradigm, and the customized design strategies at the whole genome scales. Finally, we conclude with our perspectives on the development of novel strategies for the engineering of MCFs.



**Keywords:** Bio-manufacturing; microbial cell factories; metabolic engineering; genotype-phenotype associations; data-driven

以石油等不可再生化石资源为原料的传统制造业面临着“高能耗、高污染、高排放”等严峻挑战，亟需通过生产原料、加工过程和产品创制的绿色变革，实现未来的可持续发展。绿色生物制造以可再生资源为原料，利用生物催化和转化过程，实现高附加值产品的绿色制造，由于其能够实现工业制造模式从“末端治理”到“源头控制”、从“低端产品”到“高值化转化”的转变，成为推动生物经济发展的重要路径之一。

微生物细胞工厂 (microbial cell factories,

MCFs) 是绿色生物制造的核心环节。利用微生物强大且多样的生化反应网络，通过对代谢路径的重塑和工程化，可以将微生物细胞改造为能够以低价值可再生资源为原料生产各类产品的MCFs。从青霉素<sup>[1]</sup>、谷氨酸<sup>[2]</sup>、乙醇、1,3-丙二醇<sup>[3]</sup>、法尼烯<sup>[4]</sup>到阿片类药物<sup>[5]</sup>等生物合成MCFs的成功开发，展示了MCFs合成高附加值化学品的巨大潜力。迄今为止，MCFs已能够生产抗生素、氨基酸、重组蛋白、生物能源<sup>[6]</sup>、生物塑料乃至“人造肉”，被广泛应用于制药、食品、能源和农业等

领域<sup>[7-12]</sup>。

MCFs的构建策略经历了不同的历史阶段。在20世纪90年代之前,主要通过天然微生物的筛选和非理性诱变育种技术获得目标产物高产菌株。这种将随机突变和定向筛选相结合的策略在工业菌株的开发上已经有诸多经典成功案例<sup>[13-15]</sup>。然而,由于突变过程的随机性,这种策略往往花费时间长、工作量大,是一种典型的“以时间(人力)换水平”的策略。尽管如此,由于其具有应用历史悠久、操作简单、适用范围广、属于非转基因操作等优势,非理性诱变育种策略至今仍然是微生物育种研究和产业应用的常用平台技术。

20世纪90年代以来,随着分子生物学、基因工程技术的逐步引入,代谢工程学科正式创立<sup>[16-17]</sup>。代谢工程利用重组DNA技术对生物体中已知的代谢途径进行有目的的设计,以更好地理解 and 利用细胞途径,并对细胞内的基因网络和调节过程进行调控和优化,构建具有特定功能的MCFs,例如,提高现有产品的产率,生产新的产品和扩大可用的底物范围<sup>[16-18]</sup>。按照代谢工程的理论框架,所构建的MCFs需要满足<sup>[19]</sup>:第一,通过对代谢网络物质流和能量流的设计,实现底物消耗和目标产物生产的最优化;第二,改造后的MCFs能够适应工业生产环境,满足工业生产需求。

随着分子生物学和反向遗传工程手段的发展,人们对于微生物代谢网络及其调控机制的认识得到了巨大的进步,然而,由于微生物系统的复杂性,仍然不能完全理性地设计满足工业需求的MCFs。目前,利用代谢工程策略开发MCFs的流程主要依靠“设计-构建-测试-学习”(design-build-test-learning, DBTL)循环<sup>[20]</sup>,通常从MCFs概念设计(proof of concept)到满足实际应用,需要50~300人年和数亿美元的投入<sup>[21]</sup>,耗时耗力、投入高。在代谢工程MCFs创制的DBTL循环中,“理性设计”始终是人们的理想目标,也由此应运而生诸多代谢网络模型和设计策略。然而,经典代谢工程指导的设计方法大多都基于已知的生物学知识,由于微生物代谢网络中存在诸多可能对目标表型产生影响的未知因素,或称为“生命暗物质”,这一手段仍然存在诸多局限<sup>[22-25]</sup>。生物信息学和各种组学技术的快速发展,使得研究者逐

渐有能力从系统代谢工程的层次思考MCFs的快速创制,例如,基于全基因组改组技术<sup>[26]</sup>、新型诱变技术<sup>[27]</sup>、单敲除库筛选技术<sup>[25]</sup>、基因组文库共存/共表达(coexisting/coexpressing genomic libraries, CoGeLs)<sup>[28]</sup>等技术,结合组学和生物信息学手段获取生物学知识,以进一步指导MCFs的设计。这一理性/半理性结合的策略为突破知识局限,获取更为高效的MCFs提供了重要的手段,然而,系统代谢工程手段获取知识的效率不高,MCFs改造过程仍然需要耗费大量的时间和精力。

近年来,合成生物学的进步使得MCFs构建和测试的能力得到显著提升,为提高构建效率以满足市场快速变化和多样的需求提供了重要的机遇。另一方面,二代测序(next generation sequencing, NGS)和基因组编辑的技术飞跃,使得从全基因组层次设计和构建MCFs成为可能。利用高通量研究技术,目前已经可以从全基因组层次并行研究微生物特定表型与基因型的关系,从而获得大规模的基因型-表型关联(genotype phenotype associations, GPA)数据集<sup>[29]</sup>。如果能够利用这些大规模GPA数据集,基于数据科学手段从全基因组范围深度挖掘传统分子生物学手段无法发现的未知关联基因及其位点,将有可能从数据(学习)的角度绕开理性设计的知识瓶颈,为提高MCFs设计和创建效率提供全新的研究范式。此外,由于上述数据驱动的全基因组规模定制工程策略基因型搜索范围更宽(全基因组),不依赖于现有知识(数据驱动),将有可能探索之前理性/半理性所无法达到的表型“高地”,获得生产效率更为高效、生产性能更加优越的下一代定制化MCFs。

本文将结合实际案例对MCFs的设计及构建策略进行综述,首先回顾传统诱变育种和代谢工程指导的理性/半理性设计策略,接着探讨如何突破代谢工程经典框架限制,利用全基因组规模定制工程实现全基因组水平定制化MCFs的快速构建,最后将对这一新的MCFs构建范式的未来进行展望(图1)。

## 1 诱变育种技术在MCFs构建中的应用

诱变育种(mutation breeding)是在人为条件下,利用物理、化学、生物因素,诱发生物体产

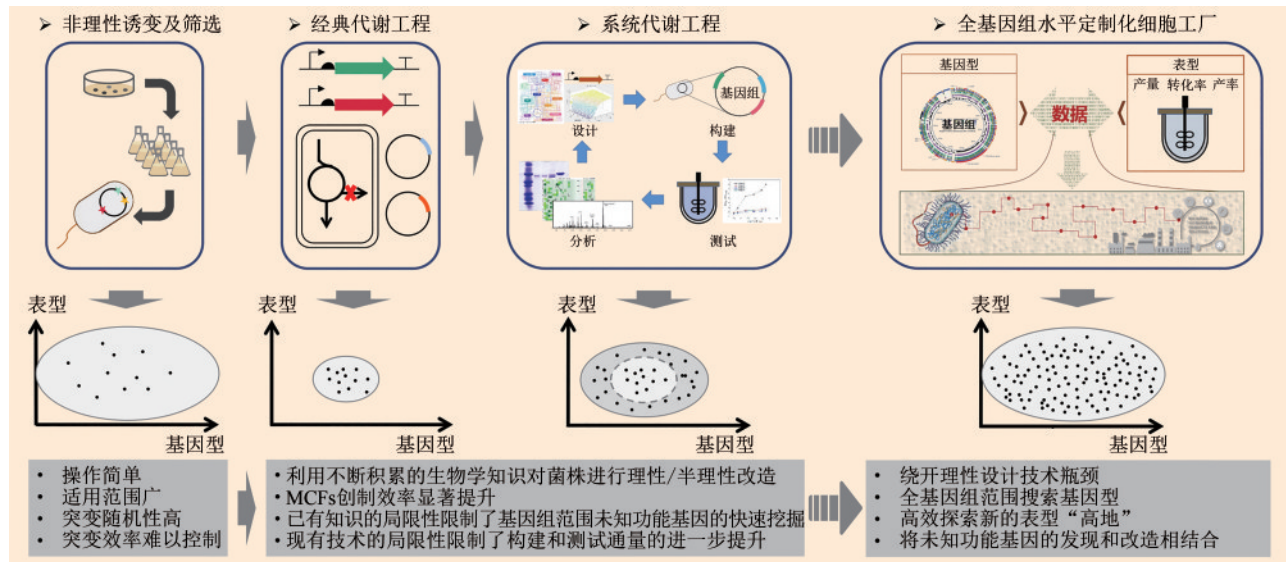


图1 微生物细胞工厂 (MCFs) 设计和构建发展历程及未来展望

Fig. 1 Development and future prospects of MCFs design and construction

生突变，从中选择、培育植物和微生物新品种的方法，长期以来在科学研究和生物产业中得到广泛应用。诱变育种手段通常包括物理诱变、化学诱变和生物诱变<sup>[30]</sup>（表1）。物理诱变主要采用电离辐射和非电离辐射等物理因素诱发变异；化学诱变主要利用烷化剂、碱基类似物、移码诱变剂、脱氨剂和羟化剂等化学物质诱发变异；生物诱变主要包括噬菌体、质粒、DNA转座子诱变和原生质体融合、DNA改组、基因组重排等能够显著提高基因重组频率的诱变技术。诱变产生的突变随机性大、且在全基因组范围分布稀疏，存在大量的无义突变，因此，采用诱变育种技术获得理想性状的微生物往往花费时间长、工作量大，是一种典型的“以时间（人力）换水平”的策略。尽管如此，由于诱变育种技术采用非理性的手段，理论上不需要任何先验的微生物代谢网络结构及其调控的知识，对于没有基因操作手段的微生物，或者具有非转基因（non-GMO）需求的应用领域（如食品），在20世纪90年代之前一直被作为MCFs开发的最重要的手段，目前也仍然是微生物技术科研和发酵产业最为常用的育种手段之一。一个典型的例子是青霉素菌株的选育，1943年，研究者从霉甜瓜中分离得到一株产黄青霉 NRRL-1951，其青霉素产量为60 mg/L，在长达50年的人工选育后，产黄青霉的青霉素产量已经达到70 g/L<sup>[31]</sup>。

近年来，多种新型诱变技术的发展也为诱变育种注入了新的活力。例如离子注入诱变、等离子体诱变等<sup>[30]</sup>（表1），这些新的诱变技术具有突变率高、变异范围广、变异稳定等优点。离子注入诱变要求严格的真空环境和安全防护措施<sup>[30]</sup>，其推广使用受到一定程度的限制。近年来新发展起来的常压室温等离子体诱变技术（ARTP）作为新型物理诱变方法，具有突变率高、突变速度快、突变库容大、安全性高、可控性好、操作简单等优点，在微生物诱变育种中得到广泛的应用<sup>[27]</sup>。

## 2 代谢工程指导的理性、半理性设计策略构建MCFs

20世纪90年代，随着分子生物学的发展，微生物代谢网络的物质流、能量流以及复杂的调控机制得到了更为深入的研究，基于生物学知识的理性设计作为代谢工程指导的经典设计策略逐渐被应用于MCFs构建领域。James E. Bailey<sup>[16]</sup>最早提出对代谢网络进行理性设计，并利用基因工程的手段对其进行改造以实现代谢流量分布的优化和目标产物产量的提升的设想。Stephanopoulos<sup>[17-18]</sup>进一步指出，微生物的代谢网络中具有诸多“刚性”节点，能够通过复杂的调控机制抵抗遗传扰动带来的通量

表1 常见物理、化学、生物诱变技术汇总

Tab. 1 Summary of mutagenesis technologies

分类	诱变技术/诱变剂	描述	参考文献
物理诱变	电离辐射(X射线、 $\gamma$ 射线等)	引起DNA双链或单链断裂,实现DNA的删除或结构改变	[30]
	非电离辐射(紫外线)	使嘧啶形成二聚体,实现GC的删除、移码突变以及GC $\rightarrow$ AT的转换	[32]
化学诱变	烷化剂(烷基磺酸盐、芥子气等)	使DNA碱基发生烷化,导致DNA复制时发生配对错误	[32]
	碱基类似物 (嘧啶类似物、嘌呤类似物)	与DNA碱基结构类似,在DNA复制时掺入并引发配对错误	[30]
	移码诱变剂(原黄素、吖啶橙等)	与DNA结合导致碱基增添或缺失	[30]
	脱氨剂(亚硝酸)	引起A、C、G碱基的脱氨,实现GC与AT的相互转换;引起DNA交联作用,引发突变	[33]
	羟化剂(羟胺)	引起胞嘧啶脱氨,实现GC $\rightarrow$ AT的转换	[35]
	生物诱变	噬菌体、质粒和DNA转座子	通过碱基取代和DNA链断裂实现碱基的删除、重复和插入
原生质体融合		将两个亲株的原生质体进行融合,形成杂合二倍体,使两个亲株发生基因组重组	[31]
DNA改组		对多个同源序列组成的文库进行随机片段化,再利用PCR使其发生随机的重组,实现多亲本的基因组重组	[34]
基因组重排		先利用传统诱变手段获得多个表型改进的菌株,然后模拟DNA改组的反应条件,对原生质体进行多次递推式融合,实现正向突变的富集	[26]
复合诱变		结合多种诱变方式提高诱变效率	[30]
新型诱变技术	离子注入诱变	离子注入细胞导致DNA损伤,细胞修复损伤的过程中出现突变	[30]
	等离子体诱变	等离子体作用于细胞造成DNA损伤,细胞修复损伤的过程中出现突变	[27,30]

变化。从代谢网络分析的角度看,代谢网络的理性设计面临三个挑战<sup>[18]</sup>:

(1) 识别代谢网络中流量分配显著影响产品合成的关键节点。多数工作仅调控位于产物合成途径末端的反应,取得的成就十分有限。事实上,来源于中心代谢的合成前体、辅因子和能量供应对产物合成也具有重要的作用,识别其中的关键节点是显著提高优化效率的基础。

(2) 确定关键节点处最合适的基因改造类型。包括过表达目标产物合成途径、抑制副产物生成酶活性、解除刚性关键节点的调控机制等。代谢网络的流量分布需要严格而小心的调控,确定合适的流量调控类型和程度,是优化MCFs的必要步骤。

(3) 优化中心代谢流量分布之后。需要识别产品合成途径中所有的代谢控制位点并对其进行精准调控,同时对潜在竞争途径中的酶活性加以平衡,以获得产量提升、性能稳定的MCFs。

在经典代谢工程领域,多种代谢网络分析和设计方法的不断提出和完善为代谢网络的理性设计提供了理论和技术支撑(表2)。这些分析方法主要包括:动力学分析法,可用于细胞内酶促反应动力学模型的建立和分析;代谢网络分析方法,可用于确定某一目标函数下代谢通量的唯一解或进行途径分析,包括代谢通量分析、通量平衡分析和代谢途径分析三类分析方法;整合了热力学的代谢网络分析,即以热力学第二定律对所有代

谢反应加以约束,例如网络嵌入的热力学分析(NET)。随着计算技术的进步和更多代谢模型的建立,多种应变优化算法及软件得到逐步开发和应

用,例如基于约束的重构与分析(COBRA)。在此基础上,研究者逐步阐明了大肠杆菌、谷氨酸棒杆菌、酿酒酵母等微生物的核心代谢网络,并提出了一系列可用于通量平衡分析的化学计量学模型。早期代谢工程领域的代谢途径设计几乎都基于核心代谢网络模型,随着系统生物学的发展,基因组规模代谢重构(genome-scale metabolic reconstruction, GSMR)的概念最先在大肠杆菌中被提出<sup>[50]</sup>,随后,基因组规模代谢模型(genome-scale metabolic model, GSMM)的概念进一步推动了大量基因组规模实验数据与数学建模的结合,增进了人们对于基因组规模代谢的理解<sup>[51-52]</sup>,MCFs的代谢模型构建规模也逐渐从核心代谢规模上升到基因组规模。Nielsen和Palsson课题组首次提出了酿酒酵母的代谢模型,这也是真核生物第一个基因组规模代谢模型<sup>[53]</sup>。随着NGS技术、组学分析技术的快速发展,越来越多的微生物代谢模型被不断提出和完善(表3)。

代谢网络全基因模型及其计算机优化模拟算法的开发极大地提高了人们对代谢网络的设计和分析能力,这种从系统层次对微生物代谢网络进行理性设计的系统代谢工程方法论逐渐成为MCFs改造领域的研究热点。2011年,Becker等<sup>[73]</sup>通过基因组

表2 用于代谢网络设计的分析方法

Tab. 2 Analytical methods for metabolic network design

类别	分析方法	描述	参考文献
动力学分析	ODE&米氏方程	利用常微分方程(ODE)和米氏方程,能够描述胞内代谢物浓度随时间的变化趋势,从而建立代谢网络的动力学模型	[36]
	代谢控制分析(MCA)	MCA可用于估算动力学模型中各个通量的控制系数,从而确定代谢途径中需要过表达的目标基因,增加通过途径的通量	[37]
代谢网络分析	代谢通量分析(MFA)	MFA根据胞内代谢物的质量平衡确定线性方程组并求解,能够计算特定培养条件下细胞内的实际代谢通量分布。MFA需要依赖大量实验数据来增加可测量通量的数量,从而计算出不可测量通量向量	[38]
	通量平衡分析(FBA)	FBA可用于确定细胞代谢网络中每个反应的最佳通量,该方法基于凸分析,通过对系统施加最大化(最小化)的目标函数来确定代谢通量矢量	[39-40]
	代谢途径分析(MPA)	MPA可用于识别代谢网络中存在的所有代谢通量向量,该方法仅以化学计量学和反应热力学为约束条件,不需要动力学参数进行计算(例如基元模式分析)	[41-42]
整合了热力学的代谢网络分析	网络嵌入的热力学分析(NET)	以热力学第二定律为依据,可用于检测代谢组学数据和假定的通量方向是否符合热力学一致性	[43]
	基于热力学的代谢通量分析(TMFA)	TMFA同时使用热力学方向性约束和质量守恒约束计算代谢通量分布	[44]
计算机应变优化算法	基于约束的重构与分析(COBRA)	COBRA采用基因组规模的计算机模拟,可用于代谢途径预测和优化,从而改善生产速率和产量	[45]
	最小化代谢调节(MOMA)	MOMA用于使野生型菌株和缺失突变体之间代谢通量分布的差别最小化,能够预测基因操作对代谢网络的影响	[46]
	开关最小化调节(ROOM)	与MOMA的优化目标相同	[47]
	OptKnock	以产品产率为优化目标的基因敲除分析工具	[48]
	生化网络集成计算浏览器(BNICE)	使用广义酶反应规则发现新代谢途径的计算工具	[49]

规模的计算模拟,仅通过12处基因改造就将谷氨酸棒杆菌内的代谢流重新定向为合成L-赖氨酸的最优途径,该基因工程改造菌LYS-12生产的L-赖氨酸滴度达到120 g/L,转化率为0.55 g氨基酸/g葡萄糖,产率达到4.0 g/(L·h),以上指标均达到了过去50多年通过诱变育种获得的最高水平,是第一个可与已有工业生产菌株竞争的基因工程设计的L-赖氨酸高产菌。然而由于模型预测准确性还亟待提高,基于该方法论的菌株开发策略仍然依赖于“设计-构建-测试-学习”的循环迭代(试错)流程<sup>[20]</sup>(图2):首先利用系统生物学工具建立微生物的代谢模型,确定MCFs改进目标;接着利用基因工程手段进行菌株构建;对菌株进行表征,并结合高通量分析或组学分析等手段对目标参数进行评估;根据分析结果,对模型进行改进;重复迭代,直至获得满足需求性状的目的MCFs。需要注意的是由于生物系统的多层次网络结构及复杂性,基于系统层次的代谢网络模型进行MCFs开发必须依赖已有的生物学知识。Sang Yup Lee等<sup>[74]</sup>主张将进化代谢工程

方法论也纳入系统代谢工程方法论体系,通过将理性设计和非理性的随机突变或适应性进化相结合,利用比较组学手段可以获取与微生物代谢网络结构及其调控相关的未知信息,从而进一步提高MCFs设计效率和性能。

理性/非理性设计指导的MCF代谢工程改造策略在实际应用中已经取得巨大成功,表4汇总了其中一些代表性案例。然而,由于生物系统的复杂性和“生命暗物质”的普遍存在,目前代谢工程主流采用的“设计-构建-测试-学习”迭代试错流程,通常从MCFs概念设计(proof of concept)到满足实际应用需求,需要50~300人年和数亿美元的投入。近年来,市场快速变化且多种多样的需求给MCFs的开发效率提出了更高的要求。另外,基于理性/非理性的设计策略主要基于已知生物学知识,难以达到某些未知的表型“高地”,从而制约MCFs性能的进一步提升。因此,亟需开发更为高效的工程化策略以满足未来快速定制化创制MCFs的需求。

表3 常见的微生物代谢模型汇总

Tab. 3 Summary of microbial metabolism models

模型范围	模型名称	菌株	反应/代谢物	参考文献	
核心代谢模型	未报道	大肠杆菌	95/94	[54]	
	未报道	谷氨酸棒杆菌	未报道	[55]	
	未报道	酿酒酵母	70/83	[56]	
	未报道	酿酒酵母	78/98	[57]	
	未报道	酿酒酵母	37/27	[58]	
全基因组代谢模型(GSMM)	iFF708	酿酒酵母	1145/825 +708 基因	[53]	
	iND750	酿酒酵母	1149/646 +750 基因	[59]	
	Yeast 4.0	酿酒酵母	1865/1319 +932 基因	[60]	
	iJO1366	大肠杆菌	2583/1805 +1366 基因	[61]	
	未报道	谷氨酸棒杆菌	495/408 +411 基因	[62]	
	未报道	乳酸菌	621/509 +358 基因	[63]	
	未报道	枯草芽孢杆菌	754/637 +614 基因	[64]	
	全基因组代谢模型 +基因表达水平	<i>T. maritima</i> iOL1650-ME	海栖热袍菌 大肠杆菌	17535/18209 76414/56902	[65] [66]
	全基因组代谢模型 +蛋白质结构特性	<i>T. maritima</i> <i>E. coli</i> GEM-PRO	海栖热袍菌 大肠杆菌	645/503 2583/1805 +1268 酶结构	[67] [68-69]
全基因组代谢模型 +转录调控	iAF1260 PROM	大肠杆菌	2583/1805 +1773 转录调控作用	[70]	
	iMH805/837	酿酒酵母	1489/972 +805 基因+837 转录调控作用	[71]	
全细胞模型	'Whole-cell' <i>M. genitalium</i>	生殖支原体	28 个细胞过程子模块 +525 基因	[72]	

### 3 全基因组规模定制工程策略构建 MCFs

微生物代谢及其控制是一个非线性复杂网络系统, 细胞依靠其精巧的基因线路以及严格的调控机制来维持各项代谢活动的稳定。按照“设计-构建-测试-学习”的框架, 基于已有的知识, 利用基因、启动子、核糖体结合位点等分子元件能够自下而上地在微生物细胞中进行基因线路的设计和构建, 从而可望开发出具有特定功能的 MCFs。然而, 由于已有生物学知识的局限性, 生物系统存在极高的不可预测性, 导致设计的基因线路在实际体系中难以达到理想的目标。

事实上, 随着生物信息学和各种组学技术的快速发展, 基因组范围内大量的未知功能基因位点逐渐被人们所认知, 包括潜在的别构调节区域<sup>[22]</sup>、复杂的转录调节网络<sup>[23]</sup>、功能未知的基因<sup>[24-25]</sup>等。这些未知因素及其背后的生物学机制限制了现有基于知识的理性/半理性设计的 MCFs 工程化策略的进一步发展。为了在 MCFs 基因组范围内挖掘对目标表型有重要影响的基因位点, 系统代谢工程领域常利用非理性诱变及筛选手段获得缺陷型菌株, 再人为选定目的基因并对其测序, 从而找到关键位点。同时, 随着分子生物学技术的发展, 亦可通过单敲除库筛选<sup>[25]</sup>、基因组文库共存/共表达<sup>[28]</sup>、全基因组 ORF 表达水平调节<sup>[86]</sup>等技

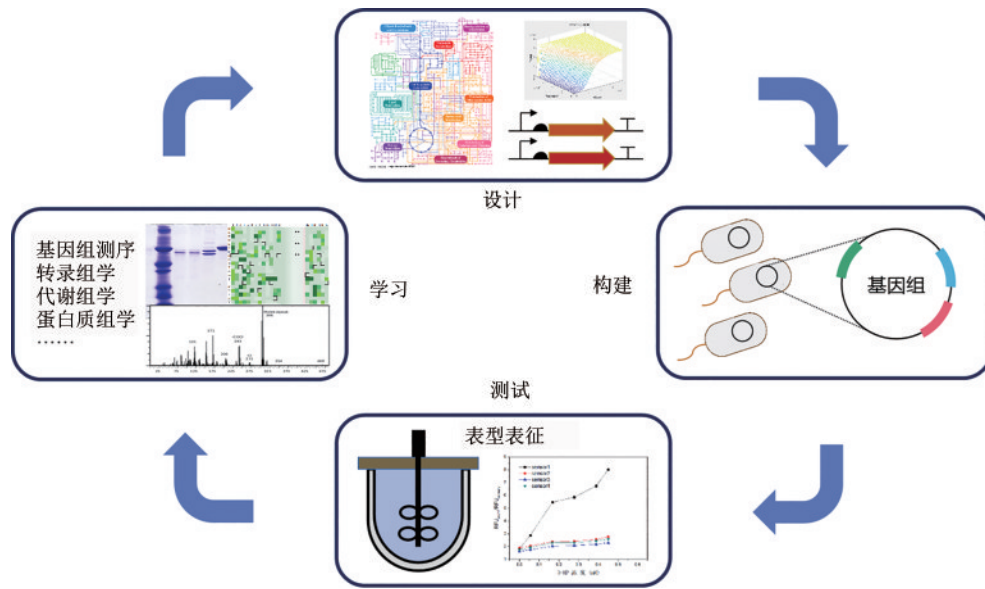


图2 菌株理性工程化的试错流程

Fig. 2 Iterative trial-and-error cycle of rational engineering of strains

(The construction of MCFs based on systematic metabolic engineering relies on the 'Design-Build-Test-Learn' iterative cycle. Firstly, metabolic models are used to design the metabolic network of MCFs. Secondly, synthetic biology tools are used to build the target MCFs. Thirdly, the MCFs are characterized to evaluate performance. Finally, the results are analyzed and the metabolic model will be modified to further improve the performance)

表4 代谢工程指导的经典设计策略的商业化应用案例

Tab. 4 Commercial application of classic design strategies guided by metabolic engineering

产物	宿主	原料	公司	参考文献
琥珀酸	大肠杆菌	玉米糖	BioAmber	[75]
	大肠杆菌	蔗糖	Myriant(现名 GC Innovation America)	
	克鲁斯假丝酵母			
	酿酒酵母	淀粉、糖类	Reverdia	
	巴斯夫产琥珀酸菌	甘油、糖类	Succinity	
1,4-丁二醇	大肠杆菌	糖类	Genomatica 和 DuPont Tate & Lyle	[76-77]
1,3-丙二醇	大肠杆菌	糖类	DuPont Tate & Lyle	[77-78]
聚羟基链烷酸酯(PHA)	大肠杆菌	糖类	Metabolix(现名 Yeild10 science)	[79]
3-羟基丙酸	大肠杆菌	未报道	OPXbio & Dow Chemical	[80]
	未报道	未报道	Perstorp	
乙醇	酿酒酵母	蔗糖、玉米糖、木质纤维素		[81]
	运动发酵单胞菌			
	马克斯克鲁维酵母			
异丁醇	酿酒酵母	糖类	Gevo	[82]
			Butalco	
			Butamax	
法尼烯	酿酒酵母	未报道	Amyris	[74,83]
青蒿素(半合成)	酿酒酵母	未报道	Amyris	[84-85]

术的发展使得人们能够挖掘基因组上功能未知的基因位点。例如，目前已经在大肠杆菌<sup>[87]</sup>、枯草芽孢杆菌<sup>[88]</sup>等模式微生物中对所有基因进行逐一敲除，形成包含数千个不同突变体的单基因敲除菌株阵列

(single-gene knockout mutants array)，为全基因组范围的功能基因组学研究提供了黄金标准。然而，这些技术往往耗时耗力，成本高昂，且仅能研究全基因组范围内少数基因与目标表型的关联 (genotype

phenotype associations, GPA), 亟需通过新技术的引入大幅提升未知功能基因挖掘的效率。

为了更为高效地设计和构建MCFs, 必须快速高效地挖掘基因组范围内与目标表型相关联的未知基因及其位点。近年来, 高通量基因编辑技术和表型筛选技术的发展有可能为未来MCFs的构建范式带来革命性的变革。在基因型方面, 首先利用基因型高通量构建技术实现模式微生物全基因组范围内基因的高效编辑、突变和表达水平调节; 在表型方面, 利用表型高通量表征及筛选技术实现目标表型的高效表征和筛选; 结合基因型高通量构建和表型表征/筛选获得的大容量混合文库样本, 利用新一代测序和数据统计分析手段, 建立特定表型关联基因及其位点GPA数据集; 最后, 利用大规模GPA数据集, 结合深度学习等数据科学手段, 基于数据科学手段从全基因组范围深度挖掘传统分子生物学手段无法发现的未知关联基因及其位点, 将有可能从数据(学习)的角度绕开现有理性设计的知识局限性, 进一步指导高效MCFs的构建, 为提高MCFs设计和创建效率提供全新的研究范式。以下将对这一研究范式涉及的不同技术环节进行概要介绍。

### 3.1 基因组高通量编辑技术

微生物的基因组改造(敲入、敲除和引入突变)最早是基于同源重组技术实现的。最为广泛应用的Red/ET同源重组技术利用单链DNA(ssDNA)的重组实现基因组的改造, 然而由于重组效率的限制, 这一方法仅限于对单个基因进行串行操作, 通量低, 远不能满足全基因组范围大量基因编辑的需求。2009年, George M. Church团队提出了多重自动化基因组工程(multiplex automated genome engineering, MAGE)技术<sup>[89]</sup>, 通过对Red/ET同源重组的参数进行优化并提出一套自动化装置, 将单次ssDNA重组效率提高到约30%。利用这一技术同时将靶向基因组上多个位点的ssDNA文库转入细胞中, 能够实现基因型的大规模并行改造。2010年, Ryan T. Gill团队提出了可追踪多重重组技术(trackable multiplex recombineering, TRMR)<sup>[90]</sup>, 通过在引入的DNA

中插入特异性标签(barcode), 在实现基因组改造的同时, 还可通过后继的标签测序反向追踪特定基因型。尽管MAGE和TRMR的出现极大地提高了微生物功能基因组学研究的通量, 然而, 高效同源重组工具只存在于少数模式微生物中, 难以推广到更多具有应用价值的非模式微生物中。

近年来, CRISPR基因编辑技术的飞速发展更为便捷的基因组编辑带来了新的机遇。CRISPR-Cas系统中的sgRNA能够特异性靶向目标DNA序列, 根据不同需要引入Cas蛋白(及其突变体或具有特定功能的融合蛋白), 就能够实现目标基因的切割<sup>[91]</sup>、抑制<sup>[92]</sup>、激活<sup>[93]</sup>、编辑<sup>[94]</sup>和突变<sup>[95]</sup>。这类技术的可推广性远远高于前述的同源重组技术, 目前已经被证明适用于包括古细菌在内的十余种重要的模式微生物<sup>[96-104]</sup>。由于sgRNA只有20 bp长, 其本身就可以作为目标基因的特异性标签, 实现混合文库中特定基因型的反向追踪。利用这一特性, 结合DNA合成和NGS技术, 能够实现sgRNA文库的定制化合成和后续基于NGS的低成本扩增子测序, 从而极大地提高了功能基因组学研究的通量和便捷性。表5列举了近年来开发的不同类型基因组高通量编辑技术及其应用特性。

### 3.2 表型高通量表征/筛选技术

尽管基于CRISPR-Cas系统的高通量基因组编辑技术极大地提高了基因型的构建能力, 在表型关联方面, 现有研究主要以生长限制因素、耐药性、溶剂耐受性等为筛选条件, 基于不同基因型微生物生长适应性(fitness)的差异, 通过连续传代实现目的表型的筛选<sup>[29, 86, 90, 108-110, 114]</sup>。由于采用传统的摇瓶或孔板培养进行传代培养, 将会耗费大量的时间和成本, 限制了基因型-表型关联研究的效率。近年新发展起来的高通量微生物进化培养系统大幅提高了传代效率, 例如, Wong等<sup>[115]</sup>开发的eVOLVER设备能够同时大规模培养16种不同的微生物, 其硬件、软件和湿件的高度模块化能够实现配置的快速更改, 以适应更多高通量、自动化连续培养实验。本团队自主开发的基于液滴微流控技术的“微生物微液滴培养仪”

表5 基因组高通量编辑技术

Tab. 5 High-throughput genotype construction technologies

分类	名称	描述	参考文献
不可追踪技术	MAGE	基于重组的基因组编辑技术,可使用多个 ssDNA 同时对多个目标位点进行修饰。与其他基因编辑工具(如 CRISPR/Cas9)联用可进一步提高基因编辑效率。主要用于原核基因组编辑	[89]
	YOGE	原理与 MAGE 类似,主要用于真核基因组编辑	[105]
可追踪技术	TRMR	基于同源重组的基因组编辑技术,能够同时对基因组上千个基因位点进行修饰	[90]
	CREATE	该技术基于同源重组和 CRISPR/Cas9 基因编辑技术,能够在全基因组范围内实现可追踪编辑	[106]
	Prime Editor	使用融合了工程逆转录酶的催化活性受损的 Cas9 和 pegRNA,以更高的效率、更低的脱靶率在全基因组范围内实现所有 12 种单碱基的自由转换以及多碱基的精准插入和删除	[107]
	Target AID	激活诱导的胞嘧啶脱氨酶(AID)可实现 C 到 T 的突变,Target AID 技术以核酸酶活性缺失的 CRISPR/Cas9 系统作为 AID 的 DNA 靶向模块,实现定点诱变	[108]
	TAM	该系统将 AID-P182X 与 dCas9 蛋白融合,可将 G 或 C 突变为另外三个碱基,从而在目标位点处产生大量突变	[109]
	CRISPR-X	使用 dCas9 募集更高活性的 AID $\Delta$ 和 MS2 修饰的 sgRNA 变体,能以较低的脱靶率同时实现多个靶基因的特异性突变	[110]
	EVOLVR	该系统由一个 CRISPR 引导的切口酶和一个易错 DNA 聚合酶组成,可在靶位点处可调窗口长度内实现所有核苷酸的突变	[95]
	CRISPRi	使用 dCas9 蛋白及其 sgRNA 阻断靶基因转录,实现基因表达水平的下调	[111-112]
CRISPRa	使用与转录激活因子融合的 dCas9 蛋白实现靶基因表达水平的上调	[93, 113]	

(microdroplet microbial culture, MMC), 结合微流控技术和光电传感与控制及自动化技术, 可以实现微升级微生物液滴平行培养、生长曲线测定和适应性进化 (通量  $10^2 \sim 10^4$ ), 对于典型的模式微生物如大肠杆菌、酿酒酵母、乳酸杆菌等, 可实现稳定传代 100 代, 大大提升了微生物底物利用能力、溶剂耐受性等表型样本获取的效率<sup>[116]</sup>。

除了基于耐受性的表型高通量筛选模型, 近年来, 还发展了一系列代谢物产量高通量筛选技术与装备。代谢物细胞传感器作为一类重要的合成生物学工具, 能够通过特定的转录因子、核糖体开关等识别元件响应细胞内特定代谢物的浓度, 并将其转化为荧光、抗逆生长等特定的输出信号, 在代谢物浓度表型表征方面得到广泛应用。通过对天然识别元件进行鉴定或工程化改造, 目前已经开发了不同的生物传感器<sup>[117-120]</sup>。代谢物细胞传感器与荧光激活细胞分选技术 (FACS) 相结合, 可以将胞内代谢物浓度转换为易于检测的荧光信号, 利用流式细胞仪可实现细胞的高通量表型及分选, 目前已有诸多成功案例。然而, 基于 FACS 的高通量代谢物检测技术仅局限于胞内代谢物浓度的检测, 无法实现胞外代谢物的高通量检测。为了解决这些问题, 研究者提出了基于液滴微流控的高通量表型筛选方法, 将微生物包裹在一个

液滴之中进行培养, 再以液滴为单位进行胞外代谢物高产菌株的高通量筛选, 表 6 列举了具有代表性的代谢物高通量表型表征/筛选技术, 包括荧光激活液滴分选 (FADS)、基于液滴的荧光激活细胞分选 (Droplet-FACS)、基于凝胶微液滴的荧光激活细胞分选 (Gel FACS)、拉曼活化液滴分选 (RADS) 等。

需要注意的是, 目前的表型高通量表型/筛选技术仍然局限于特定时间点的单一表型快照 (snapshot)。如果能够利用近年来快速发展的单细胞 RNA 测序<sup>[128]</sup>、微生物原位高精度荧光显微定位追踪<sup>[129-131]</sup>等技术, 使表型端能够输出高密度的时间序列信息, 进一步结合合成基因线路<sup>[132]</sup>, 将输出表型扩展为单细胞水平异质性等更为重要和丰富的类型, 将进一步推动表型高通量表型表征和筛选技术的发展, 大幅度提升 MCFs 的工程化能力。

### 3.3 基因型-表型关联技术

不同于阵列方法 (array) 对每个基因进行逐一敲除或过表达并进行单独研究的策略, 基于混合文库筛选的方法通过高度平行化的实验, 对基因组范围内的基因位点进行大规模的扰动, 并检

表6 微生物代谢物高通量表征/筛选技术

Tab. 6 High-throughput selection/screening technologies in single-cell level

技术	通量	功能	案例	参考文献
FACS	约 10 <sup>5</sup> 细胞/s	分选	利用色氨酸传感器,实现紫色杆菌素高产大肠杆菌筛选 利用甲羟戊酸细胞传感器,实现甲羟戊酸高产甲基杆菌筛选 利用赖氨酸传感器,实现赖氨酸高产谷氨酸棒杆菌筛选	[119] [120] [121]
FADS	约 10 <sup>4</sup> 液滴/s	培养+分选	将胞外重组酶活性与液滴荧光强度耦合,实现酶活测定及胞外重组酶高产菌株筛选 将传感器菌与 MCF 菌株共培养,实现对香豆酸高产酵母的筛选 将 $\beta$ -半乳糖苷酶产量与液滴荧光强度耦合,实现 $\beta$ -半乳糖苷酶高产大肠杆菌筛选	[122] [123] [124]
Droplet-FACS		培养+分选	通过检测核黄素的荧光,实现核黄素高产解脂耶氏酵母筛选	[125]
Gel FACS	约 3000 液滴/s	培养+分选	将胞外木聚糖酶的产量与凝胶液滴的荧光强度耦合,实现胞外木聚糖酶高产毕赤酵母筛选	[126]
RADS	260 细胞/min	培养+分选	利用虾青素的拉曼光谱,实现虾青素高产雨生红球藻的筛选	[127]

测特定筛选条件下的表型变化,能够实现功能基因的快速筛选。以基于 CRISPR 体系的混合文库筛选方法为例,其基本原理是利用定制化设计和合成的 sgRNA 文库产生混合基因型文库,并在特定的筛选条件下进行筛选,随后利用 NGS 技术对 sgRNA 进行扩增子测序,根据测序结果分析筛选前后 sgRNA 丰度的变化,从而获得特定基因型与筛选条件的关系,进而绘制全基因组规模的 GPA 图谱,实现未知功能基因位点的深度挖掘。

目前,这一策略已经被逐渐引入到代谢工程领域并用于菌株工程化改造,其中的典型代表团队是美国科罗拉多大学的 Ryan T. Gill 团队。例如,他们利用 CREATE 技术在大肠杆菌中针对 19 个基因构建了库容为 16 300 的混合文库,以赖氨酸类似物的竞争性抑制作为致死筛选条件,对赖氨酸代谢相关的基因进行了深度扫描,确定了 *lysP*、*argT* 和 *cadB* 等与类似物竞争性相关的基因<sup>[133]</sup>。利用类似的策略,该团队先后实现了异丙醇<sup>[134]</sup>、3-羟基丙酸<sup>[135]</sup>、苯乙烯<sup>[136]</sup>等高产菌株的构建。本团队利用 CRISPRi 技术首次在大肠杆菌中构建了靶向全基因组范围的 sgRNA 文库,结合混合文库筛选的策略 CRISPRi-seq,实现了大肠杆菌全基因组(4000 余个基因)范围内必需基因、代谢网络结构、与糠醛和异丁醇耐受性相关基因的快速筛选和鉴定<sup>[29]</sup>。对于 CREATE 技术,其突变位点挖掘的分辨率在碱基层次,由于数据密度极高,受限于建库和测序成本,目前只能实现部分基因的突变位点关联分析,相比之下,本团队开发的 CRISPRi-seq 方法分辨率在单基因层次,能够实现

全基因组范围功能基因的高效挖掘。

### 3.4 全基因组规模定制工程中 GPA 数据的利用

如前所述,快速发展的 NGS 技术和基因型-表型关联技术使得目前能够以较低的成本快速获得大量 GPA 数据。基于大规模并行实验表征获取的 GPA 数据集(10<sup>3</sup>以上)为机器学习提供了高质量的注释样本,如何利用数据科学手段学习基因型-表型数据背后隐藏的“机制”或“规律”成为研究者日益关注的重要方向。目前,计算生物学领域已经发展了多种对不同层次的组学数据进行分析、挖掘和学习的算法<sup>[137]</sup>。结合丰富的组学数据,机器学习算法被成功应用于通路动力学的预测<sup>[138]</sup>、酵母 5'非翻译区调控序列的研究<sup>[139]</sup>等。如果能够充分利用大规模 GPA 数据集,依托数据科学手段从全基因组范围深度挖掘传统分子生物学手段无法发现的大量的未知关联基因及其位点,将有可能从数据(学习)的角度绕开目前理性设计的知识瓶颈,实现定制化 MCFs 的快速、高效创制。

目前,机器学习手段在 GPA 数据的学习和利用方面还刚刚起步,相关研究还比较少。本团队尝试了小规模 GPA 数据在酵母 MCFs 构建中的应用。为了实现酿酒酵母中异源代谢途径表达水平的快速优化,我们提出了一种与酵母体内生物标准元件组装策略相结合的机器学习工作流程(machine-learning workflow in conjunction with YeastFab assembly, MiYA)<sup>[140]</sup>,应用人工神经网络(artificial neural network, ANN)模型,以组合空间中 2%~5% 的数据为训练集,成功预测了具有最优表达水平组合的 $\beta$ -

胡萝卜素和紫色杆菌素合成途径。类似地，机器学习策略已经被用于指导构建生产特定风味物质的啤酒酵母<sup>[141]</sup>和色氨酸高产酿酒酵母菌株<sup>[142]</sup>，展现出重要的应用潜力。此外，结合高通量基因组编辑和筛选技术，本团队还报道了迄今为止在大肠杆菌中规模最大的sgRNA活性数据集，并利用机器学习手段建立了sgRNA序列影响活性的数学模型，可用于预测并优化CRISPR-Cas系统在大肠杆菌中的工作性能<sup>[143]</sup>。

## 4 展望

MCFs的设计和构建最终目标都是实际应用，市场需求快速多样的变化对MCFs构建效率和性能的提升提出严苛的要求。如图3所示，早期的诱变育种采取非理性手段进行MCFs改造，是典型的“以时间（人力）换水平”的策略。随着生物学知识的积累，经典代谢工程的发展使得对生物代谢网络进行理性/半理性设计成为可能，以DBTL循环为基本流程，MCFs改造效率得到显著提升。系统代谢工程的建立进一步使得研究者能够结合组学和生物信息学手段获取生物学知识，从系统层次进行MCFs的设计，进一步加快MCFs的构建效率。然而，由于微生物代谢网络结构及其调控机制的复杂性和“生命暗物质”的广泛存在，基于上述策略进行MCFs设计和构建的过程仍然需要耗

费大量的时间和精力，且由于搜索的基因组空间有限，难以满足工业生产不断增长的表型“高地”需求。随着高通量研究技术的发展，由数据驱动的全基因组规模定制工程有望克服这些难题，通过将高通量技术在全基因组范围基因型空间的挖掘与改造相结合，有望以更低的开发成本、更短的研发周期获得生产效率更为高效、生产性能更加优越的下一代定制化MCFs。

全基因组规模定制工程化MCFs作为全新的研究领域和范式，目前还处于萌芽阶段。为了实现MCFs的定制化设计与构建，亟需解决如下两个重要问题：

第一，高通量高质量GPA数据的产生。基因型端的基因组编辑技术、测序技术发展较为成熟，而在表型端，由于测试目标和测试条件的复杂性，目前还无法弥补与大规模基因型样本之间的数据鸿沟。自动化技术和高通量新型表型表征技术的引入是解决这一问题的重要发展方向。近年来，国内外基于自动化机器人的BioFoundry平台建设方兴未艾，单细胞质谱、单细胞RNA测序等技术也在迅速发展，这些都将成为跨越基因型-表型数据鸿沟，实现高通量高质量GPA数据供给提供重要机遇。

第二，数据管理和学习。高通量和自动化技术在极大地提升数据获取能力的同时将不可避免地带来数据爆炸，这为数据存储、数据标准化和数据共享带来了挑战。为了实现不同数据的共享和集成，迫切需要建立可查找、可访问、可相互

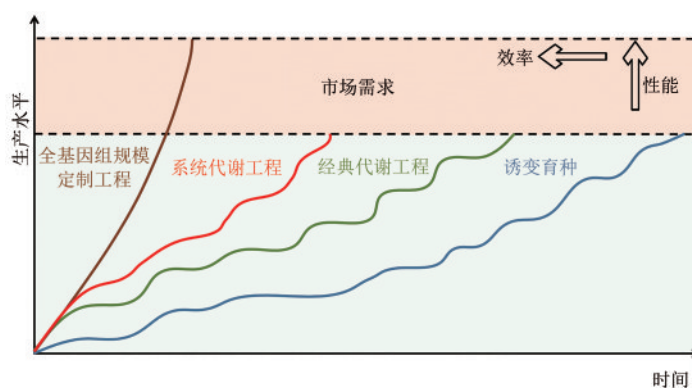


图3 微生物细胞工厂设计和构建策略效率以及性能对比

Fig. 3 Comparison of MCFs construction efficiency and performance in different stages

(Random mutagenesis of natural microbes is firstly developed but time-consuming for MCFs construction. Taking the DBTL cycle as basic process, metabolic engineering enables rational/semi-rational design of metabolic pathways which can accelerate the construction of MCFs. The development of systematic metabolic engineering further enhances the construction efficiency of MCFs. However, these strategies are difficult to meet the growing demand for phenotypic ‘high ground’ in industrial production. With the development of high-throughput technology, data-driven genome-wide customized engineering is expected to overcome these problems)

操作和可重复使用的工程生物学数据库。此外，在如何利用GPA数据集指导定制化MCFs的设计和构建方面，亟需系统创建性能更优的生物数据分析、挖掘、学习算法和工具，以充分发挥大规模数据的效力。近年来基于卷积神经网络的深度学习算法、基于生成对抗网络的生成学习算法都被逐渐应用到生物系统领域，需要加快这些算法在GPA数据集学习和利用方面的探索和应用。

生物产业是我国国民经济的重要支柱，菌种是生物产业命脉所系，然而，我国菌种长期面临国外技术垄断、专利封锁的困境，近年更是频繁出现生物发酵产业核心菌种的国际知识产权纠纷。我们期待，如果能够抓住全基因组规模定制工程化发展机遇，提升我国MCFs创新能力，将有可能在未来日益激烈的国际竞争中占领菌种创制的战略制高点，实现我国生物产业的跨越式发展。

### 符号说明

BNICE ——生化网络集成计算浏览器(biochemical network integrated computational explorer)  
 COBRA ——基于约束的重构与分析(constraint-based reconstruction and analysis)  
 CREATE ——基于CRISPR的可追踪基因组工程(CRISPR-enabled trackable genome engineering)  
 CRISPR ——规律成簇的间隔短回文重复(clustered regularly interspaced short palindromic repeats)  
 CRISPRa ——CRISPR激活(CRISPR activation)  
 CRISPRi ——CRISPR干扰(CRISPR interference)  
 FACS ——荧光激活细胞分选(fluorescence-activated cell sorting)  
 FADS ——荧光激活液滴分选(fluorescence-activated drop sorting)  
 FBA ——通量平衡分析(flux balance analysis)  
 GPA ——基因型-表型关联(genotype-phenotype associations)  
 GSMM ——全基因组代谢模型(genome-scale metabolic model)  
 GSMR ——基因组规模代谢重构(genome-scale metabolic reconstruction)  
 MAGE ——多重自动化基因组工程(multiplex automated genome engineering)  
 MCA ——代谢控制分析(metabolic control analysis)  
 MFA ——代谢通量分析(metabolic flux analysis)

MMC ——液滴微流控培养(microdroplet microbial culture)  
 MOMA ——最小化代谢调节(minimization of the metabolic adjustment)  
 MPA ——代谢途径分析(metabolic pathway analysis)  
 NET ——网络嵌入的热力学分析(network-embedded thermodynamic analysis)  
 NGS ——二代测序(next generation sequencing)  
 ODE ——常微分方程(ordinary differential equation)  
 PHA ——聚羟基链烷酸酯(polyhydroxyalkanoate)  
 RADS ——拉曼活化液滴分选(raman-activated droplet sorting)  
 ROOM ——开/关最小化调节(regulatory on/off minimization)  
 sgRNA ——向导RNA(single-guide RNA)  
 ssDNA ——单链DNA(single-strand DNA)  
 TAM ——靶向AID介导的诱变(targeted AID-mediated mutagenesis)  
 TMFA ——基于热力学的代谢通量分析(thermodynamics-based metabolic flux analysis)  
 TRMR ——可追踪多重重组(trackable multiplex recombineering)  
 YOGE ——酵母寡核苷酸介导的基因组工程(yeast oligo-mediated genome engineering)

### 参 考 文 献

- [1] NIELSEN J. Cell factory engineering for improved production of natural products[J]. *Natural Product Reports*, 2019, 36(9): 1233-1236.
- [2] SANO C. History of glutamate production[J]. *The American Journal of Clinical Nutrition*, 2009, 90(3), 728S-732S.
- [3] OTERO J M, NIELSEN J. Industrial systems biology[J]. *Biotechnology and Bioengineering*, 2010, 105(3), 439-460.
- [4] Presidential green chemistry challenge awards[EB/OL]. <https://www.epa.gov/greenchemistry/green-chemistry-challenge-winners>.
- [5] ATSUMI S, HIGASHIDE W, LIAO J C. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde[J]. *Nature Biotechnology*, 2009, 27(12), 1177-1180.
- [6] GALANIE S, THODEY K, TRENCHARD I J, et al. Complete biosynthesis of opioids in yeast[J]. *Science*, 2015, 349(6252): 1095-1100.
- [7] PARK J H, LEE K H, KIM T Y, et al. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(12), 4381-4386.

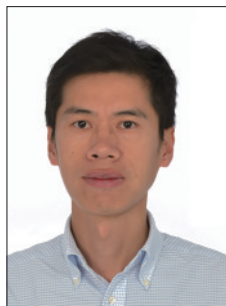
- ed States of America, 2007, 104(19): 7797-7802.
- [8] XIA X X, QIAN Z G, KI C S, et al. Native-sized recombinant spider silk protein produced in metabolically engineered *Escherichia coli* results in a strong fiber[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(32):14059-14063.
- [9] STEEN E J, KANG Y, BOKINSKY G, et al. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass[J]. Nature, 2010, 463(7280): 559-562.
- [10] LEE G, NA J. Future of microbial polyesters[J]. Microbial Cell Factories, 2013, 12(1): 54.
- [11] BROWN S W, OLIVER S G. Isolation of ethanol-tolerant mutants of yeast by continuous selection[J]. European Journal of Applied Microbiology and Biotechnology, 1982, 16(2/3): 119-122.
- [12] MEADOWS L, HAWKINS K M, TSEGAYE Y, et al. Rewriting yeast central carbon metabolism for industrial isoprenoid production[J]. Nature, 2016, 537(7622): 694-697.
- [13] WEIKERT C, SAUER U, BAILEY J. Use of a glycerol-limited, long-term chemostat for isolation of *Escherichia coli* mutants with improved physiological properties[J]. Microbiology, 1997, 143(5): 1567-1574.
- [14] SILMAN N J, CARVER M A, JONES C W. Physiology of amidase production by methylphilus methylotrophus: isolation of hyperactive strains using continuous culture[J]. Microbiology, 1989, 135(11): 3153-3164.
- [15] CORNISH A, GREENWOOD J A, JONES C W. Binding-protein-dependent sugar transport by *Agrobacterium radiobacter* and *A. tumefaciens* grown in continuous culture[J]. Microbiology, 1989, 135(11): 3001-3013.
- [16] BAILEY J. Toward a science of metabolic engineering[J]. Science, 1991, 252(5013): 1668-1675.
- [17] STEPHANOPOULOS G, VALLINO J. Network rigidity and metabolic engineering in metabolite overproduction[J]. Science, 1991, 252(5013): 1675-1681.
- [18] STEPHANOPOULOS G, SINSKEY A J. Metabolic engineering-methodologies and future prospects[J]. Trends in Biotechnology, 1993, 11(9): 392-396.
- [19] FURUSAWA C, HORINOUCHE T, HIRSAWA T, et al. Systems metabolic engineering: the creation of microbial cell factories by rational metabolic design and evolution[J]. Advances in Biochemical Engineering/Biotechnology, 2012, 131(3): 1-23.
- [20] NIELSEN J. Production of bio-pharmaceutical proteins by yeast: advances through metabolic engineering[J]. Bioengineered, 2012, 4(4): 207-211.
- [21] HONG K K, NIELSEN J. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries[J]. Cellular and Molecular Life Sciences, 2012, 69(16), 2671-2690.
- [22] CHEN Z, RAPPERT S, SUN J, et al. Integrating molecular dynamics and co-evolutionary analysis for reliable target prediction and deregulation of the allosteric inhibition of aspartokinase for amino acid production[J]. Journal of Biotechnology, 2011, 154(4): 248-254.
- [23] CHO B K, FEDEROWICZ S, PARK Y S, et al. Deciphering the transcriptional regulatory logic of amino acid metabolism[J]. Nature Chemical Biology, 2011, 8(1): 65-71.
- [24] JIN Y S, STEPHANOPOULOS G. Multi-dimensional gene target search for improving lycopene biosynthesis in *Escherichia coli*[J]. Metabolic Engineering, 2007, 9(4): 337-347.
- [25] ÖZAYDIN B, BURD H, LEE T S, et al. Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production[J]. Metabolic Engineering, 2013, 15: 174-183.
- [26] ZHANG Y X, PERRY K, VINCI V A, et al. Genome shuffling leads to rapid phenotypic improvement in bacteria[J]. Nature, 2002, 415(6872), 644-646.
- [27] ZHANG X, ZHANG X F, LI H P, et al. Atmospheric and room temperature plasma (ARTP) as a new powerful mutagenesis tool[J]. Applied Microbiology and Biotechnology, 2014, 98(12): 5387-5396.
- [28] NICOLAOU S A, GAIDA S M, PAPOUTSAKIS E T. Coexisting/Coexpressing genomic libraries (CoGeL) identify interactions among distantly located genetic loci for developing complex microbial phenotypes[J]. Nucleic Acids Research, 2011, 39(22): e152.
- [29] WANG T, CHANG G G, JIA H G, et al. Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance[J]. Nature Communications, 2018, 9(1):2475.
- [30] 李翔. 微生物诱变育种技术[J]. 现代商贸工业, 2017(34): 185-187.
- LI X. Microbial mutation breeding technology[J]. Modern Business Trade Industry, 2017(34): 185-187.
- [31] DEMAİN A L. From natural products discovery to commercialization: a success story[J]. Journal of Industrial Microbiology & Biotechnology, 2006, 33(7), 486-495.
- [32] 季宇彬, 朱兴杰, 徐昶儒, 等. 微生物诱变育种方法的研究进展[C]// 中国毒理学会环境与生态毒理学专业委员会成立大会. 2008.
- JI Y B, ZHU X J, XU C R, et al. Research progress of microbial mutation breeding methods [C] // The Founding Conference of the Professional Committee of Environmental and Ecotoxicology of the Chinese Society of Toxicology. 2008.
- [33] 微生物诱变育种编写组. 微生物诱变育种[M]. 北京: 科学出版社, 1973: 25-32.
- Microbiology Mutation Breeding Compilation Group. Microbial mutation breeding [M]. Beijing: Science Press, 1973: 25-32.

- [34] NESS J E, WELCH M, GIVER L, et al. DNA shuffling of sub-genomic sequences of subtilisin[J]. *Nature Biotechnology*, 1999, 17(9): 893-896.
- [35] 施巧琴, 吴松钢. 工业微生物育种[M]. 福州: 福建科学技术出版社, 1991: 54-55.
- SHI Q Q, WU S G. Breeding of industrial microbes[M]. Fuzhou: Fujian Science and Technology Publishing House, 1991: 54-55.
- [36] CVIJOVIC M, BORDEL S, NIELSEN J. Mathematical models of cell factories: moving towards the core of industrial biotechnology[J]. *Microbial Biotechnology*, 2010, 4(5): 572-584.
- [37] FELL D A. Metabolic control analysis: a survey of its theoretical and experimental development[J]. *Biochemical Journal*, 1992, 286(2): 313-330.
- [38] WIECHERT W. <sup>13</sup>C metabolic flux analysis[J]. *Metabolic Engineering*, 2001, 3(3): 195-206.
- [39] STEPHANOPOULOS G. Metabolic fluxes and metabolic engineering[J]. *Metabolic Engineering*, 1999, 1(1): 1-11.
- [40] VARMA A, PALSSON B O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110[J]. *Applied and Environmental Microbiology*, 1994, 60 (10): 3724-3731.
- [41] PFEIFFER T, SANCHEZ-VALDENEBRO I, NUNO J, et al. METATOOL: for studying metabolic networks[J]. *Bioinformatics*, 1999, 15(3): 251-257.
- [42] TRINH C T, WLASCHIN A, SRIENC F. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism[J]. *Applied Microbiology and Biotechnology*, 2009, 81(5): 813-826.
- [43] ZAMBONI N, KÜMMEL A, HEINEMANN M. anNET: a tool for network-embedded thermodynamic analysis of quantitative metabolome data[J]. *BMC Bioinformatics*, 2008, 9(1): 199.
- [44] HENRY C S, BROADBELT L J, HATZIMANIKATIS V. Thermodynamics-based metabolic flux analysis[J]. *Biophysical Journal*, 2007, 92(5), 1792-1805.
- [45] BECKER S A, FEIST A M, MO M L, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox[J]. *Nature Protocols*, 2007, 2(3): 727-738.
- [46] SEGRE D, VITKUP D, CHURCH G M. Analysis of optimality in natural and perturbed metabolic networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(23): 15112-15117.
- [47] SHLOMI T, BERKMAN O, RUPPIN E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(21): 7695-7700.
- [48] BURGARD A P, PHARKYA P, MARANAS C D. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization[J]. *Biotechnology and Bioengineering*, 2003, 84(6): 647-657.
- [49] HATZIMANIKATIS V, LI C IONITA J A, et al. Exploring the diversity of complex metabolic networks[J]. *Bioinformatics*, 2004, 21(8): 1603-1609.
- [50] EDWARDS J S, PALSSON B O. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2000, 97(10): 5528-5533.
- [51] COVERT M W, SCHILLING C H, FAMILI I, et al. Metabolic modeling of microbial strains in silico[J]. *Trends in Biochemical Sciences*, 2001, 26(3): 179-186.
- [52] ÖSTERLUND T, NOOKAEW I, NIELSEN J. Fifteen years of large scale metabolic modeling of yeast: developments and impacts[J]. *Biotechnology Advances*, 2012, 30(5): 979-988.
- [53] FÖRSTER J, FAMILI I, FU P, et al. Genome-Scale Reconstruction of the *Saccharomyces cerevisiae* metabolic network[J]. *Genome Research*, 2003, 13(2): 244-253.
- [54] VARMA A, BOESCH B W, PALSSON B O. Biochemical production capabilities of *Escherichia coli*[J]. *Biotechnology and Bioengineering*, 1993, 42(1): 59-73.
- [55] VALLINO J J, STEPHANOPOULOS G. Carbon flux distributions at the glucose 6-phosphate branch point in *Corynebacterium glutamicum* during lysine overproduction[J]. *Biotechnology Progress*, 1994, 10(3): 327-334.
- [56] VAN GULIK W M, HEIJNEN J J. A metabolic network stoichiometry analysis of microbial growth and product formation[J]. *Biotechnology and Bioengineering*, 1995, 48(6): 681-698.
- [57] VANROLLEGHEM P A, DE JONG-GUBBELS P, VAN GULIK W M, et al. Validation of a metabolic network for *Saccharomyces cerevisiae* using mixed substrate studies[J]. *Biotechnology Progress*, 1996, 12(4): 434-448.
- [58] NISSEN T L, SCHULZE U, NIELSEN J, et al. Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*[J]. *Microbiology*, 1997, 143(1): 203-218.
- [59] DUARTE N C. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model[J]. *Genome Research*, 2004, 14(7): 1298-1309.
- [60] DOBSON P D, SMALLBONE K, JAMESON D, et al. Further developments towards a genome-scale metabolic model of yeast[J]. *BMC Systems Biology*, 2010, 4(1): 145.
- [61] ORTH J D, CONRAD T M, NA J, et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011[J]. *Molecular Systems Biology*, 2011, 7(1): 535.
- [62] KJELDSSEN K R, NIELSEN J. In silico genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network[J]. *Biotechnology and Bioengineering*, 2009, 102(2): 583-597.
- [63] OLIVEIRA A, NIELSEN J, FORSTER J. Modeling *Lactococ-*

- cus lactis* using a genome-scale flux model[J]. BMC Microbiology, 2005, 5(1): 39.
- [64] PARK S, SCHILLING C, PALSSON B O: Compositions and methods for modeling *Bacillus subtilis* metabolism: US 20030224363[P]. 2003-12-04.
- [65] LERMAN J A, HYDUKE D R, LATIF H, et al. In silico method for modelling metabolism and gene product expression at genome scale[J]. Nature Communications, 2012, 3(1): 929.
- [66] LIU J K, O'BRIEN E J, LERMAN J A, et al. Reconstruction and modeling protein translocation and compartmentalization in *Escherichia coli* at the genome-scale[J]. BMC Systems Biology, 2014, 8:110.
- [67] ZHANG Y, THIELE I, WEEKES D, et al. Three-dimensional structural view of the central metabolic network of *Thermotoga maritima*[J]. Science, 2009, 325(5947): 1544-1549.
- [68] CHANG R L, ANDREWS K, KIM D, et al. Structural systems biology evaluation of metabolic thermotolerance in *Escherichia coli*[J]. Science, 2013, 340(6137): 1220-1223.
- [69] CHANG R L, XIE L, BOURNE P E, et al. Antibacterial mechanisms identified through structural systems pharmacology[J]. BMC Systems Biology, 2013, 7(1): 102.
- [70] CHANDRASEKARAN S, PRICE N D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*[J]. Proceedings of the National Academy Sciences of the United States of America, 2010, 107(41): 17845-17850.
- [71] HERRGARD M J. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*[J]. Genome Research, 2006, 16(5): 627-635.
- [72] KARR J R, SANGHVI J C, MACKLIN D N, et al. A whole-cell computational model predicts phenotype from genotype[J]. Cell, 2012, 150(2): 389-401.
- [73] BECKER J, ZELDER O, STEFAN HÄFNER, et al. From zero to hero-design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production[J]. Metabolic Engineering, 2011, 13(2): 159-168.
- [74] CHOI K R, JANG W D, YANG D, et al. Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering[J]. Trends in Biotechnology, 2019, 37(8): 817-837.
- [75] JANSET M L, VAN GULIK W M. Towards large scale fermentative production of succinic acid[J]. Current Opinion in Biotechnology, 2014,30: 190-197.
- [76] BURK M J. Sustainable production of industrial chemicals from sugars[J]. International Sugar Journal, 2010, 112(1333): 30-35.
- [77] DATTA J, KASPRZYK P, BŁAŻEK K, et al. Synthesis, structure and properties of poly(ester-urethane)s obtained using bio-based and petrochemical 1,3-propanediol and 1,4-butanediol[J]. Journal of Thermal Analysis and Calorimetry, 2017, 130: 261-276.
- [78] NAKAMURA C E, WHITE G M. Metabolic engineering for the microbial production of 1,3-propanediol[J]. Current Opinion in Biotechnology, 2003, 14(5): 454-459.
- [79] Yield10 bioscience agricultural company-Corporate Overview [OL]. <https://www.yield10bio.com/Corporate-Overview>.
- [80] BECKER J, LANGE A, FABARIUS J. Top value platform chemicals: bio-based production of organic acids[J]. Current Opinion in Biotechnology, 2015, 36: 168-175.
- [81] GUSTAVSSON M, LEE S Y. Prospects of microbial cell factories developed through systems metabolic engineering[J]. Microbial Biotechnology, 2016, 9(5): 610-617.
- [82] BUIJS N A, SIEWERS V, NIELSEN J. Advanced biofuel production by the yeast *Saccharomyces cerevisiae*[J]. Current Opinion in Chemical Biology, 2013, 17(3): 480-488.
- [83] GEORGE K W, ALONSO-GUTIERREZ J, KEASLING J D, et al. Isoprenoid drugs, biofuels, and chemicals-artemisinin, farnesene, and beyond[J]. Advances in Biochemical Engineering/Biotechnology, 2015, 355-389.
- [84] PADDON C J, KEASLING J D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development[J]. Nature Reviews Microbiology, 2014, 12(5): 355-367.
- [85] PADDON C J, WESTFALL P J, PITERA D J, et al. High level semi-synthetic production of the potent antimalarial artemisinin[J]. Nature, 2013, 496(7446): 528-532.
- [86] FREED E F, WINKLER J D, WEISS S J, et al. Genome-wide tuning of protein expression levels to rapidly engineer microbial traits[J]. ACS Synthetic Biology, 2015, 4(11): 1244-1253.
- [87] BABA T, ARA T, HASEGAWA M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection[J]. Molecular Systems Biology, 2006. DOI: 10.1038/msb4100050.
- [88] KOO B M, KRITIKOS G, FARELLI J D, et al. Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*[J]. Cell systems, 2017, 4(3): 291-305.
- [89] WANG H H, ISAACS F J, CARR P A, et al. Programming cells by multiplex genome engineering and accelerated evolution[J]. Nature, 2009, 460(7257): 894-898.
- [90] WARNER J R, REEDER P J, KARIMPOUR-FARD A, et al. Rapid profiling of a microbial genome using mixtures of bar-coded oligonucleotides[J]. Nature Biotechnology, 2010, 28(8): 856-862.
- [91] DOUDNA J A, CHARPENTIER E. The new frontier of genome engineering with CRISPR-Cas9[J]. Science, 2014, 346(6213): 1077.
- [92] QI L S, LARSON M H, GILBERT L A, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression[J]. Cell, 2013, 152(5): 1173-1183.

- [93] DONG C, FONTANA J, PATEL A, et al. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria[J]. *Nature Communications*, 2018, 9(1): 2489.
- [94] MANS R, VAN ROSSUM H M, WIJSMAN M, et al. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae* [J]. *FEMS Yeast Research*, 2015, 15(2). DOI: 1093/femsyr/fov004.
- [95] HALPERIN S O, TOU C J, WONG E B, et al. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window[J]. *Nature*, 2018, 560(7717): 248-252.
- [96] JIANG W, BIKARD D, COX D, et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems[J]. *Nature Biotechnology*, 2013, 31: 233-239.
- [97] OH J H, van PIJKEREN J P. CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*[J]. *Nucleic Acids Research* 2014, 42: e131.
- [98] CHOUDHARY E, THAKUR P, PAREEK M, et al. Gene silencing by CRISPR interference in mycobacteria[J]. *Nature Communications*, 2015, 6(1): 6267.
- [99] YAO L, CENGIC I, ANFELT J, et al. Multiple gene repression in cyanobacteria using CRISPRi[J]. *ACS Synthetic Biology* 2016, 5: 207-212.
- [100] XU T, LI Y, SHI Z, et al. Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. [J]. *Applied and Environmental Microbiology*, 2015, 81: 4423-4431.
- [101] PETERS J M, COLAVIN A, SHI H, et al. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria[J]. *Cell*, 2016, 165: 1493-1506.
- [102] JIANG Y, QIAN F, YANG J, et al. CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*[J]. *Nature Communications*, 2017, 8(1): 15179.
- [103] NAYAK D D, METCALF W W. Cas9-mediated genome editing in the methanogenic archaeon *Methanosarcina acetivorans* [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114: 2976-2981.
- [104] 王天民. 基于 CRISPR 干扰技术的微生物高通量功能基因组学方法[D]. 北京: 清华大学, 2018.  
WANG T M. CRISPR interference based high-throughput functional genomics platform method in bacteria[D]. Beijing: Tsinghua University, 2018.
- [105] DICARLO J E, CONLEY A J, PENTTILA M, et al. Yeast oligo-mediated genome engineering (YOGE)[J]. *ACS Synthetic Biology*, 2013, 2(12): 741-749.
- [106] GARST A D, BASSALO M C, PINES G, et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering[J]. *Nature Biotechnology*, 2016, 35(1): 48-55.
- [107] ANZALONE A V, RANDOLPH P B, DAVIS J R, et al. Search-and-replace genome editing without double-strand breaks or donor DNA[J]. *Nature*, 2019, 576(7785): 149-157.
- [108] NISHIDA K, ARAZOE T, YACHIE N, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems[J]. *Science*, 2016, 353(6305): aaf8729.
- [109] MA Y, ZHANG J, YIN W, et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells[J]. *Nature Methods*, 2016, 13(12): 1029-1035.
- [110] HESS G T, FRESARD L, HAN K, et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells[J]. *Nature Methods*, 2016, 13(12): 1036-1042.
- [111] LI Q, SEO J H, STRANGER B, et al. Integrative eQTL-Based analyses reveal the biology of breast cancer risk loci[J]. *Cell*, 2013, 152(3): 633-641.
- [112] WU J, ZHOU P, ZHANG X, et al. Efficient *de novo* synthesis of resveratrol by metabolically engineered *Escherichia coli*[J]. *Journal of Industrial Microbiology & Biotechnology*, 2017, 44(7): 1083-1095.
- [113] GILBERT L A, LARSON M H, MORSUT L, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes[J]. *Cell*, 2013, 154(2): 442-451.
- [114] HALPERIN S O, TOU C J, WANG E B, et al. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window[J]. *Nature*, 2018, 560: 248-252.
- [115] WONG B G, MANCUSO C P, KIRIAKOV S, et al. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER[J]. *Nature Biotechnology*, 2018, 36(7): 614-623.
- [116] JIAN X, GUO X, WANG J, et al. Microbial microdroplet culture system (MMC): an integrated platform for automated, high-throughput microbial cultivation and adaptive evolution[J]. *Biotechnology and Bioengineering*, 2019. DOI: 10.1101/2019.12.19.883561.
- [117] MICHENER J K, THODEY K, LIANG J C, et al. Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways[J]. *Metabolic Engineering*, 2012, 14(3): 212-222.
- [118] LIU D, EVANS T, ZHANG F Z, et al. Applications and advances of metabolite biosensors for metabolic engineering[J]. *Metabolic Engineering*, 2015, 31: 15-22.
- [119] FANG M, WANG T, ZHANG C, et al. Intermediate-sensor assisted push-pull strategy and its application in heterologous deoxyviolacein production in *Escherichia coli*[J]. *Metabolic Engineering*, 2016, 33: 41-51.
- [120] LIANG W, CUI L, CUI J, et al. Biosensor-assisted transcriptional regulator engineering for *Methylobacterium extorquens* AM1 to improve mevalonate synthesis by increasing the acetyl-CoA supply[J]. *Metabolic Engineering*, 2017, 39: 159-168.
- [121] BINDER S, SCHENDZIELORZ G, STABLER N, et al. A high-throughput approach to identify genomic variants of bac-

- terial metabolite producers at the single-cell level[J]. *Genome Biology*, 2012, 13: R40.
- [122] BENEYTON T, THOMAS S, GRIFFITHS A D, et al. Droplet-based microfluidic high-throughput screening of heterologous enzymes secreted by the yeast *Yarrowia lipolytica*[J]. *Microbial Cell Factories*, 2017, 16(1): 18.
- [123] SIEDLER S, KHATRI N K, ZSOHAR A, et al. Development of a bacterial biosensor for rapid screening of yeast *p*-coumaric acid production[J]. *ACS Synthetic Biology*, 2017, 6(10): 1860-1869.
- [124] BARET J C, MILLER O J, TALY V, et al. Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity[J]. *Lab on a Chip*, 2009, 9(13): 1850.
- [125] WAGNER J M, LIU L, YUAN S et al. A comparative analysis of single cell and droplet-based FACS for improving production phenotypes : riboflavin overproduction in *Yarrowia lipolytica*[J]. *Metabolic Engineering*, 2018, 47: 346-356.
- [126] MA C, TAN Z L, LIN Y, et al. Gel microdroplet-based high-throughput screening for directed evolution of xylanase-producing *Pichia pastoris*[J]. *Journal of Bioscience and Bioengineering*, 2019, 128(6): 662-668.
- [127] WANG X, REN L, SU Y, et al. Raman-activated droplet sorting (RADS) for label-free high-throughput screening of microalgal single-cells[J]. *Analytical Chemistry*, 2017, 89(22): 12569-12577.
- [128] DIXIT A, PARNAS O, LI B, et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens[J]. *Cell*, 2016, 167: 1853-1866.
- [129] TROPINI C, EARLE K A, HUANG K C, et al. The gut microbiome: connecting spatial organization to function[J]. *Cell Host Microbe*, 2017, 21: 433-442.
- [130] SHI H, COLAVIN A, LEE T K, et al. Strain library imaging protocol for high-throughput, automated single-cell microscopy of large bacterial collections arrayed on multiwell plates[J]. *Nature Protocols Erecipes for Researchers*, 2017, 12: 429-438.
- [131] LAWSON M J, CAMSUND D, LARSSON J, et al. *In situ* genotyping of a pooled strain library after characterizing complex phenotypes[J]. *Molecular Systems Biology*, 2017, 13: 947.
- [132] ZONG Y, ZHANG H M, LYU C, et al. Insulated transcriptional elements enable precise design of genetic circuits[J]. *Nature Communications*, 2017, 8: 52.
- [133] BASSALO M C, GARST A D, CHOUDHURY A, et al. Deep scanning lysine metabolism in *Escherichia coli*[J]. *Molecular Systems Biology*, 2018, 14(11): e8371.
- [134] LIANG L, LIU R, GARST A D, et al. CRISPR EnAbleD trackable genome engineering for isopropanol production in *Escherichia coli*[J]. *Metabolic Engineering*, 2017, 41: 1-10.
- [135] LIU R, LIANG L, GARST A D, et al. Directed combinatorial mutagenesis of *Escherichia coli* for complex phenotype engineering[J]. *Metabolic Engineering*, 2018, 47: 10-20.
- [136] LIANG L, LIU R, FOSTER K E, et al. Genome engineering of *E. coli* for improved styrene production[J]. *Metabolic Engineering*, 2020, 57: 74-84.
- [137] ANGERMUELLER C, PÄRNAMAA T, PARTS L, et al. Deep learning for computational biology[J]. *Molecular Systems Biology*, 2016, 12(7): 878.
- [138] COSTELLO Z, MARTIN H G. A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data[J]. *NPJ Systems Biology and Applications*, 2018, 4(1): 1-14.
- [139] CUPERUS J T, GROVES B, KUCHINA A, et al. Deep learning of the regulatory grammar of yeast 5' untranslated regions from 500,000 random sequences[J]. *Genome Research*, 2017, 27(12): 2015-2024.
- [140] ZHOU Y, LI G, DONG J, et al. MiYA, an efficient machine-learning workflow in conjunction with the YeastFab assembly strategy for combinatorial optimization of heterologous metabolic pathways in *Saccharomyces cerevisiae*[J]. *Metabolic Engineering*, 2018, 47: 294-302.
- [141] DENBY C M, LI R A, VU V T, et al. Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer[J]. *Nature Communications*, 2018, 9(1): 965.
- [142] ZHANG J, PSTERSEN S D, RADIVOJEVIC T, et al. Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism[J]. *Nature Communications*, 2020, 11(1): 4880.
- [143] GUO J, WANG T, GUAN C, et al. Improved sgRNA design in bacteria *via* genome-wide activity profiling[J]. *Nucleic Acids Research*, 2018, 46(14): 7052-7069.



通讯作者：张翀(1979—)，男，博士，副教授，博士生导师。主要研究方向为合成生物学、代谢工程、生物化工、系统生物学。  
E-mail: chongzhang@mail.tsinghua.edu.cn



第一作者：袁姚梦(1997—)，女，博士研究生。主要研究方向为合成生物学、代谢工程。  
E-mail: 2631825401@qq.com