

乳酸乳球菌 NZ9000 基因组规模代谢网络模型的构建与验证

孙伟康^{1,2}, 张娟^{1,2}, 堵国成^{2,3}

1 江南大学 工业生物技术教育部重点实验室, 江苏 无锡 214122

2 江南大学 生物工程学院, 江苏 无锡 214122

3 江南大学协同创新中心, 江苏 无锡 214122

孙伟康, 张娟, 堵国成. 乳酸乳球菌 NZ9000 基因组规模代谢网络模型的构建与验证. 生物工程学报, 2020, 36(8): 1629–1639.
Sun WK, Zhang J, Du GC. Construction and verification of *Lactococcus lactis* NZ9000 genome-scale metabolic model. Chin J Biotech, 2020, 36(8): 1629–1639.

摘要: 随着后基因组时代的到来, 工业微生物的代谢工程改造在工业生产上发挥着越来越重要的作用。而基因组规模代谢网络模型 (Genome-scale metabolic model, GSMM) 将生物体体内所有已知代谢信息进行整合, 为全局理解生物体的代谢状态、理性指导代谢工程改造提供了最佳的平台。乳酸乳球菌 NZ9000 (*Lactococcus lactis* NZ9000) 作为工业发酵领域的重要菌株之一, 由于其遗传背景清晰且几乎不分泌蛋白, 是基因工程改造和外源蛋白表达的理想模式菌株。文中基于基因组功能注释和比较基因组学构建了 *L. lactis* NZ9000 的首个基因组规模代谢网络模型 iWK557, 包含 557 个基因、668 个代谢物、840 个反应, 并进一步在定性和定量两个层次验证了 iWK557 的准确性, 以期为理性指导 *L. lactis* NZ9000 代谢工程改造提供良好工具。

关键词: 乳酸乳球菌 NZ9000, 基因组规模代谢网络模型, 模型验证, 系统生物学

Construction and verification of *Lactococcus lactis* NZ9000 genome-scale metabolic model

Weikang Sun^{1,2}, Juan Zhang^{1,2}, and Guocheng Du^{2,3}

1 Key Laboratory of Industrial Biotechnology, Ministry of Education, Jiangnan University, Wuxi 214122, Jiangsu, China

2 School of Biotechnology, Jiangnan University, Wuxi 214122, Jiangsu, China

3 Synergetic Innovation Center, Jiangnan University, Wuxi 214122, Jiangsu, China

Abstract: With the advent of the post-genomic era, metabolic engineering of microorganisms plays an increasingly important role in industrial production. The genome-scale metabolic model (GSMM) integrates all known metabolic information in the

Received: December 11, 2019; **Accepted:** February 24, 2020

Supported by: National Key Research and Development Program of China (No. 2017YFB0308401).

Corresponding author: Juan Zhang. Tel/Fax: +86-510-85918307; E-mail: zhangj@jiangnan.edu.cn

国家重点研发计划 (No. 2017YFB0308401) 资助。

网络出版时间: 2020-03-26

网络出版地址: <http://kns.cnki.net/kcms/detail/11.1998.q.20200325.0908.006.html>

organism to provide an optimal platform for global understanding of the metabolic state of the organism and rational guidance for metabolic engineering. As a model strain, *Lactococcus lactis* NZ9000 plays an important role in industrial fermentation, but there is still no specific genome-scale metabolic model for it. Based on genomic function annotation and comparative genomics, we constructed the first genome-scale metabolic model iWK557 of *L. lactis* NZ9000, which contains 557 genes, 668 metabolites, and 840 reactions, and further verified at both qualitative and quantitative levels, to provide a good tool for rationally guiding metabolic engineering.

Keywords: *Lactococcus lactis* NZ9000, genome-scale metabolic model, model validation, systems biology

乳酸菌发酵可以产生有机酸、抑菌素、挥发性风味物质,具有重要的益生功能,普遍应用于食品、医药、饲料等多种领域^[1]。在食品领域中乳酸菌应用最为常见,主要作为直投式发酵剂制备酸奶、奶酪等发酵制品,可以改善产品的风味和口感。同时发酵产生的抑菌素会抑制病原菌的繁殖,可作为生物防腐剂来保证食品的安全与品质^[2-3];在医药领域,乳酸菌联合给药有助于疾病的预防、治疗和术后修复^[4];饲料领域中添加乳酸菌可以有效增强动物胃肠道对营养成分的消化和吸收、提高动物机体的免疫力、减少疾病的发生,并减弱粪便的恶臭味^[2,5-6]。作为乳酸菌属的重要分支,乳酸乳球菌在工业发酵中占比超过 20%,是乳品发酵的重要菌株之一,在乳制品、肉制品、果蔬制品以及豆制品中应用广泛,是公认的食品安全级工业微生物^[7],而 *L. lactis* NZ9000 作为乳酸菌属的模式菌株,由于其生长迅速、遗传背景清晰,且几乎不分泌蛋白质,是基因工程改造以及外源蛋白表达的理想菌株^[8]。

随着基因组测序变得越来越经济和快速,系统生物学领域对于工业微生物细胞的研究也越来越深入。基因组规模代谢网络模型 (GSMM) 作为系统生物学的重要组成部分,其本质是将一个生物体所有与代谢相关的基因、酶、生物反应及代谢数据有效地组装到一起^[9],并通过不同的算法计算和预测生物体的各种代谢特性,系统地表征基因型和表型之间的关系^[10]。GSMM 被广泛应用于生长表型预测^[11]、代谢工程指导^[12]、种间进化差异分析^[13]以及多组学数据整合^[14]等 4 个方面。而工业微生物由于其胞内代谢活动的复杂性,难以通过单一的研究手段对其内部调控机制进行

系统的分析,且无法高效获得所需表型^[15]。近年来,越来越多的研究团队通过构建 GSMM 来全局分析工业微生物的代谢网络,特别在筛选代谢工程改造靶点、指导代谢工程改造方面取得了一定的进展。例如, Lee 等利用线性规划分析方法筛选出丙酮酸合成酶基因 (*ptsG*、*pykF*、*pykA*) 是大肠杆菌过量生产琥珀酸的基因敲除靶点,实验证实 *ptsG*、*pykF* 和 *pykA* 的敲除使琥珀酸的产量提高了 7 倍^[16-17]。Brochad 等在 GSMM 的指导下,通过过量表达酿酒酵母中的 *hsOMT* 基因使其香草醛产量提高了 5 倍^[18]。然而,对于乳酸菌这类与人类健康和生活紧密相连的微生物其关注度却很少,现在为数不多已建立 GSMM 的乳酸菌有乳酸乳球菌 IL1403 (*Lactococcus lactis* IL1403)^[19]、乳酸乳球菌 MG1363 (*Lactococcus lactis* MG1363)^[20]、植物乳杆菌 WCFS1 (*Lactobacillus plantarum* WCFS1)^[21]、干酪乳杆菌 LC2W (*Lactobacillus casei* LC2W)^[22]。虽同为乳酸菌属,但菌种之间个体差异较大,菌种胞内代谢调控复杂,无法直接套用。遗憾的是, *L. lactis* NZ9000 作为乳酸菌的模式菌株,现在仍没有 GSMM 供其对胞内代谢调控进行全局分析。

为系统地理解乳酸乳球菌的胞内代谢调控机制,本研究以 *L. lactis* NZ9000 作为研究对象,基于基因组功能注释和比较基因组学构建了 *L. lactis* NZ9000 的第一个基因组规模代谢网络模型 iWK557。并进一步在定性和定量两个层次验证了 iWK557 的准确性,以期为全局理解 *L. lactis* NZ9000 的胞内代谢机制和理性指导代谢工程改造提供新工具。

1 材料与方法

1.1 菌株、培养基与培养条件

目标菌株为 *Lactococcus lactis* NZ9000 (保藏编号: LLN, CP002094; 荷兰“NIZO Food Research”奶制品研究所)^[23]。*L. lactis* NZ9000 于 GM17 (M17 培养基中添加 5 g/L 的 D-葡萄糖)、Chemically defined medium (CDM)^[24]以及 Minimal medium (MS15)^[25]中培养, 培养温度为 30 °C。

1.2 实验方法

1.2.1 定性实验

将保藏在甘油管中的菌株于 GM17 固体培养基中划线活化, 将单菌落接种至 GM17 液体培养基中, 于 30 °C 静置培养 12 h 得到种子液, 随后将种子液以 1% (V/V) 的接种量转接到 MS15 培养基中培养 8 h, 取样测定吸光度 (Optical density, OD) OD_{600} 。更换 18 种底物作为唯一碳源, 底物为: 葡萄糖、淀粉、纤维二糖、果糖、半乳糖、乳糖、麦芽糖、甘露糖、蔗糖、海藻糖、半乳糖醇、肌醇、甘露醇、山梨糖醇、丙酮酸、木糖、核糖、半乳糖, 每种底物进行 6 个梯度实验 (底物浓度分别为: 1、5、10、15、20、30 g/L), 每个梯度 3 个平行样品。

1.2.2 定量实验

按 1.2.1 所述获得的种子液以 1% (V/V) 的接种量转接到 CDM 培养基中, 于 30 °C 静置培养 24 h, 每 2 h 取样。测量其生物量、细胞干重以及底物消耗。

1.3 主要仪器与分析平台

实验仪器: 恒温培养箱, 购于上海医疗器械研究所; M-100 生物传感器分析仪, 购于深圳市西尔曼科技有限公司; Agilent 1260 高效液相色谱仪, 购于安捷伦科技 (中国) 有限公司; 多功能酶标仪, 购于 Thermo Scientific 公司; 紫外可见分光光度计 (UV2450), 购于日本 Hitachi 公司。

软件: MATLAB R2018a, COBRA Toolbox^[26],

Origin 2018, Gurobi 8.0.1 optimizer。

1.4 草图模型的构建

在本研究中, 使用半自动方法构建草图模型。从 NCBI 或文献中下载 *L. lactis* NZ9000 的基因组序列并以 FASTA 格式保存到本地。通过同源比对法和手动构建法相结合的方式进行草图模型的构建^[27-28]。(1) 同源比对法: 使用本地 BLASTp, 基于 *L. lactis* NZ9000 的系统发育树, 下载亲缘相近菌株的基因组序列于本地建库比对。序列比对的筛选条件为: $e\text{-value} < 1 \times 10^{-6}$ 、 $\text{amino acid sequence identity} > 40\%$ 、 $\text{sequence matching length} > 70\%$ ^[29]。依据比对结果于亲缘相近菌株模型中进行代谢信息抓取, 得到草图模型一。(2) 手动构建法: 将 *L. lactis* NZ9000 的蛋白序列上传到 KEGG Automatic Annotation Server (KAAS)^[30]上进行基因组功能注释, 并结合 KEGG PATHWAY Database 进行代谢信息填充, 得到草图模型二。

将两种方法得到的草图模型进行整合: 基于 EC 号对草图模型间的基因进行比对, 若两者之间有相同的基因, 则看其功能注释是否一致, 若一致则保留草图模型一的结果, 若不一致则需通过文献和相关数据库进一步查证, 若无法查证, 则以草图模型一为准 (同源比对法较 KAAS 注释更准确^[31]); 若存在差异基因, 则将草图模型二中的基因-蛋白-反应 (Gene-protein-reaction, GPR) 补充进草图模型一中。通过此方法循环往复最终得到内容完整的草图模型。

1.5 模型精细化

草图模型构建完成后, 需结合各大数据库并依据 *L. lactis* NZ9000 的实际情况分别在基因水平、代谢物水平以及反应水平对其代谢信息进行填充与查证。模型精细化具体内容如下。

1.5.1 在代谢物水平进行模型精细化

代谢物格式统一: 将草图模型中的代谢物缩写形式进行统一, 格式参照 BIGG^[32]及文献中多数模型采用的小写缩写的形式。

质量电荷平衡：使用 COBRA Toolbox 中的 CheckMassChargeBalance 算法计算 GSMM 中反应的质量电荷是否平衡，并根据输出结果，对代谢物的化学式和电荷进行修正，对不平衡的反应进行配平。

1.5.2 在基因水平进行模型精细化

亚细胞定位：对基因编码蛋白质的作用区间进行亚细胞定位，将 *L. lactis* NZ9000 的基因组序列上传到 CELLO^[33]，依据预测结果和实际代谢情况进行确认。

多基因连接性区分：对于多个基因编码一个蛋白的情况进行区分，分为多酶复合体以及同工酶，并使用布尔逻辑关系中的“and”和“or”进行区分^[34]。

1.5.3 在反应水平进行模型精细化

生物量方程：依据构成 *L. lactis* NZ9000 所有前体物质及其所占的比例构建生物量方程，并依据算法‘minimum inconsistency under parsimony’ (MIP)^[35]，基于元素及质量守恒原则并结合 *L. lactis* NZ9000 的实际生理情况，对其生物量方程的化学式、电荷、分子量进行计算并对其分子量进行标准化。

反应可逆性：通过反应对应的 EC 号在 MetaCyc^[36]和 MetaNetX^[37]进行手动搜索确定反应方向性；对于没有 EC 号的反应，通过反应名称结合 BIGG 或结合乳酸菌相应文献进行确认，对于无法查证反应一律设为可逆反应^[31]。

转运反应和交换反应的添加：在 Transporter Classification Database (TCDB)^[38]下载所有的转运蛋白序列与 *L. lactis* NZ9000 的蛋白序列进行序列比对，依据序列比对结果并结合基因组功能注释和文献进行转运反应的添加；结合添加的转运反应以及“湿实验”中 *L. lactis* NZ9000 所需营养条件进行交换反应的添加。

代谢漏洞的填补：使用 gapFind 以及

gapAnalysis 算法进行代谢漏洞查找，结合基因组功能注释结果进行代谢漏洞填补。

1.6 转化为数学模型

上述 1.4 得到的草图模型以及 1.5 的模型精细化都是在 EXCEL 中进行，在进行模型的计算模拟之前需要将其转化为计算机可读的 SBML 格式。将 EXCEL 格式的模型导入 MATLAB 后使用 xls2model 算法将其转化为数学模型，随后使用 writeCbModel 算法即可将其导出为 SBML 格式。

2 结果与分析

2.1 基因组功能注释与草图模型的获得

在本研究中，基于基因组功能注释，使用半自动构建法进行草图模型的构建。

同源比对法：基于比较基因组学进行基因组功能注释，通过 *L. lactis* NZ9000 的系统发育树，选择了已有 GSMM 且亲缘相近的菌株 *L. lactis* MG1363、*L. plantarum* WCFS1、*L. casei* LC2W 作为参照菌株，进行序列比对。依据 1.4 所述的筛选条件进行双向比对，*L. lactis* NZ9000 与参照菌株之间的基因匹配数分别是 2 275 个、1 294 个和 1 153 个。而 *L. lactis* MG1363、*L. plantarum* WCFS1、*L. casei* LC2W 的序列基因数分别是 2 563 个、3 009 个和 3 003 个，由此可以发现基因匹配数的多少与基因组大小不呈正相关，与其系统发育树的亲缘远近相关。将筛选得到的结果使用自行编写的程序整合入已有 GSMM 的菌株模型中，将以上得到的代谢信息进行整合，得到整合模型，包含 548 个基因、1 290 个反应，随后通过筛选，删除重复的反应和功能注释相同的基因，保留不重复的部分，从而获得草图模型一，包含 455 个基因、739 个代谢物、874 个反应。

手动构建法：将目标菌株的蛋白序列上传到 KAAS 后，依据返回的结果，2 314 个基因中 826 个有具体的功能注释，其中仅有 69 个有明确

的 EC 号, 对应 475 个基因。依据 EC 号结合 KAAS 返回的注释链接在 KEGG 中查找对应的反应, 得到草图模型二, 包含 475 个基因、69 个反应。

依据 1.4 所述的方法将以上得到的草图模型进行整合得到最初完整的草图模型, 含有 516 个基因、739 个代谢物、876 个反应。

2.2 模型精细化

分别在代谢物水平、基因水平以及反应水平对 1.4 得到的草图模型进行了模型代谢信息填充与修正。

(1) 在代谢物水平进行模型精细化

采用不同基因组功能注释方法得到了两种不同形式的草图模型, 需要对其所有的代谢物格式进行统一。不同方法中代谢物的表现形式有所差异, 以 6-磷酸葡萄糖为例, 在模型 *L. casei* LC2W 中代谢物格式为“g6p”, 而在模型 *L. plantarum* WCFS1 中采用的则是大写缩写的格式, 为“G6P[c]”。本研究基于代谢物全称对代谢物格式进行统一, 将其统一为小写缩写的形式, 如“g6p[c]”, 同时对于不能匹配的代谢物, 按照其原有形式保留在模型中。

(2) 在基因水平进行模型精细化

对基因编码蛋白的作用区间进行亚细胞定位, 依据 CELLO 返回的结果并结合 *L. lactis* NZ9000 为原核细菌的实际情况, 将蛋白作用区间分为胞内 (Cytoplasmic) 与胞外 (Extracellular)。其中胞内 2 126 个、胞外 188 个, 总共 2 314 个基因。随后对基因关联性进行分类, 分为多酶复合体与同工酶体系, 并使用布尔逻辑关系中的“and”和“or”来进行区分, 其中多酶复合体催化了 21 个反应, 同工酶催化 151 个反应。

(3) 在反应水平进行模型精细化

生物量方程: 是人为添加用于使用模型计算细胞生长情况的虚拟反应。在拟稳态的假设下, 以生物量方程作为目标函数进行优化问题求解时, 模型能够预测细胞的最大比生长速率。其表

现形式如下:

$$\sum_i S_i^R R_i = \text{Biomass} + \sum_j S_j^B B_j$$

S_i^R 是生物量前体代谢物 R_i 的化学计量系数。 S_j^B 是副产物 B_j 的化学计量系数。由于所有生物量前体代谢物系数均基于 1 g DCW, 基于质量守恒原则, 生物量方程的分子量计算如下:

$$\left(\sum_i S_i^R M_i^R - \sum_j S_j^B M_j^B\right) / 1000 = 1 \text{ (g/mmol)}$$

M_i^R 、 M_j^B 是代谢产物 R_i 和 B_j 的分子量, 根据定义, 生物量的理论分子量为 1 g/mmol。

在本研究中, 生物量方程参考框架模型 iNF518^[20], 生长相关维持能 (GAM) 和非生长相关维持能 (NGAM) 分别确立为 39.4 mmol ATP/(g DCW·h) 和 0.92 mmol ATP/(g DCW·h), 依据算法 ‘minimum inconsistency under parsimony’ (MIP), 基于质量守恒原则并结合目标菌株的实际生理情况, 对其生物量方程的分子量、化学式以及各前体的化学计量系数进行计算和标准化, 生物量方程的分子量由原本的 825.05 g/mmol 标准化为 1 000.00 g/mmol, 化学式由 $C_{33.697509}H_{53.495976}N_{8.358924}O_{13.75044}P_{0.68944}S_{0.248079}$ 标准化为 $C_{40.6465338215}H_{64.0995743231}N_{9.9076885333}O_{16.9635199515}P_{0.8322218004}S_{0.350164004}$ 。

反应可逆性: 依据 1.5.3 所述方法对反应的方向性进行核实区分, 标准化后 GSMM 中含有 349 个可逆反应、491 个不可逆反应。

转运和交换反应: 依据 1.5.3 所述方法, 依据序列比对结果和文献挖掘进行转运反应添加, 共添加了 119 个转运反应。交换反应依据 *L. lactis* NZ9000 常用培养基中的营养物质, 并保证基础营养元素 C、H、O、N、P、S 的吸收, 共添加了 112 个交换反应。

代谢漏洞填补: 依据 1.5.3 所述方法进行漏洞查找和填补, 共填补了 105 个 gaps。

2.3 数学模型的转换

将模型精炼得到的 EXCEL 格式的模型导入

MATLAB 依据 1.6 的方法转换为 $S_{(m \times n)}$ 矩阵, 矩阵中 m 行代表代谢物的数量, n 列代表反应的数量, 矩阵由“1”、“-1”和“0”组成, 其中“1”代表该代谢物在此反应中作为底物, “-1”代表该代谢物作为产物, “0”代表该代谢物不参与该反应。这样 $S_{(m \times n)}$ 矩阵就将代谢物与反应巧妙地联系起来变为数学矩阵。 $S_{(668 \times 840)}$ 表示该 GSMM 含有 668 个代谢物和 840 个反应。模型中有 557 个基因, 依据命名规则将构建完成的 GSMM 命名为 iWK557。

2.4 模型验证

iWK557 构建完成后, 在应用之前为进一步确认其计算预测的精准性, 本研究从定性和定量两个层次对其进行验证。

2.4.1 定性验证

L. lactis NZ9000 能够代谢利用多种糖类和其他碳源以获得能量、氧化还原力和合成生物大分子的前体物质^[39]。参考文献[19, 22, 40]报道, 乳酸菌属发酵可利用的主要碳源有葡萄糖、纤维二糖、果糖、半乳糖、乳糖、麦芽糖、甘露糖、蔗糖、海藻糖、甘露醇、山梨醇。此外为了进一步考察 *L. lactis* NZ9000 的底物利用谱, 参照了大肠杆菌、酵母、谷氨酸棒状杆菌及放线菌等实验室常见菌株的可利用碳源进行了碳源类型的补充, 从而确定 18 种底物作为唯一碳源进行定性实验^[19, 40]。

使用 iWK557 利用流平衡 (Flux balance analysis, FBA) 算法对 18 种唯一碳源的利用能力及主要产物乳酸的生成进行预测, 并依据 1.2.1 的实验方法进行定性实验, 通过模型的预测和“湿实验”的验证, 进一步对模型的准确性进行评估 (表 1)。其中, 木糖在原文献中可被大肠杆菌利用, 作为本研究的备选碳源, 经实验证实不能被 *L. lactis* NZ9000 作为碳源利用。比对结果显示模型预测结果与实验结果一致, 表明模型 iWK557 能够准确预测菌体利用不同碳源进行代谢生长和产物生成。

表 1 模型与菌株底物利用能力比对

Table 1 Comparison of substrate utilization ability of the model and strain

Carbon source	Experimental results		Simulation results	
	Cell growth	Lactate production	Cell growth	Lactate production
Glucose	+	+	+	+
Starch	+	+	+	+
Cellose	+	+	+	+
Fructose	+	+	+	+
Galactose	+	+	+	+
Lactose	+	+	+	+
Maltose	+	+	+	+
Mannose	+	+	+	+
Sucrose	+	+	+	+
Trehalose	+	+	+	+
Galactitol	+	+	+	+
Inositol	+	+	+	+
Mannitol	+	+	+	+
Sorbitol	+	+	+	+
Pyruvate	+	+	+	+
Xylose	-	-	-	-
Ribose	+	+	+	+
Melibiose	+	+	+	+

2.4.2 定量验证

为进一步验证 iWK557 预测的生长值的准确性。在本研究中使用分批发酵的实验数据进行验证, 随后通过恒化实验数据进一步验证模型的预测准确性。

分批发酵验证: 使用 CDM 培养基进行分批培养, 将实验数据导入 Origin 2018 以获取 *L. lactis* NZ9000 的细胞生长、葡萄糖消耗以及乳酸生成曲线 (图 1A), 随后通过 Origin 2018“Analysis”工具栏中的“Nonlinear Curve fit”进行非线性拟合获得比生长速率、葡萄糖比消耗速率以及乳酸比生成速率曲线 (图 1B)。*L. lactis* NZ9000 在分批发酵中的最大比生长速率为 $0.017\ 004\ \text{h}^{-1}$, 葡萄糖比消耗速率为 $0.55\ \text{mmol}/(\text{g DCW} \cdot \text{h})$, 乳酸比合成速率为 $0.066\ \text{mmol}/(\text{g DCW} \cdot \text{h})$ 。在 MATLAB 中使用 FBA 进行模拟, 约束葡萄糖的吸收速率, 计算得到的最大比生长速率为 $0.017\ 1\ \text{h}^{-1}$, 乳酸比合

成速率为 $0.064\ 3\ \text{mmol}/(\text{g DCW}\cdot\text{h})$ 与实验结果的误差率分别为 0.5% 和 2.57% , 表明该模型可以准确预测 *L. lactis* NZ9000 的比生长速率和主要代谢产物的比合成速率。

恒化实验验证: 在恒定稀释率的条件下, 细胞的生长状态保持恒定, 此时, 测量发酵液中各种营养物的形成和消耗, 可以量化特定稀释率下菌株的代谢状态。将获得的恒化实验数据进行单位转换, 然后导入到模型中进行约束, 从而验证模型预测的准确性。本研究使用 Nicolas 等^[20]和 Petri-Jaan 等^[41]基于 CDM 培养基的恒化实验数据进行验证, 通过比较 4 种稀释率 ($0.05\ \text{h}^{-1}$ 、 $0.25\ \text{h}^{-1}$ 、 $0.4\ \text{h}^{-1}$ 、 $0.45\ \text{h}^{-1}$) 条件下细胞的比生长速率及主要产物乳酸和乙酸的比合成速率和模型表型预

测结果之间的差异来进行定量验证。在 4 种稀释率条件下, 碳源输入速率分别约束为 $1.20\ \text{mmol}/(\text{g DCW}\cdot\text{h})$ 、 $6.20\ \text{mmol}/(\text{g DCW}\cdot\text{h})$ 、 $11.53\ \text{mmol}/(\text{g DCW}\cdot\text{h})$ 、 $16.36\ \text{mmol}/(\text{g DCW}\cdot\text{h})$; 氮源输入速率分别约束为 $0.13\ \text{mmol}/(\text{g DCW}\cdot\text{h})$ 、 $0.55\ \text{mmol}/(\text{g DCW}\cdot\text{h})$ 、 $1.26\ \text{mmol}/(\text{g DCW}\cdot\text{h})$ 、 $1.44\ \text{mmol}/(\text{g DCW}\cdot\text{h})$ 。如表 2 所示, 比生长速率 (Specific growth rate, μ)、乳酸合成速率 (Lactate production rate, LPR) 以及乙酸合成速率 (Acetate production rate, APR) 的平均误差率分别为 4.08% 、 4.21% 、 5.47% , 最大偏差不超过 9% , 普遍低于 10% ^[42], 说明 iWK557 不仅能够准确反映细胞的生长状态, 也能准确预测细胞主要产物的生成, 真实反映细胞的生理代谢情况。

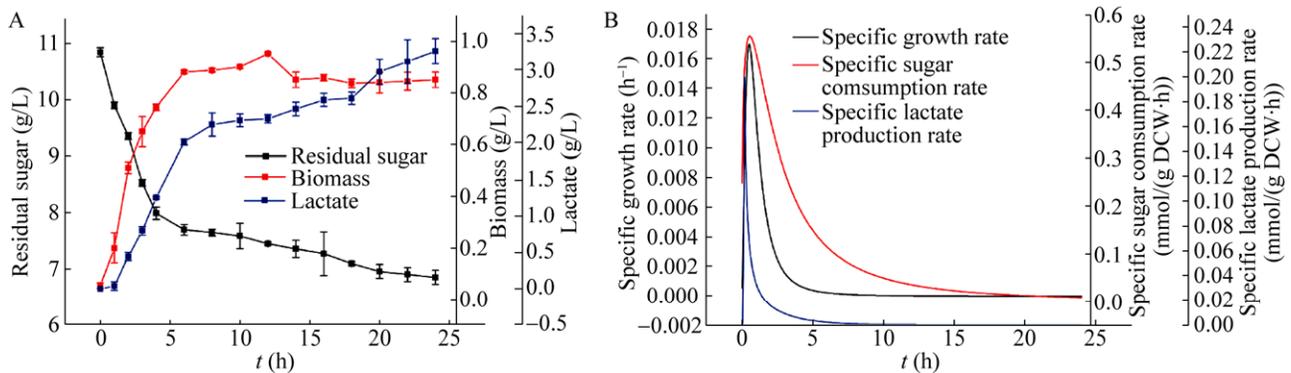


图 1 定性湿实验图 (A: *L. lactis* NZ9000 的生物量、葡萄糖消耗和乳酸生成曲线; B: *L. lactis* NZ9000 的比生长速率、葡萄糖比消耗速率和乳酸比合成速率曲线)

Fig. 1 Qualitative experiment chart. (A) Cell growth, glucose consumption and lactic acid production curve of *L. lactis* NZ9000. (B) Specific growth rate, glucose specific consumption rate and lactate specific production rate curve of *L. lactis* NZ9000.

表 2 定量模拟结果比对

Table 2 Comparison of quantitative simulation results

Project	0.05 (h^{-1})		0.25 (h^{-1})		0.40 (h^{-1})		0.45 (h^{-1})	
	<i>In silico</i>	D ^{****} (%)	<i>In silico</i>	D (%)	<i>In silico</i>	D (%)	<i>In silico</i>	D (%)
μ^*	0.050 4	0.80	0.23	6.16	0.36	7.70	0.457 6	1.68
LPR ^{**}	0.920 0	4.16	4.20	6.87	18.32	0.43	27.600 0	5.39
APR ^{***}	1.430 0	4.24	4.32	7.51	2.48	8.87	1.390 0	1.27

*: μ represents specific growth rate; **: LPR represents lactate production rate; ***: APR represents acetate production rate; ****: D represents deviation.

2.5 iWK557 的基本特征

模型 iWK557 包含 557 个基因、840 个反应 (609 个细胞质反应、119 个转运反应和 112 个交换反应), 基因覆盖率达到 23.89%。有 76% 的反应有明确的基因关联。参考 KEGG 和 RAST server, 将 iWK557 的反应分为 10 个代谢亚系统, 各代谢亚系统的反应数目如图 2 所示。在代谢亚系统中, 氨基酸代谢 (19.523%) 所占比例最高, 其次是核苷酸代谢 (14.40%), 最后是转运 (14.16%) 和交换反应 (13.33%), 这 4 个代谢亚系统所占比例超过 60%。

为进一步了解 iWK557 的详细特征, 将模型 iWK557 与 *L. lactis* MG1363、*L. lactis* I11403 的 GSMM 进行比较 (表 3), 比较结果如下所示:

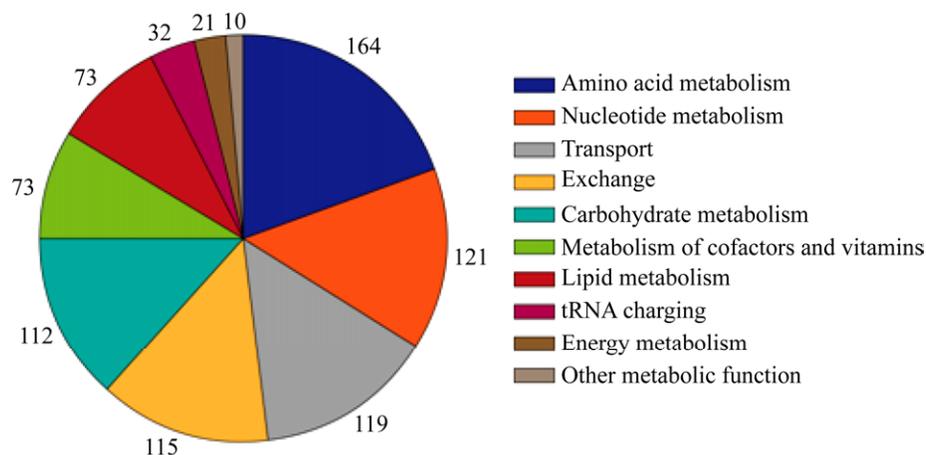


图 2 iWK557 代谢亚系统分布

Fig. 2 Metabolic subsystems distribution of iWK557.

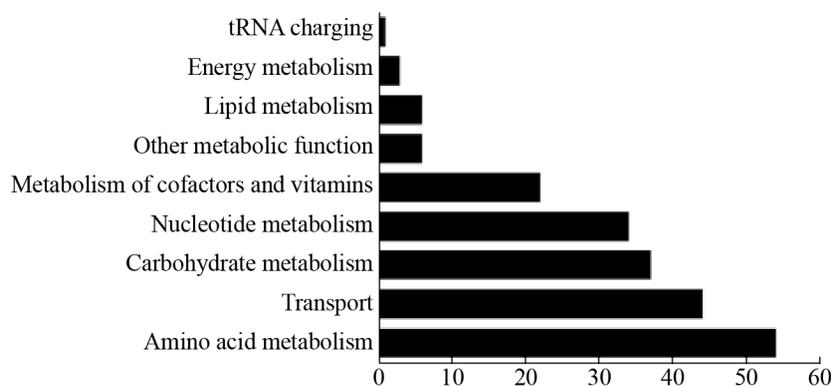


图 3 iWK557 特有反应代谢亚系统分布

Fig. 3 Metabolic subsystems distribution for specific reactions in iWK557.

1) 3 个模型的基因覆盖率分别为 23.89%、22.23%、15.72%，3 个菌株都属于乳酸乳球菌，细胞分区都为胞内和胞外两个区间。2) 3 个模型的共有反应数为 319 个，其中氨基酸代谢 (29.15%)、核苷酸代谢 (22.88%)、碳水化合物代谢 (17.24%) 以及维生素与辅因子代谢 (10.65%) 所占比例为 79.93%，表明 3 个模型在以上 4 种代谢亚系统中具有较高的相似性。iWK557 特有 207 个代谢反应，主要集中于氨基酸代谢、转运反应、碳水化合物代谢以及核苷酸代谢，特有反应的代谢亚系统分布如图 3 所示。3) 3 个模型中，*L. lactis* I11403 的代谢物是另外两个菌株模型的子集，其中，*L. lactis* MG1363 含有特有代谢物 67 个，iWK557 含有特有代谢物 86 个。

表 3 乳酸乳球菌基因组及 GSMM 特征比对

Table 3 Comparison of genome characteristics and characteristics of the models of *Lactococcus lactis*

Parameter	<i>L. lactis</i> NZ9000	<i>L. lactis</i> MG1363	<i>L. lactis</i> II1403
Genomic information			
Total length (Mb)	2.53	2.52	2.36
Protein count	2 314.00	2 563.00	2 310.00
GC (%)	35.70	35.70	35.30
Model contents			
Total reactions	840.00	754.00	621.00
Internal reactions	601.00	530.00	492.00
Transport reactions	116.00	119.00	129.00
Metabolites	668.00	650.00	422.00
Gene	557.00	518.00	358.00
Gene coverage (%)	23.89	22.23	15.72
Compartment	(c, e)	(c, e)	(c, e)

3 讨论

在本研究中, 基于基因组功能注释和比较基因组学构建了 *L. lactis* NZ9000 的首个全基因组规模代谢网络模型, 将其命名为 iWK557。通过定性和定量分析对 iWK557 进行了验证。定性验证结果显示, 模型 iWK557 对不同碳源的模拟预测结果与实验结果一致, 说明 iWK557 能够准确表征菌体生长和产物合成这两个最为重要的细胞特征; 随后在分批发酵和恒化实验中对 iWK557 对菌体生长和产物合成预测的准确性进行定量验证, 总平均偏差不超过 5%, 偏差率普遍低于 10%, 说明 iWK557 能够准确预测 *L. lactis* NZ9000 的生长和产物生成, 反映细胞的真实生理状态。

在后续的研究中, 将利用 iWK557 对 *L. lactis* NZ9000 的内部代谢调控机制进行全局分析, 并以此为基础, 通过整合多组学数据, 构建条件特异性模型进行代谢流量分析和代谢靶点预测用以理性指导代谢工程改造。

REFERENCES

[1] Zhu ZM, Ji XM, Wu ZM, et al. Improved acid-stress tolerance of *Lactococcus lactis* NZ9000 and *Escherichia coli* BL21 by overexpression of the

anti-acid component recT. *J Ind Microbiol Biotechnol*, 2018, 45(12): 1091–1101.

- [2] Zhang YW. Regulation of aspartate metabolic pathways for improving acid stress resistance in *Lactococcus lactis* NZ9000[D]. Wuxi: Jiangnan University, 2018 (in Chinese).
张彦位. 调控天冬氨酸代谢途径提高 *Lactococcus lactis* NZ9000 的酸胁迫抗性[D]. 无锡: 江南大学, 2018.
- [3] Zanirati DF, Abatamarco M Jr, De Cicco Sandes SH, et al. Selection of lactic acid bacteria from Brazilian kefir grains for potential use as starter or probiotic cultures. *Anaerobe*, 2015, 32: 70–76.
- [4] Aragón F, Carino S, Perdígón G, et al. The administration of milk fermented by the probiotic *Lactobacillus casei* CRL 431 exerts an immunomodulatory effect against a breast tumour in a mouse model. *Immunobiology*, 2014, 219(6): 457–464.
- [5] Matusiak K, Oleksy M, Borowski S, et al. The use of *Yucca schidigera* and microbial preparation for poultry manure deodorization and hygienization. *J Environ Manag*, 2016, 170: 50–59.
- [6] Kim JD, Park KM. Effectiveness of *Lactobacillus plantarum* strain KJ-10311 to remove characteristic malodorous gases in piggery slurry. *Asian-Australasian J Anim Sci*, 2006, 19(1): 144–152.
- [7] Cavanagh D, Fitzgerald GF, McAuliffe O. From field to fermentation: The origins of *Lactococcus lactis* and its domestication to the dairy environment. *Food Microbiol*, 2015, 47: 45–61.
- [8] Wu CD, Huang J, Zhou RQ. Genomics of lactic acid bacteria: current status and potential applications. *Crit Rev Microbiol*, 2017, 43(4): 393–404.
- [9] Dong FQ. Reconstruction and verification of the genome-scale metabolic models of *Bacillus Coagulans* 36D1[D]. Shanghai: East China University of Science and Technology, 2017 (in Chinese).
董风晴. 凝结芽孢杆菌 36D1 全基因组代谢网络模型的构建和验证[D]. 上海: 华东理工大学, 2017.
- [10] Feist AM, Palsson BØ. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotechnol*, 2008, 26(6): 659–667.
- [11] Motamedian E, Saeidi M, Shojaosadati SA.

- Reconstruction of a charge balanced genome-scale metabolic model to study the energy-uncoupled growth of *Zymomonas mobilis* ZM1. *Mol Biosyst*, 2016, 12(4): 1241–1249.
- [12] Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol*, 2009, 5: 320.
- [13] Pál C, Papp B, Lercher MJ, et al. Chance and necessity in the evolution of minimal metabolic networks. *Nature*, 2006, 440(7084): 667–670.
- [14] O'Brien EJ, Palsson BO. Computing the functional proteome: recent progress and future prospects for genome-scale models. *Curr Opin Biotechnol*, 2015, 34: 125–134.
- [15] Ye C, Xu N, Chen XL, et al. Application of metabolic network model to analyze intracellular metabolism of industrial microorganisms. *Chin J Biotech*, 2019, 35(10): 1901–1913 (in Chinese).
叶超, 徐楠, 陈修来, 等. 应用代谢网络模型解析工业微生物胞内代谢. *生物工程学报*, 2019, 35(10): 1901–1913.
- [16] Lee SJ, Lee DY, Kim TY, et al. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation. *Appl Environ Microbiol*, 2005, 71(12): 7880–7887.
- [17] Wang QZ, Zhao XM. The research progress of succinic acid fermentation strains. *Chin J Biotechnol*, 2007, 23(4): 570–576 (in Chinese).
王庆昭, 赵学明. 琥珀酸发酵菌种研究进展. *生物工程学报*, 2007, 23(4): 570–576.
- [18] Brochado AR, Patil KR. Overexpression of *O*-methyltransferase leads to improved vanillin production in baker's yeast only when complemented with model-guided network engineering. *Biotechnol Bioeng*, 2013, 110(2): 656–659.
- [19] Oliveira AP, Nielsen J, Förster J. Modeling *Lactococcus lactis* using a genome-scale flux model. *Bmc Microbiol*, 2005, 5: 39.
- [20] Flahaut NAL, Wiersma A, Van De Bunt B, et al. Genome-scale metabolic model for *Lactococcus lactis* MG1363 and its application to the analysis of flavor formation. *Appl Microbiol Biotechnol*, 2013, 97(19): 8729–8739.
- [21] Teusink B, Wiersma A, Molenaar D, et al. Analysis of growth of *Lactobacillus plantarum* WCFS1 on a complex medium using a genome-scale metabolic model. *J Biol Chem*, 2006, 281(52): 40041–40048
- [22] Xu N, Liu J, Ai LZ, et al. Reconstruction and analysis of the genome-scale metabolic model of *Lactobacillus casei* LC2W. *Gene*, 2015, 554(2): 140–147.
- [23] Shi WJ, Li Y, Gao XL, et al. Improvement of the respiration efficiency of *Lactococcus lactis* by decreasing the culture pH. *Biotechnol Lett*, 2016, 38(3): 495–501.
- [24] Poolman B, Konings WN. Relation of growth of *Streptococcus lactis* and *Streptococcus cremoris* to amino acid transport. *J Bacteriol*, 1988, 170(2): 700–707.
- [25] Coccagnoli M, Garrigues C, Novak L, et al. Rational development of a simple synthetic medium for the sustained growth of *Lactococcus lactis*. *J Appl Bacteriol*, 1995, 79(1): 108–116.
- [26] Heirendt L, Arreckx S, Pfau T, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protocols*, 2019, 14(3): 639–702.
- [27] Liu T, Zou W, Liu LM, et al. A constraint-based model of *Scheffersomyces stipitis* for improved ethanol production. *Biotechnol Biofuels*, 2012, 5: 72.
- [28] Ma ZB, Ye C, Deng WW, et al. Reconstruction and analysis of a genome-scale metabolic model of *Ganoderma lucidum* for improved extracellular polysaccharide production. *Front Endocrinol*, 2018, 9: 12.
- [29] Tian WD, Skolnick J. How well is enzyme function conserved as a function of pairwise sequence identity? *J Mol Biol*, 2003, 333(4): 863–882.
- [30] Moriya Y, Itoh M, Okuda S, et al. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*, 2007, 35(S2): W182–W185.
- [31] Liu T. Reconstruction and application of *Scheffersomyces stipitis* genome-scale metabolic model[D]. Jiangsu: Jiangnan University, 2013 (in Chinese).
刘婷. 树干毕赤酵母基因组规模代谢网络模型构建及应用[D]. 无锡: 江南大学, 2013.
- [32] King ZA, Lu J, Dräger A, et al. BiGG models: a platform for integrating, standardizing and sharing

- genome-scale models. *Nucleic Acids Res*, 2016, 44(D1): D515–D522.
- [33] Yu CS, Lin CJ, Hwang JK. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on *n*-peptide compositions. *Protein Sci*, 2004, 13(5): 1402–1406.
- [34] Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, 2010, 5(1): 93–121.
- [35] Chan SHJ, Cai JY, Wang L, et al. Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. *Bioinformatics*, 2017, 33(22): 3603–3609.
- [36] Caspi R, Billington R, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res*, 2018, 46(D1): D633–D639.
- [37] Moretti S, Martin O, Van Du Tran T, et al. MetaNetX/MNXref - reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Res*, 2016, 44(D1): D523–D526.
- [38] Saier MH Jr, Reddy VS, Tsu BV, et al. The transporter classification database (TCDB): recent advances. *Nucleic Acids Res*, 2016, 44(D1): D372–D379.
- [39] De Vos WM. Metabolic engineering of sugar catabolism in lactic acid bacteria. *Anton Van Leeuwenh*, 1996, 70(2): 223–242.
- [40] Castro R, Neves AR, Fonseca LL, et al. Characterization of the individual glucose uptake systems of *Lactococcus lactis*: mannose-PTS, cellobiose-PTS and the novel GlcU permease. *Mol Microbiol*, 2009, 71(3): 795–806.
- [41] Lahtvee PJ, Adamberg K, Arike L, et al. Multi-omics approach to study the growth efficiency and amino acid metabolism in *Lactococcus lactis* at various specific growth rates. *Microb Fact*, 2011, 10: 12.
- [42] Wang YL, Xu N, Ye C, et al. Reconstruction and *in silico* analysis of an *Actinoplanes* sp. SE50/110 genome-scale metabolic model for acarbose production. *Front Microbiol*, 2015, 6: 632.

(本文责编 郝丽芳)