



基于不同氮源培养条件的巴氏芽孢杆菌脲酶功能转录组分析

裴迪¹⁾ 刘志明^{1,2)*} 胡碧茹¹⁾ 吴文健¹⁾⁽¹⁾ 国防科技大学文理学院, 长沙 410073; ⁽²⁾ 军事科学院防化研究院, 北京 102205

摘要 巴氏芽孢杆菌是源于土壤的革兰氏阳性菌, 人们利用其高效的脲酶活性诱导产生碳酸钙的现象开发了多种应用场景。然而, 巴氏芽孢杆菌的生物矿化相关代谢机制还不够明确, 尤其是对在矿化作用中发挥核心作用的脲酶基因结构、表达调控机制及关联代谢等方面的研究相对较少。当前, 巴氏芽孢杆菌应用研究中面临的矿化反应不可控性及不稳定性等问题都源于脲酶代谢机制的研究匮乏。因此, 进一步揭示巴氏芽孢杆菌脲酶的基因信息、表达调控机制及相关代谢机理迫在眉睫。本文通过转录组测序, 对比了4种培养条件下巴氏芽孢杆菌的生长情况和基因表达情况, 解析了脲酶的代谢机制, 结果进一步证明ATP合成与脲酶表达及尿素水解相关联, 最终预测了脲酶的双操纵子结构。

关键词 巴氏芽孢杆菌, 转录组, 操纵子, 脲酶, ATP

中图分类号 Q819

DOI: 10.16476/j.pibb.2020.0441

微生物诱导碳酸钙沉淀 (MICP) 是指在微生物代谢活动 (包括尿素分解、光合作用、反硝化、氨化、硫酸盐还原和甲烷氧化等) 作用下形成碳酸钙沉淀的过程^[1-2]。迄今为止, 已经发现多种微生物可以通过水解尿素的方式实现MICP。该矿化方式已经得到广泛的应用, 涉及修复建筑和道路裂缝、固化沙土、制造自修复水泥、开发生物矿化材料等^[3-11]。在使用的尿素水解菌中, 巴氏芽孢杆菌是MICP中最为常见且应用最广泛的菌种^[12]。

巴氏芽孢杆菌可以产生脲酶, 将尿素分解为碳酸根和氨根, 生成的碳酸根可与胞外的钙离子反应, 形成碳酸钙沉淀^[1, 13]。虽然该矿化作用已应用于多种用途, 但它仍然存在一些应用限制, 例如对不同应用条件的适应性, 细菌生长和脲酶活性的不可控性, 可能存在的副产物以及大规模工程应用上的不稳定性, 尤其是存在着比化学法及传统方法成本更高的问题。

当前已有一些研究尝试通过优化巴氏芽孢杆菌的生长和矿化条件来解决这些问题, 包括优化尿素、氯化钙和镍离子等浓度, 以最大化巴氏芽孢杆菌的生长速率和碳酸钙沉淀能力^[14]。但已开展的

应用研究发现, 单纯增强矿化作用无法有效解决大规模工程应用中存在的问题。此外, 脲酶活性在巴氏芽孢杆菌的矿化作用中同样起到至关重要的作用, 但是巴氏芽孢杆菌却存在着复杂的脲酶表达调控机制, Whiffin等^[15]研究表明巴氏芽孢杆菌的脲酶并不是简单的恒表达型, 细菌自身的代谢废物可能会抑制脲酶的表达。尽管对影响巴氏芽孢杆菌脲酶表达变化规律的因素进行了研究^[16], 但总体而言, 各种因素对巴氏芽孢杆菌脲酶表达产生影响的作用机制尚不明确, 在特定的应用环境中, 通过人为调节以改变其脲酶表达情况仍是一项难以实现的技术方法。2019年, 东南大学向NCBI提交了巴氏芽孢杆菌基因组测序结果^[17], 为转录组分析等生物信息学研究提供了不可或缺的基础数据库。

本文通过高通量转录组测序技术对巴氏芽孢杆菌在4种培养条件下的转录组表达情况进行了研究, 解析了巴氏芽孢杆菌脲酶表达调控机制及相关代谢机理, 初步揭示了巴氏芽孢杆菌特殊的脲酶操

* 通讯联系人。

Tel: 0731-87001815, E-mail: lzm@nudt.edu.cn

收稿日期: 2020-12-17, 接受日期: 2021-03-11

纵子 (operon) 结构, 此研究有助于实现大规模工程应用中的脲酶表达人为操控, 以达到高效调节矿化反应的目的.

1 材料与方法

1.1 巴氏芽孢杆菌的培养

本研究采用的巴氏芽孢杆菌菌株为 BNCC 337394, 其冻干粉采购于 Bnbio 公司 (北京北纳创联生物技术有限公司). 该菌株全基因组测序结果在 2019 年 3 月于 NCBI (NZ_CP038012.1) 发布^[17]. 细菌培养时, 首先用预制的活化培养基 (蛋白胨 5 g/L, 酵母提取物 3 g/L, 尿素 20 g/L, pH=7.3) 对细菌冻干粉进行活化, 在 30°C 下活化培养 24 h. 然后将得到的少量菌液在 Christensen's Urea Agar^[18] 平板上进行涂板, 平板在 30°C 下培养 24 h, 挑选其中长势较好的单菌落进行扩大培养并以此进行后续研究, 扩大培养的培养基为美国菌种保藏中心 (ATCC) 推荐的 NH₄-YE 培养基.

1.2 不同培养条件下细菌的生长测试及转录组样本选择

为挑选合适的培养时间点进行转录组测序, 对 4 种水相培养基条件下的菌液进行 48 h 培养, 4 组培养基及其有效组分为: a. 对照组 (B), 酵母提取物 20 g/L; b. 铵根组 (N), 酵母提取物 20 g/L、氯化铵 10 g/L; c. 尿素组 (U), 酵母提取物 20 g/L、尿素 5.62 g/L, 氮含量与铵根组相同; d. 铵根 Tris 组 (N_T), 酵母提取物 20 g/L、氯化铵 10 g/L、Tris 15.75 g/L. 使用 1 mol/L 的 NaOH 溶液将 a~c 组培养基的 pH 调节至 9.0, d 组由于 Tris 的存在自身 pH 为 9.0, 均处于巴氏芽孢杆菌的最适 pH 范围, 并用 1 mol/L 的 NaCl 溶液调节使所有组别 Na⁺ 浓度相同. 上述各组培养基均配制 300 ml 并置于 1 L 的锥形瓶中, 经 121°C、20 min 灭菌后 (对于含有铵根或尿素的培养基, 氯化铵或尿素需预先配制溶液经过抽滤灭菌, 在培养基高温灭菌后再将尿素或氯化铵溶液加入培养基中), 待后续培养使用. 培养时, 取 A₆₀₀ 为 1 的菌液 3 ml 进行离心并分别接种至 4 种培养基中进行培养. 发现除对照组 B 中细菌因缺乏营养物质生长受限外, 其他 3 组中的细菌均在 24 h 内呈指数型增长. 经多次生长测试实验发现, 在选取转录组采样时间点时, 无法保证 4 组细菌都处于同一生长阶段和最优的生长状态. 为准确比较不同培养条件下细菌的生长情况, 需选取相同的培养时间进行采样分析. 综合考虑 4 组培养条件下细

菌的生长规律, 选取 N 组和 U 组生长活性最好的 12 h 节点为 RNA 提取采样点.

1.3 RNA 提取和测序

使用 TRIzol 试剂从 4 组菌液样品中提取总 RNA, 按照制造商的规程进行. 使用无 RNA 酶琼脂糖凝胶电泳和 Agilent Bioanalyzer 2100 系统 (Agilent Technologies, Santa Clara, CA, USA) 对提取的总 RNA 质量和数量进行验证. 然后将 4 组 RNA 样品用于文库构建. 使用 oligo-dT 珠 (Qiagen, Hilden, Germany) 分离 mRNA, 并使用 NEB Next[®] Ultra[™] RNA 文库制备试剂盒 (NEB, USA) 制备 cDNA 文库. 使用 Agilent Bioanalyzer 2100 系统评估文库质量. 最后, 在 Illumina HiSeq[™] 4000 平台上进行了高通量测序, 读数长度为 150 bp.

1.4 差异性基因表达分析

采用无生物学重复的转录组差异基因分析方法, 用 DESeq R 软件包 (1.18.0) 对 4 组数据进行差异化表达分析^[19]. DESeq 提供了基于泊松分布的基因表达定量分析的统计学方法. 计算每个样本的归一化读数. 所得的 *P* 值通过 Benjamini 和 Hochberg 方法进行调整, 以控制错误发现率 (FDR). FDR 调整后的 *P* 值 (*q* value) 阈值为 0.05, 用于筛选不同表达的基因. 本研究中, 在不同组别进行差异基因分析时, 以 $|\log_2(\text{fold change})| > 1$ 且 $P < 0.05$ 表示为两组的差异基因^[20].

1.5 基因结构与操纵子预测分析

通过 Rockhopper 软件, 根据 reads 在基因组上的分布情况, 对操纵子、转录起始位点 (transcription start site, TSS) 和转录终止位点 (transcription termination site, TTS) 进行预测. 然后, 提取基因转录起始位点上游 700 bp 序列, 通过时延神经网络 (time-delay neural network, TDNN) 方法进行启动子预测. 对于操纵子的预测, 该方法具体是结合基因间距离和基因表达量相关性 2 个特征, 用朴素贝叶斯分类器模型来预测, 该预测算法的敏感性和特异性均高达 95%^[21].

1.6 GO 富集分析和 KEGG 富集分析

Gene Ontology (简称 GO) 是基因功能国际标准分类体系. 作为基因本体联合会 (Gene Ontology Consortium) 所建立的数据库, 它旨在建立一个适用于各物种的、对基因和蛋白质功能进行限定和描述的、并能随着研究不断深入而更新的语言词汇标准. 利用 GOrse-R 软件包实现了差异表达基因的基因本体 (GO) 富集分析, 修正了基因

长度偏差^[22]. 校正后的 P 值(q value)小于0.05的GO项是差异表达基因显著富集的.

KEGG是一个数据库资源, 用于从分子水平的信息(尤其是通过基因组测序和其他高通量生成的大规模分子数据集)中了解生物系统(例如细胞、生物体和生态系统)的高级功能和生物效应(<http://www.genome.jp/kegg/>). 我们使用KOBAS (2.0)软件来计算KEGG途径中差异表达基因的Pathway富集分析^[23]. 校正后的 P 值(q value)小于0.05的KEGG途径为差异表达基因显著富集的.

2 结果与分析

2.1 细菌生长测试

4种培养条件下的细菌生长测试如图1所示, 在对照组(B)条件下巴氏芽孢杆菌无法正常生长, 而在铵根组(N)和尿素组(U)条件下细菌呈指数增长, 铵根 Tris 组(N_T)在呈现一定的生长延迟后, 也呈现指数增长特征, 但稳定期时菌液浓度低于铵根组和尿素组. 分析4种培养条件下细菌生长模式差异的原因: 在细菌可正常生长的培养条件中, N组和U组的细菌生长速度最快, 但N组的最终菌液浓度最高, 表明巴氏芽孢杆菌对铵根和尿素的吸收效率较为接近, 但代谢机制可能存在差异, 使得细菌对铵根的利用程度更高; N_T组的细菌生长呈现出迟滞现象, 生长速度受到一定抑制, 并且最终菌液浓度较低, 说明培养基中Tris的存在不利于细菌对铵根的吸收和利用, 影响细菌的正常生长. 而对照组(B)条件下巴氏芽孢杆菌无法正常生长, 表明小分子氮源是细菌生长代谢所需的基本养分. 后续对4组培养条件下的细菌转录组测序分析进一步证实了上述判断.

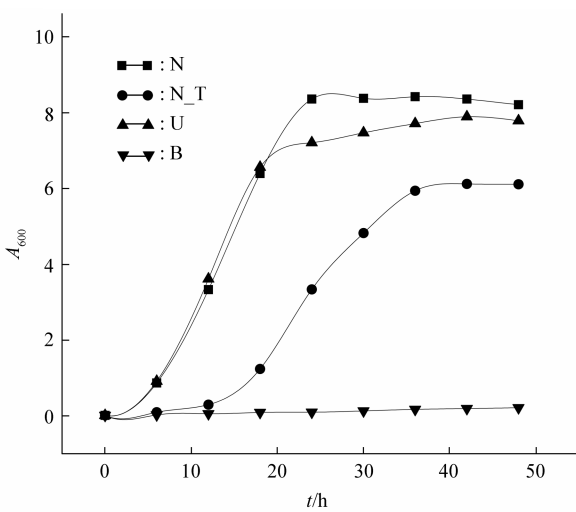


Fig. 1 Bacteria growth test under four conditions

2.2 测序质量评估与比对分析

检测结果如表1所示, 本研究中构建的4种培养条件下的巴氏芽孢杆菌样本微量文库质检合格, 满足转录组测序要求. 4个样本的Q30百分比为92.9%~93.48%, 说明测序质量和文库构建质量高. 4个样本的测序结果中碱基A-T和G-C的含量分布曲线基本对应重复, 表明碱基组成稳定且平衡. 通过对测序原始数据进行过滤, 包括去除带头的reads, 去除N(N表示无法确定碱基信息)的比例大于10%的reads, 去除低质量reads, 得到Clean reads, 后续的生物信息分析都是基于Clean reads. 通过计算4组样本测序结果的Clean reads/Raw reads数值, 可以发现高质量的测序数据占原始数据的比例均处于很高水平, 说明测序质量高, 数据可靠, 可进行后续数据分析.

Table 1 Quality assessment of the sequencing data

Sample name	Raw reads	Clean reads	Clean reads/ Raw reads	Error rate/%	Q30/%	GC content/%
N	11 398 640	11 171 416	98.01%	0.03	93.48	43.31
N_T	16 971 312	16 679 774	98.28%	0.03	93.19	41.64
U	12 988 828	12 850 322	98.93%	0.03	92.9	42.39
B	11 660 170	11 374 996	97.55%	0.03	92.9	42.26

Raw reads: statistics of raw sequence data; Clean reads: filtered sequencing data; Clean bases: number of Clean reads multiplied by length and converted into units of G; Error rate: error rate of base discrimination; Q30: percentage of bases correctly identified >99.9% of the total bases.

将测序所得数据与NCBI中巴氏芽孢杆菌的基因组数据(*Sporosarcina pasteurii*: NZ_CP038012.1)进行比对(表2), 4组样品比对到基

因组数据库的比对率在93.23%~99.18%之间, 均大于70%, 处于高水平. 在参考基因组中比对到多个位置的基因序列在0.59%~1.56%之间, 远低于

10%. 以上结果表明测序的样本不存在污染且基因组选择准确. 因此, 转录组测序产生的数据可以用

作后续分析与研究.

Table 2 The analysis of reads mapped to reference genome

Sample name	N	N_T	U	B
Total reads	11 171 416	16 679 774	12 850 322	11 374 996
Total mapped	11 079 974 (99.18%)	16 368 425 (98.13%)	12 450 579 (96.89%)	10 605 154 (93.23%)
Multiple mapped	66 133 (0.59%)	119 879 (0.72%)	118 758 (0.92%)	178 008 (1.56%)
Uniquely mapped	11 013 841 (98.59%)	16 248 546 (97.41%)	12 331 821 (95.97%)	10 427 146 (91.67%)

Total reads: clean reads; Total mapped: the number of sequenced sequences that can be localized to the reference genome; Multiple mapped: the number of sequences that match to multiple positions in the reference genome; Uniquely mapped: the number of sequences that match to a unique position in the reference genome.

2.3 差异表达基因分析

通过分析4种培养条件下巴氏芽孢杆菌转录组的基因表达变化差异, 得到了不同组别之间的基因差异表达结果. 根据基因组数据库和本次转录组测序所产生的数据, 在本研究中, 巴氏芽孢杆菌共有3 090个基因产生表达, 差异表达基因统计结果如图2所示. N v.s. B差异表达基因共1 152个, 其中

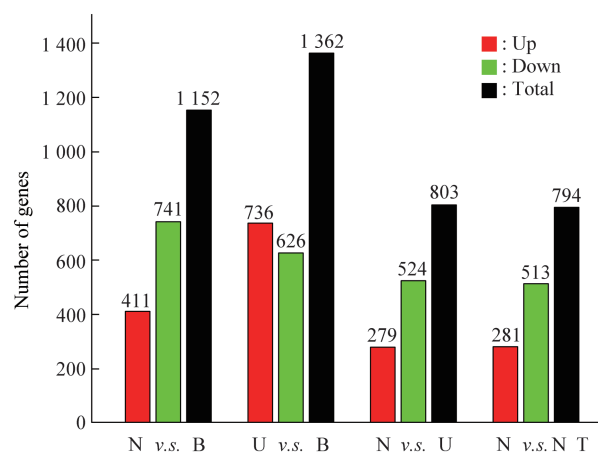


Fig. 2 Number of DEGs during bacteria under four conditions

上调的为411个, 下调的为741个; U v.s. B差异表达基因共1 362个, 其中上调的736个, 下调的626个; N v.s. U差异表达基因共803个, 其中上调的为279个, 下调的为524个; N v.s. N_T差异表达基因共794个, 其中上调的为281个, 下调的为513个.

2.3.1 差异基因GO富集分析

比较4种培养条件下的巴氏芽孢杆菌基因表达

谱, N组与B组比对的差异表达基因GO分析结果显示, 显著富集的生物过程 (biological process)、细胞组分 (cellular component) 和分子功能 (molecular function) 全部出现在下调的基因富集分析当中. 所有下调基因富集分析中最显著富集的30个条目如图3所示 (*表示该项为差异表达基因富集项). 其中最显著的包括细胞过程 (cellular process)、初级代谢过程 (primary metabolic process)、细胞氮化合物代谢 (cellular nitrogen compound metabolic process)、有机物生物合成 (organic substance biosynthetic process) 等, 可以看出N组与B组的主要差异均体现在细菌的基础生命活动及基础代谢方面, 这样的结果也符合生长测试的表现.

与N v.s. B相似, U组与B组比对的差异表达基因同样全部出现在下调的基因富集分析当中, 所有下调基因富集分析中最显著富集的30个条目如图4所示. 但U组与B组的全部差异表达基因富集当中也出现了部分富集项 (图5), 均包含于下调基因富集项中, 说明这些条目中下调基因的富集效应更加显著. 下调基因GO富集中最显著的包括细胞过程 (cellular process)、生物合成过程 (biosynthetic process)、细胞氮化合物的生物合成过程 (cellular nitrogen compound biosynthetic process)、有机氮化合物代谢过程 (organonitrogen compound metabolic process) 等, 可以看出U组与B组的主要差异同样体现在细菌的基础生命活动及基础代谢方面, 不同于N组与B组的是, U组细菌由于需要代谢尿素, 在GO分析中有机氮化合物的代谢过程体现了这一差异.

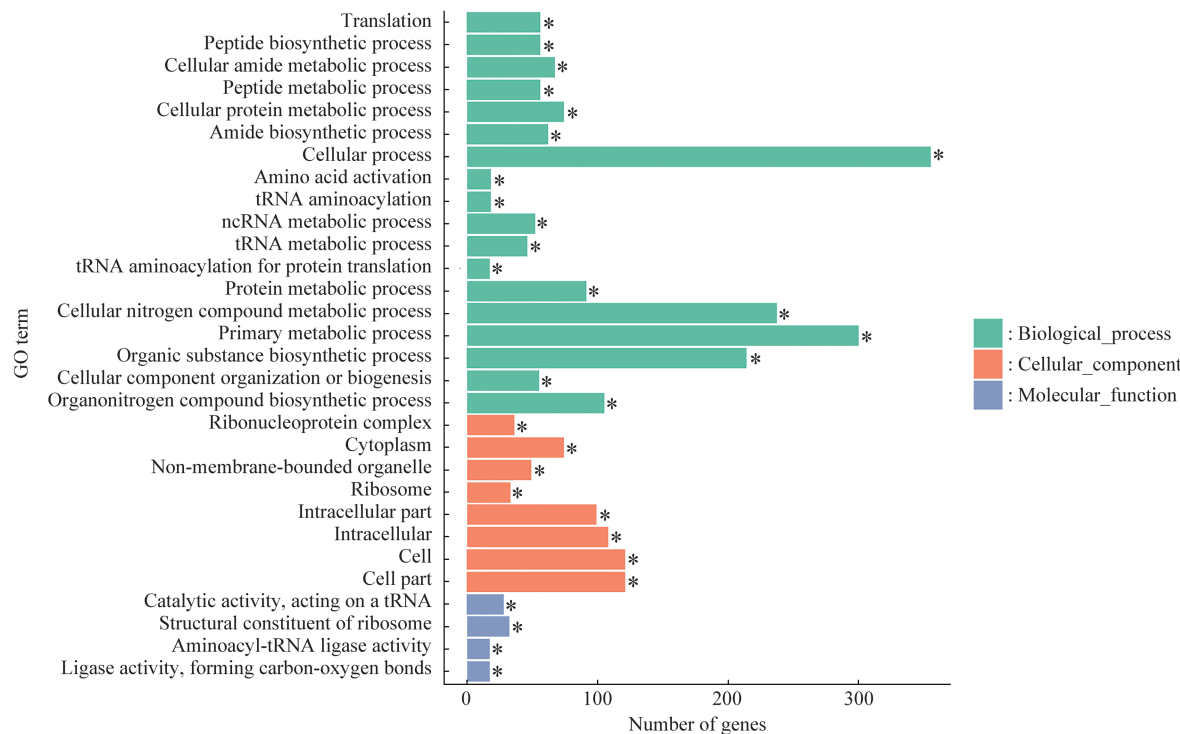


Fig. 3 The most enrich GO terms of N v.s. B down regulated

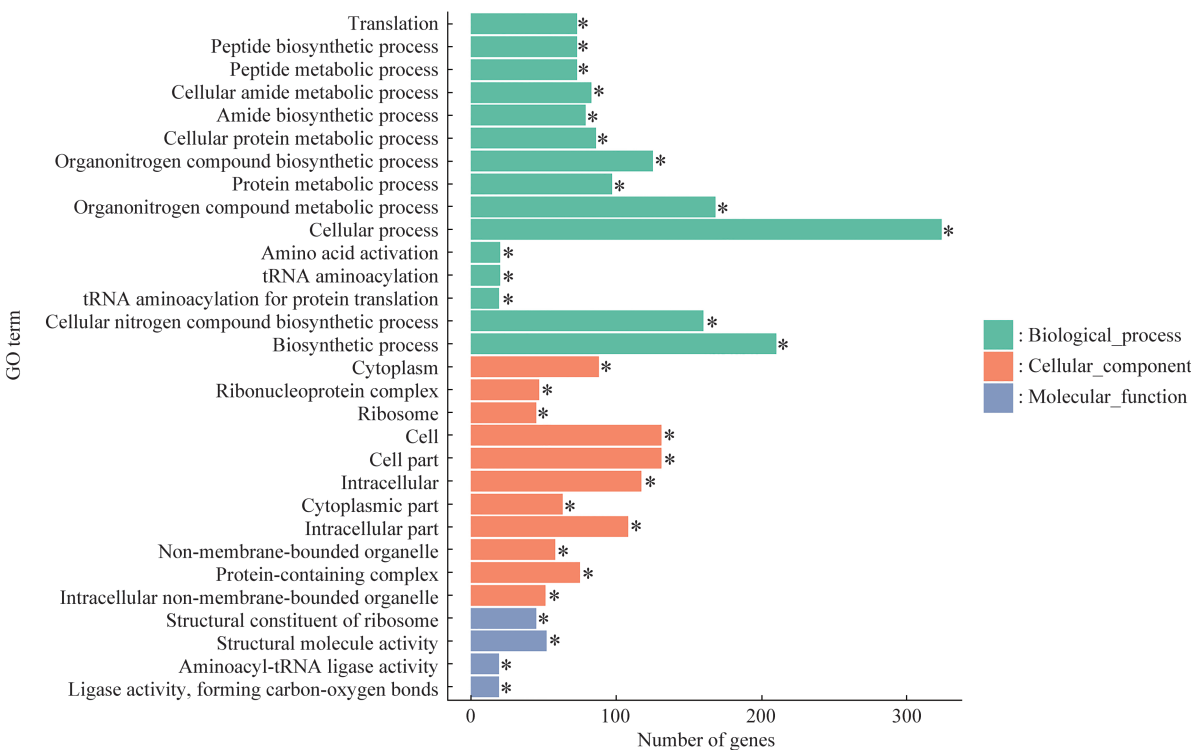


Fig. 4 The most enrich GO terms of U v.s. B down regulated

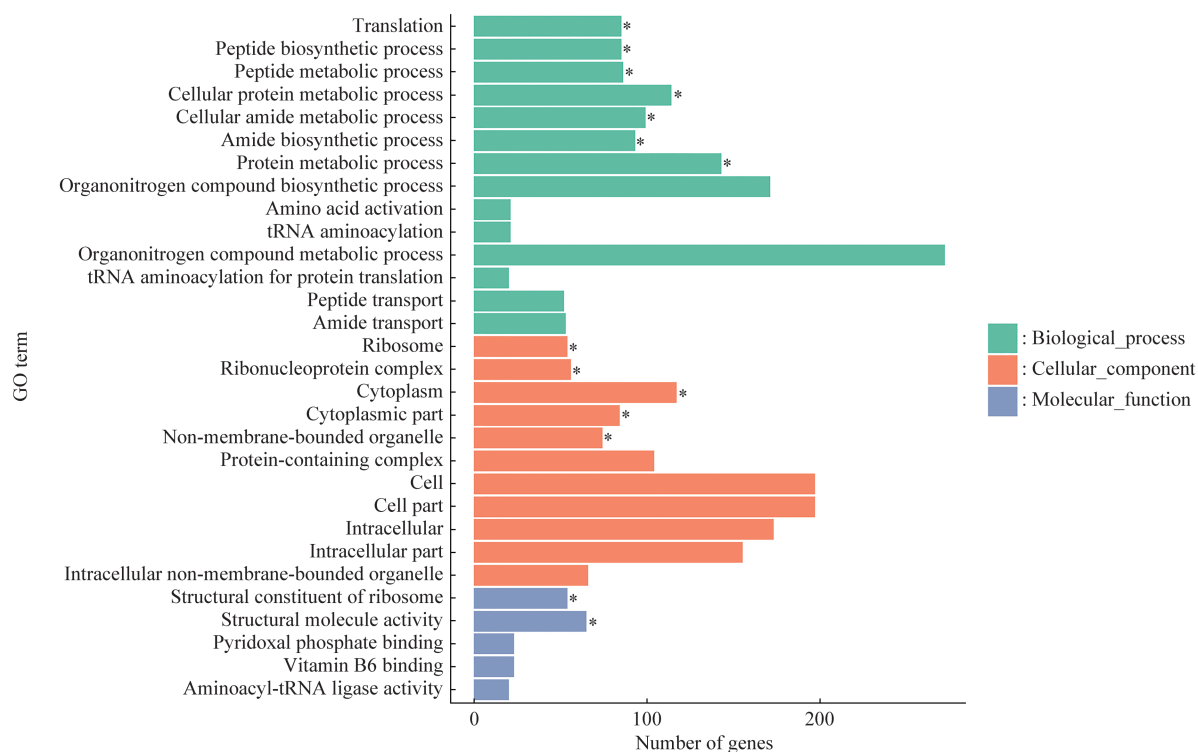


Fig. 5 The most enrich GO terms of U v.s. B

N组与U组比对的差异表达基因GO分析中，
仅在上调基因中出现了一个电子转移活性

(electron transfer activity) 富集项，所有上调基因
中最显著富集的30个条目如图6所示。从表象看，



Fig. 6 The most enrich GO terms of N v.s. U up regulated

N组与U组的差异主要在脲酶是否分解尿素. 尿素在细菌内分解, 我们推测产生的碳酸根全部被排出胞外, 而产生的铵根一部分被细菌利用, 另一部分则同样需要排出胞外. 这样便产生了电势差, 电子转移活性的差异可能体现在细菌需要平衡这部分电势差.

N组与N_T组比对的差异表达基因GO分析

中, 富集项全部出现在N组比N_T组下调的基因富集中, 所有下调基因中最显著富集的30个条目如图7所示. 其中最显著的包括翻译 (translation)、肽生物合成过程 (peptide biosynthetic process)、氨基酸活化 (amino acid activation) 等方面, 可以发现两组差异主要体现在蛋白质表达过程的各阶段, 后续分析也可对此进行印证.

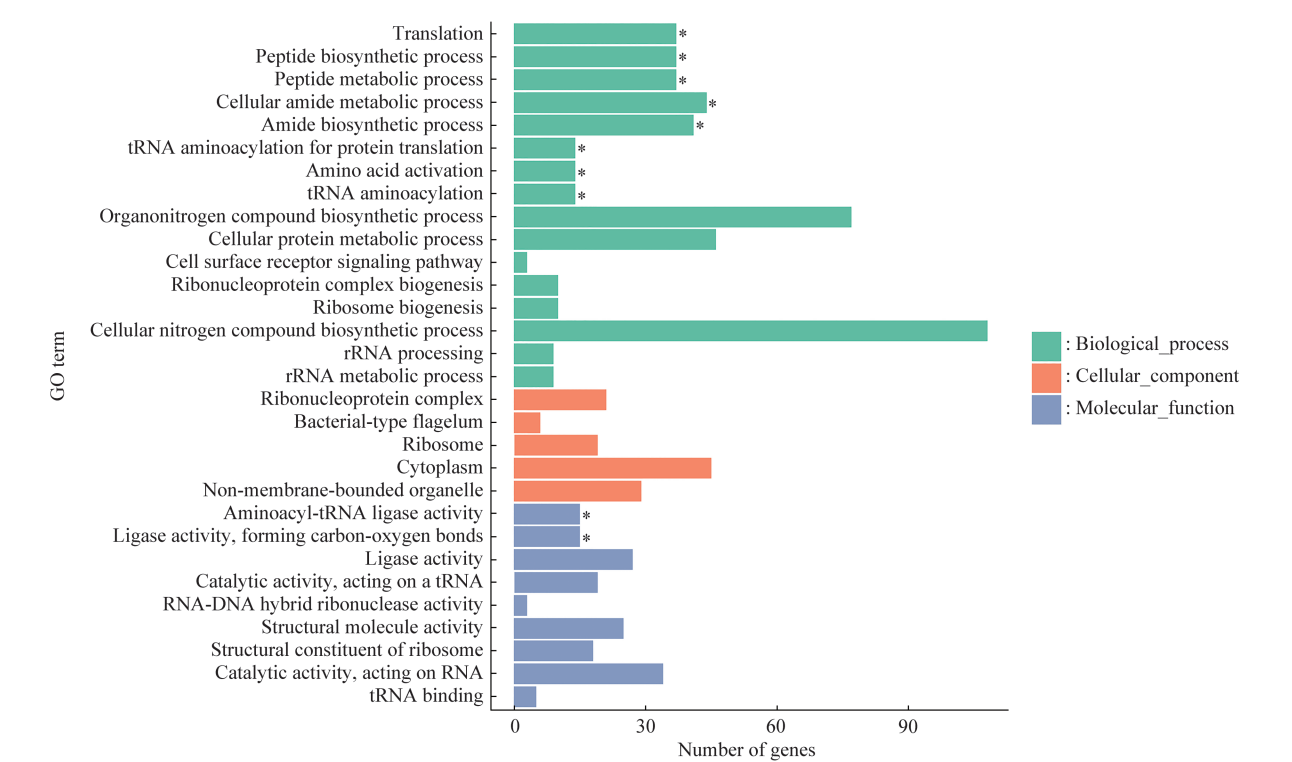


Fig. 7 The most enrich GO terms of N v.s. N_T down regulated

2.3.2 差异基因KEGG富集分析

通过对差异基因进行KEGG分析发现, N组与B组比较时, KEGG富集项中最显著的20个条目都来自下调的基因富集当中 (图8). 与GO分析相似的情况再次出现, 在KEGG分析中, U组与B组比较时, 富集项中最显著的20个条目同样都来自下调的基因富集当中. 其中, 呈统计学显著性富集的两项, 分别为核糖体 (ribosome) 和鞭毛组装 (flagellar assembly) (图9). 对比图8可以发现, N组与B组比较中富集程度最高的两项同样为核糖体和鞭毛组装. 这种差异表明, B组在维系基本生命活动中处于挣扎状态, 一方面增强鞭毛表达以增强运动能力, 从而寻找营养物质以摆脱营养匮乏, 另

一方面增强核糖体的活动以合成基础生命活动需要的蛋白质组分.

N组与U组比较时, 在KEGG分析当中仅在上调基因富集当中出现了一个氧化磷酸化 (oxidative phosphorylation) 富集项, 所有上调基因中最显著富集的30个条目如图10所示. 这项差异表明N组比U组氧化磷酸化活动显著增强, 该代谢过程承担了细菌ATP合成的主要任务, 表明N组具备更强的能量产生能力.

2.4 基因结构与操纵子预测

在原核细菌的基因组中, 功能上相关的几个基因往往串联排列在一起, 构成操纵子结构作为基因的表达单元, 脲酶的功能蛋白及辅助蛋白基因就是

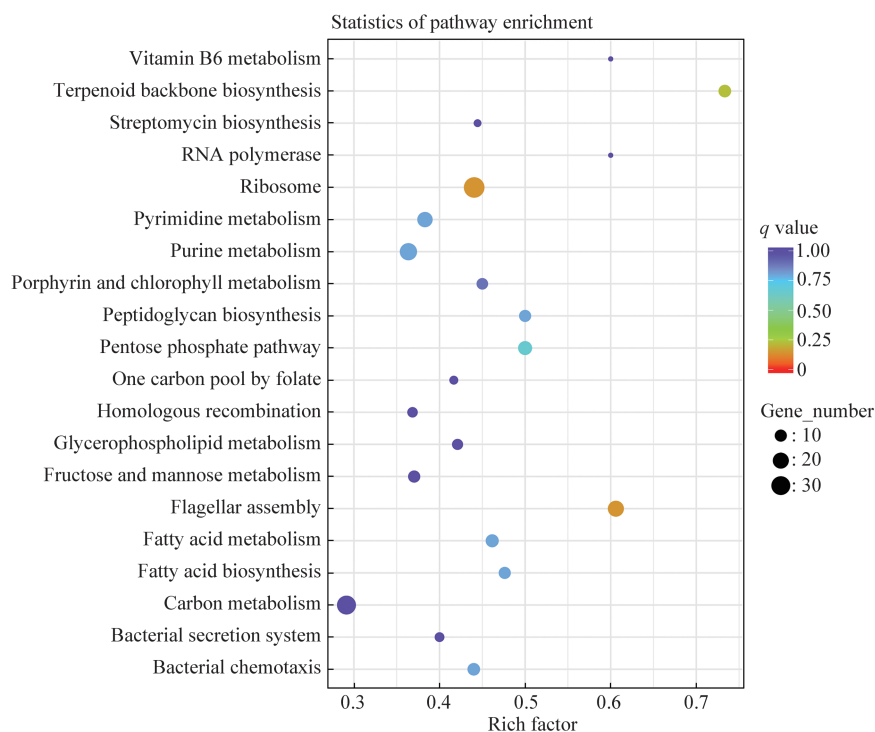


Fig. 8 The most enrich KEGG terms of N v.s. B down regulated

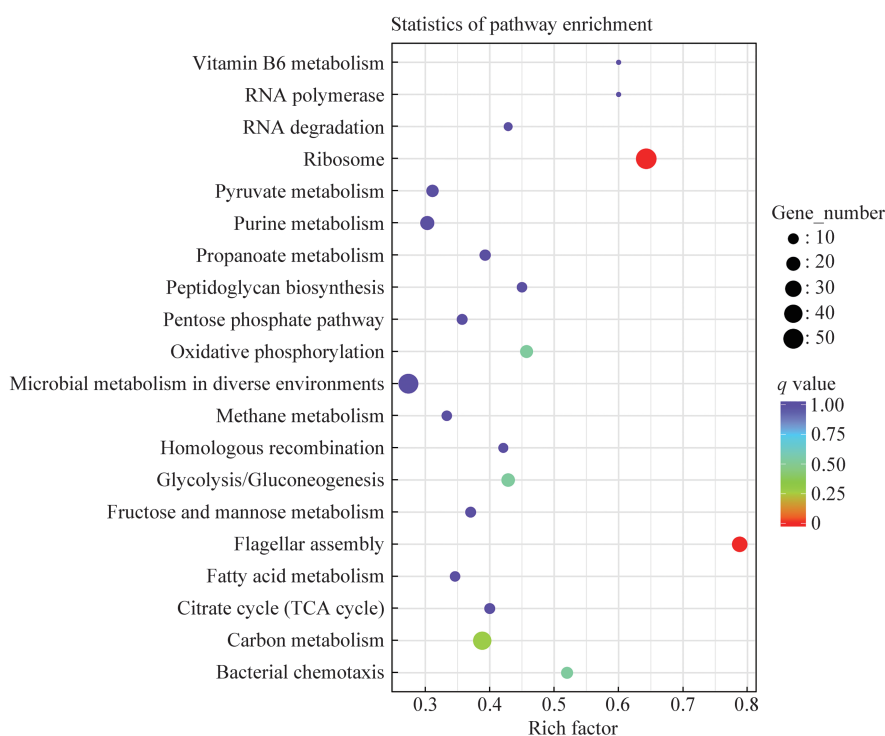


Fig. 9 The most enrich KEGG terms of U v.s. B down regulated

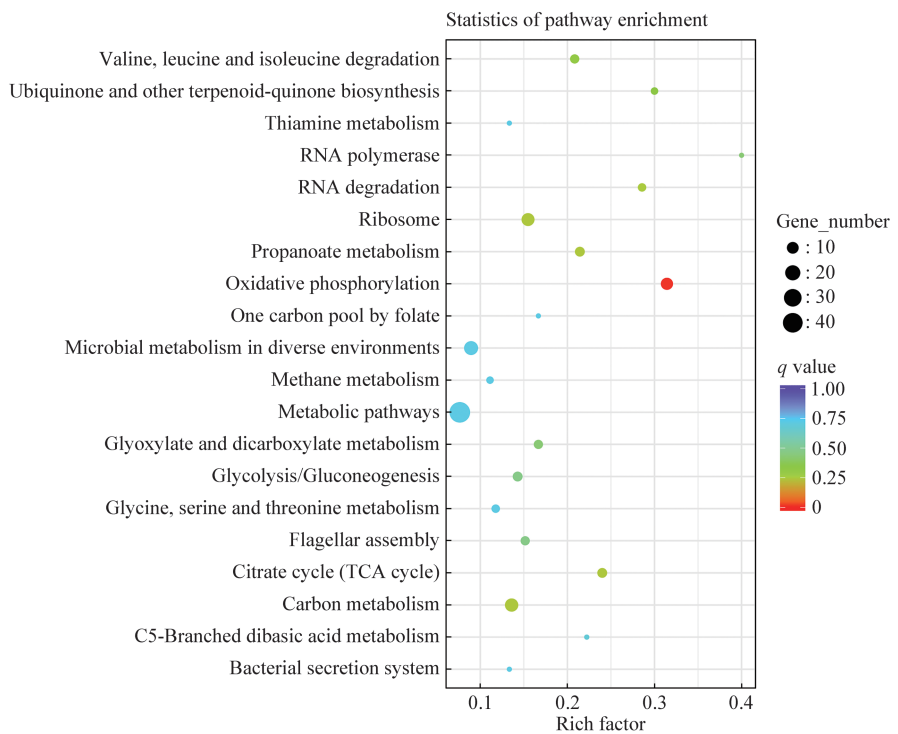


Fig. 10 The most enrich KEGG terms of N v.s. U up regulated

以操纵子结构进行表达和调控的. 我们主要关注脲酶的基因结构, 表3对脲酶操纵子中各基因的信息进行了统计, 包括基因在NCBI数据库中的ID号、基因名称、长度、产物及在基因组中的位置, 另外本文还统计了脲酶各基因之间的核酸间距. 表4对脲酶的操纵子、转录起始位点、转录终止位点等信息进行了预测. 结果显示, *ureA*、*B*、*C*、*E*、*F*为一个操纵子(起始位点为基因组中第636 643 核酸位点, 终止位置为第640 294 核酸位点), *ureG*、*D*为另一个操纵子(起始位点为第640 505 核酸位点, 终止位置为第641 942 核酸位点).

Table 3 Urease genes information

Gene ID	Gene name	Gene length	Product	Gene location	Distance from the previous gene
E2C16_RS03190	<i>ureA</i>	303 bp	subunit gamma	636 643–636 945	null
E2C16_RS03195	<i>ureB</i>	363 bp	subunit beta	637 015–637 377	69 bp
E2C16_RS03200	<i>ureC</i>	1 713 bp	subunit alpha	637 381–639 093	3 bp
E2C16_RS03205	<i>ureE</i>	444 bp	accessory protein <i>ureE</i>	639 169–639 612	75 bp
E2C16_RS03210	<i>ureF</i>	663 bp	accessory protein <i>ureF</i>	639 632–640 294	19 bp
E2C16_RS03215	<i>ureG</i>	636 bp	accessory protein <i>ureG</i>	640 505–641 140	210 bp
E2C16_RS03220	<i>ureD</i>	828 bp	accessory protein <i>ureD</i>	641 115–641 942	–26 bp

Table 4 The prediction of urease operon

Start	Stop	Strand	Number of genes	Genes
636 643	640 294	+	5	E2C16_RS03190, E2C16_RS03195, E2C16_RS03200, E2C16_RS03205, E2C16_RS03210
640 505	641 942	+	2	E2C16_RS03215, E2C16_RS03220

3 讨 论

3.1 脲酶在不同培养条件下的表达差异

由于原核生物只有在有参考基因组的条件下才能利用转录组测序技术进行生物信息分析, 这导致巴氏芽孢杆菌的转录组分析长久以来未能实现. 2019年东南大学Ma等^[17]对巴氏芽孢杆菌进行了基因组测序, 进而开展了相应的转录组数据挖掘与分析. 研究结果表明, 培养基中较高浓度的氨水(由20 g/L的尿素完成水解所产生)抑制了脲酶的合成. 结合本研究结果可见, 虽然高浓度的尿素会使巴氏芽孢杆菌生长速度加快, 但其水解产生的高浓度氨水作为副产物会使巴氏芽孢杆菌脲酶合成受到抑制. 本研究使用的铵根及尿素浓度远低于上述研究. 在本研究中, 与无尿素或铵根的对照组(B)相比, 有尿素或铵根的U组或N组在脲酶基因的表达方面也明显低于B组(图11). 与Ma等的研究结果一致, 在无尿素或铵根的条件下(B组), 巴氏芽孢杆菌无法正常生长. 因此, 虽然B组脲酶表

达量高, 但很可能是由于缺乏氮源而增强了脲酶的表达, 以寻求环境中的尿素作为自身的氮源加以利用. 关于这一点, 在N v.s. B和U v.s. B的鞭毛基因表达中也有一定的证据可寻(图8, 9). KEGG分析结果表明, 与N或U组相比, 在B组中关于鞭毛组装基因的表达呈现显著性增强. 鞭毛的存在使得细菌具有运动能力, 趋向营养物质或逃离有害物质^[24]. 那么鞭毛基因的表达增强, 本研究推测是由于巴氏芽孢杆菌在B组条件下缺乏氮源这一必要营养物质所致. 由GO分析结果可以发现, 在B组中, 有多项关于代谢与合成含氮物质的途径均呈显著性基因表达增强, 尤其是代谢和消化含氮物质的途径, 脲酶各基因也包含于此. 这进一步说明, 在B组条件下, 巴氏芽孢杆菌处于严重缺乏氮源的环境, 因此无法正常生长. 综上所述, 虽然B组条件下, 脲酶表达增强, 但对于巴氏芽孢杆菌的矿化应用或许不具有参考意义. 不过, 通过对比N组与U组可以发现, 在含氮量相同的情况下脲酶在铵根组的表达量更高(图12), 该结果可能与巴氏芽孢杆菌脲酶催化尿素分解及其他相关代谢活动相关.

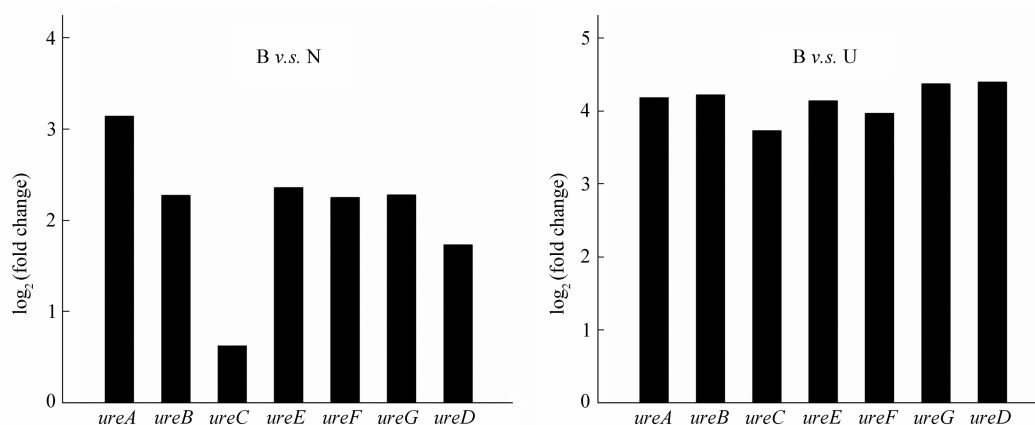


Fig. 11 The differentially expressed urease genes between the group B v.s. N, group B v.s. U

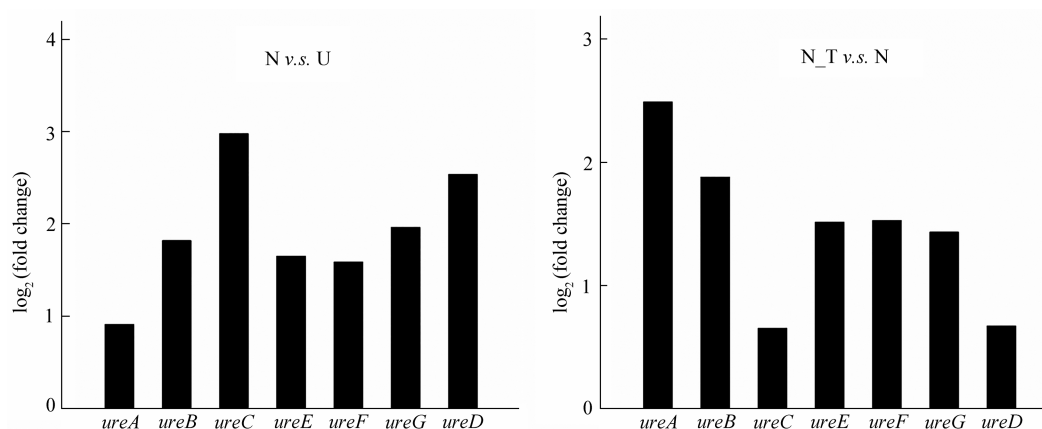


Fig. 12 The differentially expressed urease genes between the group N v.s. U, group N_T v.s. N

继续对比N组与N_T组的脲酶表达情况可以发现, N_T组的脲酶活性较N组有大幅度增强. 可以发现, 在本研究中4种培养条件下脲酶活性由强到弱的排列顺序为: B组>N_T组>N组>U组. 这一现象表明, 脲酶的表达不仅仅是为了获取氮源, 其表达调控应该是多因素共同调节的结果^[15-16]. 虽然起始pH相同, 但N_T组较N组和U组具有更稳定的pH值, 而N组随着细菌代谢废物的累计pH值会逐渐降低, U组随着尿素催化分解pH值会逐渐升高. 从表象来看, 稳定的pH值更有利于脲酶的表达, 而过高的pH可能会抑制脲酶的表达, 可以推测脲酶的表达对于细菌调节环境pH以达到自身所需环境条件具有重要作用. 进而, 根据上述脲酶活性与相应培养条件下的细菌生长情况的差异性, 可以发现脲酶活性与细菌对氮源营养环境的耐受力直接相关.

3.2 脲酶表达变化与能量代谢之间的关系

对比N组和U组可以发现, KEGG分析中两组

的差异表达基因在氧化磷酸化代谢通路中呈现显著性富集(图10). 根据之前我们的推断, 尿素或铵根是巴氏芽孢杆菌的氮源, 那么在这两种不同氮源的情况下, 两组的主要差异在于, 细菌表达的脲酶是否需要分解尿素, 即是否存在尿素分解反应. 通过比较这两组的差异, 我们可以发现脲酶水解尿素反应对巴氏芽孢杆菌造成的生理差异, 进而可以发现该反应对于巴氏芽孢杆菌的生理意义. KEGG分析结果表明, 氧化磷酸化这一代谢通路呈显著性富集. 对于巴氏芽孢杆菌而言, 氧化磷酸化是物质在其细胞质内氧化时释放的能量通过电子传递链传递供给ADP与无机磷酸合成ATP的偶联反应, 生物体内95%的ATP来自这种方式^[25], 可以说是生物体最重要的ATP生产方式. 数据结果显示, ATP合成酶和脲酶在N组的表达量更高(图13), 而N组与N_T组之间不存在显著差异, 并且3组条件下细菌都能够正常生长.

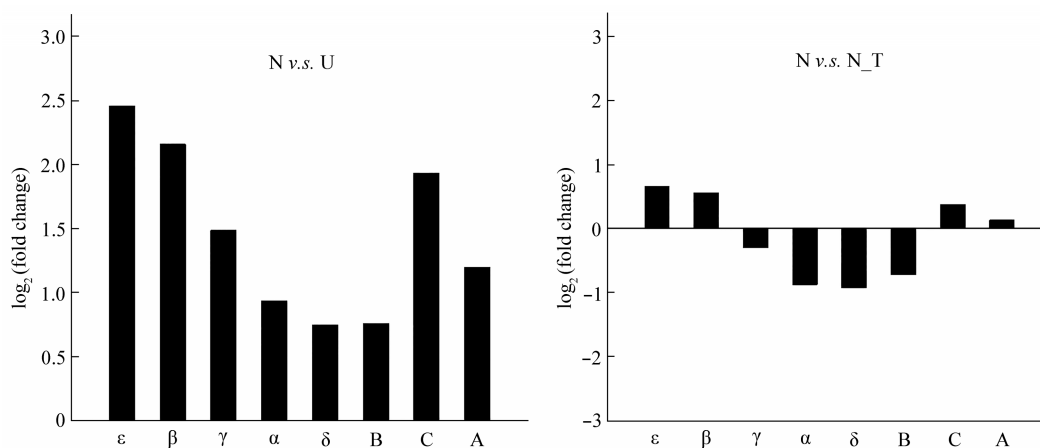


Fig. 13 The differentially expressed ATP synthase genes between the group N v.s. U, N v.s. N_T

α, β, γ, δ, ε, A, B and C represents the different subunits of F₀F₁ ATP synthase.

另外, GO分析结果表明, N组较U组电子转移活性要显著增强(图6). 根据以上结果可知, 在N组生长的巴氏芽孢杆菌, 细菌内ATP生成量要高于尿素组. Jahns^[26]的研究表明, 巴氏芽孢杆菌存在着与尿素分解相关联的ATP生成机制, 铵根参与了细菌内部pH的调节, 并导致细菌质子电动势(Dp)的增加, 从而刺激ATP的合成. 结合本实验结果, 在氮含量相同的情况下, 我们发现铵根的加入相对于尿素可以更直接更高效地刺激ATP的合成. 这一结果可能是因为铵根是刺激巴氏芽孢

杆菌ATP合成的直接因素. 但当环境中存在尿素时, 细菌特有的高效脲酶活性可以将尿素转化为其代谢所需要的铵根离子, 从而完成ATP的合成及其他代谢活动. 本文的研究结果表明, 在巴氏芽孢杆菌的能量代谢方面铵根比尿素更具优势.

3.3 脲酶的基因结构预测

操纵子是细菌中单个调控系统可以控制多个基因的一种手段, 操纵子结构的准确注释对基因功能和基因调控网络的研究具有重要意义. 研究表明, 脲酶是一种高分子质量的多聚体含镍酶, 其基因以

操纵子的结构形式存在于绝大多数物种的染色体DNA上^[27-28]。虽然不同物种之间的脲酶操纵子及基因结构不尽相同,但其核心基因是高度保守的。图14总结了几种物种的脲酶基因结构,可见包括巴氏芽孢杆菌在内的大多数物种脲酶都包含了*ureA*、*B*、*C*、*D*、*E*、*F*、*G*等7个及7个以上的基因。图中巴氏芽孢杆菌的脲酶结构为单个操纵子且包含共7个脲酶基因,即*ureA*、*B*、*C*、*E*、*F*、*G*、*D*,该操纵子结构是基于已知的克雷伯氏菌脲酶基因结构进行基因比对而来,关于脲酶基因结构的直接研究结果未见报道。

本文通过转录组测序的方法对脲酶基因结构进行了分析和预测,预测结果表明,巴氏芽孢杆菌脲酶基因结构为双操纵子结构(表4)。操纵子1包含编码脲酶功能蛋白 α 、 β 、 γ 亚基的*ureA*、*B*、*C*基因,编码辅助蛋白的*ureE*、*F*基因;操纵子2包含了编码辅助蛋白的*ureG*、*D*基因。当前,通过生物信息学方法计算是基因组操纵子结构注释的最主要来源,这些计算方法的参数依据包括基因间距离、使用密码子的相似性、种间保守性和功能注释的相似性等。在这些非实验性的特征中,同一链上连续基因之间的核酸间隔距离具有最强的预测能力^[29-31]。本文中操纵子预测是结合两个特征来估计同一条链上的连续基因作为同一操纵子被共转录的可能性,这两个特征为基因间的核酸距离和

RNA-seq数据中基因表达的相似性。基于转录组数据的分析仍是目前对于基因结构及操纵子预测最有效的方式之一,本文根据转录组数据中基因表达的相似性得出以上预测结果,该方法由Ryan等^[32]提出并经其验证敏感性范围为89%~96%,特异性范围为86%~95%。正如上文所述,依据基因间核酸距离判断操纵子结构已具备很强的预测能力。通常如果属于同一个操纵子,那么它们通常有较短的基因间距,一般操纵子内部相邻基因间距要小于不同操纵子的基因间距。北京大学吴文琪等^[33]通过建立操纵子数据库并以迭代学习的方式预测操纵子结构,该数据库以基因间核算距离为主要特征进行统计学分析,可以发现,在统计的762个基因组中,操纵子之间的基因间距平均值为82 bp,标准差为19 bp,且全部数据中基因间距都不超过180 bp。虽然以上研究的数据库并不能代表全部物种的全部基因,但其囊括的范围较为充分,一定程度上辅证了本研究中脲酶基因结构预测结果。本研究中,*ureA*、*B*、*C*、*E*、*F*基因之间的核酸距离均小于80 bp,*ureG*、*D*间更是有26 bp的核酸重合(属于正常现象),但*ureF*和*ureG*之间的距离高达210 bp,结合脲酶各基因的表达情况分析,本研究认为*ureA*、*B*、*C*、*E*、*F*和*ureG*、*D*分别属于两个操纵子,即脲酶基因基因为双操纵子结构。

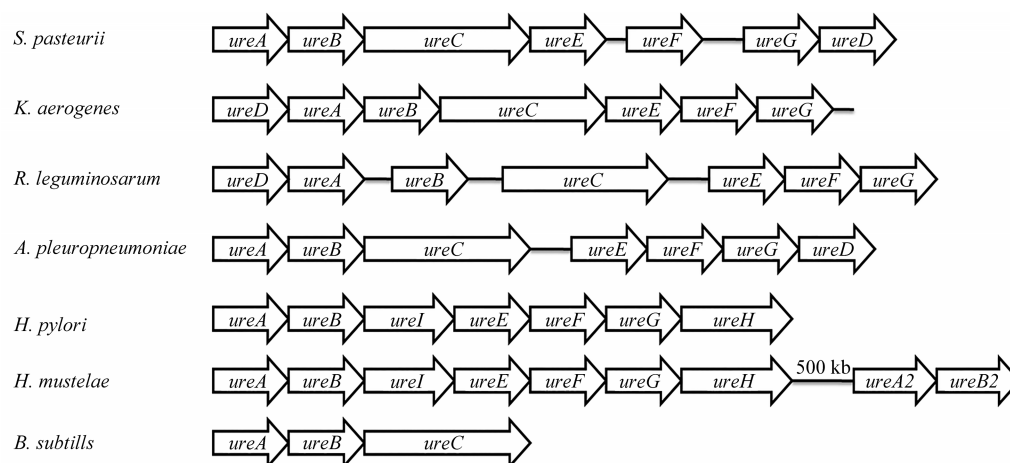


Fig. 14 Comparison of urease gene cluster of different microorganisms

本文通过高通量测序技术对巴氏芽孢杆菌在4种培养条件下的生长情况和脲酶表达差异情况进行了分析,揭示了脲酶在不同培养条件下的表达变化规律及与脲酶表达变化相关联的能量代谢机制,并

预测了脲酶基因的双操纵子结构,相关研究丰富了巴氏芽孢杆菌的转录组信息,为通过操控脲酶活性调控巴氏芽孢杆菌生长及其诱导的矿化作用奠定了理论基础。

参 考 文 献

- [1] Anbu P, Kang C H, Shin Y J, *et al.* Formations of calcium carbonate minerals by bacteria and its multiple applications. SpringerPlus, 2016, **5**:250
- [2] Tingting Z, Maria D. Carbonate precipitation through microbial activities in natural environment, and their potential in biotechnology: a review. Front Bioeng Biotechnol, 2016, **4**: 4
- [3] Ferris F G, Stehmeier L G, Kantzas A, *et al.* Bacteriogenic mineral plugging. Can Petr Technol, 1996, **35**(8): 56-61
- [4] Ramachandran S K, Ramakrishnan V, Bang S S. Remediation of concrete using microorganisms. ACI Materials Journal, 2001, **98**(1): 3-9
- [5] 王瑞兴, 钱春香, 王剑云, 等. 水泥石表面微生物沉积碳酸钙覆膜的不同工艺. 硅酸盐学报, 2008, **36**(10): 1378-1384
Wang R X, Qian C X, Wang J Y, *et al.* Journal of the Chinese Ceramic Society, 2008, **36**(10): 1378-1384
- [6] 王瑞兴, 钱春香. 微生物沉积碳酸钙修复水泥基材料表面缺陷. 硅酸盐学报, 2008, **36**(4): 37-44
Wang R X, Qian C X. Journal of The Chinese Ceramic Society, 2008, **36**(4): 37-44
- [7] Jonkers H M. Self healing concrete: a biological approach//Zwang S. Self Healing Materials. Dordrecht: Springer, 2007, **100**: 195-204
- [8] Jonkers H M, Thijssen A, Muyzer G, *et al.* Application of bacteria as self-healing agent for the development of sustainable concrete. Ecological Engineering, 2010, **36**(2): 230-235
- [9] Ednie-Brown P, Burry M, Burrow A. Biomason and the speculative engagements of biotechnical architecture. Architectural Design, 2013, **83**(1): 84-91
- [10] Larson M. Professor uses bacteria to make ecofriendly bricks. Architectural Record, 2010, **198**(8): 29
- [11] Randall D G, Naidoo V. Urine: the liquid gold of wastewater. Journal of Environmental Chemical Engineering, 2018, **6**(2): 2627-2635
- [12] 裴迪, 刘志明, 胡碧茹, 等. 巴氏芽孢杆菌矿化作用机理及应用研究进展. 生物化学与生物物理进展, 2020, **47**(6): 467-482
Pei D, Liu Z M, Hu B R, *et al.* Progress in Biochemistry and Biophysics, 2020, **47**(6): 467-482
- [13] Stocks-Fischer S, Galinat J K, Bang S S. Microbiological precipitation of CaCO₃. Soil Biology and Biochemistry, 1999, **31**(11): 1563-1571
- [14] Onal O T, Frigi R D. Optimized carbonate micro-particle production by *Sporosarcina pasteurii* using response surface methodology. Ecological Engineering, 2014, **62**: 168-174
- [15] Whiffin V S. Microbial CaCO₃ precipitation for the production of biocement[D]. Australia: Murdoch University, 2004
- [16] Huang S C, Burne R A, Chen Y Y M, *et al.* The pH-dependent expression of the urease operon in *Streptococcus salivarius* is mediated by CodY. Appl Environ Microbiol, 2014, **80**(17): 5386-5393
- [17] Ma L, Pang A P, Luo Y S, *et al.* Beneficial factors for biomineralization by ureolytic bacterium *Sporosarcina pasteurii*. Microbial Cell Factories, 2020, **19**(1):12
- [18] Kane J, Fischer J B. The differentiation of Trichophyton rubrum and T. mentagrophytes by use of Christensen's urea broth. Canadian Journal of Microbiology, 1971, **17**(7): 911-913
- [19] Love M I, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 2014, **15**: 550
- [20] Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biology, 2010, **11**(10): R106
- [21] Liu T, Wei W Y, Liu J X, *et al.* Chain specificity transcriptome analysis in different temperatures conditioned of *Yersinia ruckeri*. Acta Hydrobiologica Sinica, 2019, **43**(5): 969-976
刘韬, 魏文燕, 刘家星, 等. 水生生物学报, 2019, **43**(5): 969-976
- [22] Young M D, Wakefield M J, Smyth G K, *et al.* Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology, 2010, **11**(2): R14
- [23] Mao X, Cai T, Olyarchuk J G, *et al.* Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics, 2005, **21**(19): 3787-3793
- [24] Haiko J, Westerlund-Wikstri B. The role of the bacterial flagellum in adhesion and virulence. Biology, 2013, **2**(4): 1242-1267
- [25] Senior A E. ATP synthesis by oxidative phosphorylation. Physiological Reviews, 1988, **68**(1): 177-231
- [26] Jahns T. Ammonium/urea-dependent generation of a proton electrochemical potential and synthesis of ATP in *Bacillus pasteurii*. Journal of Bacteriology, 1996, **178**(2): 403-409
- [27] Kim S D, Spizizen J. Molecular cloning and expression of *Bacillus pasteurii* urease gene in *Escherichia coli*. Kor J Appl Microbiol Bioeng, 1985, **13**(3): 297-302
- [28] Gerlach G F, Clegg S, Nichols W A. Characterization of the genes encoding urease activity of *Klebsiella pneumoniae*. FEMS Microbiology Letters, 1988, **50**(2-3): 131-135
- [29] Westover B P, Buhler J D, Sonnenburg J L, *et al.* Operon prediction without a training set. Bioinformatics, 2005, **21**(7): 880-888
- [30] Joseph B, Mark C, David P, *et al.* A Bayesian network approach to operon prediction. Bioinformatics, 2003, **19**(10): 1227-1235
- [31] Dam P, Olman V, Harris K, *et al.* Operon prediction using both genome-specific and general genomic information. Nucleic Acids Research, 2006, **35**(1): 288-298
- [32] Ryan M, Divya B, Yan S, *et al.* Computational analysis of bacterial RNA-Seq data. Nucleic Acids Research, 2013, **41**(14): e140
- [33] 吴文琪, 郑晓斌, 刘永初, 等. 基于迭代自学习的操纵子结构预测. 生物化学与生物物理进展, 2012, **38**(7): 642-651
Wu W Q, Zheng X B, Liu Y C, *et al.* Progress in Biochemistry and Biophysics, 2012, **38**(7): 642-651

Transcriptome Analyses Reveal The Urease Function of *Sporosarcina pasteurii* Based on Different Nitrogen Source Culture Conditions

PEI Di¹⁾, LIU Zhi-Ming^{1,2)*}, HU Bi-Ru¹⁾, WU Wen-Jian¹⁾

¹⁾College of Science, National University of Defense Technology, Changsha 410073, China;

²⁾Institute of Chemical Defense, Academy of Military Science, Beijing 102205, China)

Abstract *Sporosarcina pasteurii* (*S. pasteurii*) is a kind of Gram-positive bacteria from soil. Various applications have been developed based on the efficient urease activity that can induce the precipitation of calcium carbonate. However, the metabolic mechanism related to biomineralization of *S. pasteurii* has not been clearly elucidated. Especially, there are few studies on the gene structure of urease, regulation mechanism of expression and associated metabolism, which play key roles in biomineralization. Nowadays, the uncontrollability and instability of biomineralization reactions in the applications of *S. pasteurii* root in the lack of research on urease metabolisms. Therefore, it is urgent to further reveal the gene information, expression regulation and related metabolism of urease in *S. pasteurii*. In this paper, we compared the growth and gene expression of *S. pasteurii* BNCC 337394 under four different culture conditions through high-throughput transcriptome analyses. The four medium conditions were: (1) control group with yeast extract of 20 g/L; (2) ammonium group with yeast extract of 20 g/L and ammonium chloride of 10 g/L; (3) urea group with yeast extract of 20 g/L and urea of 5.62 g/L; (4) ammonium Tris group with yeast extract of 20 g/L, ammonium chloride of 10 g/L and Tris of 15.75 g/L. Transcriptome data were analyzed by bioinformatics methods of differential gene expression analysis, GO enrichment analysis, KEGG enrichment analysis and operon prediction. The results showed that there were significant differences in the growth of *S. pasteurii* among the four conditions. The bacteria could not grow and reproduce normally in the control group. There was a growth delay in the ammonium Tris group. The growth rate and trend of the ammonium group and the urea group with the same nitrogen content were similar. A total of 3 090 genes were generated and expressed in *S. pasteurii* in the experiment. There were 1 152 differentially expressed genes (DEGs) between the ammonium group and the control group, of which 411 were up-regulated and 741 were down-regulated. There were 1 362 DEGs between the urea group and the control group, of which 736 were up-regulated and 626 were down-regulated. There were 803 DEGs between the ammonium group and the urea group, of which 279 were up-regulated and 524 were down-regulated. There were 794 DEGs between the ammonium group and the ammonium Tris group, of which 281 were up-regulated and 513 were down-regulated. Furthermore, it was found that the expression of urease was significantly enhanced in the control group which was short of a nitrogen source comparing with the ammonium group and the urea group. GO and KEGG analyses revealed that the control group needed to enhance its basal metabolism and express more flagellum in order to survive. The pathways of electron transfer activity and oxidative phosphorylation were different in the ammonium group and the urea group. It indicated that the synthesis of ATP was associated with the hydrolysis of urea in *S. pasteurii*. Meanwhile, ammonium stimulated the expression of urease and ATP synthase in the ammonium group which was more significant than that of the urea group. The results of the ammonium Tris group indicated that a stable and moderate alkaline pH environment favored a high level expression of urease. Finally, the double operon gene structure of urease was predicted based on the two key characteristics of gene expression similarity and gene spacing.

Key words *Sporosarcina pasteurii*, transcriptome, operon, urease, ATP

DOI: 10.16476/j.pibb.2020.0441

* Corresponding author.

Tel: 86-731-87001815, E-mail: lzm@nudt.edu.cn

Received: December 17, 2020 Accepted: March 11, 2021