



谷氨酸棒杆菌不同生长时期非编码 sRNA 的预测与鉴定

王弋, 范志豪, 饶志明, 徐美娟*

江南大学生物工程学院 工业生物技术教育部重点实验室, 江苏 无锡 214122

王弋, 范志豪, 饶志明, 徐美娟. 谷氨酸棒杆菌不同生长时期非编码 sRNA 的预测与鉴定[J]. 微生物学报, 2023, 63(10): 3876-3890.

WANG Yi, FAN Zhihao, RAO Zhiming, XU Meijuan. Prediction and identification of non-coding sRNAs in *Corynebacterium glutamicum* at different growth stages[J]. Acta Microbiologica Sinica, 2023, 63(10): 3876-3890.

摘要: 【目的】探究与挖掘不同生长时期谷氨酸棒杆菌(*Corynebacterium glutamicum*)中潜在的小的非编码 RNA (small non-coding RNA, sRNA)。【方法】检测不同时期和不同工业改造的谷氨酸棒杆菌菌株的 RNA-Seq 数据构建表达文库, 通过 sRNA-Detect 和 APERO 两种方法构建潜在的 sRNA 数据库。以逆转录 PCR 检测(real-time reverse transcription PCR, RT-PCR)方法检测 sRNA 的表达, 并通过生物信息学方法分析潜在 sRNA 的调控位点、二级结构及保守结构域。【结果】构建了包含 2 653 条潜在 sRNA 的数据库, 依据其相较于编码序列(coding sequence, CDS)区域的位置进行分类、预测功能。发现了在高 GC 革兰氏阳性菌中普遍存在的 sRNA00130, 随生长时期表达量变化的 sRNA00257, 常时高表达的 sRNA02036、sRNA02037 等 sRNA。依据自由结合能预测潜在 sRNA 可能的结合位点、二级结构, 预测结果显示潜在的 sRNAs 同多种细胞复制、代谢通路基因相关。大多数潜在 sRNA 仅在谷氨酸棒杆菌中具有保守性。【结论】谷氨酸棒杆菌潜在 sRNA 的发掘对于填补谷氨酸棒杆菌生长调控机制的空缺具有重要作用, 提供了新的可能的调控工具与改造位点。

关键词: 谷氨酸棒杆菌; 小的非编码 RNA (sRNA); 生物信息学; 全基因组分析; sRNA-Detect

资助项目: 国家自然科学基金(32270036); 国家重点研发计划(2021YFC2100900); 中央高校基本科研专项资金(JUSRP221012)

This work was supported by the National Natural Science Foundation of China (32270036), the National Key Research and Development Program of China (2021YFC2100900), and the Fundamental Research Funds for the Central Universities (JUSRP221012).

*Corresponding author. E-mail: xumeijuan@jiangnan.edu.cn

Received: 2023-02-24; Accepted: 2023-04-27; Published online: 2023-05-10

Prediction and identification of non-coding sRNAs in *Corynebacterium glutamicum* at different growth stages

WANG Yi, FAN Zhihao, RAO Zhiming, XU Meijuan*

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

Abstract: [Objective] To investigate the small non-coding RNAs (sRNAs) in *Corynebacterium glutamicum* at different growth stages. [Methods] We examined the RNA-Seq data of *C. glutamicum* strains at different growth stages and transformed for industrial production and then employed sRNA-Detect and APERO approaches to construct the prospective sRNA library. real-time reverse transcription PCR (RT-PCR) was carried out to determine the expression of sRNAs. The functions, secondary structures, and conserved domains of potential sRNAs were predicted by bioinformatics tools. [Results] We constructed a library containing 2 653 potential sRNAs and performed classification and functional prediction based on their positions relative to the coding sequence (CDS) region. We identified sRNA00130 ubiquitous in Gram-positive bacteria with high GC content, sRNA00257 with varied expression during the growth period, and sRNA02036 and sRNA02037 with high expression. We predicted the possible binding sites and secondary structures of potential sRNAs based on free binding energy. The prediction results showed that potential sRNAs were associated with a variety of genes involved in cell replication and metabolic pathways. Most potential sRNAs were conserved only in *C. glutamicum*. [Conclusion] The discovery of potential sRNAs in *C. glutamicum* helps to fill the vacancy in the growth regulation mechanism of *C. glutamicum* and provides a new possible regulation tool and modification site.

Keywords: *Corynebacterium glutamicum*; small non-coding RNAs (sRNA); bioinformatics; genome-wide analysis; sRNA-Detect

谷氨酸棒杆菌(*Corynebacterium glutamicum*)是一种非致病性、无芽孢的革兰氏阳性土壤细菌,属于放线菌目,基因组中GC含量较高,为53.8%^[1]。在工业中,它被广泛用于生产各种高值产品,包括氨基酸、有机酸、聚合物前体、芳香化合物和蛋白质^[2-5]。谷氨酸棒杆菌通过基因组和转录组测序^[1,6-9]以及代谢通量^[6]的分析,构建了较为清晰的遗传背景,为工业改造提供了基础。基于完整的基因组序列,已鉴定出一个复杂的调控网络。迄今为止,该代谢网络中已包含107种转录调节蛋白及1641种调控相互

作用^[10]。然而,目前谷氨酸棒杆菌来源的小的非编码RNA (small non-coding RNA, sRNA)及其调控作用还知之甚少。

sRNA是一种非编码短链调节RNA,通常为50–500 nt^[11],参与多种基因表达,大多数细菌sRNA通过同靶基因mRNA进行不完全碱基互补配对作用调节靶基因的转录、翻译等过程^[1],在细菌适应多种环境变化中发挥关键作用^[12-13]。由于这种不完全互补配对的调控机制,通过基因组预测的生物信息学方法挖掘sRNA并鉴定其调节作用是较为困难的。以往的sRNA的预测

方法包括基因间区域转录信号的预测、RNA 结构基序的系统发育保守性预测或基于鸟枪克隆的 RNomics 方法^[14]等。基于深度 RNA-Seq 技术的发展, 开拓了新的 sRNA 挖掘方式, 大量未被注释的转录本被发现, 搜索确定了许多新的 sRNA^[15]。现已开发出多种基于 RNA-Seq 检测 sRNA 的预测方法, 例如 APERO^[16]、ANNOgesic^[17]和 Srna-Detect^[18], 其中 sRNA-Detect 具有最高的召回率。到目前为止确定的 RNA 均收集在 Rfam^[19]在线数据库中, 该数据库可以搜索、浏览和下载信息, 并使用协方差模型对 RNA 序列进行注释。Rfam^[19]数据库中现已收集了 2 200 多个家族, 有 600 多万条序列。

目前, sRNA 的研究主要集中在大肠杆菌、沙门氏菌等革兰氏阴性菌或致病菌中^[20], 现已表征了超过 200 种 sRNA^[21]。然而, 谷氨酸棒杆菌同其他放线菌等高 GC 革兰氏阳性菌中缺少革兰氏阴性菌中 sRNA 发挥调控作用的重要 RNA 伴侣蛋白 Hfq 及其同系物^[22]。这表明 sRNA 在谷氨酸棒杆菌中可能采用同革兰氏阴性菌不同的调控策略。*ArnA*^[23]和 6C RNA^[24-25]是目前在谷氨酸棒杆菌中通过实验确定的 2 条 sRNA, *ArnA* 位于 CGL_RS11480 的上游反义链上, 6C RNA 则位于 CGL_RS01570 上游的正义链上, 它们不需要伴侣蛋白, 可直接同靶基因 mRNA 相互作用。

本研究通过对不同生长时期的谷氨酸棒杆菌进行深度测序, 结合多个 RNA sequencing (RNA-Seq) 实验建立测序文库, 通过基于 RNA-Seq 新开发的生物信息学工具对谷氨酸棒杆菌进行搜索, 检测与鉴定新的与谷氨酸棒杆菌生长时期相关的 sRNA, 并对其功能进行初步的分析与预测, 提供一种新的代谢改造工具。

1 材料与方法

1.1 菌株培养

谷氨酸棒杆菌(*Corynebacterium glutamicum*)

ATCC 13032、E01 (专利公开号为 103215198B)、G01 (中国典型培养物保藏中心, 保藏号: M2013418)均为本实验保存菌株。在脑心浸出液肉汤(brain heart infusion broth, BHI)培养基(磷酸氢二钠 2.5 g/L、胰蛋白胨 10 g/L、氯化钠 5 g/L、葡萄糖 2 g/L、牛脑浸粉 12.5 g/L、牛心浸粉 5 g/L)中培养 18 h 和 26 h 后进行 RNA-Seq 检测。

1.2 RNA 提取

使用 TIANGEN 公司培养细胞/细菌总 RNA 提取试剂盒, 通过溶菌酶裂解细胞, 使裂解产物中的核酸同离心吸附柱结合, 添加试剂盒中的 Dnase I 去除吸附柱上的 DNA, 最后使用 DEPC 酶的水洗脱, RNA 悬浮在无 RNA 酶的水中, 并通过分光光度法测定其浓度。

1.3 转录组测序分析

将谷氨酸棒杆菌接种于 BHI 培养基中, 30 °C 培养至 18 h 和 26 h 后, 4 °C、12 000 r/min 离心 2 min, 收集菌体沉淀, 用液氮保存。经 NanoDrop ND-1000 核酸检测仪(LabTech)检测 RNA 的浓度与纯度。基于 Illumina 二代高通量测序平台(HiSeq 4000)测序, 由金唯智科技有限公司完成。

1.4 sRNAs 预测

新 sRNA 通过 sRNA-Detect^[18]进行预测, 是一种专门应用于从 RNA-Seq 数据文件中检测 sRNAs 的 python 脚本。将每个 RNA-Seq 数据文件输入 sRNA-Detect^[18]中同谷氨酸棒杆菌 ATCC 13032 基因组(NC_003450.3)进行映射。参数如下: 长度在 20–500 nt 之间的转录本, 每个样本的最小覆盖率为 10 次读取。使用 Bed-Tools intersect^[26]进行过滤步骤, 以排除与基因组注释区重叠的转录物。为了确定其转录起始位点, 还通过第二种生物信息工具 APERO^[16]以相同的参数进行了搜索。

两种方法进行处理比较, 针对表达量使用每千个碱基的转录每百万映射读取的片段 (fragments per kilobase of exon model per million mapped reads, FPKM)^[27]方法(公式 1)进行归一化处理仅保留 FPKM 高于 10 的表达值。在验证以调整预测的略微不同的基因组位置和大小之后, 获得了新的推定 sRNAs 的最终列表。

$$\text{FPKM} = \frac{\text{Exon mapped fragments} \times 10^9}{\text{Total mapped fragments} \times \text{Exon length}} \quad \text{公式(1)}$$

1.5 sRNAs 分析

通过 Tertools^[28]提取谷氨酸棒杆菌 ATCC 13032 全基因组转录物序列, 使用 IntaRNA 进行 sRNAs 结合位点预测, 采用默认参数。创建了包含预测的 ncRNAs 序列的 FASTA 文件, 并将其提交至 Rfam^[19]数据库。通过 eggNOG^[29]定位分析 sRNA 可能相关的基因及其调控通路, 利用 IntaRNA^[30]、ViennaRNA^[31]工具包分析挖掘 sRNA 的潜在靶位点和保守二级结构。

1.6 实时荧光定量 PCR 分析

使用诺唯赞 HiScript II ReverseTranscriptase

试剂盒制备 cDNA。选择了 10 个 sRNA 表达基因, 以高水平、恒定表达的 16S rRNA 基因为内参基因, 设计特异性引物(表 1)。以合成的 cDNA 为模板, 使用 ChanQ Universal SYBR qPCR Master Mix 进行实时逆转录 PCR 检测 (quantitative real-time PCR, qRT-PCR)分析, 检测 mRNA 相对表达量。qPCR 结果与转录组测序数据结果均以 $|\log_2\text{Fold Change}|$ 表示。

2 结果与分析

2.1 使用 sRNA-Detect 挖掘预测谷氨酸棒杆菌中潜在 sRNA

根据前期工作, 谷氨酸棒杆菌在 18 h 位于生长的指数期, 也是谷氨酸快速积累的起始时期; 而 26 h 是指数期转换为平台期的节点, 谷氨酸积累速度进一步增加^[32]。为挖掘同谷氨酸棒杆菌生长、谷氨酸代谢途径相关的潜在 sRNA, 基于 Illumina HiSeq 二代测序技术对培养至 18 h 和 26 h 的谷氨酸棒杆菌原始菌株 ATCC 13032、工业菌株 G01 和 E01 进行全基因组测序构建文库。使用 sRNA-Detect^[18]默认参数获得预测出 3 467 条潜在 sRNA 序列, 其中包含

表 1 qPCR 引物

Table 1 Primers of qRT-PCR

Gene	Forward primer (5'→3')	Reverse primer (5'→3')
sRNA2114	GCCGGTCAAGCCTCCACGTACTCA	TCTTCAGCTGCCTTTTCTGCTTCTT
sRNA0257	TCTGTGCGTCTACTGACTGGC	CAGTAGGGTGCATGGGAGAAGAATTC
sRNA2518	GCCTGCTACCGCGGAGTAACCGTCGAAC	GGACGTCGGGTTTGGGGAATTGGTA
sRNA1040	CGGATTCTGTACCAGTTGC	GTATGTCTGTGCAGGTGTTGAAG
sRNA2046	GGCGAGGCGGGAAGAGTTGGCAAGC	CTTTAGCCAAGGACACCAAGGTCGCATCC
sRNA2515	GCGGCGTCAAAGTCCGACTC	CGGTGTGGCCGTCGATGATG
sRNA2037	TTAAGCCCGCAGCGGGCAACAGC	GGGCTTATAAGGGTCGTTGTGTAG
sRNA2036	GGCAGTAAGTGAATGAAAAAGAAG	GAGTGTGGGTTGTTGAAGAAG
sRNA1314	GCGATGATCTTTGCCATGTGTTTTG	GAGGTTGGTCCACACTCACTAACC
sRNA1816	GGTGGGACTTTAAGGAAACAGTG	GTGCTCTGCTTTGGGAAAGTTATC
16S rRNA	CGGCTAATTCACAGCAGATTCCACG	GGCAATAACGTTTCATATTCTGCCGGATG

部分 rRNA、tRNA 以及长基因中的部分小片段，通过 bedtools^[26]排除掉已注释序列。sRNA-Detect 这种检测方式具有较高的覆盖广度以及更高的召回率，但是预测结果往往短于 sRNA 本身的长度。因此，选择了通过 APERO 进行 sRNA 起止位点预测，以较高的精度定位分析转录本片段末端，同

时能够识别单个 sRNA 的前体和成熟 sRNA 亚型，确定了 sRNA 转录起止位点。2 653 个潜在 sRNAs (表 2，附表 1 已提交国家微生物科学数据中心，编号：NMDCX0000197)分布在整个基因组的正义链(1 506 个 sRNAs)和反义链(1 147 个 sRNAs)上(图 1)。

表 2 sRNA-Detect 预测谷氨酸棒杆菌中潜在 sRNA

Table 2 sRNA-Detect predicts potential sRNA in *Corynebacterium glutamicum*

sRNA_id	Start	Stop	Strand	Length (bp)	Reads
sRNA00001	7 494	7 577	+	83	31.810
sRNA00002	7 773	7 918	+	145	166.802
sRNA00003	9 862	9 978	+	116	9.679
sRNA00004	11 083	11 170	+	87	47.309
sRNA00005	11 630	11 720	+	90	23.955
sRNA00006	15 338	15 390	+	52	51.395
sRNA00007	19 165	19 310	+	145	9.843
sRNA00008	19 375	19 481	+	106	16.642
sRNA00009	23 642	23 779	+	137	19.623
sRNA00010	25 302	25 447	+	145	10.812
sRNA00011	26 293	26 494	+	201	9.833
sRNA00012	27 043	27 188	+	145	350.222
sRNA00013	27 299	27 504	+	205	1 456.682
sRNA00014	28 263	28 354	+	91	10.290
sRNA00015	28 646	28 845	+	199	9.992
sRNA00016	29 562	29 707	+	145	10.177
sRNA00017	35 796	35 996	+	200	69.437
		⋮			
sRNA02642	3 269 012	3 269 187	-	175	20.259
sRNA02643	3 269 813	3 269 938	-	125	312.594
sRNA02644	3 270 002	3 270 152	-	150	68.258
sRNA02645	3 270 702	3 270 775	-	73	10.684
sRNA02646	3 271 483	3 271 629	-	146	185.068
sRNA02647	3 272 356	3 272 503	-	147	51.745
sRNA02648	3 275 360	3 275 478	-	118	16.615
sRNA02649	3 283 256	3 283 398	-	142	5 074.972
sRNA02650	3 288 753	3 288 896	-	143	537.392
sRNA02651	3 293 156	3 293 220	-	64	35.277
sRNA02652	3 299 289	3 299 432	-	143	11.222
sRNA02653	3 299 916	3 299 966	-	50	9.955

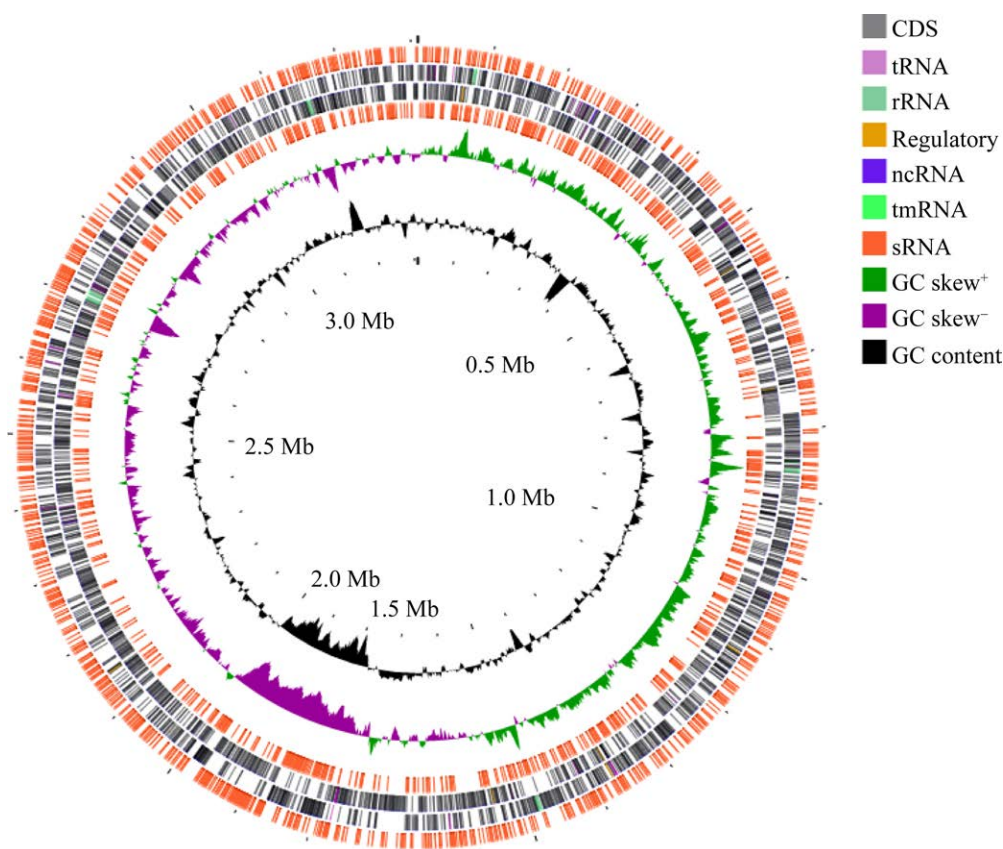


图 1 2 653 个潜在 sRNA 所在谷氨酸棒状杆菌基因组中的相对位置

Figure 1 Distribution of the 2 653 putative sRNAs in the genome of *Corynebacterium glutamicum* is represented in red.

2.2 通过相对于注释蛋白编码序列的位置对潜在 sRNA 基因进行分类

将 188 个 3' 端距下游编码序列 (coding sequence, CDS) 小于 100 nt 的转录物定义为 mRNA 前导序列 (mRNA-leader)。同 Rfam^[19] 进行对比, 在棒杆菌属中较为保守的 *atpB*^[33] 基因、多磷酸激酶 *ppk2*、脱氧核糖磷酸醛缩酶 *deoC* 和参与氨基酸代谢的 4-氨基丁酸氨基转移酶 *gabT*^[34] 等多个代谢相关基因前均发现了 mRNA 前导序列。

反义 RNA (antisense RNA, asRNAs) 位于已注释序列的反义链上, 将其分为 4 个子类型。(1) asCDSs, 位于 CDS 区域的反义链上, 共 1 478 条潜在 sRNA; (2) as5'UTRs, 位于基因 5'UTR 区

域 (基因 5' 端前 100 nt) 反义链, 该分类包含 256 条潜在 sRNA; (3) as3'UTRs, 位于基因 3'UTR 区域 (基因 3' 端后 60 nt), 有 351 条潜在 sRNA 属于该类型; (4) asRNAs, 由于 CDS 的特殊排列 100 条潜在 sRNA 同时符合多种分类标准。471 个位于基因间区域的转录物定义为顺式编码 sRNA (trans-encoded sRNAs) (图 2)。

为了检测 sRNAs 是否具有同系物, 在 Rfam^[19] 数据库中搜索与预测出的 sRNAs 相似的转录物。其中 13 个 sRNAs 预测与其他物种 tRNA 具有同源性, 18 种 sRNAs 与 rRNA 具有类似的序列, 6 种 sRNA 与 5S RNA 相似, 还检测到了高 GC 革兰氏阳性菌中保守的 6C sRNA。大部分 sRNAs 缺乏保守性, 未能检测出其同系物。

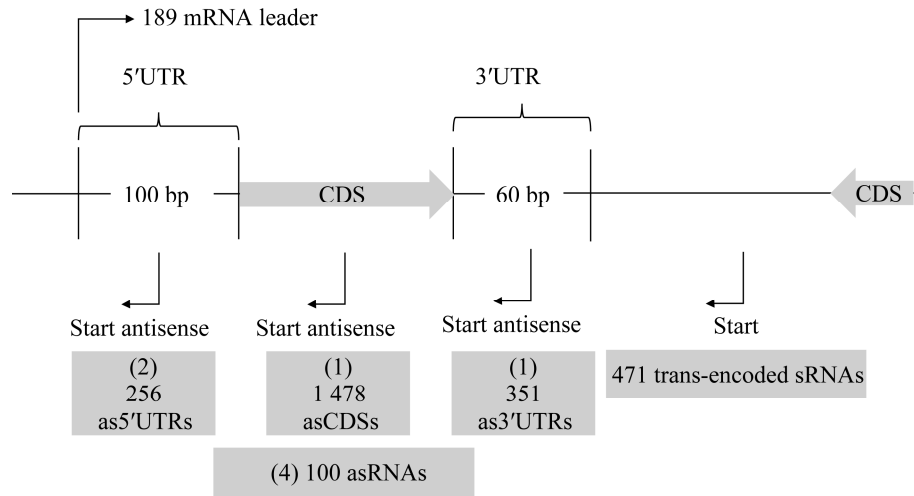


图 2 根据其相对于注释的蛋白质编码序列的位置对潜在的 sRNA 基因进行分类

Figure 2 Classification of potential sRNA genes by their positions relative to annotated protein-coding sequences. asCDSs: Located on the antisense chain of the coding sequence (CDS) region; as5'UTR: Located on the 5'-untranslated regions (5'UTR); as3'UTR: Located on the 3'-untranslated regions (3'UTR); asRNAs: Located on the multiple regions on the same time; trans-encoded sRNAs: Located on the intergenic regions (IGR).

2.3 谷氨酸棒杆菌潜在 sRNA 不同生长时期的表达量差异分析

检测所有测试条件下潜在 sRNA 的表达水平，为了获得更为准确的对比值，通过 FPKM^[27]方法对每个 sRNAs 进行归一化。在热图(图 3)中绘制所有分析条件下每个 sRNAs 的归一化表达值，并通过 K-means 方法进行聚类分析^[35]。多种 sRNA 在模式菌株 ATCC 13032 中具有最高的表达量，例如 sRNA01974、sRNA01240 等潜在 sRNA 与 E01、G01 相比，无论什么时期均具有较高的表达量。

sRNA01853、sRNA01423 在 E01 与野生型菌株中具有相似的表达趋势，在生长至平台期后相较于指数期具有更高的表达量，sRNA00189、sRNA00191 等潜在 sRNA 则在指数期具有更高的表达量。编码转录调控蛋白的 *whcB*^[36]基因反义链上的 sRNA00229 在 E01 与 G01 中随着生长时间到达平台期表达量下降，而在原始菌株中其表达量不随生长时期的变化波动。基因

porB^[37]反义链上的 sRNA0036、sRNA00397；三羧酸循环(tricarboxylic acid cycle, TCA)中的醛脱氢酶 *thcA*^[38]反义链上的 sRNA001355 在高产菌株到达平台期后具有更高的表达量，而在原始菌株中出现了下调。sRNA 在工业菌株与野生型菌株不同生长时期展现出的表达量变化差异可能是菌株间产量差异的原因之一。

2.4 蛋白质编码基因内的反义 sRNA 分析

预测结果中，超过 80%的 sRNA (2 185/2 653)为 asRNAs，这些 asRNAs 分布于 1 008 个不同的 CDS 区域，部分 CDS 具有 1 个以上的 asRNAs。为了将编码功能蛋白的基因同 asRNAs 相关联，根据 eggNOG 功能分类系统对蛋白进行分类^[28]。经定位，2 185 个 asRNAs 位于 1 008 个 eggNOG 分类的已知功能 CDS 反义链上，233 个 eggNOG 分类下未知功能 CDS 反义链上(图 4)。asRNAs 广泛分布于各类已注释 CDS 上，在绝大多数分类中 40%–60%的基因具有 asRNAs。细胞响应外界环境应激的蛋白具有 asRNAs 的比例较高

fiuQ (CGL_RS02560) 3种铁复合物转运系统渗透蛋白, 转运金属阳离子蛋白的 *asRNA*。在细菌生长进入平台期后, 3种 *asRNAs* 表达量均出现了不同程度的下降(图 3), 可能参与了细胞金属阳离子的转运调节。另一个出现了表达量明显变化 *asRNAs* 的组是翻译后修饰、蛋白质转换、伴侣 (“O”), *sRNA01314* 和 *sRNA01816* 是 CGL_RS13545 和 *groS* (CGL_RS03000) 的反义 *asRNAs*。CGL_RS13545 具有防止错误折叠促进在压力条件下产生的未折叠多肽的折叠和正确组装的功能、*groS*^[41] 是一种分子伴侣蛋白, 在 Mg-ATP 的作用下同 *Cpn60* 结合并抑制 *Cpn60* 的 ATP 活性。*sRNA01314* 在生长进入平台期后, 在 3 种菌株中均出现了表达量的下降, 在 E01 中尤为明显, *sRNA01816* 则有所不同, 在 E01 与 G01

进入生长平台期后, 出现了不同程度的表达量上升, 其中 G01 在生长至 26 h 时, *sRNA01816* 具有最高的表达量。

2.5 发掘不同生长时期高表达 *sRNA* 的潜在调控位点

IntaRNA^[30] 可以通过 RNA 序列的碱基互补配对、自由结合能预测 *sRNA* 可能的靶位点。绝大多数 *sRNAs* 通过同靶基因 5'UTR 区域结合调控靶基因位点。除了可能抑制自己所在对链基因的 5'端 UTR, 3'UTR 区域外仍可能抑制远端基因的表达。提取出谷氨酸棒杆菌全基因的 5'UTR 前 100 nt 序列, 选择高度保守的 *sRNA00130* (6C RNAs) 及表达量恒定且高表达的 4 个潜在 *asRNAs* 进行靶位点预测分析(表 3)。*sRNA02036*、*sRNA02037* 预测同多种生长、ATP

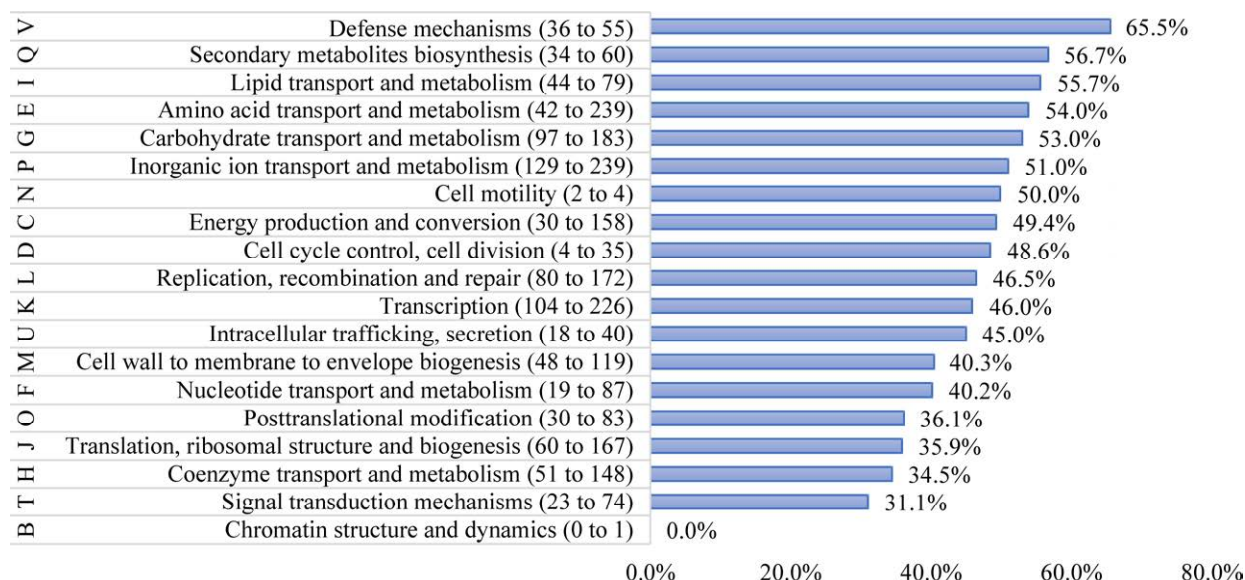


图 4 根据 eggNOG 对具有 *asRNAs* 的谷氨酸棒杆菌基因进行功能分类

Figure 4 Functional classification of *Corynebacterium glutamicum* genes having *asRNAs* according to eggNOG. B: Chromatin structure and dynamics; C: Energy production and conversion; D: Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure and biogenesis; K: Transcription; L: Replication, recombination and repair; M: Cell wall to membrane to envelope biogenesis; N: Cell motility; O: Posttranslational modification; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion; V: Defense mechanisms.

表 3 潜在 sRNA 预测调控靶基因位点及其功能

Table 3 sRNAs regulate target gene sites and their functions

Id1	Start1	End1	Id2	Start2	End2	E	Product
sRNA00130							
<i>zupT</i>	109	181	sRNA00130	35	101	-17.51	Zinc transporter <i>ZupT</i>
<i>scpA</i>	127	138	sRNA00130	40	50	-17.43	Methylmalonyl-CoA mutase
<i>catA</i>	55	69	sRNA00130	75	88	-16.86	Catechol 1% 2C2-dioxygenase
<i>nrdH</i>	22	35	sRNA00130	76	89	-15.25	Glutaredoxin-like protein <i>NrdH</i>
<i>mraZ</i>	48	58	sRNA00130	76	86	-14.74	Division/cell wall cluster transcriptional repressor <i>MraZ</i>
<i>nrdF</i>	124	135	sRNA00130	78	88	-14.59	Class 1b ribonucleoside-diphosphate reductase subunit beta
<i>rsmH</i>	286	292	sRNA00130	40	46	-14.28	16S rRNA (cytosine(1402)-N(4))-methyltransferase <i>RsmH</i>
<i>rplN</i>	85	100	sRNA00130	75	87	-14.19	50S ribosomal protein L14
<i>rimP</i>	174	184	sRNA00130	76	85	-14.17	Ribosome maturation factor <i>RimP</i>
<i>mgo</i>	32	42	sRNA00130	78	87	-14.17	Malate dehydrogenase (quinone)
sRNA02036							
<i>idi</i>	41	100	sRNA02036	113	181	-26.92	Isopentenyl-diphosphate delta-isomerase
<i>glx</i>	2	31	sRNA02036	142	169	-18.48	Glutamate-tRNA ligase
<i>secD</i>	38	92	sRNA02036	72	123	-17.70	Protein translocase subunit <i>SecD</i>
<i>tisS</i>	1	97	sRNA02036	21	116	-16.99	tRNA lysidine(34) synthetase <i>TisS</i>
<i>fudC</i>	31	50	sRNA02036	146	178	-16.01	Furfural detoxification alcohol dehydrogenase <i>FudC</i>
<i>rpmF</i>	77	94	sRNA02036	148	165	-15.73	50S ribosomal protein L32
<i>rph</i>	32	95	sRNA02036	53	121	-15.61	Ribonuclease PH
<i>xseA</i>	79	90	sRNA02036	154	165	-15.13	Exodeoxyribonuclease VII large subunit
<i>carR</i>	31	53	sRNA02036	1	22	-14.92	MarR family transcriptional regulator <i>CarR</i>
<i>lexA</i>	84	97	sRNA02036	149	165	-14.33	Transcriptional repressor <i>LexA</i>
sRNA02037							
<i>yaaA</i>	3	61	sRNA02037	83	147	-19.83	Peroxide stress protein <i>YaaA</i>
<i>dtd</i>	59	70	sRNA02037	10	21	-16.55	D-aminoacyl-tRNA deacylase
<i>aroP</i>	38	46	sRNA02037	12	20	-16.19	Aromatic amino acid transport protein <i>AroP</i>
<i>orn</i>	21	49	sRNA02037	8	31	-15.49	Oligoribonuclease
<i>serC</i>	88	95	sRNA02037	13	20	-13.72	Phosphoserine transaminase
<i>ftsE</i>	22	47	sRNA02037	121	149	-13.02	Cell division ATP-binding protein <i>FtsE</i>
<i>iolB</i>	23	58	sRNA02037	95	133	-12.94	5-deoxy-glucuronate isomerase
<i>gap</i>	40	52	sRNA02037	123	135	-12.90	Type I glyceraldehyde-3-phosphate dehydrogenase
<i>hisG</i>	40	48	sRNA02037	12	20	-12.75	ATP phosphoribosyltransferase
<i>mnhG</i>	68	77	sRNA02037	11	20	-12.53	Monovalent cation/H ⁺ antiporter subunit G
<i>cysE</i>	4	32	sRNA02037	111	142	-12.23	Serine O-acetyltransferase
sRNA02046							
<i>tatB</i>	16	50	sRNA02046	2	39	-17.46	Sec-independent protein translocase subunit <i>TatB</i>
<i>urtE</i>	70	100	sRNA02046	104	132	-16.67	Urea ABC transporter ATP-binding subunit <i>UrtE</i>
<i>qsuB</i>	56	75	sRNA02046	14	33	-15.55	Dehydroshikimate dehydratase <i>QsuB</i>
<i>secD</i>	35	84	sRNA02046	39	95	-15.47	Protein translocase subunit <i>SecD</i>
<i>sucD</i>	43	73	sRNA02046	2	36	-15.37	Succinate-CoA ligase subunit alpha
<i>lldD</i>	1	11	sRNA02046	154	164	-15.24	Quinone-dependent L-lactate dehydrogenase
<i>brnE</i>	26	42	sRNA02046	17	33	-14.62	Branched-chain amino acid exporter <i>BrnE</i>
<i>pheA</i>	9	52	sRNA02046	2	58	-14.30	Prephenate dehydratase
<i>aceA</i>	28	95	sRNA02046	110	171	-13.20	Isocitrate lyase
<i>menE</i>	22	39	sRNA02046	22	41	-12.72	o-succinylbenzoate-CoA ligase

依赖性转运基因存在结合位点。而参与 SOS 系统的 sRNA0130 预测结果中包含了 *mraZ* 等细胞自我复制相关蛋白。sRNA2046 则可能与能量代谢和辅酶 A 相关。

2.6 sRNA00130、sRNA02037 的保守性分析

选取全部条件下均具有极高表达量的 sRNA2037 进行二级结构以及保守性分析。同大多数 sRNAs 相同，sRNA02037 缺乏保守性，仅能在谷氨酸棒杆菌属中检测到 sRNA02037 的基

因(图 5B)，6C RNA 则作为特例具有更广泛的保守性^[25]，在高 GC 革兰氏阳性菌中均有表达。通过 RNafold 及 LocARNA^[31]分析不同种属来源的 sRNA02037 及 sRNA00130 (6C sRNA)，可以生成二级结构预测并突出显示保守区域。sRNA00130 的 2 个富含 C 的茎环具有保守性，尤其 100 bp 附近的含 C 茎环在多个菌株中均展现了良好的保守性。sRNA02037 预测 2 段位于茎环双链部分富含 GC 的区域及 2 个茎环的单链部分具有较高的保守性。

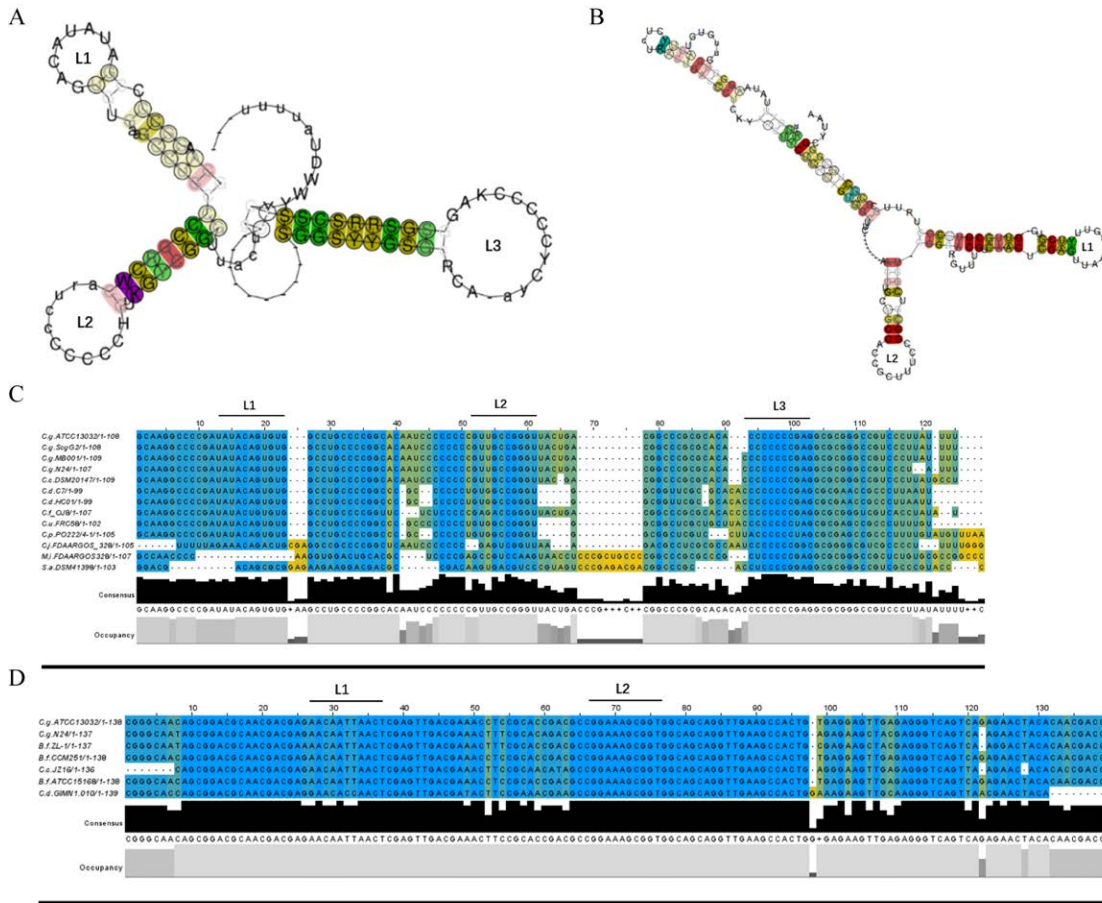


图 5 sRNA00130、sRNA02037 的保守性、保守结构域分析

Figure 5 Conservation analysis of sRNA00130 and sRNA02037. A: The secondary structure of sRNA02037. B: The secondary structure of sRNA00130. C: Sequence alignment of sRNA0130. C.g: *Corynebacterium glutamicum*; C.f: *Corynebacterium flavescens*; C.c: *Corynebacterium callunae*; C.d: *Corynebacterium diphtheriae*; C.u: *Corynebacterium ulcerans*; C.p: *Corynebacterium pseudotuberculosis*; C.j: *Corynebacterium jeikeium*; M.i: *Micromonospora inositola*; S.a: *Streptomyces albus*. D: Sequence alignment of sRNA02037. C.g: *Corynebacterium glutamicum*; C.c: *Corynebacterium callunae*; C.d: *Corynebacterium diphtheriae*; B.f: *Brevibacterium flavum*.

2.7 通过 RT-PCR 检验不同菌株、不同生长时期潜在 sRNA 表达量变化

为了验证 sRNA 数据库的准确性, 选择了 11 种在多种条件下均保持高表达、或具有明显差异表达的 sRNA 进行 qRT-PCR 实验验证(图 6A)。qRT-PCR 测得的 sRNA 大小约为 150–200 nt。sRNA2036 与 sRNA2037 有着最高的表达水平, 而且不受菌株、生长时间的限制, 一直维持在一个较高的水平上。sRNA00257 具有最高表达差异, 在不同菌株的生长后期均出现了不同程度的上调, 可能是主要调节谷氨酸棒杆菌生长相关途

径的 sRNA 之一。sRNA00257 位于编码 O-乙酰高丝氨酸(硫醇)裂解酶的负链上, 可能参与了半胱氨酸与蛋氨酸代谢途径的调控。sRNA00130 在高 GC 革兰氏阳性菌中广泛存在, 目前已经验证同生长和 SOS 系统相关^[25], 其过表达会抑制细胞的生长, 一般在指数期转换至平台期时会有显著的表达量上调, 在进入平台期后其表达量下降。sRNA01314 和 sRNA01816 展现了较为明显的菌株差异; sRNA01314 在 E01 中生长后期几乎不存在; sRNA01816 在 G01 与 E01 的生长后期均具有较为明显的表达量上调, 在 G01 中增幅更为明显(图 6B)。

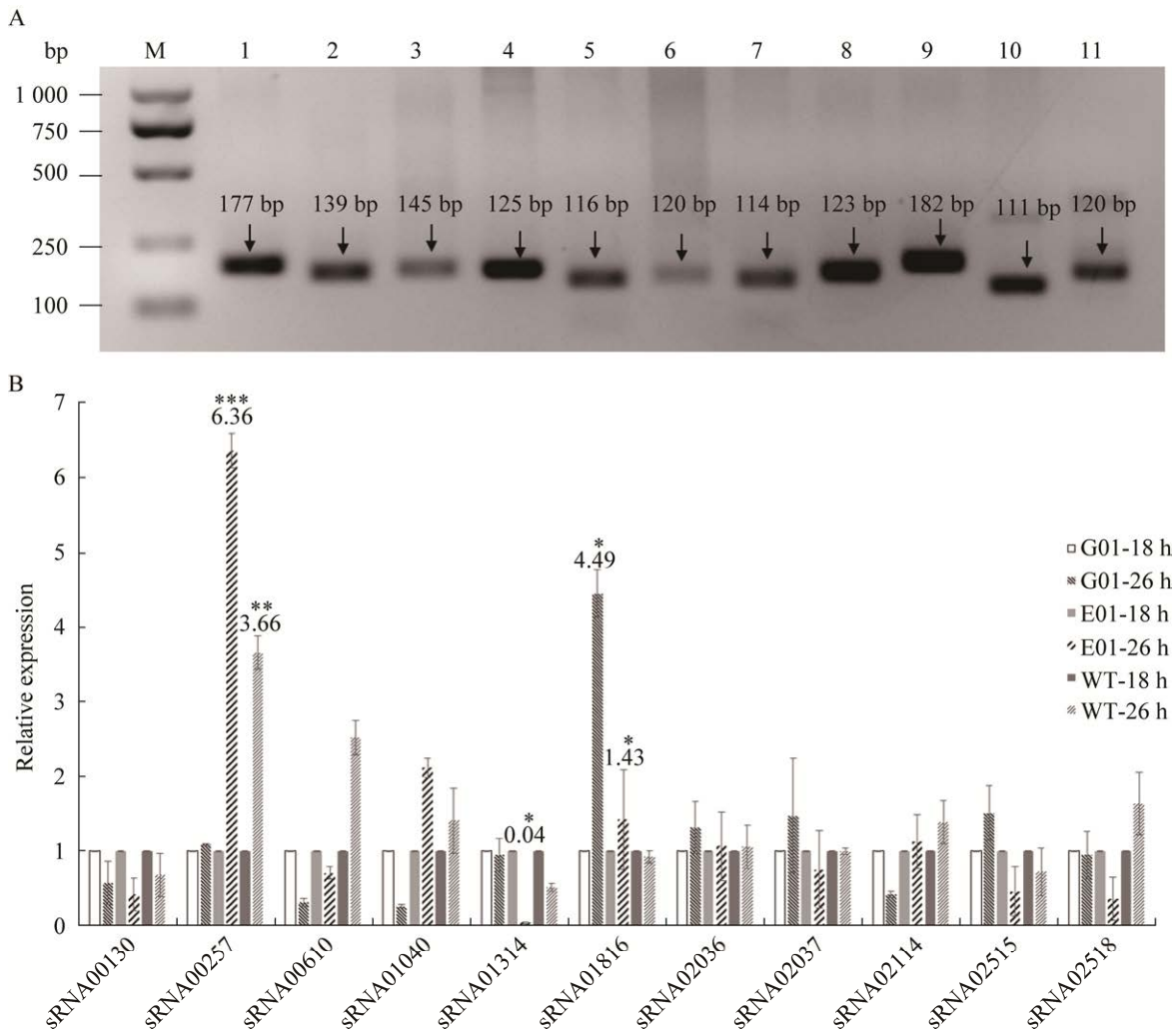


图 6 潜在 sRNA 的 RT-PCR 分析

Figure 6 RT-PCR analysis of putative sRNA. A: RT-PCR detection of 11 sRNAs. 1: sRNA2114; 2: sRNA0257; 3: sRNA2518; 4: sRNA1040; 5: sRNA2046; 6: sRNA2515; 7: sRNA1030; 8: sRNA2037; 9: sRNA2036; 10: sRNA1314; 11: sRNA1798. B: Detection of expression of sRNA change by qRT-PCR. ***: $P < 5 \times 10^{-4}$; **: $P < 0.005$; *: $P < 0.05$.

3 讨论与结论

sRNA 是微生物响应环境变化, 参与应激反应通路的重要调节机制。目前, 在工业上已开发出部分针对大肠杆菌的 sRNA 代谢调控工具, 具有可以快速高效的调控多个靶点、对细胞代谢小等多种优点, 已广泛用于代谢改造中^[2,30]。高 GC 革兰氏阳性菌因为缺乏 Hfq 及其同系物, 其特有的功能结构域尚不明确, 发掘数量较少, 对谷氨酸棒杆菌潜在 sRNA 的发掘将有助于开发适用于高 GC 革兰氏阳性菌的 sRNA 代谢改造工具。迄今为止, 已经证明谷氨酸棒杆菌中的 sRNA ArnA 同热应激相关^[23], 在 40 °C 热休克条件下表达量显著上升; 6C RNA 同 SOS 系统具有相关性, 缺失 6C RNA 会导致谷氨酸棒杆菌修复基因损伤的性能下降^[24-25]。本研究选择多种工业用菌株, 在生长指数期、平台期进行 RNA-Seq 检测, 构建潜在的 sRNA 数据库, 通过位置、自由结合能等方法预测分析可能的调控位点。发现其中超过一半的 sRNA 位于已标注基因的反义链上, 部分 sRNA 展现出了随生长时间表达量变化的特征, 少部分 sRNA 在各个时期均具有强转录效果。本研究详细地分析了谷氨酸棒杆菌不同生长时期的 sRNA, 研究结果可为更好地理解谷氨酸棒杆菌遗传背景及相关分子调控机制, 为开发新型 sRNA 代谢改造工具提供理论基础。

参考文献

- [1] IKEDA M, NAKAGAWA S. The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes[J]. *Applied Microbiology and Biotechnology*, 2003, 62(2): 99-109.
- [2] 徐美娟, 上官春雨, 陈鑫, 张显, 杨套伟, 饶志明. 谷氨酸棒杆菌耐受胁迫机制及工业鲁棒性合成生物学研究进展[J]. *生物工程学报*, 2021, 37(3): 831-845. XU MJ, SHANGGUAN CY, CHEN X, ZHANG X, YANG TW, RAO ZM. Advances in stress tolerance mechanisms and synthetic biology for the industrial robustness of *Corynebacterium glutamicum*[J]. *Chinese Journal of Biotechnology*, 2021, 37(3): 831-845 (in Chinese).
- [3] WANG YY, XU JZ, ZHANG WG. Metabolic engineering of l-leucine production in *Escherichia coli* and *Corynebacterium glutamicum*: a review[J]. *Critical Reviews in Biotechnology*, 2019, 39(5): 633-647.
- [4] HERMANN T. Industrial production of amino acids by coryneform bacteria[J]. *Journal of Biotechnology*, 2003, 104(1/2/3): 155-172.
- [5] BECKER J, ROHLES CM, WITTMANN C. Metabolically engineered *Corynebacterium glutamicum* for bio-based production of chemicals, fuels, materials, and healthcare products[J]. *Metabolic Engineering*, 2018, 50: 122-141.
- [6] KALINOWSKI J, BATHE B, BARTELS D, BISCHOFF N, BOTT M, BURKOVSKI A, DUSCH N, EGGELING L, EIKMANNS BJ, GAIGALAT L, GOESMANN A, HARTMANN M, HUTHMACHER K, KRÄMER R, LINKE B, MCHARDY AC, MEYER F, MÖCKEL B, PFEFFERLE W, PÜHLER A, et al. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins[J]. *Journal of Biotechnology*, 2003, 104(1/2/3): 5-25.
- [7] REIS-DENNIS S. Review of rethinking health care ethics by Stephen scher and kasia kozlowska[J]. *Monash Bioethics Review*, 2020, 38(1): 83-86.
- [8] PFEIFER-SANCAR K, MENTZ A, RÜCKERT C, KALINOWSKI J. Comprehensive analysis of the *Corynebacterium glutamicum* transcriptome using an improved RNA-seq technique[J]. *BMC Genomics*, 2013, 14: 888.
- [9] TAUCH A, HOMANN I, MORMANN S, RÜBERG S, BILLAULT A, BATHE B, BRAND S, BROCKMANN-GRETZA O, RÜCKERT C, SCHISCHKA N, WRENGER C, HOHEISEL J, MÖCKEL B, HUTHMACHER K, PFEFFERLE W, PÜHLER A, KALINOWSKI J. Strategy to sequence the genome of *Corynebacterium glutamicum* ATCC 13032: use of a cosmid and a bacterial artificial chromosome library[J]. *Journal of Biotechnology*, 2002, 95(1): 25-38.
- [10] TEIXEIRA DORNELLES PARISE M, PARISE D, BENTES KATO R, PAULING JK, TAUCH A, de CARVALHO AZEVEDO VA, BAUMBACH J. CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks[J]. *Scientific Data*, 2020, 7: 142.

- [11] LIVNY J, TEONADI H, LIVNY M, WALDOR MK. High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs[J]. *PLoS One*, 2008, 3(9): e3197.
- [12] 刘含梅, 钱心怡, 唐雨婷, 黄炜莉, 张鹭. 结核分枝杆菌 sRNA Mpr5 的功能验证[J]. *微生物学报*, 2022, 62(7): 2850-2858.
LIU HM, QIAN XY, TANG YT, HUANG WL, ZHANG L. Functional verification of sRNA Mpr5 in *Mycobacterium tuberculosis*[J]. *Acta Microbiologica Sinica*, 2022, 62(7): 2850-2858 (in Chinese).
- [13] STORZ G, ALTUVIA S, WASSARMAN KM. An abundance of RNA regulators[J]. *Annual Review of Biochemistry*, 2005, 74: 199-217.
- [14] ROMBY P. Small RNAs in bacteria and archaea: who they are, what they do, and how they do it[J]. *Advances in Genetics*, 2015, 90: 133-208.
- [15] DÜHRING U, AXMANN IM, HESS WR, WILDE A. An internal antisense RNA regulates expression of the photosynthesis gene *isiA*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103(18): 7054-7058.
- [16] LEONARD S, MEYER S, LACOUR S, NASSER W, HOMMAIS F, REVERCHON S. APERO: a genome-wide approach for identifying bacterial small RNAs from RNA-Seq data[J]. *Nucleic Acids Research*, 2019, 47(15): e88.
- [17] YU SH, VOGEL J, FÖRSTNER KU. ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes[J]. *GigaScience*, 2018, 7(9): giy096.
- [18] PEÑA-CASTILLO L, GRÜELL M, MULLIGAN ME, LANG AS. Detection of bacterial small transcripts from RNA-seq data: a comparative assessment[J]. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*, 2016, 21: 456-467.
- [19] DAUB J, EBERHARDT RY, TATE JG, BURGE SW. Rfam: annotating families of non-coding RNA sequences[J]. *Methods in Molecular Biology (Clifton, N J)*, 2015, 1269: 349-363.
- [20] VOGEL J, LUISI BF. Hfq and its constellation of RNA[J]. *Nature Reviews Microbiology*, 2011, 9(8): 578-589.
- [21] HÖR J, GORSKI SA, VOGEL J. Bacterial RNA biology on a genome scale[J]. *Molecular Cell*, 2018, 70(5): 785-799.
- [22] HANING KT, CHO SH, CONTRERAS LM. Small RNAs in mycobacteria: an unfolding story[J]. *Frontiers in Cellular and Infection Microbiology*, 2014, 4: 96.
- [23] ZEMANOVÁ M, KADERÁBKOVÁ P, PÁTEK M, KNOPPOVÁ M, SILAR R, NESVERA J. Chromosomally encoded small antisense RNA in *Corynebacterium glutamicum*[J]. *FEMS Microbiology Letters*, 2008, 279(2): 195-201.
- [24] PAHLKE J, DOSTÁLOVÁ H, HOLÁTKO J, DEGNER U, BOTT M, PÁTEK M, POLEN T. The small 6C RNA of *Corynebacterium glutamicum* is involved in the SOS response[J]. *RNA Biology*, 2016, 13(9): 848-860.
- [25] MAI JT, RAO CT, WATT J, SUN X, LIN C, ZHANG L, LIU J. *Mycobacterium tuberculosis* 6C sRNA binds multiple mRNA targets via C-rich loops independent of RNA chaperones[J]. *Nucleic Acids Research*, 2019, 47(8): 4292-4307.
- [26] QUINLAN AR, HALL IM. BEDTools: a flexible suite of utilities for comparing genomic features[J]. *Bioinformatics*, 2010, 26(6): 841-842.
- [27] CONESA A, MADRIGAL P, TARAZONA S, GOMEZ-CABRERO D, CERVERA A, MCPHERSON A, SZCZEŚNIAK MW, GAFFNEY DJ, ELO LL, ZHANG XG, MORTAZAVI A. A survey of best practices for RNA-Seq data analysis[J]. *Genome Biology*, 2016, 17: 13.
- [28] CHEN CJ, CHEN H, ZHANG Y, THOMAS HR, FRANK MH, HE YH, XIA R. TBtools: an integrative toolkit developed for interactive analyses of big biological data[J]. *Molecular Plant*, 2020, 13(8): 1194-1202.
- [29] JENSEN LJ, JULIEN P, KUHN M, von MERING C, MULLER J, DOERKS T, BORK P. eggNOG: automated construction and annotation of orthologous groups of genes[J]. *Nucleic Acids Research*, 2008, 36(suppl_1): D250-D254.
- [30] LORENZ R, BERNHART SH, HÖNER ZU SIEDERDISSEN C, TAHER H, FLAMM C, STADLER PF, HOFACKER IL. ViennaRNA package 2.0[J]. *Algorithms for Molecular Biology*, 2011, 6:26.
- [31] LEISTRA AN, CURTIS NC, CONTRERAS LM. Regulatory non-coding sRNAs in bacterial metabolic pathway engineering[J]. *Metabolic Engineering*, 2019, 52: 190-214.
- [32] 乔郅钠, 徐美娟, 龙梦飞, 杨套伟, 张显, 中西秀树, 饶志明. TCA 循环关键节点对 L-谷氨酸合成的影响[J]. *生物工程学报*, 2020, 36(10): 2113-2125.
QIAO ZN, XU MJ, LONG MF, YANG TW, ZHANG X, ZHONG XXS, RAO ZM. Effect of key nodes of TCA

- cycle on L-glutamate production[J]. *Chinese Journal of Biotechnology*, 2020, 36(10): 2113-2125 (in Chinese).
- [33] BARRIUSO-IGLESIAS M, BARREIRO C, SOLA-LANDA A, MARTÍN JF. Transcriptional control of the F₀F₁-ATP synthase operon of *Corynebacterium glutamicum*: SigmaH factor binds to its promoter and regulates its expression at different pH values[J]. *Microbial Biotechnology*, 2013, 6(2): 178-188.
- [34] ZHU LF, MACK C, WIRTZ A, KRANZ A, POLEN T, BAUMGART M, BOTT M. Regulation of γ -aminobutyrate (GABA) utilization in *Corynebacterium glutamicum* by the PucR-type transcriptional regulator GabR and by alternative nitrogen and carbon sources[J]. *Frontiers in Microbiology*, 2020, 11: 544045.
- [35] ARTHUR D, VASSILVITSKII S. K-means++: the advantages of careful seeding[C]//Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. January 7-9, 2007, New Orleans, Louisiana. New York: ACM, 2007: 1027-1035.
- [36] LEE JY, PARK JS, KIM HJ, KIM Y, LEE HS. *Corynebacterium glutamicum whcB*, a stationary phase-specific regulatory gene[J]. *FEMS Microbiology Letters*, 2012, 327(2): 103-109.
- [37] BUHR F, LAHROUSSI A, SPRINGER A, RUSTGI S, von WETTSTEIN D, REINBOTHE C, REINBOTHE S. NADPH: protochlorophyllide oxidoreductase B (PORB) action in *Arabidopsis thaliana* revisited through transgenic expression of engineered barley PORB mutant proteins[J]. *Plant Molecular Biology*, 2017, 94(1): 45-59.
- [38] Bott M. Offering surprises: TCA cycle regulation in *Corynebacterium glutamicum*[J]. *Trends in Microbiology*, 2007, 15(9): 417-425.
- [39] RÜCKERT C, PÜHLER A, KALINOWSKI J. Genome-wide analysis of the L-methionine biosynthetic pathway in *Corynebacterium glutamicum* by targeted gene deletion and homologous complementation[J]. *Journal of Biotechnology*, 2003, 104(1/2/3): 213-228.
- [40] SCHULTZ-HAUSER G, van HOVE B, BRAUN V. 8-azido-ATP labelling of the FecE protein of the *Escherichia coli* iron citrate transport system[J]. *FEMS Microbiology Letters*, 1992, 95(2/3): 231-234.
- [41] XU GC, XIAO L, WU AN, HAN RZ, NI Y. Enhancing *n*-butanol tolerance of *Escherichia coli* by overexpressing of stress-responsive molecular chaperones[J]. *Applied Biochemistry and Biotechnology*, 2021, 193(1): 257-270.