

大肠杆菌在葡萄糖和甘油碳源培养条件下的基因表达谱分析^{*}

朱梦黎, 张玉娟

(重庆师范大学 昆虫与分子生物学研究所 媒介昆虫重庆市重点实验室, 重庆 401331)

摘要:【目的】探讨葡萄糖和甘油碳源对大肠杆菌(*Escherichia coli*)基因表达谱的影响。【方法】通过基因表达数据库 GEO 下载基因芯片数据集 GSE2037 和高通量测序数据集 GSE156143, 分别对两个数据集中的葡萄糖和甘油碳源样本进行差异表达基因筛选, 对筛选到的差异表达基因进行并集后, 进行 GO 功能富集分析、KEGG 通路富集分析以及蛋白质相互作用网络构建。【结果】筛选出 701 个差异表达基因, 差异表达基因主要富集于小分子分解代谢过程、碳水化合物运输、趋化性等生物学过程, 细胞器、细菌型鞭毛、甲基接受趋化蛋白复合物等细胞组分, 离子跨膜转运蛋白活性、碳水化合物结合、肌动活动等分子功能以及不同环境中的微生物代谢和 ABC 转运体相关通路。此外, 筛选出的 14 个核心基因在鞭毛合成和趋化中发挥作用。【结论】大肠杆菌在葡萄糖和甘油碳源培养条件下具有不同的基因表达模式, 所确定的 14 种与趋化和鞭毛合成有关的基因有助于进一步揭示大肠杆菌在应对不同营养环境变化时采取的分子调控机制。

关键词: 大肠杆菌; 葡萄糖; 甘油; 基因表达谱; 生物信息学分析

中图分类号: Q175

文献标志码: A

文章编号: 1672-6693(2022)02-0135-06

大肠杆菌(*Escherichia coli*)能在最佳生长条件下快速生长, 是被广泛研究的细菌之一^[1]。该菌种具有精细的调节机制, 不同调节因子可相互作用于它复杂的调控网络中以协调基因表达^[2]。在饥饿胁迫(如生长速度受限、营养条件受限)下, 大肠杆菌的基因表达会发生特有变化^[3]。在碳限制^[4]、缺氧^[5]、铁限制^[6]、硫和氮限制^[7]等各种营养限制条件下对大肠杆菌进行广泛的基因表达谱研究, 可以确定参与相应过程的主要调节因子。大量研究对大肠杆菌适应养分消耗的机制进行了全面探索, 且通常侧重于碳限制条件^[8-9]。在大肠杆菌的碳限制期间, 可以观察到参与完全由碳原子或混合原子组成的化合物分解代谢的基因表达上调, 而表达下调的基因相对较少^[10]。在碳限制条件下, 碳分解代谢基因表达量随生长率的降低呈线性增长; 而生物合成基因表达量与碳分解代谢基因表达量的线性增长呈依赖性相反^[11]。单个底物的代谢对大肠杆菌的定植和适应性产生了深远的影响, 且不同的底物为细菌细胞提供的碳和能量是不同的^[8]。除了这些高度特异性的机制外, 大肠杆菌还具有多个层面的全局调控。在碳和能量限制生长过程中, 全局调节因子 RpoS 和 Cyclic-AMP/CRP 对大肠杆菌的分解代谢和转录组产生了影响^[12]。大肠杆菌对碳稀缺的反应不仅限于通过 cAMP 信号维持足够的碳代谢, 而更主要的反应是调整代谢以应对缓慢的生长速度^[13]。

微生物的首选碳源长期以来被认为是葡萄糖^[14]。然而许多微生物也可能通过呼吸代谢使用其他碳源, 如甘油^[15]。目前有关不同碳源培养对大肠杆菌基因表达调控影响的研究还较为少见, 因此本研究聚焦于葡萄糖和甘油这两种糖酵解底物碳源, 通过生物信息学分析方法探究两者对大肠杆菌基因表达谱的影响, 为揭示大肠杆菌如何通过分子调控机制应对营养环境变化提供参考资料。

1 材料与方法

1.1 数据来源

从基因表达数据库 GEO(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>)下载大肠杆菌在不同碳源条件下的转录组测序数据 GSE2037 和 GSE156143。GSE2037 是利用 Affymetrix 公司的基因芯片测定的大肠

^{*} 收稿日期: 2021-10-25 修回日期: 2021-12-01 网络出版时间: 2022-03-29 10:29

资助项目: 国家自然科学基金(No. 31871274); 重庆市自然科学基金和前沿研究计划项目(No. cstc2018jcyjA2487)

第一作者简介: 朱梦黎, 女, 研究方向为生物信息学, E-mail: 3074770654@qq.com; 通讯作者: 张玉娟, 女, 教授, 博士, E-mail: zhangyj@cqnu.edu.cn

网络出版地址: <https://kns.cnki.net/kcms/detail/50.1165.N.20220328.1409.033.html>

杆菌转录组数据。GSE156143 是利用 Illumina 公司的 HiSeq 4000 测得的大肠杆菌转录组数据。

1.2 主成分分析

分别对 GSE2037 和 GSE156143 提取葡萄糖和甘油相关样本。GSE2037 数据集共包含了以葡萄糖、甘油、琥珀酸、丙氨酸、醋酸盐和脯氨酸为碳源共 15 个样本的大肠杆菌基因表达谱数据。GSE156143 包括以葡萄糖六磷酸、乳糖、葡萄糖、琥珀酸盐、甘油、山梨糖醇和醋酸盐共 7 种碳源以及氯化铵为氮源的高通量测序大肠杆菌基因表达谱数据。利用 R 软件(4.0.3 版, <https://www.r-project.org>)中的 FactoMineR 包^[16], 分别对两个数据集中的以葡萄糖和甘油为碳源的大肠杆菌基因表达谱数据进行主成分分析。

1.3 差异表达基因筛选

利用 R 软件中 Limma 包^[17]和 edgeR 包^[18]对从 GSE2037 和 GSE156143 提取到的葡萄糖和甘油样本基因表达谱数据分别进行差异基因表达分析。选择同一标准即差异倍数(Fold change, V_{FC})大于 1 或者小于 1 且统计学检验 p 值小于 0.05 来筛选差异表达基因。利用 R 软件中 ggvenn 包^[19]展示两组数据组间差异表达基因的数目以及组间差异表达基因的重叠关系。本研究中, 以甘油碳源作为实验组, 葡萄糖碳源作为对照组, 由此比较两种碳源培养下大肠杆菌的基因表达差异。

1.4 GO 功能富集和 KEGG 通路富集

对从 GSE2037 和 GSE156143 筛选到的差异表达基因取并集, 之后对并集的差异表达基因进行 GO 功能富集分析^[20]和 KEGG 通路富集分析^[21]。GO 功能富集分为生物学过程(Biological process)、细胞组分(Cellular component)和分子功能(Molecular function)共 3 部分, 利用 R 软件中 clusterProfiler 包^[22]进行分析, 选取 p 值最小的 10 个 GO 条目进行绘图。KEGG 通路可系统分析基因在生物体内的代谢、膜运输、信号转导、细胞周期等功能, 利用 R 软件中 GSEA 包^[23]进行。 $p < 0.05$ 的基因被认为是显著富集的基因。

1.5 蛋白质相互作用关系网络构建

对从 GSE2037 和 GSE156143 筛选到的差异表达基因取并集, 利用 R 软件中的 STRINGdb 包^[24]对差异表达基因进行蛋白质相互作用关系网络构建, 从而预测蛋白质之间的相互作用关系, 并将网络度值大于 52 的基因定义为核心基因。

2 结果与分析

2.1 主成分分析结果

图 1a, b 分别显示了 GSE2037 和 GSE156143 中葡萄糖和甘油样本的主成分分析结果。由两图可知, 同一碳源的大肠杆菌基因表达谱数据聚集, 而不同碳源的大肠杆菌基因表达谱数据较分散, 因而样本辨识度较好, 具有进行后续分析的价值。

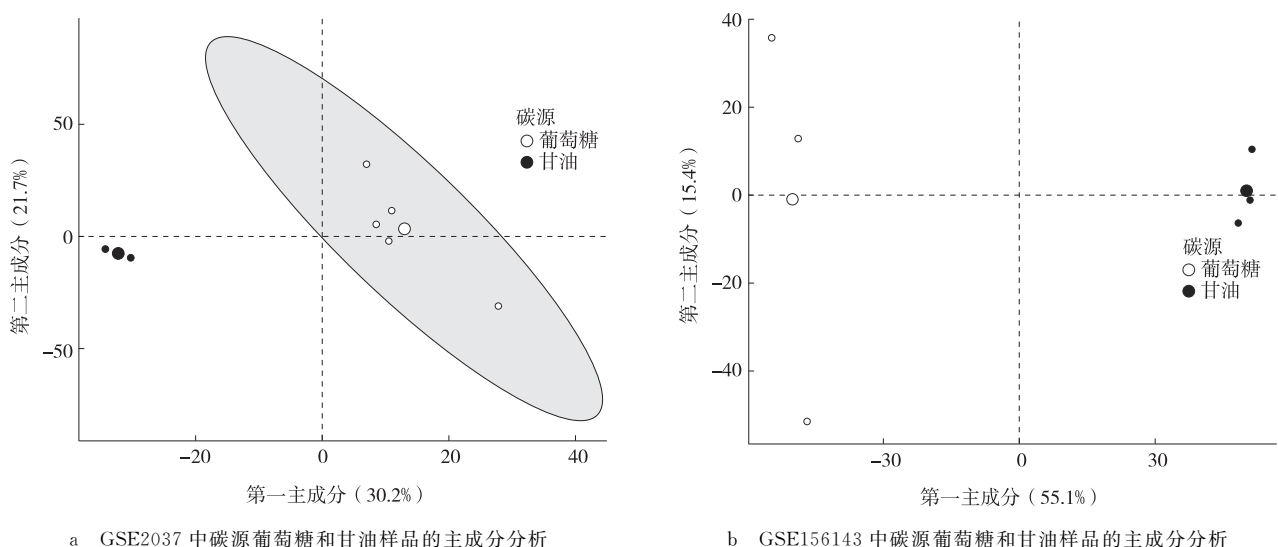


图 1 GSE2037 和 GSE156143 中碳源葡萄糖和甘油样品的主成分分析

Fig. 1 Principal component analysis of samples cultured by glucose and glycerol in GSE2037 and GSE156143

2.2 差异表达基因的筛选结果

如封三彩图 2a 所示,通过差异表达基因筛选分别从 GSE2037 和 GSE156143 数据集得到大肠杆菌差异表达基因 116 和 599 个。两组数据集交集基因为 14 个,并集基因为 701 个。选择并集基因作为后续分析的差异表达基因,并筛选出差异表达基因中 p 值最小的 20 个基因。如封三彩图 2b 所示,701 个差异表达基因中有 551 个基因表达上调,150 个基因表达下调。在差异表达基因 p 值最小的 20 个基因中,有 19 个基因表达上调,它们分别为:*glpA*、*glpF*、*glpQ*、*glpB*、*fliD*、*glpC*、*tar*、*cstA*、*acs*、*cheZ*、*ptsG*、*flgK*、*lhgO*、*pck*、*yjcH*、*actP*、*exuT*、*flgA*、*gpr* 和 *fliL*。仅 1 个基因 *ptsG* 表达下调。这些差异表达基因对葡萄糖和甘油两种底物具有较高的辨识度,提示部分基因表达模式的改变与不同碳源作为底物有关,通过对这部分差异表达基因的富集分析,可为有关不同碳源作为底物的大肠杆菌基因表达模式改变的分子机制研究提供线索。

2.3 GO 富集分析结果

将从 GSE2037 和 GSE156143 数据集中筛选出的差异表达基因进行并集后进行 GO 功能富集,所得结果见封三彩图 3,其中圆圈颜色越偏红色表示对应的 GO 功能富集越显著,圆圈尺寸越大表示对应的 GO 功能富集到的基因数量越大。研究结果显示:小分子分解代谢过程、碳水化合物运输、趋化性等生物学过程显著富集(封三彩图 3a);细胞器、细菌型鞭毛、甲基接受趋化蛋白复合物等细胞组分显著富集(封三彩图 3b);离子跨膜转运蛋白活性、碳水化合物结合、肌动活动等分子功能显著富集(封三彩图 3c)。总体上看,小分子分解代谢过程、碳水化合物运输、离子跨膜转运蛋白活性等功能的改变可能与大肠杆菌以葡萄糖和甘油两种不同碳源作为生长底物有关。

2.4 KEGG 富集分析结果

将从 GSE2037 和 GSE156143 数据集中筛选出的差异表达基因进行并集后,进行 KEGG 通路富集。封三彩图 4a 展示了差异表达基因在基因排名列表中的位置。KEGG 富集结果如封三彩图 4b 所示,其中圆圈颜色越偏红色表示此 KEGG 通路越富集,圆圈尺寸越大表示此 KEGG 通路富集到的基因数目越多。由封三彩图 4b 可知,不同环境中的微生物代谢和 ABC 转运体两条通路被显著富集。已有研究发现,酿酒酵母(*Saccharomyces cerevisiae*)在葡萄糖、麦芽糖等不同碳源培养条件下蛋白质组学发生改变,KEGG 分析发现有关基因表达差异最大的途径为“次生代谢产物的生物合成”和“不同环境中的微生物代谢”,表明这些代谢的变化是特定碳源生长的直接结果^[25]。因此,不同环境中的微生物代谢和 ABC 转运体功能的改变可能与大肠杆菌以葡萄糖和甘油这两种不同碳源作为生长底物有关。

2.5 蛋白质相互作用关系网络构建结果

对从 GSE2037 和 GSE156143 数据集中筛选出的差异表达基因进行并集后,进一步预测差异表达基因的蛋白相互作用关系网络,其中节点数为 88 个;再将网络度值大于 52 的基因定义为核心基因,最后筛选出核心基因 14 个。图 5 中的圆圈代表节点,圆圈尺寸越大代表网络度值越大,连接到该节点的边数相应越多;标示了基因名称的圆圈即为筛选出的 14 个核心基因。将这些基因按照它们的功能可分为两类:趋化和鞭毛合成。编码趋化相关蛋白的基因共有 6 个,分别为:趋化蛋白(Chemotaxis protein)基因 *cheZ*、趋化反应调节蛋白-蛋白质谷氨酸甲酯酶/蛋白质谷氨酰胺脱酰胺酶(Chemotaxis response regulator protein-protein glutamate methylesterase/protein glutamine deamidase)基因 *cheB*、趋化蛋白甲基转移酶(Chemotaxis protein methyltransferase)基因 *cheR*、趋化蛋白基因 *cheW*、甲基接受趋化蛋白 Tap(Methyl-accepting chemotaxis protein Tap)基因 *tap* 和甲基接受趋化蛋白 Tar(Methyl-accepting chemotaxis protein Tar)基因 *tar*。编码鞭毛合成相关蛋白的基因共有 8 个,分别为:鞭毛生物合成蛋白(Flagellar biosynthesis protein)基因 *flgN*、鞭毛生物合成及钩组装起始((Flagellar biosynthesis, initiation of hook assembly)基因 *flgD*、鞭毛钩丝连接蛋白 1(Flagellar hook-filament junction protein 1)基因 *flgK*、鞭毛钩丝连接蛋白 2(Flagellar hook-filament junction protein 2)基因 *flgL*、鞭毛丝加帽蛋白(Flagellar filament capping protein)基因 *fliD*、FliA 的反 sigma 因子(sigma(28))(Anti-sigma factor for FliA (sigma(28)))基因 *flgM*、鞭毛运动开关蛋白(Flagellar motor switch protein)基因 *fliG* 和 *fliM*。

3 讨论

目前,有关细菌如何控制基因表达以适应营养环境变化的问题仍然为有关学者所重视。根据过去的研究来

看,随着限制生长的营养物质变化,大肠杆菌基因表达谱也会发生相应的变化^[9]。本研究通过分析葡萄糖和甘油这两种不同碳源培养条件下的大肠杆菌基因表达数据来对这一问题进行探讨。利用基因芯片数据集 GSE2037 和高通量测序数据集 GSE156143 中的数据,分别得到有关的差异表达基因,然后对差异表达基因进行并集,对并集的基因进行 GO 功能富集、KEGG 通路富集和蛋白质相互作用关系网络分析。

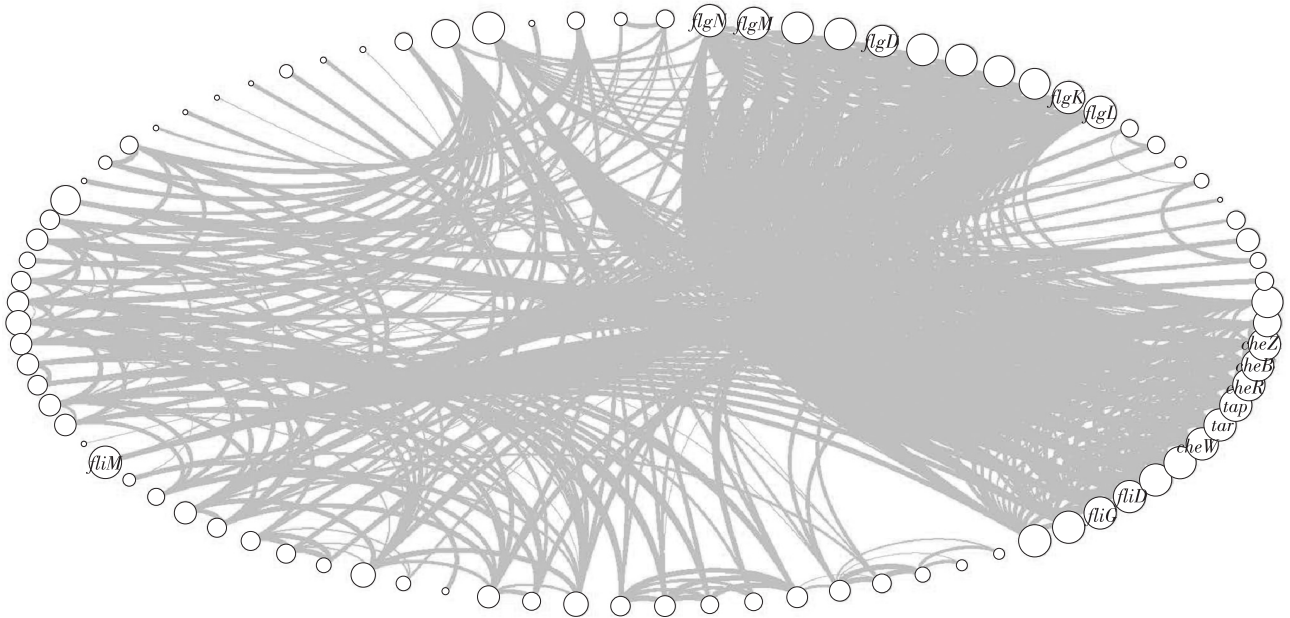


图 5 差异表达基因蛋白互作网络

Fig. 5 Protein-protein interaction network of differentially expressed genes

在总共被调查的 5 731 个基因中,本研究共发现 701 个基因在两种碳源中有差异化表达,其中涉及不同环境中的微生物代谢和 ABC 转运体通路,小分子分解代谢过程、碳水化合物运输、趋化性等生物学过程,细胞器、细菌型鞭毛、甲基接受趋化蛋白复合物等细胞组分以及离子跨膜转运蛋白活性、碳水化合物结合、肌动活动等分子功能。此外,筛选出差异常表达基因中 p 值最小的 20 个基因,其中 1 个基因表达下调而其余 19 个基因表达上调。这个表达显著下调的基因为 *ptsG*,而大肠杆菌控制自身葡萄糖摄取的一个候选机制可能就是葡萄糖转运基因 *ptsG* 的表达^[26]。

通过蛋白质相互作用关系网络的构建,筛选出了 14 个核心基因:*flgN*、*flgM*、*flgD*、*flgK*、*flgL*、*cheZ*、*cheB*、*cheR*、*cheW*、*tap*、*tar*、*fliD*、*fliG* 和 *fliM*。大肠杆菌作为一种运动的革兰氏阴性菌,当分别以葡萄糖和甘油作为碳源时,它的基因表达谱发生了相应变化,功能注释显示 14 个核心基因与鞭毛合成和趋化相关。之前已有研究表明,细菌鞭毛的运动性和趋化性帮助细胞到达最有利的环境,在适应多种碳源的过程中,运动细菌通过增加鞭毛操作、限制鞭毛合成来对碳源作出反应^[27]。本研究结果证实了上述观点,并表明大肠杆菌也是通过调节与鞭毛合成和趋化相关基因的表达来应对碳限制环境变化,以实现最大限度的生存。另外,本研究所确定的这 14 种与趋化和鞭毛合成有关的基因,有助于进一步揭示大肠杆菌在应对不同营养环境变化过程中所采取的分子调控机制。

参考文献:

- [1] JANG J, HUR H G, SADOWSKY M J, et al. Environmental *Escherichia coli*: ecology and public health implications: a review [J]. Journal of Applied Microbiology, 2017, 123(3): 570-581.
- [2] MARTINEZ-ANTONIO A, JANGA S C, THIEFFRY D. Functional organisation of *Escherichia coli* transcriptional regulatory network [J]. Elsevier Sponsored Documents, 2008, 381(1): 238-247.
- [3] FERENCI T. Hungry bacteria: definition and properties of a nutritional state [J]. Environmental Microbiology, 2001, 3(10): 605-611.
- [4] LIU M Z, DURFEE T, CABRERA J E, et al. Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli* [J]. Journal of Biological Chemistry, 2005, 280(16): 15921-15927.

- [5] KANG A, TAN M H, HUA L, et al. Systems-level characterization and engineering of oxidative stress tolerance in *Escherichia coli* under anaerobic conditions[J]. *Molecular Biosystems*, 2013, 9(2): 285-295.
- [6] FOLSOM J P, PARKER A E, CARLSON R P. Physiological and proteomic analysis of *Escherichia coli* iron-limited chemostat growth[J]. *Journal of Bacteriology*, 2014, 196(15): 2748-2761.
- [7] GYANESHWAR P, PALIY O, Mc AULIFFE J, et al. Sulfur and nitrogen limitation in *Escherichia coli* K-12: specific homeostatic responses[J]. *Journal of Bacteriology*, 2005, 187(3): 1074-1090.
- [8] BERTIN Y, CHAUCHEYRAS-DURAND F, ROBBE-MASSELOT C, et al. Carbohydrate utilization by enterohaemorrhagic *Escherichia coli* O157:H7 in bovine intestinal content[J]. *Environmental Microbiology*, 2013, 15(2): 610-622.
- [9] HUA Q, YANG C, OSHIMA T, et al. Analysis of gene expression in *Escherichia coli* in response to changes of growth-limiting nutrient in chemostat cultures[J]. *Applied and Environmental Microbiology*, 2004, 70(4): 2354-2366.
- [10] DEUTSCHER J, FRANCKE C, POSTMA P W. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria[J]. *Microbiology and Molecular Biology Reviews*, 2006, 70(4): 939-1031.
- [11] YOU C H, OKANO H, HUI S, et al. Coordination of bacterial proteome with metabolism by cyclic AMP signalling[J]. *Nature*, 2013, 500(7462): 301-306.
- [12] FRANCHINI A G, IHSEN J, EGLI T. Effect of global regulators RpoS and cyclic-AMP/CRP on the catabolome and transcriptome of *Escherichia coli* K12 during carbon- and energy-limited growth[J]. *PLoS One*, 2015, 10(7): e0133793.
- [13] LI Z J, PAN Q, XIAO Y Z, et al. Deciphering global gene expression and regulation strategy in *Escherichia coli* during carbon limitation[J]. *Microbial Biotechnology*, 2019, 12(2): 360-376.
- [14] VICHÍ J, SALAZAR E, JACINTO V J, et al. High-throughput transcriptome sequencing and comparative analysis of *Escherichia coli* and *Schizosaccharomyces pombe* in respiratory and fermentative growth[J]. *PLoS One*, 2021, 16(3): e0248513.
- [15] MARTINEZ-GÓMEZ K, FLORES N, CASTAÑEDA H M, et al. New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol[EB/OL]. (2012-07-04) [2021-10-25]. <https://link.springer.com/content/pdf/10.1186%2F1475-2859-11-46.pdf>.
- [16] LÉ S, JOSSE J, HUSSON F. FactoMineR: an R package for multivariate analysis[J]. *Journal of Statistical Software*, 2008, 25(1): 1-18.
- [17] RITCHIE M E, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. *Nucleic Acids Research*, 2015, 43(7): e47.
- [18] ROBINSON M D, Mc CARTHY D J, SMYTH G K. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data[J]. *Bioinformatics*, 2010, 26(1): 139-140.
- [19] YAN L L. Ggvenn: draw venn diagram by 'ggplot2'[EB/OL]. (2021-06-29) [2021-10-25]. <https://cran.r-project.org/web/packages/ggvenn/index.html>.
- [20] ASHBURNER M, BALL C A, BLAKE J A, et al. Gene ontology: tool for the unification of biology. the gene ontology consortium[J]. *Nature Genetics*, 2000, 25(1): 25-29.
- [21] KANEHISA M, GOTO S. KEGG: kyoto encyclopedia of genes and genomes[J]. *Nucleic Acids Research*, 2000, 28(1): 27-30.
- [22] YU G C, WANG L G, HAN Y Y, et al. ClusterProfiler: an R package for comparing biological themes among gene clusters[J]. *Omics: a Journal of Integrative Biology*, 2012, 16(5): 284-287.
- [23] SUBRAMANIAN A, TAMAYO P, MOOTHA V K, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(43): 15545-15550.
- [24] SZKLARCZYK D, FRANCESCHINI A, WYDER S, et al. STRINGv10: protein-protein interaction networks, integrated over the tree of life[J]. *Nucleic Acids Research*, 2015, 43(Database Issue): D447-D452.
- [25] PAULO J A, O'CONNELL J D, EVERLEY R A, et al. Quantitative mass spectrometry-based multiplexing compares the abundance of 5000 *S. cerevisiae* proteins across 10 carbon sources[J]. *Journal of Proteomics*, 2016, 148: 85-93.
- [26] CHUBUKOV V, SAUER U. Environmental dependence of stationary-phase metabolism in *Bacillus subtilis* and *Escherichia coli*[J]. *Applied and Environmental Microbiology*, 2014, 80(9): 2901-2909.
- [27] ZHAO K, LIU M Z, BURGESS R R. Adaptation in bacterial flagellar and motility systems: from regulon members to 'foraging'-like behavior in *E. coli*[J]. *Nucleic Acids Research*, 2007, 35(13): 4441-4452.

Gene Expression Profiling Analysis of *Escherichia coli* under the Conditions of Glucose and Glycerol Carbon Source Culture

ZHU Mengli, ZHANG Yujuan

(Chongqing Key Laboratory of Vector Insects, Institute of Insects and Molecular Biology,

Chongqing Normal University, Chongqing 401331, China)

Abstract: [Purposes] To investigate the effects of glucose and glycerol carbon sources on the gene expression profiles of *Escherichia coli*. [Methods] The gene chip dataset GSE2037 and the high-throughput sequencing dataset GSE156143 were downloaded from the gene expression database GEO, and the differentially expressed genes were screened for the glucose and glycerol carbon source samples in the two datasets, respectively. After merging, GO (gene ontology) functional enrichment analysis, KEGG (kyoto encyclopedia of genes and genomes) pathway enrichment analysis, and protein-protein interaction networks (PPI) were performed. [Findings] 701 differentially expressed genes were screened out. The differentially expressed genes are mainly enriched in the biological processes related to small molecule catabolic process, carbohydrate transport, chemotaxis, cellular component related to organelle, bacterial-type flagellum, methyl accepting chemotaxis protein complex, molecular function related to ion transmembrane transporter activity, carbohydrate binding, motor activity. And microbial metabolism in diverse environments, ABC transporters pathways. In addition, 14 core genes were screened out, which played roles in flagella synthesis and chemotaxis. [Conclusions] *E. coli* has different gene expression patterns under the conditions of glucose and glycerol carbon sources. The identified 14 genes related to chemotaxis and flagella synthesis are helpful to further reveal the molecular regulatory mechanism adopted by *E. coli* in response to different nutrient environment changes.

Keywords: *Escherichia coli*; glucose; glycerol; gene expression profiling; bioinformatics analysis

(责任编辑 方 兴)

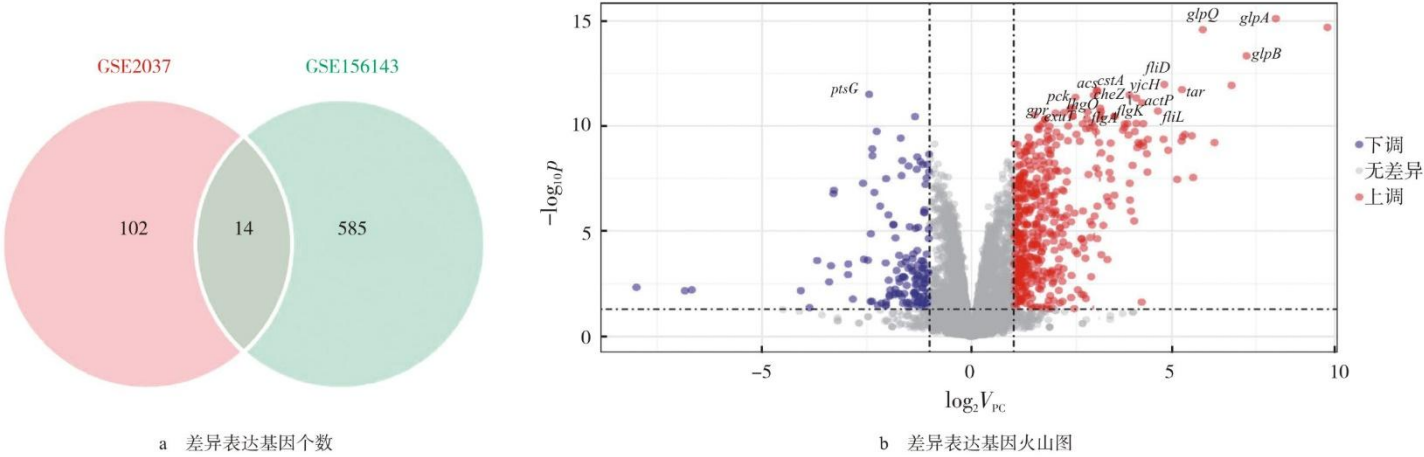


图2 GSE2037和GSE156143中碳源甘油和葡萄糖样品的差异表达基因个数及差异表达基因火山图
Fig. 2 The number of differentially expressed genes of samples cultured by glycerol and glucose in GSE2037 and GSE156143 and volcano map of differentially expressed genes

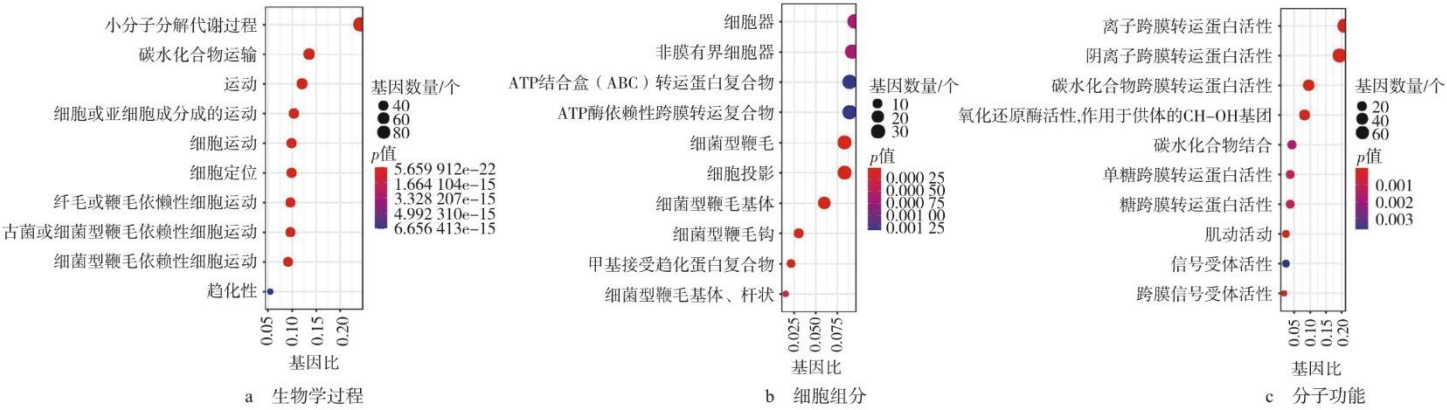


图3 GO 功能注释
Fig. 3 GO functional annotation

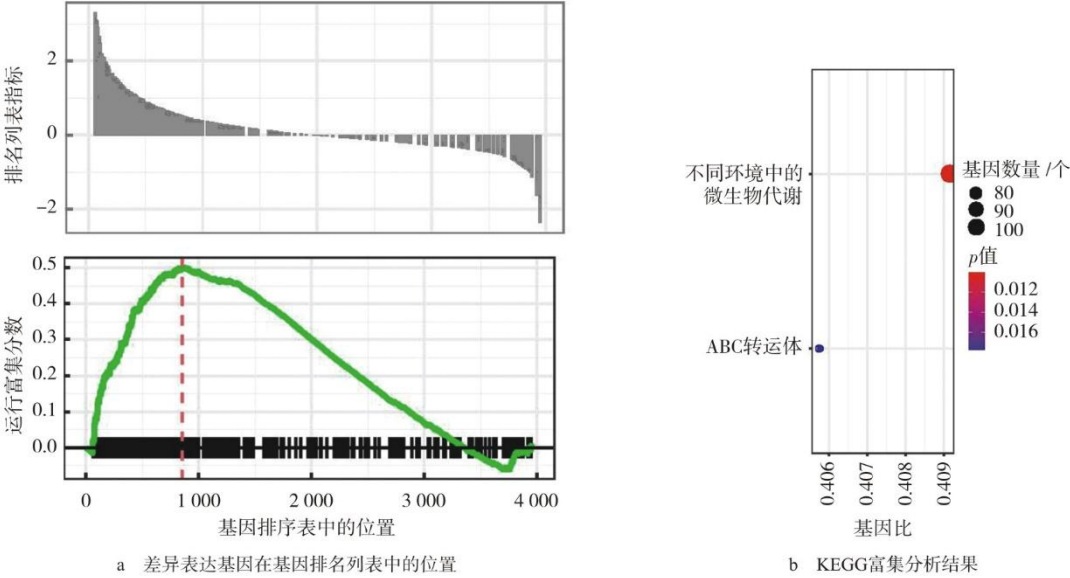


图4 KEGG通路富集分析
Fig. 4 KEGG pathway enrichment analysis