

阮先乐. 基于 GEO 数据库分析水稻低温胁迫关键基因[J]. 江苏农业科学, 2024, 52(3): 61–66.

doi:10.15889/j.issn.1002-1302.2024.03.009

基于 GEO 数据库分析水稻低温胁迫关键基因

阮先乐

(周口师范学院生命科学与农学院, 河南周口 466001)

摘要:为了筛选水稻在低温胁迫下的关键基因,从 GEO 数据库下载水稻 4 个数据集中的 70 个样本。利用在线分析程序 GEO2R 进行共同差异表达基因分析,并对这些差异表达基因进行 GO、KEGG 分析,构建蛋白质互作网络,对关键基因构建热图。结果表明,获得共同差异表达基因 51 个,其中上调表达基因 1 个,下调表达基因 50 个。上述基因的 GO 分析结果表明,其细胞组成主要集中在细胞、细胞要素和细胞器上;在分子功能上,上述基因的功能主要集中在结合、催化活性上;在生物过程中,上述基因的功能主要集中在细胞过程、代谢过程和生物调控上。KEGG 信号通路分析结果表明,上述基因主要参与植物激素信号转导等通路。在构建的共同差异表达基因的蛋白质网络中,有 29 个节点。另外,得到 10 个关键基因、2 个关键子网络。研究结果为进一步研究水稻低温胁迫关键基因奠定了基础,也有利于水稻低温育种。

关键词:水稻;GEO 数据库;低温胁迫;共同差异表达基因;GO 功能分析;KEGG 信号通路分析

中图分类号:S511.01;S126 **文献标志码:**A **文章编号:**1002-1302(2024)03-0061-06

水稻(*Oryza sativa* L.)起源于热带与亚热带,是低温敏感型作物。低温严重影响了水稻的产量和品质,也限制了水稻向高海拔、高纬度地区扩展^[1]。从全球范围来看,目前有 24 个国家约 1 500 万 hm^2 的水稻受到低温影响,在亚洲南部、东南部,约 700 万 hm^2 的土地由于受到低温影响而无法种植水稻^[2]。在我国的东北地区,由于纬度较高、温度偏低,每 3~5 年就会发生 1 次冷害^[3]。因此,培育耐冷水稻品种具有十分重要的现实意义。

植物的耐冷性是多基因效应的,其机理复杂,而利用传统育种方式在改良植物耐冷性方面有限制。如果采用基因工程技术进行作物育种,是提高植物耐冷性比较有效的途径^[4]。水稻与低温胁迫相关的基因可以分 2 类:第 1 类是功能基因,其编码产物在水稻受到低温胁迫时直接起到保护作用;第 2 类是调控基因表达的信号因子^[5]。闫凌月等研究发现,*OsMADS25* 通过提高水稻在低温胁迫下对活性氧的清除能力,进而提高其对低温的耐受性^[6]。Hur 等研究发现,水稻中合成脯氨酸的关键基因 *OsP5CS2* 对提高水稻的耐冷性十分重要^[7]。Li 等研究发现,*OsTPS1* 基因的超表达能够提高转基因水

稻对于干旱、盐和低温胁迫的耐受性^[8]。陈能刚等研究发现,在孕穗期、开花期,转异戊烯基转移酶基因(isopentenyl transferase *ipt*)水稻的相对电导率的变化量和叶绿体受低温的危害较小^[9]。张明星构建的 *OsWRKY63* 水稻突变体具有较强的耐冷性,为水稻耐冷品种的选育提供了潜在基因资源和中间育种材料^[10]。Wang 等研究发现,含有 *OsDREB1F* 基因的转基因水稻对盐、干旱和低温的耐受性有所提高^[11]。另有研究也发现,转录因子 MYB、SNAC、TCP、PHD、bZIP 与水稻的耐冷性有关^[12-16]。

GEO 数据库(<https://www.ncbi.nlm.nih.gov/geo/>)是一个公共功能基因组数据存储库,目前储存了 100 多个物种的约十亿个基因表达数据,科研人员借助基于 WEB 的工具,有效地探索、查询和下载这些海量数据,从而更好地分析和设计自己的试验。本研究采用生物信息学技术分析 GEO 数据库中与水稻低温胁迫相关的一些数据,筛选关键差异基因及其功能、信号通路,以期研究低温胁迫对水稻生长发育影响的分子机制,从而为今后水稻耐冷性育种奠定良好基础。

1 材料与方法

1.1 基因芯片数据的获取与筛选

水稻低温胁迫基因表达谱芯片数据来源于 NCBI(<https://www.ncbi.nlm.nih.gov/>)的 GEO 数

收稿日期:2023-04-10

作者简介:阮先乐(1977—),男,河南淮阳人,硕士,讲师,主要从事植物育种和生物信息学研究。E-mail:ruanxianle@126.com。

数据库。基于对获得的初始数据的分析与试验要求，采用 GSE6901、GSE31874、GSE37940 和 GSE71680 共 4 个数据集，选择其中 70 个样本进行数据分析

(表 1)。试验分析时间是 2022 年 10—12 月，地点在周口师范学院。

表 1 试验数据采集与分组

数据集	对照组	处理组
GSE6901	GSM159259、GSM159260、GSM159261	GSM159268、GSM159269、GSM159270
GSE31874	GSM790485、GSM790486	GSM790491、GSM790492
GSE37940	GSM930404、GSM930405、GSM930406、GSM930422、GSM930423、GSM930424	GSM930389、GSM930395、GSM930396、GSM930390、GSM930397、GSM930398、GSM930391、GSM930399、GSM930400、GSM930392、GSM930401、GSM930402、GSM930393、GSM930403、GSM930407、GSM930394、GSM930408、GSM930409、GSM930410、GSM930411、GSM930412、GSM930413、GSM930414、GSM930415、GSM930416、GSM930417、GSM930418、GSM930419、GSM930420、GSM930421
GSE71680	GSM1843159、GSM1843162、GSM1843165、GSM1843168、GSM1843171、GSM1843174、GSM1843179、GSM1843182	GSM1843160、GSM1843161、GSM1843163、GSM1843164、GSM1843166、GSM1843167、GSM1843169、GSM1843170、GSM1843172、GSM1843173、GSM1843175、GSM1843176、GSM1843177、GSM1843178、GSM1843180、GSM1843181

1.2 方法

1.2.1 共同差异表达基因的筛选 首先利用 GEO 数据库中的在线分析程序 GEO2R 对不同数据集的处理组、对照组进行分析，并以 $P < 0.05$ 、 $|\log_2 FC| \geq 2$ 作为差异表达基因筛选的标准。然后，利用 Evenn (<http://www.ehbio.com/test/venn/#/>) 获得 4 个数据集的共同差异表达基因。

1.2.2 共同差异表达基因的 GO 功能分析和 KEGG 信号通路分析 首先在 UniProt 数据库 (<https://www.uniprot.org>) 中查询各个共同差异表达基因的 GO 号，然后利用 WEGO 2.0 (<https://wego.genomics.cn/>) 进行共同差异表达基因的 GO 功能分析。KEGG 信号通路分析利用 DAVID (<https://david.ncifcrf.gov/home.jsp>)、KOBAS (<http://kobas.chi.pku.edu.cn/>) 2 个在线网站。

1.2.3 蛋白互作网络的构建 用 STRING (<https://cn.string-db.org/>) 和 Cytoscape 3.9.1 软件构建共同差异表达基因的蛋白质互作网络关系图，用 Cytoscape 3.9.1 Cytohubba 插件筛选关键基因，用 Cytoscape 3.9.1 MCODE 筛选关键子网络。

2 结果与分析

2.1 共同差异表达基因的筛选结果

利用 4 个数据集的 70 个样本进行研究分析，按照规定的筛选标准，从 GSE6901 中获得 1 107 个基因，从 GSE31874 中获得 2 537 个基因，从 GSE71680 中获得 184 个基因，从 GSE37940 中获得 498 个基因。之后，经 Evenn 在线分析，获得 51 个共同差异

表达基因(图 1)，其中上调表达基因 1 个，是 AKI08642，下调表达基因 50 个(表 2)。

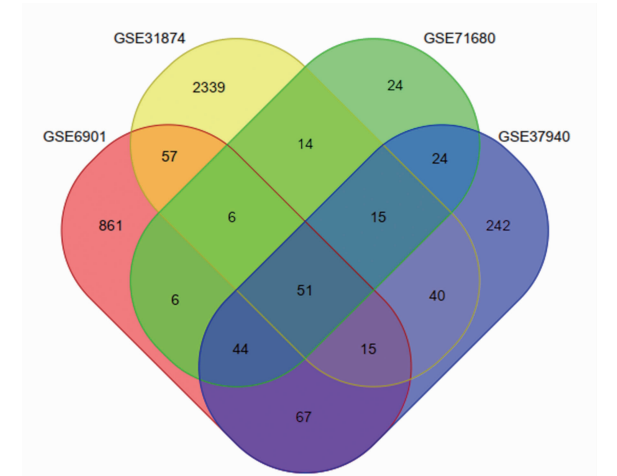


图 1 共同差异表达基因的维恩图

2.2 共同差异表达基因的 GO 功能分析和 KEGG 信号通路分析

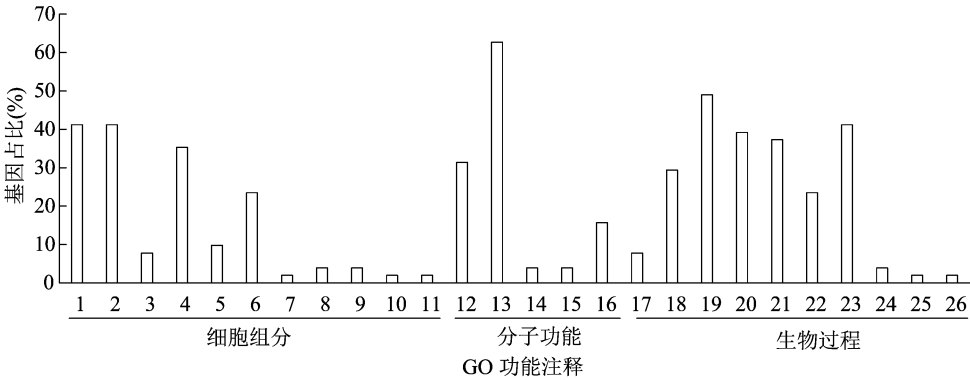
对共同差异表达基因的 GO 分析结果(图 2)表明，细胞组成主要集中在细胞、细胞要素、细胞器上，分别达到 41.2%、41.2%、35.3%。在分子功能上，这些基因的功能主要集中在结合、催化活性上，分别达到 62.7%、31.4%。在生物过程中，这些基因的功能主要集中在细胞过程、代谢过程、生物调控上，分别达到 49.0%、41.2%、39.2%。KEGG 信号通路分析结果(表 3、图 3)表明，这些差异表达基因主要参与植物激素信号转导、植物病原相互作用、类胡萝卜素生物合成、脂肪酸延伸、RNA 降解和次生代谢产物的生物合成等通路中。

表 2 共同差异表达基因信息

编号	基因号	log ₂ FC	蛋白质编号	蛋白质性质
AB045120	8698832	-2.591 97	EL5.1	泛素蛋白连接酶
AK058518	32968536	-2.882 22	XP_025876068.1	丝裂原激活蛋白激酶
AK061011	32971029	-2.229 31	XP_015614860.1	含 CBS 结构域蛋白质
AK061038	32971056	-3.133 64	XP_015614650.1	菱形蛋白质
AK061293	32971311	-4.271 88	OsJ_10535	不清楚
AK062671	32972689	-2.694 11	CML31	钙结合蛋白
AK062882	32972900	-4.083 68	XP_015649367.1	乙烯响应转录因子
AK063605	32973623	-3.304 40	XP_015623373.1	EXORDIUM 蛋白质
AK064287	32974305	-4.539 32	OS12T0150200-01	细胞色素 P450
AK064767	32974785	-2.332 42	XP_015630799.1	nst1 蛋白
AK067007	32977025	-3.159 12	XP_015627713.1	脱落酸 8' 羟化酶
AK067195	32977213	-5.530 76	OS09T0457900-01	乙烯响应转录因子
AK067214	32977232	-2.606 19	CML15	钙结合蛋白
AK067907	32977925	-2.558 63	OS02T0205500-01	KCS 酶
AK068115	32978133	-2.899 82	P0505D12.1	YLS9 蛋白
AK069247	32979271	-2.698 76	XP_015630503.1	不清楚
AK069398	32979422	-2.853 85	rf1b	Rfl 蛋白
AK070609	32980633	-2.671 18	XP_015624439.1	EXO70A1 蛋白
AK070649	32980673	-3.334 13	TIFY11A	TIFY 蛋白
AK070731	32980754	-2.777 66	XP_015631125.1	AAA-ATPase
AK071274	32981297	-2.529 25	OsJ_04163	NAC 结构蛋白
AK071585	32981608	-3.701 46	ACH99697.1	NPKL3 蛋白
AK071689	32981712	-2.842 38	XP_015610640.1	不清楚
AK099776	32984985	-2.707 10	XP_015638006.1	HSPRO1 蛋白
AK101209	32986418	-3.387 50	XP_015636284.1	不清楚
AK101655	32986864	-3.555 29	XP_015623938.1	EXORDIUM 蛋白
AK102357	32987566	-3.281 31	XP_015628852.1	重金属相关异戊二烯化植物蛋白
AK103556	32988765	-2.428 01	XP_015649506.1	不清楚
AK105196	32990405	-4.224 38	XP_015639022.1	丝裂原激活蛋白激酶
AK105946	32991155	-4.241 99	XP_015638678.1	丝裂原激活蛋白激酶
AK106041	32991250	-3.100 69	DREB1G	DREB 转录因子
AK106404	32991613	-4.929 72	XP_015615638.1	细胞色素 P450 94C1
AK106901	32992110	-3.395 07	ABA95756.1	GRAS 转录因子
AK107146	32992355	-3.036 62	XP_015624274.1	乙烯响应转录因子
AK107168	32992377	-2.345 44	XP_015639001.1	丝裂原激活蛋白激酶
AK107854	32993063	-4.168 95	TIFY11E	TIFY 蛋白
AK108001	32993210	-3.349 14	OS03T0734500-01	不清楚
AK108331	32993540	-3.889 19	XP_015624026.2	不清楚
AK108642	32993851	2.557 39	XP_015613474.1	不清楚
AK109037	32994246	-3.714 91	XP_015613068.1	钙离子转运 ATP 酶
AK109382	42821697	-2.688 26	OsJ_15331	还原酶
AK111266	32996475	-2.389 10	XP_015624276.1	不清楚
AK111442	32996651	-2.229 45	OsJ_00269	SARD1 蛋白
AK111712	37988375	-3.154 76	XP_015635906.1	CCR4 因子
AK111775	37988438	-3.147 41	OS01T0313300-01	乙烯响应转录因子

表 2(续)

编号	基因号	log ₂ FC	蛋白质编号	蛋白质性质
AK120087	37989710	-3.761 05	XP_015633147.1	TIFY 蛋白
AK120226	37989849	-2.704 71	XP_015628606.1	DLC 蛋白
AY327040	32966056	-3.474 73	XP_015642451.1	DREB 转录因子
AY342322	33694267	-2.779 41	AAQ24872.1	单糖转运蛋白
AY345234	33637699	-3.097 33	NP_001359120.1	DREB 转录因子
D64038	1841319	-2.630 85	EL2	EL2 转录因子



1—细胞; 2—细胞要素; 3—细胞器要素; 4—细胞器; 5—蛋白复合物; 6—细胞膜; 7—超分子复合物; 8—细胞间区域; 9—细胞间区域部分; 10—共质体; 11—细胞连接; 12—催化活性; 13—结合; 14—转运因子活性; 15—分子功能调节因子; 16—转录调节活性; 17—定位; 18—刺激应答; 19—细胞过程; 20—生物调控; 21—生物调控过程; 22—信号传递; 23—代谢过程; 24—生物过程的负调控; 25—发育过程; 26—多细胞生物过程

图2 共同差异表达基因的 GO 功能注释

表 3 共同差异表达基因参与的 KEGG 通路

条目	KEGG 编号	<i>P</i> 值	基因编号
植物激素信号传导	osa04075	0.006 249	AK107854 AK120087 AK070649
植物病原互作	osa04626	0.028 305	AK062671 AK067214
类胡萝卜素生物合成	osa00906	0.045 196	AK067007
脂肪酸延伸	osa00062	0.061 262	AK067907
RNA 降解	osa03018	0.160 072	AK111712
次生代谢产物的生物合成	osa01110	0.842 826	AK067907
代谢途径	osa01100	0.974 441	AK067907

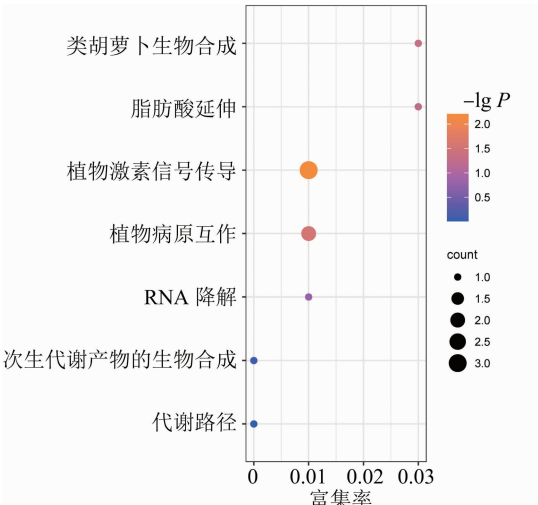


图3 KEGG 通路分析

2.3 共同差异表达基因的蛋白质网络构建与关键基因的筛选

利用 STRING 数据库与 Cytoscape 软件对共同差异表达基因构建蛋白质网络,最终得到 29 个节点(图 4-a)。另外,得到 10 个关键基因(图 4-b)、2 个关键子网络(图 4-c、图 4-d)。这 10 个关键基因的热图见图 5。

3 讨论和结论

本研究基于 GPL2025、GPL7344 等 2 个平台上的 4 个数据集的 70 个样本进行基因表达数据的分析。最终筛选出共同差异表达基因 51 个,其中上调表达的基因只有 1 个,下调表达的基因 50 个,并用 Cytoscape

低温和干旱诱导了 EL2 mRNA 的表达^[28]。

GO 功能分析结果也表明,这些共同差异基因的功能主要集中在细胞过程、代谢过程和生物调控上。利用 Cytoscape 3.9.1 – MCODE 筛选得到 2 个关键子网络,这 2 个关键子网络所涉及的 7 个基因都属于重点关注的 10 个关键基因。结合以上分析,在水稻的耐冷育种中,可以重点考虑上面提到的 10 个关键基因。本研究为水稻的耐冷育种提供了一定的理论依据,奠定了良好的基础。

参考文献:

- [1] 吉凌霄. *OsTILI* 和 *OsCBP60* 调控水稻苗期耐冷性信号转导机制的研究[D]. 武汉:华中农业大学,2022:1–2.
- [2] Pradhan S K, Pandit E, Nayak D K, et al. Genes, pathways and transcription factors involved in seedling stage chilling stress tolerance in indica rice through RNA – Seq analysis[J]. BMC Plant Biology, 2019,19(1):352.
- [3] 王海涛. CHS1 调控水稻耐冷性的机制研究[D]. 长春:吉林大学,2022:1–5.
- [4] 段俊枝,李莹,周雷,等. 利用基因工程技术提高水稻耐冷性的研究进展[J]. 浙江农业学报,2015,27(4):705–712.
- [5] 徐青山,黄晶,孙爱军,等. 低温影响水稻发育机理及调控途径研究进展[J]. 中国水稻科学,2022,36(2):118–130.
- [6] 闫凌月,张豪健,郑雨晴,等. 转录因子 *OsMADS25* 提高水稻对低温的耐受性[J]. 遗传,2021,43(11):1078–1087.
- [7] Hur J H, Jung K H, Lee C H, et al. Stress – inducible *OsP5CS2* gene is essential for salt and cold tolerance in rice[J]. Plant Science, 2004,167(3):417–426.
- [8] Li H W, Zang B S, Deng X W, et al. Overexpression of the trehalose – 6 – phosphate synthase gene *OsTPSI* enhances abiotic stress tolerance in rice[J]. Planta,2011,234(5):1007–1018.
- [9] 陈能刚,余显权,赵德刚,等. 转 *ipt* 基因水稻植株耐冷性研究[J]. 西南农业学报,2006,19(2):255–259.
- [10] 张明星. *OsWRKY63* 调控水稻耐冷性的分子机制研究[D]. 长春:吉林大学,2022:21–48.
- [11] Wang Q Y, Guan Y C, Wu Y R, et al. Overexpression of a rice *OsDREB1F* gene increases salt, drought, and low temperature tolerance in both *Arabidopsis* and rice[J]. Plant Mol Biol,2008,67(6):589–602.
- [12] Yang A, Dai X Y, Zhang W H. A R2R3 – type MYB gene, *OsMYB2*, is involved in salt, cold, and dehydration tolerance in rice[J]. Journal of Experimental Botany,2012,63(7):2541–2556.
- [13] Hu H H, You J, Fang Y J, et al. Characterization of transcription

- factor gene *SNAC2* conferring cold and salt tolerance in rice[J]. Plant Mol Biol,2008,67:169–181.
- [14] 王孙婷. 水稻 *OsPCF6* 与 *OsTCP21* 基因冷胁迫应答功能研究[D]. 哈尔滨:东北农业大学,2014:37–57.
- [15] 马奔. 水稻 PHD – finger 家族低温诱导基因的鉴定与启动子功能分析[D]. 合肥:安徽农业大学,2014:30–34.
- [16] 关可兴. 水稻耐冷性相关转录因子 *OsbZIP32* 的功能分析[D]. 长春:吉林大学,2015:1–9.
- [17] 张明菊,朱莉,夏启中. 植物激素对胁迫反应调控的研究进展[J]. 湖北大学学报(自然科学版),2021,43(3):242–253.
- [18] Daie J, Campbell W F. Response of tomato plants to stressful temperatures increase on abscisic acid concentrations[J]. Plant Physiol,1981,67(1):26–29.
- [19] Eze J M O, Dumbroff E B, Thompson J E. Effects of temperature and moisture stress on the accumulation of abscisic acid in bean[J]. Physiologia Plantarum,1983,58(2):179–183.
- [20] Lalk I, Dörffling K. Hardening, abscisic acid, proline and freezing resistance in two winter wheat varieties[J]. Physiologia Plantarum, 1985,63(3):287–292.
- [21] Flores A, Grau A, Laurich F, et al. Effect of new terpenoid analogues of abscisic acid on chilling and freezing resistance[J]. Journal of Plant Physiology,1988,132(3):362–369.
- [22] Mohapatra S S, Poole R J, Dhindsa R S. Absciscic acid – regulated gene expression in relation to freezing tolerance in alfalfa[J]. Plant Physiology,1988,87(2):468–473.
- [23] Ye H Y, Du H, Tang N, et al. Identification and expression profiling analysis of TIFY family genes involved in stress and phytohormone responses in rice[J]. Plant Mol Biol,2009,71(3):291–305.
- [24] 徐佳宁. 苹果 UPL 基因家族及转录组和蛋白质组响应非生物胁迫的研究[D]. 泰安:山东农业大学,2017:42–94.
- [25] 陈文烨. 小麦钙调素/类钙调素基因的特征分析、克隆及功能研究[D]. 石家庄:河北科技大学,2019:1–2.
- [26] Hu X J, Zhang Z B, Fu Z Y, et al. Significance of a β – ketoacyl – CoA synthase gene expression for wheat tolerance to adverse environments[J]. Biologia Plantarum,2010,54(3):575–578.
- [27] Divi U K, El – Tahchy A, Vanhercke T, et al. Transcriptional and biochemical responses of monoacylglycerol acyltransferase – mediated oil synthesis and associated senescence – like responses in *Nicotiana benthamiana* [J]. Frontiers in Plant Science,2014,5:204–216.
- [28] Peres A, Churchman M L, Hariharan S, et al. Novel plant – specific cyclin – dependent kinase inhibitors induced by biotic and abiotic stresses[J]. Journal of Biological Chemistry, 2007, 282 (35): 25588–25596.