

辣椒 *FAR1* 基因家族的全基因组分析

夏承东, 方大伟 (安徽徽大农业有限公司, 安徽合肥 230088)

摘要 *FAR-RED IMPAIRED RESPONSE 1 (FAR1)* 基因在植物生长发育中起着重要的调控作用。该研究采用生物信息学的方法, 从辣椒基因组中鉴定得到 20 个 *CaFAR1* 转录因子家族成员, 系统进化分析表明, 20 个 *CaFAR1* 基因分为 6 个亚家族, 其中 18 个 *CaFAR1* 基因分布在 8 条染色体上, 8 号染色体上有 3 个, 2 号、6 号、7 号、12 号染色体各分布 1 个, 6 号和 10 号染色体上各分布 2 个, 3 号染色体上有 6 个。GO 功能注释表明, 20 个 *CaFAR1* 蛋白序列基于氨基酸的相似性被分为三大类: 生物过程 (BP)、细胞成分 (CC) 和分子功能 (MF)。

关键词 辣椒; *FAR* 基因家族; 生物信息学; 全基因组分析

中图分类号 Q943.2 文献标识码 A

文章编号 0517-6611(2020)14-0097-06

doi: 10.3969/j.issn.0517-6611.2020.14.027



开放科学(资源服务)标识码(OSID):

Analysis of *FAR1* Gene Family from Whole Genome of Pepper

XIA Cheng-dong, FANG Da-wei (Anhui Huida Agricultural Co., Ltd., Hefei, Anhui 230088)

Abstract *FAR-RED IMPAIRED RESPONSE 1 (FAR1)* gene plays an important regulatory role in plant growth and development. In this study, bioinformatics was used to identify 20 *CaFAR1* transcription factor family members from the pepper genome. Phylogenetic analysis showed that 20 *CaFAR1* genes were divided into 6 subfamilies. Of all 20 *CaFAR1* genes, 18 genes were mapped to the 8 chromosomes. Three genes were mapped to chromosome 3, four genes were mapped to chromosome 2, 6, 7 and 12 with one for each. Two genes were mapped to the chromosome 6 and 10. There are 6 genes were mapped to the chromosome 3. GO analysis showed that the functions of 20 *CaFAR1* genes were enriched in biological process, cell component and molecular function.

Key words Pepper; *FAR* gene family; Bioinformatics; Whole genome analysis

FAR-RED IMPAIRED RESPONSE 1 (FAR1) 是转座酶衍生的 一类转录因子, 是能够在远红光下直接激活 *FHY1/FHL* 的基因表达, 其在植物淀粉合成代谢和碳饥饿引发的能量匮乏过程中起着重要调控作用^[1-2]。目前, 已在拟南芥中分离得到了 14 个 *FAR1* 家族成员被鉴定, 编码的蛋白长度为 531~851 个氨基酸^[3]。而在棉花中鉴定出 88 个 *FAR1* 基因, 其中亚洲棉 27 个、雷蒙氏棉 35 个、陆地棉 26 个, 编码的蛋白长度为 158~1 134 个氨基酸^[4]。辣椒是我国一种重要的蔬菜作物, 光照时间、光照强度以及不同光质对辣椒产量和品质都有很大的影响。然而关于辣椒光敏色素调控相关基因的报道鲜见。该研究利用生物信息学方法, 筛选、鉴定辣椒 *FAR1* 基因家族成员, 通过分析其系统进化关系、基因结构, 以及染色体定位, 探索辣椒 *FAR1* 基因功能及其应用。

1 材料与方法

1.1 植物序列来源 辣椒 (*Capsicum annuum* L.) *FAR1* 转录因子家族数据来源于 NCBI (<https://www.ncbi.nlm.nih.gov/genome/?term=Capsicum+annuum>) 与 PlantTFDB (<http://planttfdb.cbi.pku.edu.cn>) 等数据库。拟南芥 (*Arabidopsis thaliana*) 与番茄 (*Solanum lycopersicum*) *FAR1* 转录因子氨基酸序列来源于 PlantTFDB 数据库。

1.2 辣椒 *FAR1* 转录因子家族基因的筛选 从 TAIR (<https://www.arabidopsis.org>) 数据库中下载拟南芥 *FAR1* 转录因子序列, 利用 BlastP 筛选出辣椒的 *FAR1* 转录因子, 得到候选 *CaFAR1* 蛋白序列。将这些序列导入 CD-HIT Suite (http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi) 在线网站中去除冗余序列, 再经过 SMART (<http://smart.embl-heidelberg>).

[de/opennewwindow](#)) 在线预测, 得到具有 *FAR1* 保守结构域的蛋白序列。参考 Lin 等^[5] 的分类方法, 将辣椒 *FAR1* 蛋白序列归为 6 类。

1.3 辣椒 *FAR1* 转录因子家族理化分析与进化树构建 利用在线 ProtParam 网站 (<https://web.expasy.org/protparam/>) 对鉴定出的辣椒 *FAR1* 蛋白进行理化性质分析。使用 MEGA 7.0 软件内置的 Clustal W 程序对鉴定出的 *FAR1* 类蛋白序列和拟南芥 *FAR1* 家族、番茄 *FAR1* 家族蛋白序列进行序列比对, 采用邻接 (neighbor-joining, N) 法构建系统发育树, 选用 Poisson 模型并设置 Bootstrap value 为 1 000, 缺失值处理方式 为配对删除 (pairwise deletion), 其他参数使用默认值^[5]。

1.4 辣椒 *FAR1* 转录因子家族氨基酸序列比对、保守结构域分析 将辣椒 *FAR1* 转录因子家族的氨基酸序列上传至 DNAMAN8.0 中进行比对, 得到氨基酸序列比对结果。利用 MEME 在线软件 (<http://meme.suit.org>) 对辣椒 *FAR1* 转录因子保守元件进行预测分析, 将 motif 数量设为 10。对辣椒 *FAR1* 转录因子保守元件进行图形化显示。

1.5 辣椒 *FAR1* 转录因子家族基因结构与染色体定位分析 利用本地软件 UltraEdit 得到辣椒 *FAR1* 转录因子家族基因结构信息与染色体信息, 将辣椒 *FAR1* 转录因子家族基因结构信息上传至在线软件 GSDS (<http://gsds.cbi.pku.edu.cn/>) 中得到基因结构示意图。用在线软件 MC2C (http://mg2c.iask.in/mg2c_v2.0/) 对辣椒 *FAR1* 转录因子家族的染色体信息进行图形化显示。

1.6 辣椒 *FAR1* 转录因子家族基因 GO 注释分析 使用 Blast2GO 的非冗余蛋白质数据库软件与默认参数, 首先将 *CaFAR1* 氨基酸序列与 NCBI 进行 BlastP 比对。然后对每一个 *CaFAR1* 都获取 GO 注释, 得到 GO 注释的 ID, 导出数据之后使用 GraphPad 作图。

基金项目 安徽省科技重大专项 (17030701002)。

作者简介 夏承东 (1978—), 男, 安徽肥东人, 从事蔬菜品种选育工作。

收稿日期 2020-05-27; 修回日期 2020-06-02

2 结果与分析

2.1 CaFAR1 转录因子家族的鉴定及特征 利用本地 Blast 软件对辣椒全基因组蛋白序列进行检索,同时手动剔除重复序列,只保留一个 FAR1 结构域的蛋白序列。最后得到 20 个只具有一个 FAR1 结构域的转录因子家族成员^[6],利用 ExPASy(<https://web.expasy.org/cgi-bin/protparam/protparam>)在线软件对 CaFAR1 转录因子家族进行分析,得到 CaFAR1 转录因子的氨基酸大小、分子量和等电点等信息(表 1)。辣椒 FAR1 转录因子家族成员蛋白序列含有氨基酸数目为 220

(XP_016567405.1)~877(XP_016580413.1),利用 ExPASy 软件预测 CaFAR1 分子量大小为 25 546.68(XP_016567405.1)~100 799.3 Da (XP_016580413.1),等电点为 5.12(XP_016553207.1)~9.39(XP_016567405.1),除 CaFAR1-14 之外,其余 CaFAR1 蛋白都不稳定。从以上理化性质来分析,辣椒 FAR1 转录因子家族成员之间的各项指标跨度较大,具有差异性,在筛选出的 20 个转录因子家族成员中,有 13 个转录因子的等电点小于 7,占 65%,呈弱酸性。

表 1 CaFAR1 转录因子家族的鉴定及特征
Table 1 Identification and characteristics of CaFAR1 transcription factor family

NCBI 登陆号 Accession	基因编号 Gene No.	氨基酸 AA	分子量 MW//Da	等电点 pI	稳定性 Stability index	亲水性 (GRAVY)
XP_016538323.1	CaFAR1-1	862	97 136.00	6.24	不稳定	-0.354
XP_016538964.1	CaFAR1-2	672	77 063.79	8.26	不稳定	-0.391
XP_016540073.1	CaFAR1-3	771	88 593.40	6.66	不稳定	-0.444
XP_016580413.1	CaFAR1-4	877	100 799.30	6.62	不稳定	-0.617
XP_016543349.1	CaFAR1-5	807	92 459.31	8.43	不稳定	-0.556
XP_016544432.1	CaFAR1-6	386	43 798.68	8.50	不稳定	-0.489
XP_016562564.1	CaFAR1-7	773	88 314.03	6.45	不稳定	-0.477
XP_016549426.1	CaFAR1-8	313	37 205.48	6.76	不稳定	-0.396
XP_016553207.1	CaFAR1-9	572	65 858.07	5.12	不稳定	-0.482
XP_016553844.1	CaFAR1-10	554	63 147.67	6.39	不稳定	-0.421
XP_016558493.1	CaFAR1-11	360	41 760.66	6.72	不稳定	-0.456
XP_016562565.1	CaFAR1-12	773	88 314.03	6.45	不稳定	-0.477
XP_016562968.1	CaFAR1-13	687	79 864.92	7.72	不稳定	-0.435
XP_016567405.1	CaFAR1-14	220	25 546.68	9.39	稳定	-0.171
XP_016578018.1	CaFAR1-15	873	99 427.65	6.14	不稳定	-0.435
XP_016562564.1	CaFAR1-16	773	88 314.03	6.45	不稳定	-0.477
XP_016580466.1	CaFAR1-17	694	80 191.06	6.26	不稳定	-0.390
XP_016565037.1	CaFAR1-18	827	95 047.82	6.64	不稳定	-0.543
XP_016565829.1	CaFAR1-19	837	95 690.33	7.50	不稳定	-0.305
XP_016568196.1	CaFAR1-20	734	83 504.30	7.04	不稳定	-0.446

2.2 CaFAR1 转录因子家族系统进化树 PlantTFDB v5.0(<http://planttfdb.cbi.pku.edu.cn>)在线软件中下载拟南芥与番茄的 FAR1 转录因子家族,利用 MEGA7 系统进化树构建软件构建拟南芥、番茄与辣椒的系统进化树。系统进化树的构建使用最大近似法,MEGA7 参数采用系统默认值,最终得到图 1 所示的进化树。根据 Lin 等^[5]对拟南芥的分类方法对得到的系统进化树进行亚家族分类,共分为 6 个亚家族,命名为 I~VI。

2.3 氨基酸序列比对 使用 DANMAN8.0 序列分析软件对辣椒 FAR1 基因保守结构域进行分析,结果表明,20 个蛋白序列均含有 FAR1 保守结构域(图 2,3)。所有 FRS 亚组的大多数成员都具有 N 末端 C₂H₂ 锌指结构域,N 末端 FAR1 DNA 结合结构域是 WRKY-神经胶质细胞缺失(WRKY-GCM1)超家族中的一种 C₂H₂ 锌指结构域,它与多种靶标的启动子区域中的特定顺式元件结合。

2.4 CaFAR1 保守结构域与基因结构分析 将 CaFAR1 氨

基酸序列上传至在线软件 MEME(<http://meme.suit.org>)对辣椒 FAR1 序列的保守元件进行预测分析,将 motif 数量设为 10,分析结果如图 4B 所示。发现,IV 亚家族中除 CaFAR1-10 与 CaFAR1-20 外均含有 10 个 motif;亚家族 I 中所有 CaFAR1 都含有 10 个 motif,亚家族 III 中缺少 motif 9;亚家族 V 中均含有 motif 3;亚家族 VI 中均含有 motif 1 与 motif 8。

通过辣椒 GFF 注释文件中获得 CaFAR1 的内含子外显子的数据,上传至在线软件 GSDS(<http://gsds.cbi.pku.edu.cn/>)对辣椒 FAR1 基因家族的基因结构进行分析。由图 4C 可知,不同亚组成员的内含子数量存在差异。亚家族 I 中基因内含子较为相似,内含子数量较多,CaFAR1-16 内含子有 10 个,在所有 CaFAR1 中数量最多;亚家族 II 的成员含有 1~3 个内含子;亚家族 III 的 CaFAR1-5 内含子有 1 个,CaFAR1-8 有 2 个内含子;亚家族 IV 成员含有 2~5 个内含子;亚家族 V 成员含有 2~5 个内含子;亚家族 VI 成员含有 2~6 个内含子。

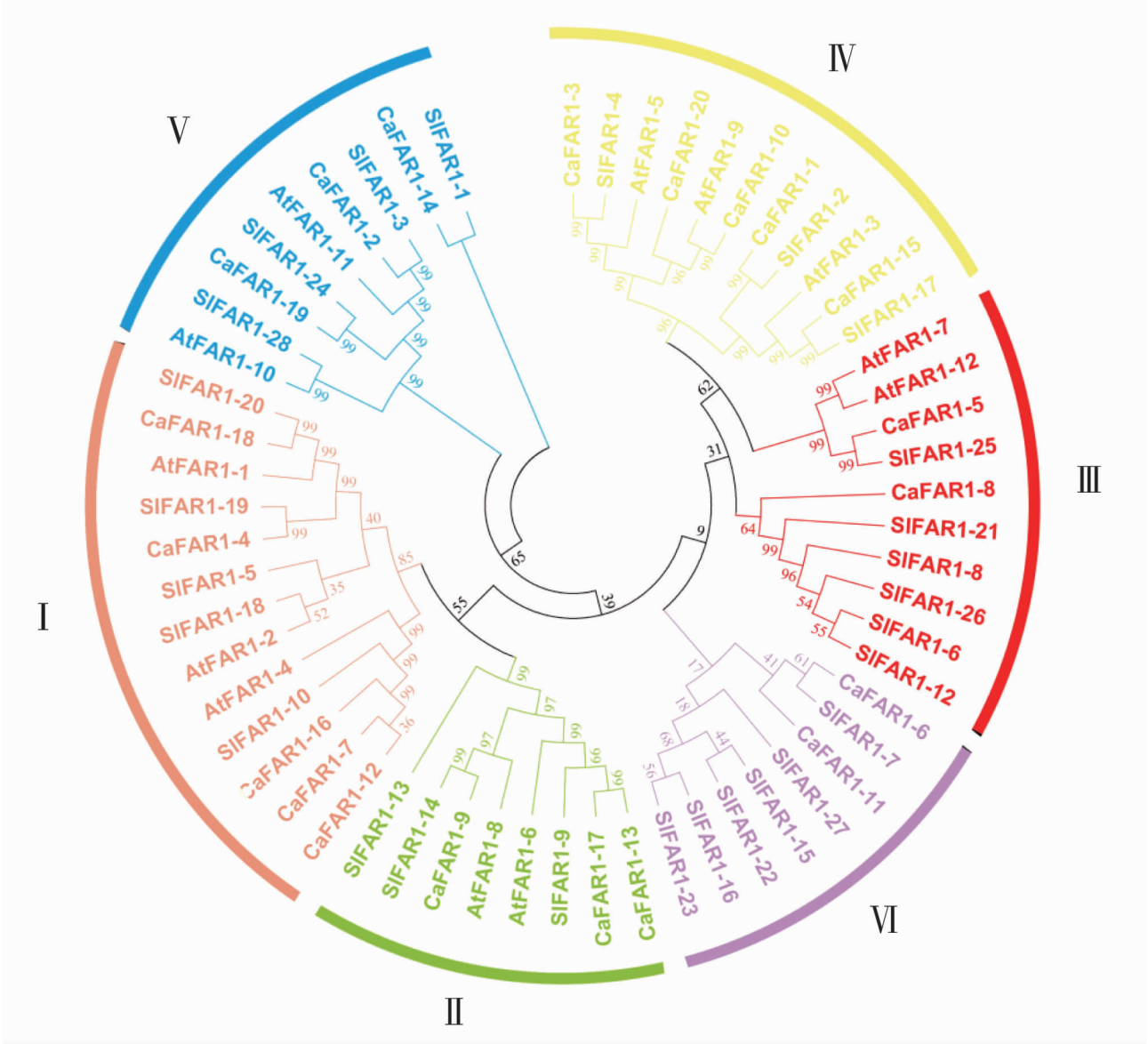


图 1 辣椒、拟南芥与番茄 FAR1 转录因子家族系统发育进化树

Fig.1 Phylogenetic evolutionary tree of FAR1 transcription factor family in pepper, Arabidopsis and Solanum lycopersicum



图 2 FAR1 保守结构域 logo

Fig.2 FAR1 conservative domain sequence logo

2.5 染色体定位 通过辣椒 GFF 注释文件中获得 *CaFAR1* 的位置信息,上传至在线软件 mg2c (http://mg2c.iask.in/mg2c_v2.0/)中,将 18 个 *CaFAR1* 基因定位到辣椒染色体上,有 2 个 *CaFAR1* 基因未被定位到染色体上。如图 5 所示,辣椒的 18 个 *CaFAR1* 基因分布在 8 条染色体上,8 号染色体上有 3 个,2 号、6 号、7 号、12 号染色体各分布 1 个,6 号和 10 号染色体上各分布 2 个,3 号染色体上有 6 个。

2.6 GO 分析 使用 Blast2GO 的非冗余蛋白质数据库软件与默认参数,首先将 *CaFAR1* 氨基酸序列与 NCBI 进行 BlastP

比对。然后对每一个 *CaFAR1* 都获取 GO 注释,得到 GO 注释的 ID,导出数据之后使用 GraphPad 作图(图 6)。可以发现,20 个 *CaFAR1* 蛋白序列基于氨基酸的相似性被分为三大类:生物过程(BP)、细胞成分(CC)和分子功能(MF),大部分 *CaFAR1* 与离子结合相关,显然与 *CaFAR1* 的锌指结构域功能吻合。*CaFAR1* 在细胞成分(CC)和分子功能(MF)这两方面发挥了很大的作用,由此可以推断 *CaFAR1* 在辣椒的生长发育与信号传递中发挥了很大的作用。

CaFAR1-1	HNSVETDI	8
CaFAR1-2		0
CaFAR1-3		0
CaFAR1-4		0
CaFAR1-5		0
CaFAR1-6		0
CaFAR1-7		0
CaFAR1-8		0
CaFAR1-9		0
CaFAR1-10		0
CaFAR1-11		0
CaFAR1-12		0
CaFAR1-13		0
CaFAR1-14		0
CaFAR1-15		0
CaFAR1-16		0
CaFAR1-17		0
CaFAR1-18		0
CaFAR1-19		0
CaFAR1-20		0
CaFAR1-21		0
CaFAR1-22		0
CaFAR1-23		0
CaFAR1-24		0
CaFAR1-25		0
CaFAR1-26		0
CaFAR1-27		0
CaFAR1-28		0
CaFAR1-29		0
CaFAR1-30		0
CaFAR1-31		0
CaFAR1-32		0
CaFAR1-33		0
CaFAR1-34		0
CaFAR1-35		0
CaFAR1-36		0
CaFAR1-37		0
CaFAR1-38		0
CaFAR1-39		0
CaFAR1-40		0
CaFAR1-41		0
CaFAR1-42		0
CaFAR1-43		0
CaFAR1-44		0
CaFAR1-45		0
CaFAR1-46		0
CaFAR1-47		0
CaFAR1-48		0
CaFAR1-49		0
CaFAR1-50		0
CaFAR1-51		0
CaFAR1-52		0
CaFAR1-53		0
CaFAR1-54		0
CaFAR1-55		0
CaFAR1-56		0
CaFAR1-57		0
CaFAR1-58		0
CaFAR1-59		0
CaFAR1-60		0
CaFAR1-61		0
CaFAR1-62		0
CaFAR1-63		0
CaFAR1-64		0
CaFAR1-65		0
CaFAR1-66		0
CaFAR1-67		0
CaFAR1-68		0
CaFAR1-69		0
CaFAR1-70		0
CaFAR1-71		0
CaFAR1-72		0
CaFAR1-73		0
CaFAR1-74		0
CaFAR1-75		0
CaFAR1-76		0
CaFAR1-77		0
CaFAR1-78		0
CaFAR1-79		0
CaFAR1-80		0
CaFAR1-81		0
CaFAR1-82		0
CaFAR1-83		0
CaFAR1-84		0
CaFAR1-85		0
CaFAR1-86		0
CaFAR1-87		0
CaFAR1-88		0
CaFAR1-89		0
CaFAR1-90		0
CaFAR1-91		0
CaFAR1-92		0
CaFAR1-93		0
CaFAR1-94		0
CaFAR1-95		0
CaFAR1-96		0
CaFAR1-97		0
CaFAR1-98		0
CaFAR1-99		0
CaFAR1-100		0

注：* 表示潜在锌指结构域的保守的 Cys 和 His 残基

Note: * indicated the conserved Cys and His residues of the potential zinc finger domain

图 3 CaFAR1 氨基酸序列比对

Fig.3 CaFAR1 amino acid sequence alignment

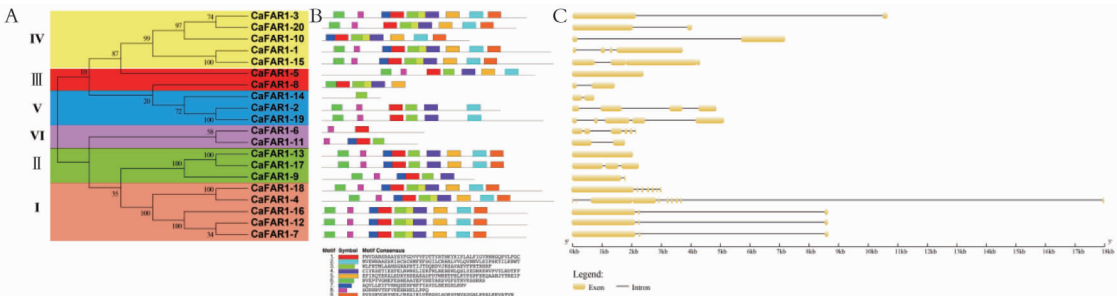


图 4 CaFAR1 保守序列与基因结构分析

Fig.4 CaFAR1 conserved sequence and gene structure analysis

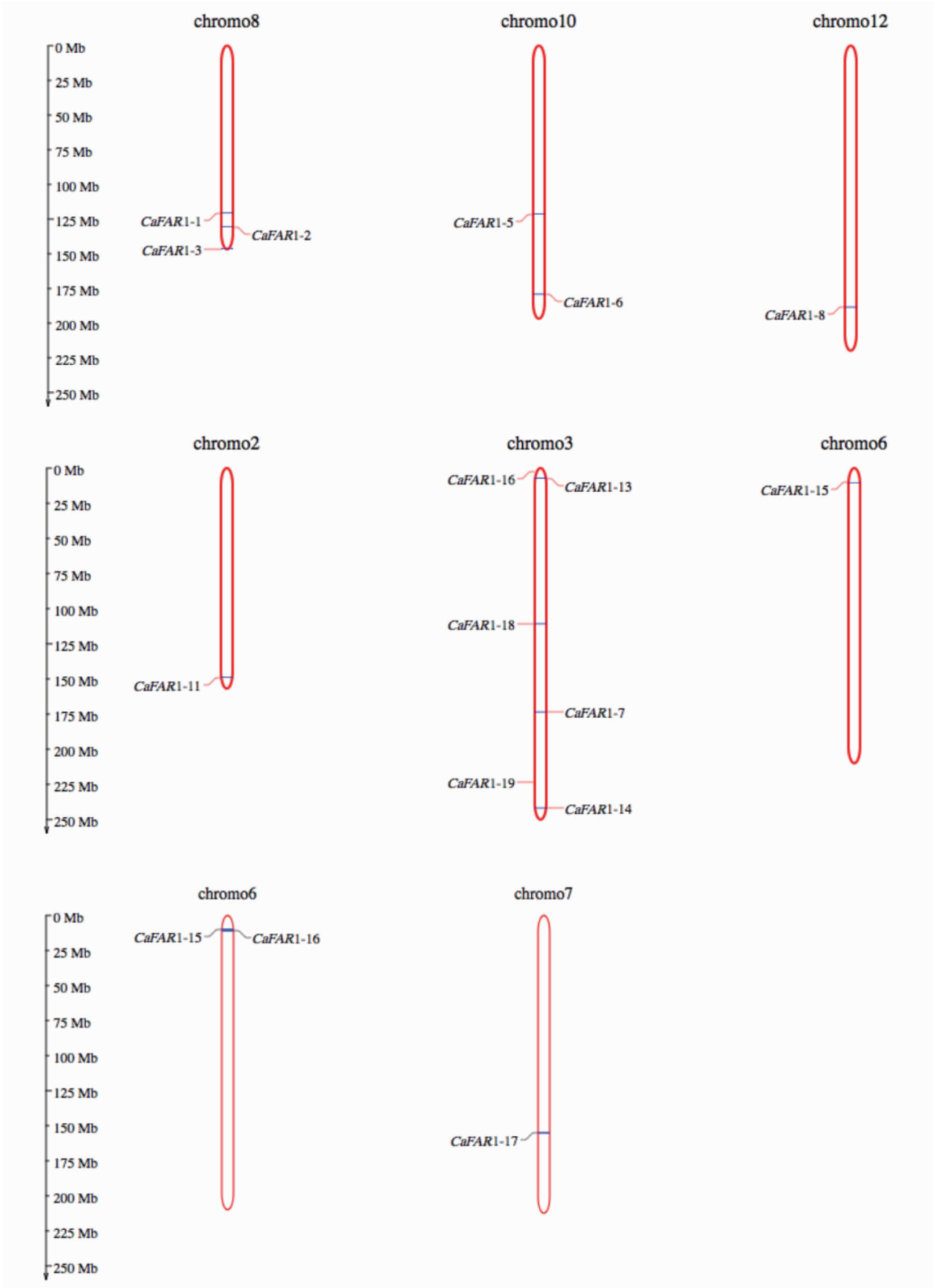


图 5 *CaFAR1* 基因在染色体上的位置

Fig.5 The location of *CaFAR1* gene on chromosome

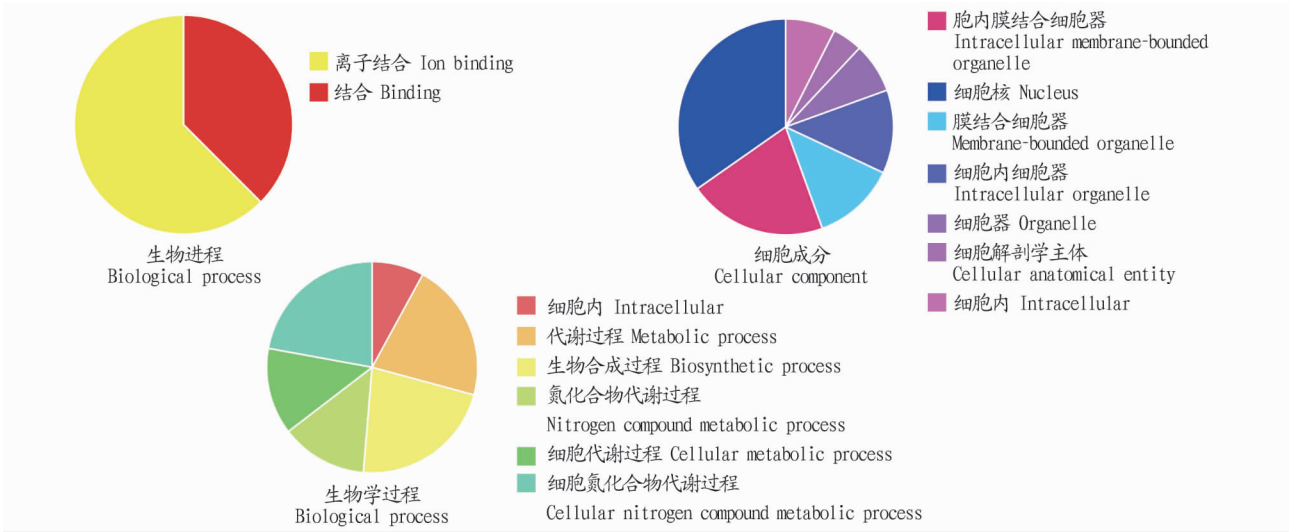


图 6 *CaFAR1* 基因 GO 注释分析
Fig.6 *CaFAR1* gene GO annotation analysis

3 讨论与小结

光是植物生长和发育的能量来源,光敏色素(phytochromes)是高等植物光受体网络中不可或缺的一个,其通过感受红光和远红光,直接或间接调控下游的转录因子,从而控制下游基因表达,影响植物的种子萌发、幼苗去黄化、叶绿体运动等生长发育进程^[4,7-10]。FAR1 是 phyA 信号通路中重要的正向调控因子之一。已在拟南芥、棉花中分离得到了多个 FAR1 家族成员^[3-4]。该研究通过对辣椒全基因组扫描,从中鉴定得到了 20 个 *CaFAR1* 转录因子家族成员,结合基因结构分析与系统进化分析,20 个 *CaFAR1* 基因分为 6 个亚家族。其中,18 个 *CaFAR1* 基因分布在 8 条染色体上。而对其功能注释表明,20 个 *CaFAR1* 蛋白序列基于氨基酸的相似性被分为三大类:生物过程(BP)、细胞成分(CC)和分子功能(MF)。

参考文献

[1] 陈芳,邓兴旺.远红光受体伴侣蛋白 FHY1 在介导基因表达和植物生长发育过程中的独特作用[J].遗传,2014,36(9):958.

[2] CASAL J J. Phytochromes, cryptochromes, phototropin: Photoreceptor interactions in plants[J]. Photochemistry & photobiology, 2000, 71(1): 1-11.

[3] LIN R C, WANG H Y. Arabidopsis *FHY3/FAR1* gene family and distinct roles of its members in light control of Arabidopsis development[J]. Plant physiology, 2004, 136(4): 4010-4022.

[4] 袁娜,王彤,刘廷利,等.棉花 FAR1/FHY3 基因家族的全基因组分析[J].棉花学报,2018,30(1): 1-11.

[5] LIN M, GANG L. FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) family proteins in arabidopsis growth and development[J]. Frontiers in plant science, 2018, 9: 692.

[6] 惠甜,沈兵琪,王连春,等.桑树 bHLH 转录因子家族全基因组鉴定与分析[J].分子植物育种,2019,17(17): 5624-5637.

[7] HEIJDE M, ULM R. UV-B photoreceptor-mediated signalling in plants[J]. Trends in plant science, 2012, 17(4): 230-237.

[8] 刘圈伟,何云,齐胜利,等.光敏色素研究进展[J].中国农学通报,2005, 21(5): 237-241.

[9] FURUYA M. Phytochromes: Their molecular species, gene families, and functions[J]. Annual review of plant physiology and plant molecular biology, 1993, 44(1): 617-645.

[10] TEPPERMAN J M, ZHU T, CHANG H S, et al. Multiple transcription-factor genes are early targets of phytochrome A signaling[J]. Proceedings of the national academy of sciences of the United States of America, 2001, 98(16): 9437-9442.