

葡萄 *Actin* 基因家族的鉴定及进化和表达分析

崔力文, 郑 婷, 张克坤, 张 川, 上官凌飞, 房经贵^①

(南京农业大学园艺学院, 江苏 南京 210095)

摘要: 采用生物信息学方法从葡萄(*Vitis vinifera* Linn.)全基因组中鉴定 *Actin* 基因家族,并对各基因的染色体定位和结构特征,编码蛋白质的理化性质、亚细胞定位、二级结构、三级结构和系统进化,以及不同组织的基因表达进行研究。结果表明:葡萄 *Actin* 基因家族 16 个基因分布在 12 条染色体上。16 个基因的结构特征及其编码蛋白质的理化性质差异较大。16 个基因的长度及其内含子总长度的变化范围较大,编码序列(CDS)和外显子总长度的变化范围较小。除登录号 GSVIVG01008254001 和 GSVIVG01014035001 的基因外,其他 14 个基因的 GC 含量均低于其 CDS 的 GC 含量。除登录号 GSVIVG01008254001 的基因外,其他 15 个基因编码的蛋白质的理论相对分子质量为 12 534.54~82 612.33,理论等电点为 pI 4.92~pI 9.13。16 个基因编码蛋白质的消光系数为 14 105~73 645,脂肪族氨基酸指数为 65.54~92.06,其中 9 个为稳定蛋白,7 个为不稳定蛋白。除登录号 GSVIVG01014035001 的基因外,其他 15 个基因编码的蛋白质均为亲水性蛋白。登录号 GSVIVG01016517001 的基因编码的蛋白质定位于细胞质和细胞核,其他 15 个基因编码的蛋白质定位于细胞质。二级结构和三级结构显示:葡萄 *Actin* 基因家族 16 个基因编码的蛋白质均由 α 螺旋、无规则卷曲和延伸链构成,且总体以无规则卷曲为主。系统进化分析和不同组织的基因表达分析结果显示:与拟南芥[*Arabidopsis thaliana* (Linn.) Heynh.]相似,葡萄 *Actin* 基因家族 16 个基因编码的蛋白质分为 3 个亚家族,Class II 亚家族(营养型)包括登录号 GSVIVG01003099001 和 GSVIVG01026580001 的基因编码的蛋白质,这 2 个基因在所有组织中的表达均较高;Class III 亚家族(生殖型)包括登录号 GSVIVG01033494001、GSVIVG01024980001 和 GSVIVG01016550001 的基因编码的蛋白质,这 3 个基因在花粉、雄蕊和花中的表达均较高;Class I 亚家族包括其他 11 个基因编码的蛋白质,这 11 个基因在各组织中的表达总体上较低。研究结果显示:葡萄 *Actin* 基因家族的表达具有组织特异性。

关键词: 葡萄; *Actin* 基因家族; 系统进化; 基因表达

中图分类号: Q943.2; S663.1 文献标志码: A 文章编号: 1674-7895(2017)03-0001-10

DOI: 10.3969/j.issn.1674-7895.2017.03.01

Identification, evolution and expression analyses of *Actin* gene family of *Vitis vinifera* CUI Liwen, ZHENG Ting, ZHANG Kekun, ZHANG Chuan, SHANGGUAN Lingfei, FANG Jinggui^① (College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China), *J. Plant Resour. & Environ.*, 2017, 26(3): 1-10

Abstract: *Actin* gene family was identified from the whole genome of *Vitis vinifera* Linn. by bioinformatics method, and chromosome localization and structural characteristics of each gene, physicochemical property, subcellular localization, secondary structure, tertiary structure and phylogeny of proteins encoded, and gene expression in different tissues were researched. The results show that sixteen genes in *Actin* gene family of *V. vinifera* are distributed on twelve chromosomes. There are great differences in structural characteristics of sixteen genes and physicochemical property of their proteins encoded. Variation ranges of length of sixteen genes and their total length of introns are large, and those of total length of coding sequences (CDSs) and exons are small. Except genes with accession number of

收稿日期: 2017-03-22

基金项目: 国家自然科学基金资助项目(31672131)

作者简介: 崔力文(1991—),女,山东莒县人,硕士研究生,主要从事果树遗传育种与分子生物学方面的研究。

^①通信作者 E-mail: fanggg@njau.edu.cn

GSVIVG01008254001 and GSVIVG01014035001, GC contents in other fourteen genes are lower than those in their CDSs. Except gene with accession number of GSVIVG01008254001, theoretical relative molecular mass of proteins encoded by other fifteen genes is 12 534.54–82 612.33, and their theoretical isoelectric point is pI 4.92–pI 9.13. Extinction coefficient of proteins encoded by sixteen genes is 14 105–73 645, their aliphatic amino acid index is 65.54–92.06, in which, there are nine stable proteins and seven instable proteins. Except gene with accession number of GSVIVG01014035001, proteins encoded by other fifteen genes are hydrophilic proteins. Protein encoded by gene with accession number of GSVIVG01016517001 is located in cytoplasm and nucleus, and those encoded by other fifteen genes are located in cytoplasm. The secondary and tertiary structures show that proteins encoded by sixteen genes in *Actin* gene family of *V. vinifera* are composed with α helix, random coil and extended strand, and in general, random coil is the main component. Analysis results of phylogeny and gene expression in different tissues show that being similar to *Arabidopsis thaliana* (Linn.) Heynh., proteins encoded by sixteen genes in *Actin* gene family of *V. vinifera* can be divided into three subfamilies. Class II subfamily (vegetative type) includes proteins encoded by genes with accession number of GSVIVG01003099001 and GSVIVG01026580001, and the expression of these two genes in all tissues is higher. Class III subfamily (reproductive type) includes proteins encoded by genes with accession number of GSVIVG01033494001, GSVIVG01024980001 and GSVIVG01016550001, and the expression of these three genes in pollen, stamen and flower is higher. Class I subfamily includes proteins encoded by the rest of eleven genes, and the expression of these eleven genes in all tissues is generally low. The results of gene expression analysis show that there is tissue specificity in expression of *Actin* gene family of *V. vinifera*.

Key words: *Vitis vinifera* Linn.; *Actin* gene family; phylogeny; gene expression

自从在脊椎动物骨骼细胞中发现肌动蛋白 (actin) 以来,陆续在真菌、高等植物^[1]以及藻类中发现了肌动蛋白的存在。肌动蛋白广泛存在于植物界,是微丝的主要组分。肌动蛋白参与许多对植物发育至关重要的生理活动,如细胞形状的维持,胞质流动以及细胞的分裂分化、重力感应和信号传导等^[2];在尖端生长的花粉管中,膜泡和生殖细胞基于肌动蛋白的运动而移动^[3]。有关多核糖体与微丝共定位的研究认为,微丝骨架影响多核糖蛋白在细胞内的分布^[4]。植物肌动蛋白丰富的功能和作用由 *Actin* 基因家族的多样性决定。多细胞植物具有 10 个及以上 *Actin* 基因,如拟南芥 [*Arabidopsis thaliana* (Linn.) Heynh.]^[5] 有 10 个 *Actin* 基因,碧冬茄 (*Petunia hybrida* Vilm.)^[6] 的基因组中有 100 多条 *Actin* 基因序列,从豌豆 (*Pisum sativum* Linn.)^[7] 卷须中克隆出 18 个 *Actin* 基因。高等植物的 *Actin* 基因家族成员具有多样性和表达差异。McDowell 等^[8] 研究了拟南芥 *Actin* 基因家族的结构及进化关系,认为拟南芥 *Actin* 基因家族至少包含 2 个古代基因类型,在陆地植物进化早期分化产生生殖和营养 2 种类型,然后进一步分化产生 6 种肌动蛋白亚类,其中 5 种肌动蛋白亚类具有不同的组织特异性表达模式。拟南芥肌动蛋白的 8 个异型体中有大量的非保守氨基酸替换,所有氨基

酸替换都发生在分子表面,除 1 个 *Actin* 基因外,在其余 *Actin* 基因的核苷酸序列中的同一位置均含有 3 个内含子,且每个 *Actin* 基因都含有 1 个 180~220 nt 的 3'-非翻译区 (untranslated region, UTR)^[9-10]。豌豆肌动蛋白异型体可分为 3 种类型^[11],这 3 类肌动蛋白异型体基因在根、茎、叶、卷须、花粉和幼嫩果实中均可表达,但在发育时间以及表达强度上存在明显差异,而且具有组织特异性^[12-13]。对豌豆 II 类肌动蛋白异型体 3 个基因的全序列测定结果显示:3 个基因的核苷酸序列完全相同,差别仅在于 3'-UTR 的长度不同。此外,豌豆卷须中 3 类肌动蛋白异型体基因的核苷酸序列同源率为 80%,其氨基酸序列同源率为 94%^[11]。

葡萄 (*Vitis vinifera* Linn.) 是世界最古老的果树树种之一,也是世界性的重要水果。2007 年,葡萄基因组草图测序完成^[14],为后续通过生物信息学方法挖掘、鉴定和分析葡萄基因组奠定了基础。本研究采用生物信息学方法对葡萄全基因组中的 *Actin* 基因家族进行鉴定,对 *Actin* 基因家族中各基因的结构特征,编码蛋白质的理化性质、亚细胞定位、二级结构、三级结构和系统进化,以及各基因的组织表达模式进行分析,以期为进一步研究葡萄 *Actin* 基因的功能提供基础资料。

1 材料和方法

1.1 材料

葡萄全基因组信息由 Phytozome v12.0 (<https://phytozome.jgi.doe.gov/pz/portal.html>) 数据库提供。葡萄芯片表达数据来源于 NCBI 的 GEO 数据库,编号 GSE36128,包括葡萄 54 个不同组织样品,覆盖了葡萄不同生长发育阶段的大部分器官^[15]。

1.2 方法

1.2.1 *Actin* 基因家族数据集的建立和染色体定位分析 首先利用葡萄全基因组序列,构建本地 BLAST 数据库,以拟南芥 *Actin* 基因家族序列为靶序列执行本地 BLAST 搜索;同时,利用 Pfam 数据库工具建立葡萄全基因组蛋白质结构域模型,利用 HMMER 程序筛选含有 *Actin* 典型结构域(PF00022)的蛋白质序列^[16]。采用 E 值对筛选结果进行可靠性评价,E 值越低越可靠。本研究选用 E 值小于等于 10^{-5} 的序列作为候选蛋白质序列,删除重复序列。

从 Phytozome v12.0 数据库下载的葡萄全基因组信息中提取 *Actin* 基因家族各基因的登录号和染色体定位信息。

1.2.2 基因结构特征分析 根据 Phytozome v12.0 数据库提供的信息,分析基因长度以及编码序列(coding sequence, CDS)、外显子、内含子、3'-非翻译区(UTR)和 5'-UTR 的数量和总长度,采用 DNASTAR 软件计算 GC 含量。

1.2.3 编码蛋白质的理化性质分析 利用 ProtParam tool (<http://web.expasy.org/protparam/>) 分析 *Actin* 基因家族中各基因编码蛋白质的基本理化性质,包括氨基酸残基数量、理论相对分子量、理论等电点、平均亲水系数、消光系数、不稳定系数及脂肪族氨基酸指数^[17]。

1.2.4 编码蛋白质的亚细胞定位、二级结构及三级结构分析 参照董晨等^[18]的方法利用 Plant-mPLOC sever 程序进行 *Actin* 基因家族各基因编码蛋白质的亚细胞定位;利用 SOPMA 程序分析蛋白质二级结构;利用 Phyre² (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) 在线绘制蛋白质三级结构图。

1.2.5 系统发育树的构建 通过 Clustal Omega 程序对拟南芥和葡萄 *Actin* 基因家族中各基因编码的蛋白质进行多序列联配比对分析并构建系统发育树。

1.2.6 *Actin* 基因家族的组织表达分析 在葡萄芯片表达数据库中选取 *Actin* 基因家族各基因的表达值,将原值以 2 为底取对数后采用 MeV 软件绘制热图。

2 结果和分析

2.1 葡萄 *Actin* 基因家族的全基因组鉴定

葡萄 *Actin* 基因家族的信息见表 1。由表 1 可见:葡萄 *Actin* 基因家族共包括 16 个基因,且各基因在染色体上的分布比较分散,分别分布在 01、04、06、08、11、12、13、14、16、17、18 和 19 号共 12 条染色体上,其中,13 号染色体分布的基因数量最多(3 个),11 和 17 号染色体均分布 2 个基因,其他染色体均分布 1 个基因。

表 1 葡萄 *Actin* 基因家族的信息
Table 1 Information of *Actin* gene family of *Vitis vinifera* Linn.

编号 No.	登录号 Accession number	染色体定位 Chromosome localization
1	GSVIVG01000587001	ch01:6782564-6810457
2	GSVIVG01003099001	ch12:8678866-8681135
3	GSVIVG01008254001	ch17:4031362-4061477
4	GSVIVG01008255001	ch17:4013614-4030933
5	GSVIVG01014035001	ch16:9532118-9532444
6	GSVIVG01014840001	ch19:10434342-10444022
7	GSVIVG01015049001	ch11:713295-714860
8	GSVIVG01015556001	ch11:4862914-4867036
9	GSVIVG01016517001	ch13:3197303-3207623
10	GSVIVG01016550001	ch13:2880916-2883022
11	GSVIVG01024980001	ch06:5548568-5552191
12	GSVIVG01026580001	ch04:21427667-21430968
13	GSVIVG01032120001	ch13:22480789-22496776
14	GSVIVG01033049001	ch14:25325845-25334371
15	GSVIVG01033494001	ch08:20249818-20252600
16	GSVIVG01036362001	ch18:4225053-4229977

2.2 葡萄 *Actin* 基因家族的结构特征分析

葡萄 *Actin* 基因家族的结构特征见表 2。由表 2 可见:葡萄 *Actin* 基因家族 16 个基因含 1~16 个编码序列(CDS)和外显子以及 0~15 个内含子,登录号 GSVIVG01008254001 的基因中外显子和内含子数量最多,登录号 GSVIVG01014035001 的基因中外显子和内含子数量最少。葡萄 *Actin* 基因家族 16 个基因长度的变化范围较大,为 327~30 116 bp。除登录号 GSVIVG01014035001 的基因不含内含子外,其他 15 个基因中内含子总长度的变化范围较大,为 314~28 412 bp。葡萄 *Actin* 基因家族 16 个基因中 CDS 和

外显子总长度的变化范围均较小,分别为327~2 178和326~2 544 bp。

由表2还可见:葡萄 *Actin* 基因家族16个基因中GC含量的变化范围为30.48%~43.68%,CDS中GC含量的变化范围为34.09%~50.23%。除登录号GSVIVG01008254001和GSVIVG01014035001的基因外,其他14个基因的GC含量均低于其CDS的GC含量。

由表2还可见:葡萄 *Actin* 基因家族16个基因中

登录号GSVIVG01008255001、GSVIVG01014035001和GSVIVG01015556001的基因不含3'-非翻译区(UTR),其他13个基因均含1个3'-UTR,且这13个基因中3'-UTR的总长度为120~1 171 bp。登录号分别为GSVIVG01008254001、GSVIVG01014035001、GSVIVG01015049001、GSVIVG01015556001、GSVIVG01016550001和GSVIVG01036362001的基因不含5'-UTR,登录号GSVIVG01024980001和GSVIVG01033494001的基因含2个5'-UTR,其他8个

表2 葡萄 *Actin* 基因家族的结构特征¹⁾

Table 2 Structural characteristics of *Actin* gene family of *Vitis vinifera* Linn.¹⁾

登录号 Accession number	基因长度/bp Gene length	GC含量/% GC content	CDS			外显子 Exon	
			数量 Number	总长度/bp Total length	GC含量/% GC content	数量 Number	总长度/bp Total length
GSVIVG01000587001	27 894	38.23	11	1 149	45.87	11	2 110
GSVIVG01003099001	2 270	41.98	5	867	48.10	5	1 315
GSVIVG01008254001	30 116	34.39	16	1 329	34.09	16	1 655
GSVIVG01008255001	17 320	37.88	8	531	41.81	8	644
GSVIVG01014035001	327	41.28	1	327	41.28	1	326
GSVIVG01014840001	9 681	30.48	12	1 422	45.08	12	1 772
GSVIVG01015049001	1 566	43.68	4	1 131	46.77	4	952
GSVIVG01015556001	4 123	39.19	7	1 302	50.23	7	1 302
GSVIVG01016517001	10 321	35.94	12	2 178	44.21	12	2 544
GSVIVG01016550001	2 107	42.76	4	1 134	47.18	4	1 440
GSVIVG01024980001	3 624	38.27	4	1 131	47.66	5	1 611
GSVIVG01026580001	3 302	39.92	6	1 227	46.21	6	2 118
GSVIVG01032120001	15 988	37.40	7	1 086	44.11	7	1 464
GSVIVG01033049001	8 527	34.58	15	1 170	43.50	15	1 680
GSVIVG01033494001	2 783	40.32	5	870	48.51	6	1 327
GSVIVG01036362001	4 925	43.29	3	741	45.88	3	1 913

登录号 Accession number	内含子 Intron		3'-UTR		5'-UTR	
	数量 Number	总长度/bp Total length	数量 Number	总长度/bp Total length	数量 Number	总长度/bp Total length
GSVIVG01000587001	10	25 802	1	586	1	373
GSVIVG01003099001	4	955	1	436	1	10
GSVIVG01008254001	15	28 412	1	325	—	—
GSVIVG01008255001	7	16 746	—	—	1	112
GSVIVG01014035001	0	—	—	—	—	—
GSVIVG01014840001	11	7 909	1	273	1	75
GSVIVG01015049001	3	314	1	120	—	—
GSVIVG01015556001	6	2 821	—	—	—	—
GSVIVG01016517001	11	7 777	1	320	1	44
GSVIVG01016550001	3	667	1	305	—	—
GSVIVG01024980001	3	1 093	1	356	2	121
GSVIVG01026580001	5	1 185	1	675	1	213
GSVIVG01032120001	6	14 340	1	255	1	121
GSVIVG01033049001	14	6 837	1	338	1	180
GSVIVG01033494001	4	831	1	376	2	78
GSVIVG01036362001	2	3 013	1	1 171	—	—

¹⁾ CDS: 编码序列 Coding sequence; UTR: 非翻译区 Untranslated region. —: 无数据 No datum.

基因均含1个5'-UTR,且这10个基因中5'-UTR总长度的变化范围较小,为10~373 bp。

2.3 葡萄 *Actin* 基因家族16个基因编码蛋白质的理化性质分析

葡萄 *Actin* 基因家族16个基因编码蛋白质的理化性质见表3。结果显示:葡萄 *Actin* 基因家族16个基因编码蛋白质的氨基酸残基数量为108~725,登录号 GSVIVG01016517001 的基因编码蛋白质的氨基酸残基数量最多,登录号 GSVIVG01014035001 的基因编码蛋白质的氨基酸残基数量最少。除登录号 GSVIVG01008254001 的基因外,葡萄 *Actin* 基因家族中其他15个基因编码蛋白质的理论相对分子质量为

12 534.54 ~ 82 612.33,理论等电点为 pI 4.92 ~ pI 9.13,其中,12个基因编码蛋白质的理论等电点小于 pI 6.5,显酸性,3个基因编码蛋白质的理论等电点大于 pI 8.0,显碱性。葡萄 *Actin* 基因家族16个基因编码蛋白质的消光系数为14 105~73 645,脂肪族氨基酸指数为65.54~92.06。葡萄 *Actin* 基因家族16个基因中,9个基因编码蛋白质的不稳定指数小于40,为稳定蛋白;其他7个基因编码蛋白质的不稳定指数大于40,为不稳定蛋白。除登录号 GSVIVG01014035001 的基因编码的蛋白质为疏水性蛋白外,葡萄 *Actin* 基因家族中其他15个基因编码的蛋白质为亲水性蛋白。

表3 葡萄 *Actin* 基因家族16个基因编码蛋白质的理化性质¹⁾

Table 3 Physicochemical property of proteins encoded by sixteen genes in *Actin* gene family of *Vitis vinifera* Linn.¹⁾

登录号 Accession number	氨基酸残基数量 Number of amino acid residues	理论相对分子质量 Theoretical relative molecular mass	理论等电点 Theoretical isoelectric point	平均亲水系数 Grand average of hydropathicity	消光系数 Extinction coefficient	不稳定指数 Instability index	脂肪族氨基酸指数 Aliphatic amino acid index
GSVIVG01000587001	382	42 106.90	5.59	-0.132	39 685	37.07	79.08
GSVIVG01003099001	288	31 542.22	5.85	-0.107	38 515	37.82	86.01
GSVIVG01008254001	442	—	—	-0.164	24 785	32.50	65.54
GSVIVG01008255001	176	19 488.90	4.94	-0.044	18 825	40.03	69.20
GSVIVG01014035001	108	12 534.54	8.71	0.014	14 105	34.95	87.31
GSVIVG01014840001	473	53 012.41	9.13	-0.038	70 985	51.69	85.14
GSVIVG01015049001	376	42 098.45	5.23	-0.157	47 245	42.36	88.88
GSVIVG01015556001	432	48 477.53	5.68	-0.204	35 005	44.96	89.10
GSVIVG01016517001	725	82 612.33	5.77	-0.638	73 645	51.75	73.56
GSVIVG01016550001	377	41 814.81	5.31	-0.199	46 090	35.53	82.52
GSVIVG01024980001	375	41 436.47	5.31	-0.174	44 600	36.96	85.33
GSVIVG01026580001	408	45 329.11	5.59	-0.126	47 580	39.73	88.43
GSVIVG01032120001	360	39 635.10	4.92	-0.134	30 285	42.47	92.06
GSVIVG01033049001	389	44 158.94	5.52	-0.202	43 820	36.19	91.65
GSVIVG01033494001	289	31 675.33	5.85	-0.135	38 515	38.39	84.36
GSVIVG01036362001	246	27 642.63	8.67	-0.178	36 245	48.80	78.46

¹⁾—: 不明确 Undefined.

2.4 葡萄 *Actin* 基因家族16个基因编码蛋白质的亚细胞定位、二级结构及三级结构分析

葡萄 *Actin* 基因家族16个基因编码蛋白质的亚细胞定位和二级结构见表4,16个基因编码蛋白质的三级结构见图1。

2.4.1 亚细胞定位 由表4可见:葡萄 *Actin* 基因家族中仅登录号 GSVIVG01016517001 的基因编码的蛋白质定位于细胞质和细胞核,其他15个基因编码的蛋白质均定位于细胞质。

2.4.2 二级结构 由表4还可见:葡萄 *Actin* 基因家族16个基因编码蛋白质的二级结构均由 α 螺旋、无

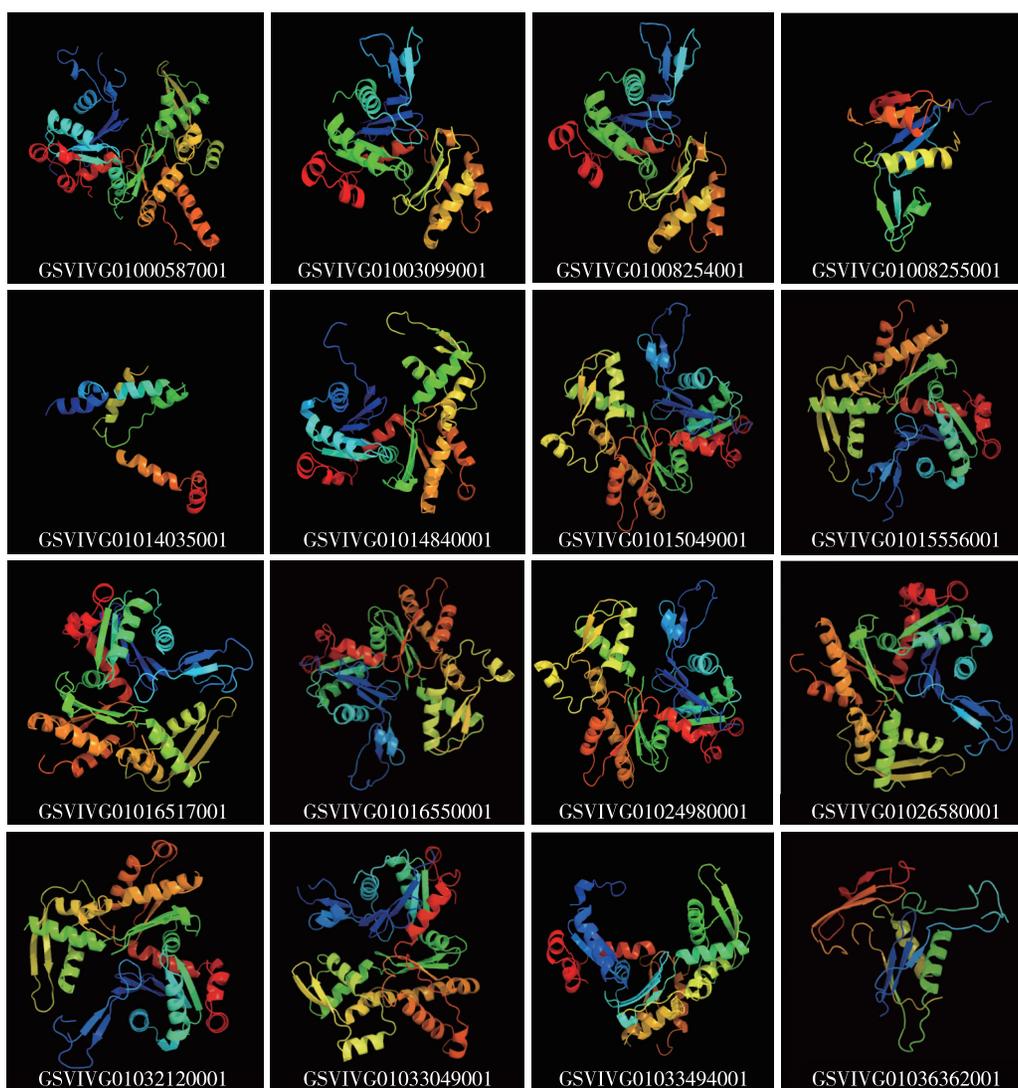
规则卷曲和延伸链组成,且总体以无规则卷曲为主。登录号 GSVIVG01000587001、GSVIVG01003099001、GSVIVG01008255001、GSVIVG01014035001、GSVIVG01033494001和 GSVIVG01036362001 的基因编码蛋白质的二级结构中,各构成元件所占比例由高到低依次为无规则卷曲、延伸链、 α 螺旋;登录号 GSVIVG01008254001 的基因编码蛋白质的二级结构中,各构成元件所占比例由高到低依次为延伸链、无规则卷曲、 α 螺旋;其余9个基因编码蛋白质的二级结构中,各构成元件所占比例由高到低依次为无规则卷曲、 α 螺旋、延伸链。

表 4 葡萄 *Actin* 基因家族 16 个基因编码蛋白质的亚细胞定位及二级结构Table 4 Subcellular localization and secondary structure of proteins encoded by sixteen genes in *Actin* gene family of *Vitis vinifera* Linn.

登录号 Accession number	亚细胞定位 ¹⁾ Subcellular localization ¹⁾	比例/% ²⁾ Percentage ²⁾			登录号 Accession number	亚细胞定位 ¹⁾ Subcellular localization ¹⁾	比例/% ²⁾ Percentage ²⁾		
		H	RC	ES			H	RC	ES
GSVIVG01000587001	C	20.42	57.85	21.73	GSVIVG01016517001	C,N	36.46	45.58	17.96
GSVIVG01003099001	C	22.57	50.35	27.08	GSVIVG01016550001	C	29.71	48.28	22.02
GSVIVG01008254001	C	23.98	35.97	40.05	GSVIVG01024980001	C	34.31	45.48	20.21
GSVIVG01008255001	C	5.11	65.91	28.98	GSVIVG01026580001	C	27.21	46.81	25.98
GSVIVG01014035001	C	12.04	62.04	25.93	GSVIVG01032120001	C	37.95	46.81	15.24
GSVIVG01014840001	C	33.19	42.71	24.10	GSVIVG01033049001	C	33.93	47.81	18.25
GSVIVG01015049001	C	37.23	42.82	19.95	GSVIVG01033494001	C	22.49	50.17	27.34
GSVIVG01015556001	C	26.64	52.19	22.17	GSVIVG01036362001	C	13.41	57.72	28.86

¹⁾ C: 细胞质 Cytoplasm; N: 细胞核 Nucleus.

²⁾ H: α 螺旋 α helix; RC: 无规则卷曲 Random coil; ES: 延伸链 Extended strand.



图中编号为基因登录号 Numbers in figures are accession numbers of genes.

图 1 葡萄 *Actin* 基因家族 16 个基因编码蛋白质的三级结构Fig. 1 Tertiary structure of proteins encoded by sixteen genes in *Actin* gene family of *Vitis vinifera* Linn.

2.4.3 三级结构 由图1可见:葡萄 *Actin* 基因家族中,由 α 螺旋数量来看,登录号 GSVIVG01016517001 的基因编码的蛋白质中 α 螺旋数量最多,为26个;登录号 GSVIVG01008254001 和 GSVIVG01024980001 的基因编码的蛋白质均包含14个 α 螺旋;登录号分别为 GSVIVG01000587001、GSVIVG01003099001 和 GSVIVG01003099001 的基因编码的蛋白质均包含8个 α 螺旋;登录号 GSVIVG01008255001 的基因编码的蛋白质中 α 螺旋数量最少,仅2个。

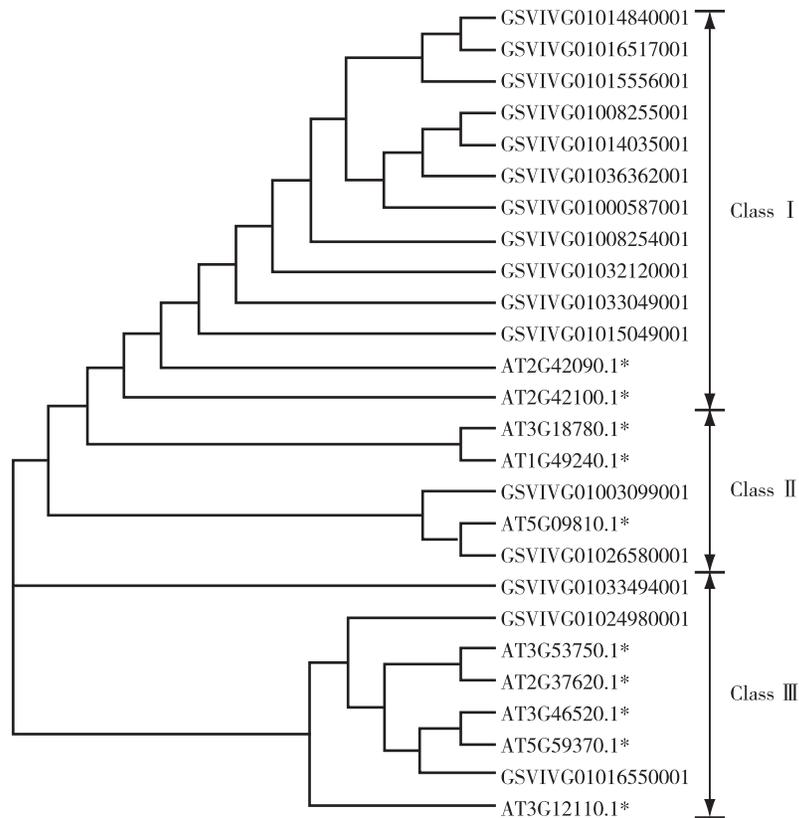
由延伸链数量来看,登录号 GSVIVG01016517001 的基因编码的蛋白质中延伸链数量最多,为31个;登录号 GSVIVG01000587001、GSVIVG01008254001、GSVIVG01033049001 和 GSVIVG0100309900 的基因编码的蛋白质均包含17个延伸链;登录号分别为 GSVIVG01036362001、GSVIVG01024980001 和 GSVIVG01033494001 的基因编码的蛋白质均包含16个延伸链;登录号 GSVIVG01014035001 的基因编码

的蛋白质中延伸链最少,仅6个。

此外,虽然登录号 GSVIVG01000587001 和 GSVIVG01003099001 的基因编码的蛋白质含有相同的 α 螺旋和延伸链数量,但二者的三级结构存在一定差异,推测这些差异导致二者在功能上有所区别。

2.5 葡萄 *Actin* 基因家族 16 个基因编码蛋白质的系统进化分析

基于葡萄 *Actin* 基因家族 16 个基因编码蛋白质的全长序列构建的系统发育树见图2。参照拟南芥 *Actin* 基因家族的分类方法^[18],葡萄 *Actin* 基因家族 16 个基因编码蛋白质分为 Class I、Class II 和 Class III 3 个亚家族,其中,Class II 亚家族属于营养型,包括登录号 GSVIVG01003099001 和 GSVIVG01026580001 的基因编码的蛋白质;Class III 亚家族属于生殖型,包括登录号 GSVIVG01033494001、GSVIVG01024980001 和 GSVIVG01016550001 的基因编码的蛋白质;Class I 亚家族包括其他 11 个基因编码的蛋白质。



*: 来自拟南芥 From *Arabidopsis thaliana* (Linn.) Heynh.

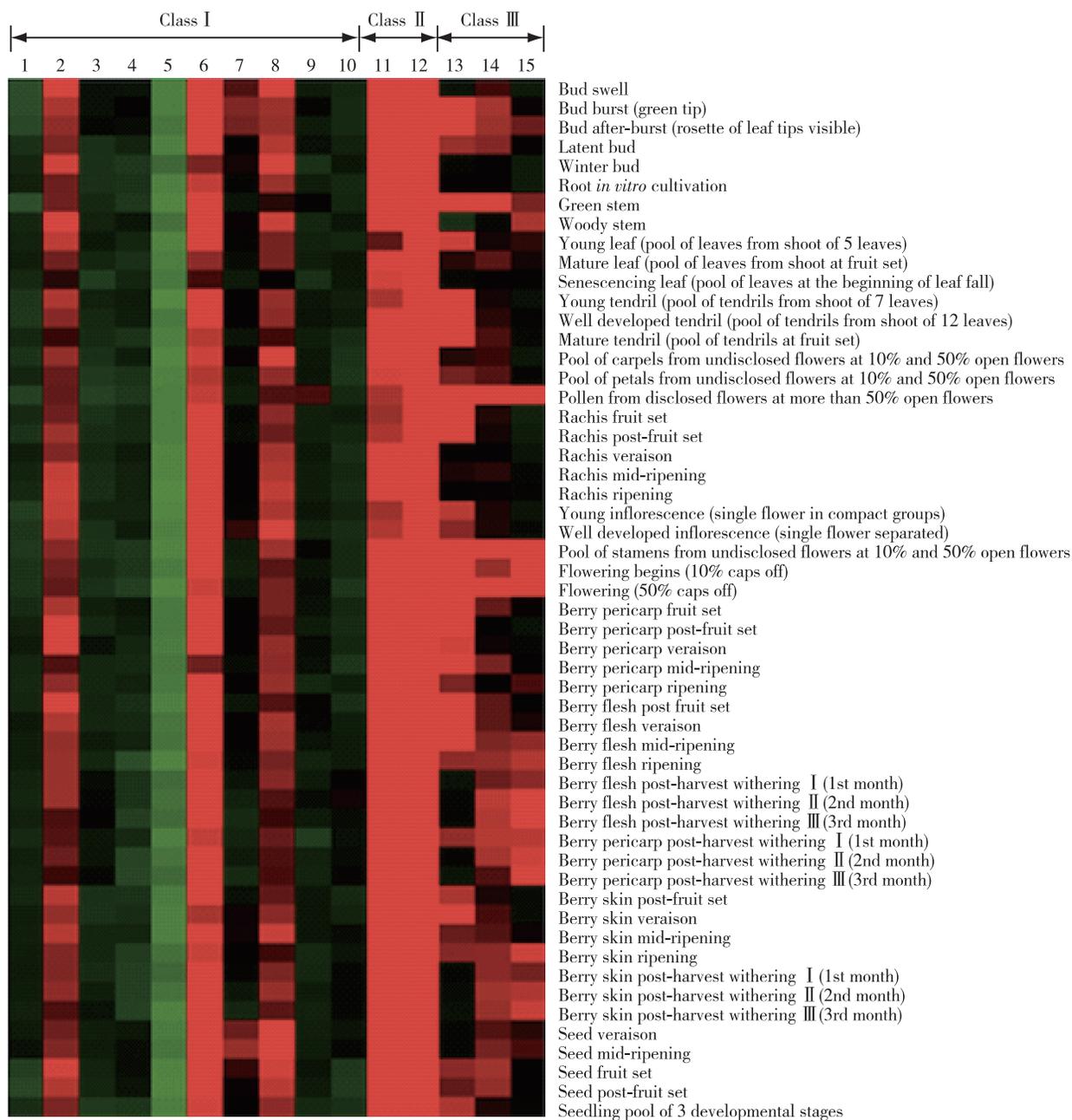
图2 葡萄与拟南芥 *Actin* 基因家族中各基因编码蛋白质的系统发育树

Fig. 2 Phylogenetic tree of proteins encoded by each gene in *Actin* gene family of *Vitis vinifera* Linn. and *Arabidopsis thaliana* (Linn.) Heynh.

2.6 葡萄 *Actin* 基因家族的组织表达分析

通过葡萄基因芯片平台,对葡萄 *Actin* 基因家族(登录号 GSVIVG01036362001 的基因在 GSE36128

数据集未出现)进行组织表达分析(图 3)。结果显示:Class II 亚家族中登录号 GSVIVG01003099001 和 GSVIVG01026580001 的基因在所有组织中的表达均



颜色越绿表示基因表达越弱,颜色越红表示基因表达越强 The more green the color, the weaker the gene expression, and the more red the color, the stronger the gene expression.

1-15: 基因登录号 Accession number of genes. 1: GSVIVG01014840001; 2: GSVIVG01016517001; 3: GSVIVG01015556001; 4: GSVIVG01008255001; 5: GSVIVG01014035001; 6: GSVIVG01000587001; 7: GSVIVG01008254001; 8: GSVIVG01032120001; 9: GSVIVG01033049001; 10: GSVIVG01015049001; 11: GSVIVG01003099001; 12: GSVIVG01026580001; 13: GSVIVG01033494001; 14: GSVIVG01024980001; 15: GSVIVG01016550001.

图 3 葡萄 *Actin* 基因家族的组织表达模式
Fig. 3 Tissue expression pattern of *Actin* gene family of *Vitis vinifera* Linn.

较高;Class III 亚家族中登录号 GSVIVG01033494001、GSVIVG01024980001 和 GSVIVG01016550001 的基因在花粉、雄蕊和花中的表达均较高,其中,登录号 GSVIVG01033494001 的基因在果皮中的表达也较高;Class I 亚家族中除登录号 GSVIVG01016517001、GSVIVG01000587001 和 GSVIVG01032120001 的基因在各组织中均有一定表达之外,Class I 亚家族中其他基因在各组织中的表达均较低,尤其是登录号 GSVIVG01014035001 的基因在所有组织中几乎不表达。

3 讨论和结论

肌动蛋白是存在于真核生物细胞质中最丰富的蛋白质,其氨基酸序列也是保守性最高的蛋白质氨基酸序列之一,在整个进化过程中几乎没有改变。肌动蛋白的氨基酸序列在不同种类藻类和人类中的差异不超过 20%^[19]。拟南芥中 10 个 *Actin* 基因显示出不同的表达模式,并涉及一系列不同细胞功能^[20]。基于拟南芥中 10 个 *Actin* 基因表达类型,其中 2 个 *Actin* 基因为假基因,其他 8 个 *Actin* 基因可以分为营养型和生殖型 2 类,表现出不同的表达模式,并且同一类中各基因间紧密相关^[21]。目前,已有关于拟南芥、大豆和芜菁 (*Brassica rapa* Linn.) *Actin* 基因家族的全基因组分析^[19],并根据其功能分为不同的类型,为进一步研究葡萄 *Actin* 基因家族提供了重要的参考价值。

本研究通过对葡萄 *Actin* 基因家族进行全基因组生物信息学分析,共鉴定出 16 个基因,这 16 个基因编码蛋白质的亚家族分类与拟南芥和芜菁^[19]相似,说明葡萄 *Actin* 基因家族是植物进化中非常保守的家族。虽然 *Actin* 基因家族在葡萄全基因组中只有 16 个基因,但这 16 个基因的结构特征及其编码蛋白质的理化性质差异较大。其中,基因长度和内含子总长度的差异较大,分别相差 29 789 和 28 098 bp;内含子数量变化范围也较大,在 0~15 之间。内含子不同程度的插入和缺失构成了基因结构的多样性,在一定程度上,内含子的进化历程反映了基因结构的进化^[22]。除登录号 GSVIVG01014035001 的基因外,葡萄 *Actin* 基因家族其他 15 个基因编码的蛋白质为亲水性蛋白,其理化性质存在较大差异,体现在氨基酸残基数相差 549,理论相对分子质量相差 63 123.43,理论等电点相差 pI 4.21。

本研究中,葡萄 *Actin* 基因家族中 Class II 和 Class III 亚家族分别属于营养型和生殖型,这 2 个亚家族中的基因在葡萄生长发育过程中具有重要作用^[23]。Class II 和 Class III 亚家族 *Actin* 基因中,基因和编码序列(CDS)的 GC 含量以及 CDS、外显子、内含子、3'-非翻译区(UTR)和 5'-UTR 的数量接近,其中,Class II 和 Class III 亚家族基因的 GC 含量分别相差 2.06%和 4.49%,CDS 的 GC 含量相差更少,分别仅 1.89%和 1.33%。Class II 和 Class III 亚家族基因编码蛋白质的理论等电点均小于 pI 6.5,显酸性;并且,不稳定指数均小于 40,为稳定蛋白,在一定程度上反映出其功能的稳定性。

与其他植物相比,拟南芥的芯片数据更加全面,并已得到广泛验证。本研究通过聚类分析比较了葡萄和拟南芥 *Actin* 基因家族中各基因编码蛋白质间的进化关系,认为这 2 种植物间的进化关系具有保守性,同一亚家族中的 *Actin* 基因可能具有相似的基因功能。葡萄 *Actin* 基因家族中,Class II 亚家族基因在所有组织中的表达均较高,Class III 亚家族基因在花粉、雄蕊和花中的表达均较高,Class I 亚家族中多数基因在各组织中的表达均较低,这与拟南芥 *Actin* 基因家族的表达情况相似,因此,可根据拟南芥中已有功能注释的 *Actin* 基因,对葡萄 *Actin* 基因进行功能预测。拟南芥中 *ACT7* 基因(登录号 AT5G09810.1)不仅在愈伤组织形成过程中发挥着重要作用^[5],还可能对外界刺激反应最强烈的 *Actin* 基因^[24]。葡萄 *Actin* 基因家族中登录号 GSVIVG01026580001 和 GSVIVG01003099001 的基因与拟南芥中 *ACT7* 基因共同聚在 Class II 亚家族中,推测这 2 个基因可能具有相似的功能。拟南芥中 *ACT12* (登录号 AT3G46520.1)和 *ACT4* (登录号 AT5G59370.1)基因主要在花粉管伸长过程中表达^[23],推测与二者共同聚在 Class III 亚家族中登录号 GSVIVG01016550001 的基因也可能在花粉管伸长过程中起重要作用。说明葡萄 *Actin* 基因家族的表达在调控葡萄生长发育过程中发挥着重要作用。3 个亚家族中基因的表达具有组织特异性,暗示 *Actin* 基因家族中不同基因具有特异性功能,但每个亚家族中的基因对葡萄生长发育进程的具体调控机制尚有待进一步研究。

本研究采用生物信息学方法,对葡萄 *Actin* 基因家族进行全基因组分析,为验证其基因功能提供了重要参考信息,但基因的具体功能还需进一步研究。

参考文献:

- [1] 阎龙飞, 石德权. 高等植物中的收缩蛋白[J]. 生物化学与生物物理学报, 1963, 3(4): 490-496.
- [2] 刘曦, 张少斌, 汪澈. 植物肌动蛋白功能的研究进展[J]. 生物技术通报, 2010(3): 13-16.
- [3] MASCARENHAS J P. Molecular mechanisms of pollen tube growth and differentiation[J]. The Plant Cell, 1993, 5: 1303-1314.
- [4] ITO Y, ABE S, DAVIES E. Co-localization of cytoskeleton proteins and polysomes with a membrane fraction from peas[J]. Journal of Experimental Botany, 1994, 45: 253-259.
- [5] KANDASAMY M K, MCKINNEY E C, MEAGHER R B. Functional nonequivalency of actin isoforms in *Arabidopsis* [J]. Molecular Biology of the Cell, 2002, 13: 251-261.
- [6] BAIRD W V, MEAGHER R B. A complex gene superfamily encodes actin in petunia [J]. The EMBO Journal, 1987, 6: 3223-3231.
- [7] 曹晓风, 王荣臣, 阎隆飞, 等. 豌豆卷须 cDNA 文库构建及肌动蛋白基因序列分析[J]. 科学通报, 1993, 38(19): 1804-1808.
- [8] MCDOWELL J M, HUANG S, MCKINNEY E C, et al. Structure and evolution of the actin gene family in *Arabidopsis thaliana* [J]. Genetics, 1996, 142: 587-602.
- [9] AN Y Q, HUANG S, MCDOWELL J M, et al. Conserved expression of the *Arabidopsis ACT1* and *ACT3* actin subclass in organ primordia and mature pollen[J]. The Plant Cell, 1996, 8: 15-30.
- [10] HUANG S, AN Y Q, MCDOWELL J M, et al. The *Arabidopsis thaliana ACT4/ACT12* actin gene subclass is strongly expressed throughout pollen development[J]. The Plant Journal, 1996, 10: 189-202.
- [11] 胡松年, 阎隆飞. 豌豆卷须肌动蛋白 II 类异型体 cDNA 克隆的序列分析[J]. 中国生物化学与分子生物学报, 1999, 15(6): 857-860.
- [12] 凌毅, 赵武玲. 豌豆肌动蛋白异型体基因的特异性表达[J]. 植物学通报, 2001, 18(1): 76-80.
- [13] JIANG Y Q, ZHAO W L. Expression and phylogenetic analysis of pea actin isoforms [J]. 植物学报(英文版), 2002, 44(12): 1456-1461.
- [14] JAILLON O, AURY J M, NOEL B, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla[J]. Nature, 2007, 449: 463-468.
- [15] FASOLI M, SANTO S D, ZENONI S, et al. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program [J]. The Plant Cell, 2012, 24: 3489-3505.
- [16] 郭景康, 陈青云, 戢茜, 等. 拟南芥、水稻和杨树 ACTIN 家族全基因组分析[J]. 上海大学学报(自然科学版), 2009, 15(4): 426-431.
- [17] ARTIMO P, JONNALAGEDDA M, ARNOLD K, et al. ExPASy: SIB bioinformatics resource portal [J]. Nucleic Acids Research, 2012, 40: W597-W603.
- [18] 董晨, 贾利强, 决登伟, 等. 香蕉 α -淀粉酶基因家族的系统进化分析[J]. 中国南方果树, 2016, 45(5): 12-17, 22.
- [19] GUNNING P W, GHOSH DASTIDER U, WHITAKER S, et al. The evolution of compositionally and functionally distinct actin filaments [J]. Journal of Cell Science, 2015, 128: 2009-2019.
- [20] ŠLAJCHEROVÁ K, FIŠEROVÁ J, FISCHER L, et al. Multiple actin isoforms in plants: diverse genes for diverse roles? [J]. Frontiers in Plant Science, 2012, 3: 226.
- [21] MEAGHER R B, MCKINNEY E C, KANDASAMY M K. Isovariant dynamics expand and buffer the responses of complex systems: the diverse plant *actin* gene family [J]. The Plant Cell, 1999, 11: 995-1005.
- [22] 魏潇, 刘威生, 刘宁, 等. 果实软化相关 PG 基因的进化分析和基因组定位[J]. 园艺学报, 2011, 38(9): 1791-1799.
- [23] 张少斌, 刘国琴. 植物肌动蛋白异型体研究进展[J]. 植物学通报, 2006, 23(3): 242-248.
- [24] MCDOWELL J M, AN Y Q, HUANG S, et al. The *Arabidopsis ACT7* actin gene is expressed in rapidly developing tissues and responds to several external stimuli [J]. Plant Physiology, 1996, 111: 699-711.

(责任编辑: 张明霞)