

doi:10.11937/bfyy.20170746

甘蓝 *ARF* 基因家族的鉴定与生物信息学分析

卫聪聪¹, 江定², 吴根堃², 陈长明²

(1. 黄河水利职业技术学院, 河南 开封 475000; 2. 华南农业大学 园艺学院, 广东 广州 510642)

摘要:生长素响应因子(auxin response factor, ARF)基因家族是植物特有的转录因子家族,在调控植物生长发育过程中起重要作用。以公布的甘蓝基因组数据库为基础,采用生物信息学方法鉴定出 29 个 *ARF* 基因,命名为 *BoARFs*,研究分析了其基因结构、染色体分布、蛋白质保守结构域、系统进化树及表达模式,以期对甘蓝 *ARF* 基因家族功能的深入研究提供依据。结果表明:*BoARF* 家族基因结构相对复杂,含有 2~14 个外显子;在染色体上呈不均匀分布,除 *BoARF29* 基因未定位到染色体上外,其它基因均定位在不同的染色体上;所有 *BoARF* 蛋白均具有保守的 B3 结构域和 Auxin_resp 结构域,但只有 21 个 *BoARF* 蛋白包含 1~2 个 AUX/IAA 结构域;转录组数据表达模式分析表明,*ARF* 基因具有不同的组织表达模式,部分基因表现出组织特异性。

关键词:甘蓝;*ARF*;生物信息学

中图分类号:S 635.601 **文献标识码:**A **文章编号:**1001-0009(2017)19-0016-09

生长素响应因子(auxin response factor, ARF)家族是一类调控生长素响应基因表达的转录因子^[1],能够特异地与生长素响应基因启动子区域的生长素响应元件(auxin response element AuxREs)TGTCTC 结合、激活或抑制基因的表达^[1-4]。绝大多数 ARF 蛋白具有 3 个保守的结构域,在 N-端有植物特异的 B3 DNA 结合结构域(DNA-binding domain, DBD),该结构域是植物特有的而且保守性强,通常 DBD 不能结合生长素应答基因,但具有 DNA 结合活性,能直接与生长素调控的下游基因启动子区的 AuxREs 结合;中间区域是 Auxin_resp 结构域,具有激活结构域(AD)或抑制结构域(RD)活性,具有激活或抑制转录子活性的作用,如果 ARF 结合生长素应答元件就起到激活作用,如果结合生长素就起到抑

制作用,具有激活作用的结构域含有丰富的谷氨酰胺,具有抑制作用的结构域富含丝氨酸、脯氨酸、甘氨酸和色氨酸;C 末端是 AUX/IAA 结构域,介导 AUX/IAAs 和 ARFs 的同源二聚化和异源二聚化的区域,调节 ARFs 与 AuxREs 的结合并调控下游生长素响应基因的表达^[1-2,4-5]。

甘蓝(*Brassica oleracea* L.)属十字花科芸苔属一二年生草本植物,是我国重要蔬菜之一。2014 年,甘蓝全基因组测序完成,为甘蓝分子生物学研究奠定了良好的基础^[6]。*ARF* 基因在植物体内都是以家族的形式存在,在拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)、玉米(*Zea mays* L.)、番茄(*Solanum lycopersicum*)、葡萄(*Vitis vinifera*)、杨树(*Populus trichocarpa*)、普通烟草(tobacco)、谷子(*Setaria italica*)、大麦(*Hordeum vulgare*)中分别鉴定出了 23、25、31、17、20、39、50、24、17 个 *ARF* 成员^[7-14]。目前,甘蓝中还没有 *ARF* 基因家族分析的报道,也缺少对该基因家族的生物信息学分析和表达分析,因此有必要在甘蓝基因组测序完成的基础上对 *ARF* 基因家族进行系统分析。该研究利用生物信息学方法在全基因组水平上鉴定甘蓝 *ARF* 基因,并对其基因结构、染色体分布、蛋白保守结构

第一作者简介:卫聪聪(1983-),女,硕士,讲师,研究方向为观赏园艺。E-mail:weicongcong@yrceti.edu.cn.

责任作者:陈长明(1984-),男,博士,副教授,现主要从事蔬菜遗传育种与生物技术教学与科研工作。E-mail:cmchen@scau.edu.cn.

基金项目:广东省科技计划资助项目(2015A020209117)。

收稿日期:2017-04-06

域、分子进化及表达模式进行分析,以期为进一步探讨甘蓝 *ARF* 家族成员的生物学功能奠定理论基础。

1 材料与方法

1.1 数据来源

甘蓝全基因组数据下载来源: <http://plants.ensembl.org/info/website/ftp/index.html>; 23 个拟南芥 *ARF* 基因数据下载来源: <https://www.ncbi.nlm.nih.gov/>; 芥蓝转录组数据来自课题组测得的转录数据^[15]。

1.2 试验方法

1.2.1 甘蓝 *ARF* 基因家族成员的鉴定及其编码蛋白质的理化性质分析

首先用甘蓝全基因组序列 (<http://plants.ensembl.org/info/website/ftp/index.html>) 构建本地 BLAST 数据库,然后以已报道的 23 个拟南芥 *ARF* 基因序列为 Query 序列,利用 BLAST 程序 ($1e-10$) 搜索本地基因组数据库,搜索这些序列的相似序列,删除重复序列,得到无重复的基因。然后把甘蓝 *ARF* 候选基因的蛋白质序列利用在线网站 CDD (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) 和 SMART (<http://smart.embl-heidelberg.de/>) 进行保守域结构分析^[16-17],筛选含有 B3、AUX_I AA、AUX_I AA 结构域的序列。最后利用 ExPASyProtParam (http://web.expasy.org/compute_pi/) 对所有 *ARF* 蛋白序列进行分子量、等电点预测。

1.2.2 甘蓝 *ARF* 基因的结构和染色体定位分析

从 Gramene 数据库中检索甘蓝 *ARF* 基因的信息,包括内含子、外显子、染色体上的物理位置信息。利用在线网站 GSDS (<http://gsds.cbi.pku.edu.cn/>) 分析甘蓝 *ARF* 基因编码序列和完整序列得到内含子-外显子结构。利用 MapDraw 软件构建染色体定位图。按照甘蓝 *ARF* 基因在染色体上的排列顺序进行基因命名。

1.2.3 分子进化树的构建

利用 MEGA 7.0 软件对甘蓝和拟南芥的 *ARF* 基因 cDNA 序列进行翻译,将翻译的蛋白序列进行 phylogeny 分析,采用邻接法 (neighbor-joining, NJ) 构建甘蓝和拟南芥的 *ARF* 蛋白系统进化树,其中,校验参数 (bootstrap) 设置为 1 000,

其余均为默认参数。

1.2.4 甘蓝 *ARF* 基因家族的表达模式分析

因为芥蓝 (*Brassica oleracea* var. *alboglabra* Bailey) 是甘蓝的变种,它们的基因同源性非常高,表达模式相似,因此利用课题组测定的芥蓝转录组数据初步分析 29 个 *BoARF* 基因的表达模式。找到 *BoARF* 在芥蓝中的同源基因,从芥蓝转录组数据库调取芥蓝中同源 *ARF* 基因不同组织部位的表达量数据^[15],得到 *ARF* 基因在甘蓝的老叶、成熟叶、嫩叶、叶脉、叶柄、嫩茎杆肉、中间茎杆肉、主茎外皮、花蕾、根茎部和根中的特异性表达情况。

2 结果与分析

2.1 甘蓝 *ARF* 基因家族成员鉴定及染色体定位

利用已报道的 23 个拟南芥 *ARF* 基因序列与本地建立好的甘蓝全基因组数据库中进行本地 BLASTP 比对,得到近 50 个甘蓝 *ARF* 候选基因,通过使用 SMART 和 CDD 检测蛋白序列结构域,检查缺失并除去冗余,得到确定含有 *ARF* 结构域的基因,最终获得 29 个甘蓝 *ARF* 基因序列,按照甘蓝 *ARF* 基因在染色体上的排列顺序分别命名为 *BoARF1*~*BoARF29* (表 1)。分析甘蓝 29 个 *ARF* 蛋白序列发现不同蛋白差异很大:编码 *ARF* 蛋白的氨基酸长度范围为 543 (*BoARF18*)~1 174 aa (*BoARF4*、*BoARF28*),编码区的长度为 1 632 (*BoARF18*)~3 525 bp (*BoARF4*、*BoARF28*),分子量范围为 59 727.75 (*BoARF18*)~129 754.34 Da (*BoARF28*),等电点范围为 5.26 (*BoARF18*)~7.59 (*BoARF1*)。

甘蓝共含有 9 条染色体,全部染色体上均含有 *ARF* 基因 (图 1),其中第 4 条染色体上含的 *ARF* 基因最多,为 6 个;第 1、3、5、7、8、9 条染色体上含 3 个 *ARF* 基因;第 2、6 条染色体含量最少,为 2 个 *ARF* 基因;还有一个 *ARF* 基因没有定位到染色体上,为 *BoARF29*。

2.2 甘蓝 *ARF* 基因的结构和系统进化树

利用 29 个甘蓝 *ARF* 蛋白全长序列构建了系统进化树 (图 2A),更好地了解甘蓝 *ARF* 基因系统进化关系。29 个 *ARF* 基因形成了 11 个旁系同源基因对,其中 5 对基因步长值 (bootstrap values) 为 99,表明这 5 对 *ARF* 基因的亲缘关系非常近,基因序列非常相近,可能这些基因存

表 1
Table 1
甘蓝 ARF 家族基因的基本信息
Basic information of ARF gene family in *Brassica oleracea*

基因名 Gene name	基因号 Gene ID	染色体 Chromosome distribution	位置 Location	外显子数 No. of extron	编码区长度 Coding sequence length/bp	氨基酸 Amino acid/aa	分子量 Molecular weight/Da	等电点 Isoelectric point	拟南芥的同源基因 <i>Arabidopsis</i> homologous gene
BoARF1	Bo1g014600.1	1	4 832 143 ~4 834 545	3	1 956	651	71 906.42	7.59	AtARF16
BoARF2	Bo1g037360.1	1	11 279 286~11 282 533	12	1 830	609	69 110.45	5.95	AtARF9
BoARF3	Bo1g090350.1	1	26 703 351 ~26 707 114	14	1 989	662	73 478.58	5.97	AtARF1
BoARF4	Bo2g023110.1	2	5 831 679 ~5 836 716	13	3 525	1 174	129 795.23	6.36	AtARF7
BoARF5	Bo2g161810.1	2	51 054 759~51 058 451	14	2 499	832	92 897.96	6.12	AtARF2
BoARF6	Bo3g039330.1	3	15 697 709~15 700 496	12	1 692	563	63 229.87	6.07	AtARF11
BoARF7	Bo3g108130.1	3	38 649 794~38 653 509	14	2 547	848	94 489.99	6.43	AtARF2
BoARF8	Bo3g148950.1	3	52 865 577~52 869 734	13	2 550	849	93 716.25	6.11	AtARF6
BoARF9	Bo4g008150.1	4	952 313~955 297	12	1 752	583	65 667.65	6.88	AtARF11
BoARF10	Bo4g041340.1	4	9 444 755~9 447 390	9	1 827	608	66 607.46	6.52	AtARF3
BoARF11	Bo4g071620.1	4	15 124 906~15 127 289	4	2 112	703	77 829.81	6.56	AtARF10
BoARF12	Bo4g083590.1	4	19 696 761~19 700 366	13	2 049	682	76 161.90	6.29	AtARF1
BoARF13	Bo4g136370.1	4	35 291 685~35 295 432	14	2 463	820	90 502.65	5.96	AtARF8
BoARF14	Bo4g182310.1	4	48 098 381~48 100 997	10	1 785	594	65 639.33	6.73	AtARF3
BoARF15	Bo5g027250.1	5	9 419 604~9 423 812	11	3 159	1 052	116 626.96	6.19	AtARF19
BoARF16	Bo5g027930.1	5	9 951 648~9 955 879	12	2 643	880	97 244.45	5.84	AtARF5
BoARF17	Bo5g062810.1	5	19 869 627~19 873 407	14	2 703	900	99 001.81	5.87	AtARF6
BoARF18	Bo6g081980.1	6	25 651 558~25 653 293	2	1 632	543	59 727.75	5.26	AtARF17
BoARF19	Bo6g099440.1	6	31 625 877~31 629 680	14	2 334	777	87 012.40	5.92	AtARF8
BoARF20	Bo7g062090.1	7	23 050 812~23 054 435	12	2 583	860	95 380.51	5.71	AtARF5
BoARF21	Bo7g109250.1	7	43 195 773~43 199 130	13	1 896	631	71 227.49	6.15	AtARF9
BoARF22	Bo7g114200.1	7	45 274 317~45 277 353	4	1 965	654	72 137.63	6.61	AtARF16
BoARF23	Bo8g068380.1	8	22 374 342~22 378 596	11	3 063	1 020	113 085.42	6.20	AtARF19
BoARF24	Bo8g069820.1	8	22 709 952~22 713 372	12	2 541	846	93 808.66	5.60	AtARF5
BoARF25	Bo8g097320.1	8	33 079 230~33 081 871	12	1 686	561	62 937.75	6.02	AtARF18
BoARF26	Bo9g014770.1	9	4 562 333~4 565 949	14	2 490	828	92 567.60	6.34	AtARF2
BoARF27	Bo9g143570.1	9	43 403 705~43 407 113	12	2 310	769	85 178.70	6.29	AtARF4
BoARF28	Bo9g151530.1	9	44 981 231~44 985 966	13	3 525	1 174	129 754.34	6.63	AtARF7
BoARF29	Bo01141s010.1	未知	7 927~10 651	11	1 668	555	62 529.32	6.12	AtARF18

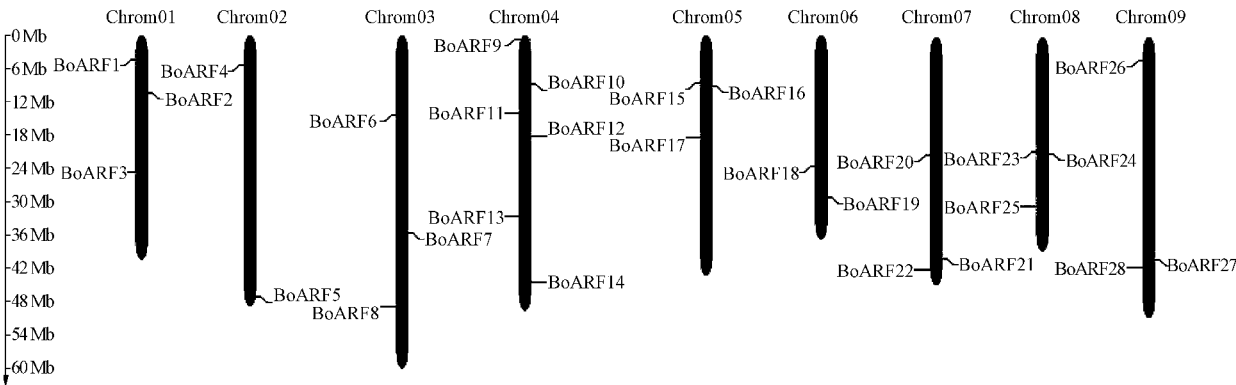


图 1 甘蓝 ARF 基因的染色体定位
Fig. 1 Chromosome location of the ARF genes in *Brassica oleracea*

在冗余功能。为进一步研究甘蓝 *ARF* 基因的特征,使用 GSDS 2.0 在线软件对 *ARF* 基因结构进行分析(图 2B),结果表明,*ARF* 基因家族大多数成员的结构比较复杂,含有外显子从 2 个到 14 个不等,大多数基因的外显子在 10 个以上,其中只有 *BoARF21* 基因含有 5' 上游序列。在 Gramene 数据库中通过 Gbrowse 查看成簇分布的 *ARF* 基因家族成员与侧翼蛋白质编码基因的位置关系,同一家族的不同成员如果位于同一个或相邻的基

因间区域,则成员间为串联重复关系,按照此标准,甘蓝 *ARF* 基因家族的扩增不是通过串联重复实现的。甘蓝 *ARF* 基因的扩增分为 2 种模式:分散复制与片段复制。从进化树的聚类分析可知,同一个基因的扩增或者复制的蛋白聚在一起,如 *BoARF20* 和 *BoARF24* 是片段复制聚在一起,*BoARF15* 和 *BoARF23* 是分散复制聚在一起。从基因的结构特征可以初步分析基因的复制和扩增方式。

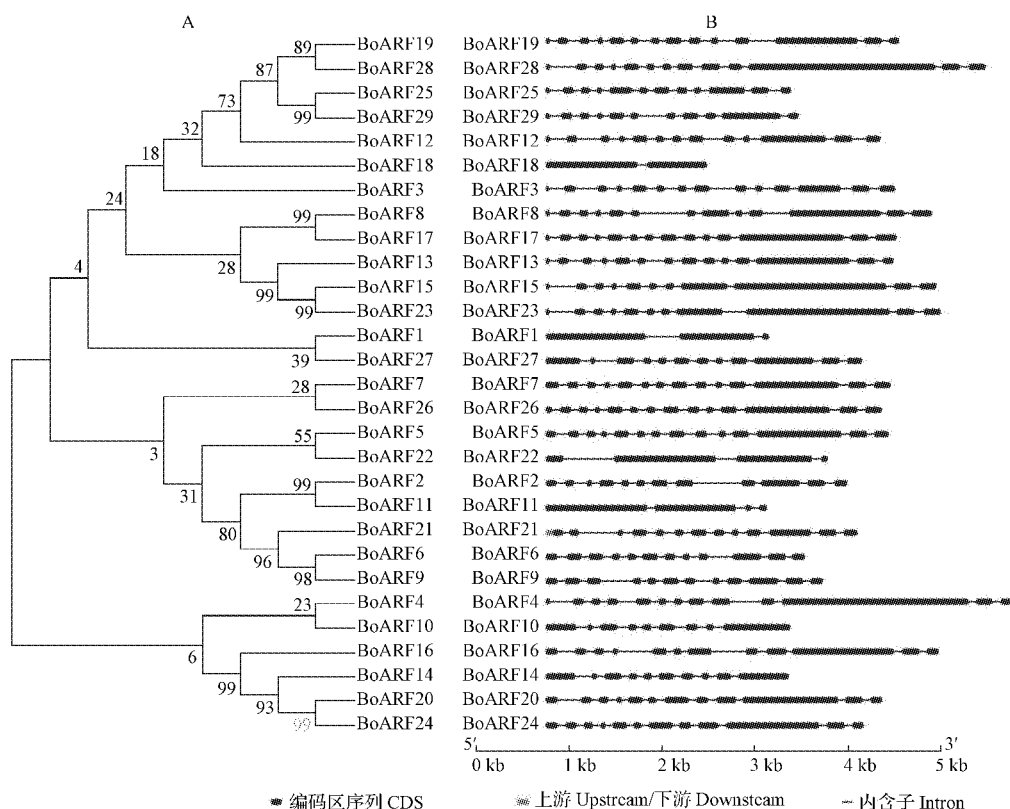


图 2 甘蓝 *ARF* 家族系统进化树(A)及基因结构(B)

Fig. 2 Phylogenetic tree(A) and gene structure of *ARF* gene family(B) in *Brassica oleracea*

构建了甘蓝与拟南芥 *ARF* 基因的系统进化树,进一步分析甘蓝 *ARF* 基因与拟南芥同源基因的进化关系(图 3)。根据甘蓝与拟南芥构建进化树中基因之间的亲缘关系,将 *BoARF* 基因分为 Class I、II 和 III(IIIa、IIIb、IIIc)这 5 类,其中甘蓝在 Class I 有 7 条 *BoARF* 基因,Class II 含有 7 个 *BoARF* 基因,8 个 *BoARF* 基因属于 IIIa,6 个 *BoARF* 基因属于 IIIb,1 个 *BoARF* 基因属于 IIIc。

将甘蓝的 *ARF* 成员与模式植物拟南芥进行了同源比对分析,寻找在拟南芥中的直系同源基

因,并借鉴拟南芥中基因功能的研究成果预测甘蓝 *ARF* 基因功能,其中拟南芥中研究的 *ARF* 基因功能包括促进顶端优势、侧根形成、花器官发育、响应逆境胁迫等。

2.3 甘蓝 *ARF* 蛋白的结构域分析

利用 SMART (<http://smart.embl-heidelberg.de/>)在线网站分析了甘蓝 *ARF* 蛋白的结构域(表 2)。根据这些信息利用 IBS 1.0.2(Illustration for Biological Sequences)软件分析了甘蓝 29 个 *ARF* 基因蛋白保守结构域,由图 4 可知,所有甘蓝 *ARF* 蛋白均含有 N 端 DNA 结合域(B3

结构域),并具有激活或抑制活性的中间区域(Auxin_resp 结构域),21 个 ARF 蛋白含有 C 末端二聚化结构域(AUX/IAA 结构域),其中 BoARF2、BoARF3、BoARF9、BoARF21 和 BoARF29 分别含有 2 个 AUX/IAA 结构域,但

BoARF1、BoARF8、BoARF10、BoARF11、BoARF14、BoARF18、BoARF22 和 BoARF27 含有 2 个结构域,即只含有 B3 DNA 结合结构域和 Auxin_resp 结构域,不含 AUX/IAA 结构域。

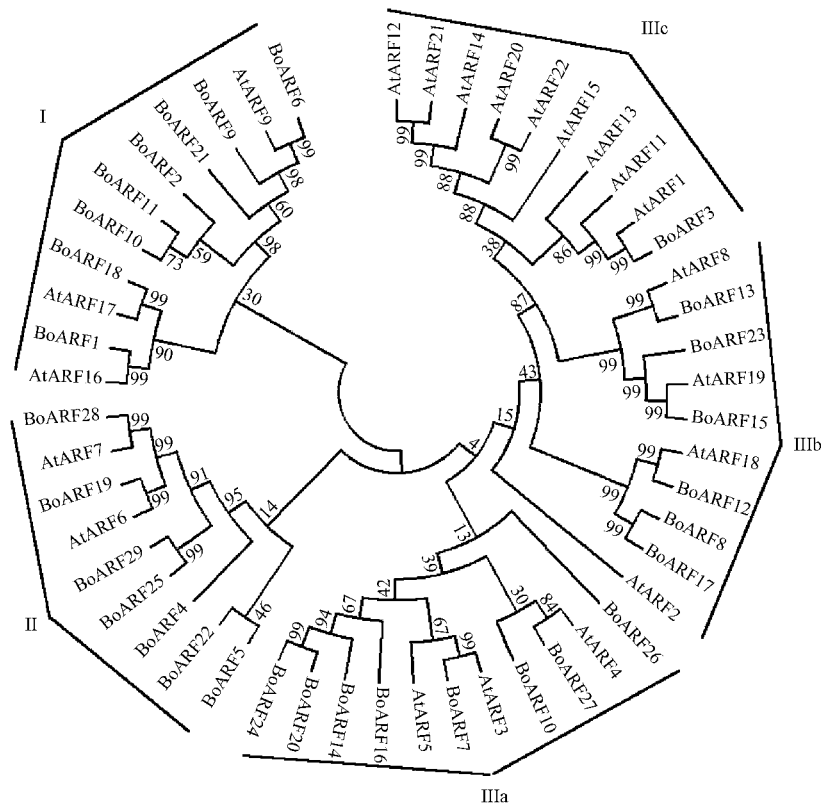


图 3 甘蓝与拟南芥的 ARF 转录因子的系统进化树
Fig. 3 Phylogenetic tree of ARF transcription factors in *Brassica oleracea* and *Arabidopsis thaliana*

表 2 甘蓝 ARF 蛋白结构域的基本信息
Table 2 Basic information of the *Brassica oleracea* ARF protein structure domain

基因名 Gene name	基因号 Gene ID	氨基酸数 Amino acid/aa	Start	B3			Auxin_resp			AUX/IAA		
				End	E-value		Start	End	E-value	Start	End	E-value
BoARF1	Bo1g014600.1	651	118	220	7.82e-22		276	359	4.4e-31			
BoARF2	Bo1g037360.1	609	112	214	9.12e-23		238	317	2.1e-33	465	541	3.2e-7
BoARF3	Bo1g090350.1	662	124	226	2.30e-19		250	332	2.5e-34	534	587	0.000016
BoARF4	Bo2g023110.1	1174	145	247	1.02e-20		271	353	3.0e-33	507	578	0.000081
BoARF5	Bo2g161810.1	832	149	251	1.90e-20		275	357	2.7e-37	575	630	0.000029
BoARF6	Bo3g039330.1	563	116	218	1.15e-19		242	321	1.5e-27	992	1092	1.9e-9
BoARF7	Bo3g108130.1	848	159	261	6.16e-20		285	367	5.6e-37	750	805	0.000066
BoARF8	Bo3g148950.1	849	129	231	2.37e-21		255	338	8.2e-34	423	498	2.6e-8
BoARF9	Bo4g008150.1	583	115	218	3.94e-18		242	321	5.0e-29	681	813	1.1e-10
BoARF10	Bo4g041340.1	608	160	262	1.09e-20		286	368	2.3e-31	440	512	3.5e-8
BoARF11	Bo4g071620.1	703	112	214	3.08e-20		280	363	3.4e-32	508	562	6.4e-7

表 2(续)
Table 2(Continued)

基因名 Gene name	基因号 Gene ID	氨基酸数 Amino acid/aa	B3			Auxin_resp			AUX/IAA		
			Start	End	E-value	Start	End	E-value	Start	End	E-value
BoARF12	Bo4g083590.1	682	128	229	1.64e-15	253	333	3.4e-32	531	654	2.6e-10
BoARF13	Bo4g136370.1	820	126	228	4.13e-21	252	335	2.3e-31	685	804	8.0e-8
BoARF14	Bo4g182310.1	594	156	258	1.34e-20	282	364	2.0e-30			
BoARF15	Bo5g027250.1	1 052	126	228	2.09e-22	252	334	8.2e-34	907	1 014	8.5e-10
BoARF16	Bo5g027930.1	880	155	257	5.32e-24	281	364	7.4e-31	740	860	1.0e-8
BoARF17	Bo5g062810.1	900	129	231	9.12e-23	255	338	9.3e-34	726	848	4.1e-8
BoARF18	Bo6g081980.1	543	119	221	1.97e-18	253	333	3.1e-18			
BoARF19	Bo6g099440.1	777	127	229	1.90e-20	253	336	2.8e-31	650	766	5.6e-8
BoARF20	Bo7g062090.1	860	156	258	5.32e-24	282	365	8.6e-31	721	841	1.4e-9
BoARF21	Bo7g109250.1	631	115	217	6.01e-23	241	320	4.8e-32	483	563	4.4e-8
BoARF22	Bo7g114200.1	654	117	219	8.25e-21	276	359	2.1e-30	552	605	0.0 000 035
BoARF23	Bo8g068380.1	1 020	126	228	1.85e-23	252	334	2.0e-32	878	982	2.8e-9
BoARF24	Bo8g069820.1	846	156	258	5.32e-24	282	365	1.5e-30	710	830	5.7e-9
BoARF25	Bo8g097320.1	561	127	229	6.50e-19	253	332	1.7e-29	432	547	3.9e-13
BoARF26	Bo9g014770.1	828	137	239	1.90e-20	263	345	7.4e-37	589	794	4.4e-11
BoARF27	Bo9g143570.1	769	175	277	1.80e-21	300	382	1.6e-33			
BoARF28	Bo9g151530.1	1 174	127	229	2.21e-21	253	335	1.2e-33	1 027	1 137	3.4e-9
BoARF29	Bo01141s010.1	555	127	229	5.28e-19	253	332	1.2e-29	428	491	2.9e-7
									487	540	0.0 000 036

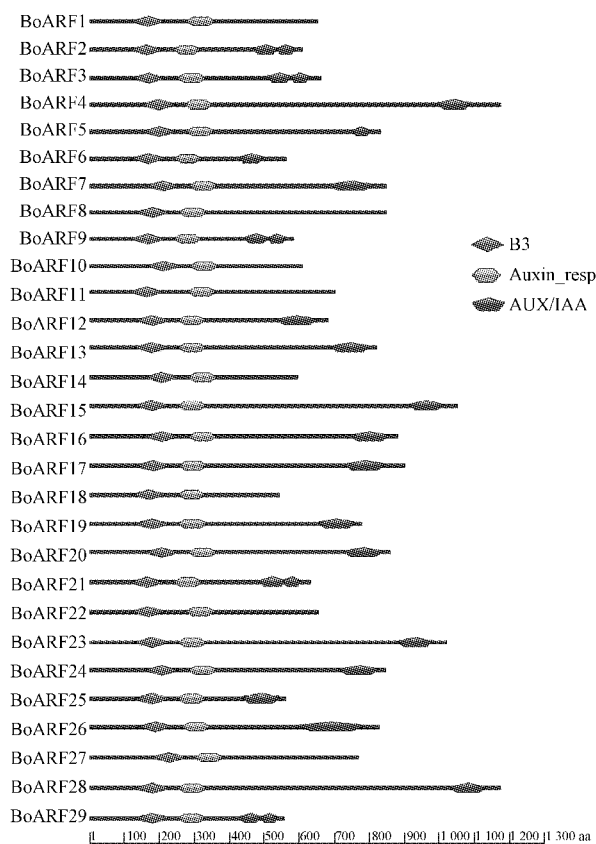
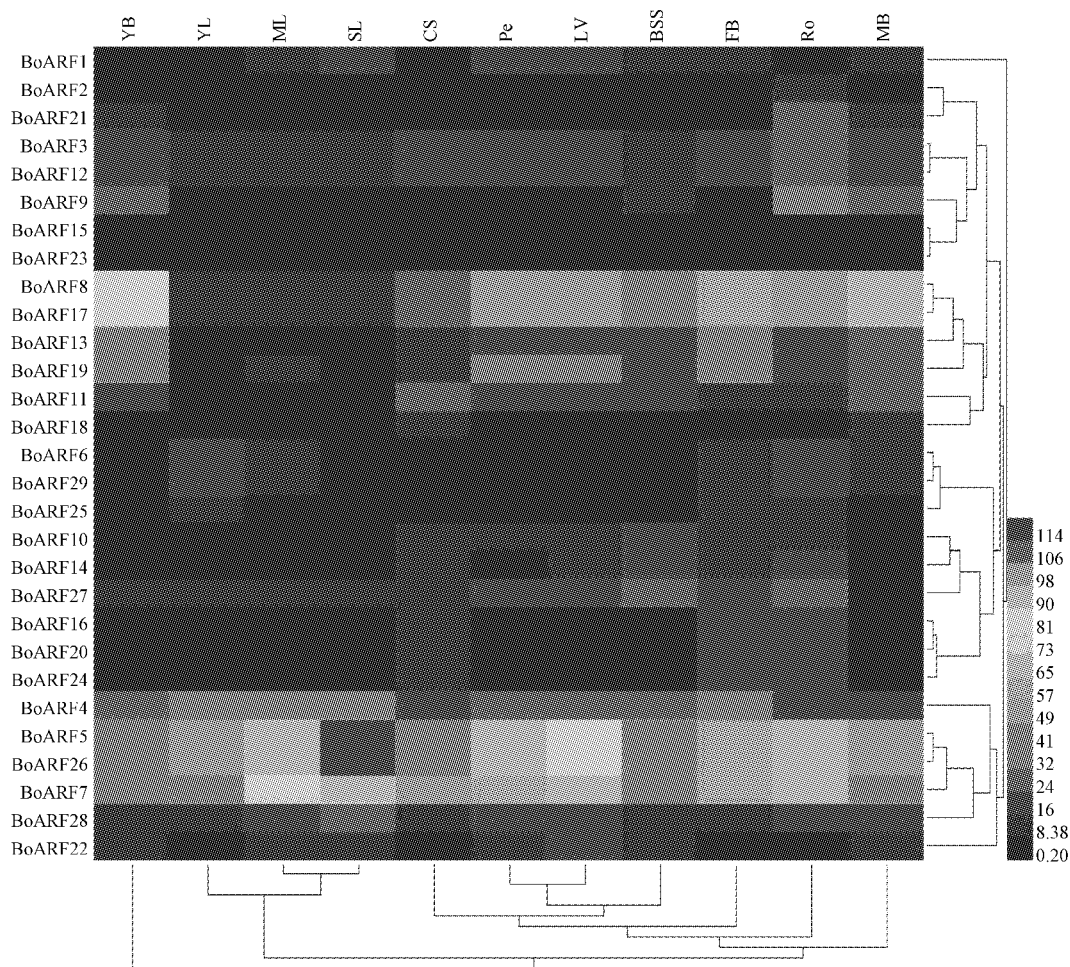


图 4 甘蓝 ARF 蛋白的结构域
Fig. 4 Domain structure of ARF proteins in *Brassica oleracea*

2.4 甘蓝 ARF 基因组织表达分析

为进一步了解甘蓝 *BoARF* 基因的特性和功能作用,对基因的组织特异表达情况进行分析,获得甘蓝 *BoARF* 基因在 11 个组织中的表达量值,并利用 HemI 1.0 软件^[18]构建聚类热图(图 5)。可知,多数基因是组成型表达,在不同的组织中均检测到基因的表达, *BoARF2*、*BoARF15*、*BoARF18*、*BoARF23* 和 *BoARF25* 基因在 11 个组织中的表达量比较低。 *BoARF5*、*BoARF7* 和

BoARF26 基因在 11 个组织中的表达量普遍都比较高,这可能暗示着这 3 个基因有可能发挥多方面的作用。不同的基因在不同组织部位表达量也是存在差异的, *BoARF8* 和 *BoARF17* 在嫩茎秆肉、中间茎秆肉中表达量是最高的, *BoARF5* 和 *BoARF26* 在老叶、成熟叶、嫩叶、叶脉、叶柄、主茎外皮、花蕾、根茎部、根表达量是最高的。同一基因在不同部位的表达量有些基因差异比较大、有些基因差异不大,还需要进一步研究。



注:SL,老叶;ML,成熟叶;YL,嫩叶;LV,叶脉;Pe,叶柄;YB,嫩茎秆肉;MB,中间茎秆肉;BS,主茎外皮;FB,花蕾;CS,根茎部;Ro,根。

Note:SL,senescent leaf;ML,mature leaf;YL,young leaf;LV,leaf vein;Pe,petiole;YB,young bolting stem flesh;MB,middle bolting stem flesh;BS,bolting stem skin;FB,flower buds;CS,combining sites;Ro,roots.

图 5 甘蓝 ARF 基因在不同组织部位的表达热图

Fig. 5 Heatmap of ARF gene expression in different tissues of *Brassica oleracea*

3 讨论

该研究通过对甘蓝全基因组进行生物信息学分析,共鉴定出 29 个 *BoARF* 编码基因。把甘蓝 *ARF* 基因家族分成 3 类 Class I、II、III(IIIa、IIIb、IIIc)。*ARF* 基因结构复杂,外显子数目多(一般在 10 个以上)。通常,大多数 *BoARF* 蛋白都具有 3 个保守的区域,即在 N-端 B3 结构域,中间的 Auxin_resp 结构域以及在 C 端的 AUX/IAA 结构域^[2]。甘蓝的 29 个 *ARF* 蛋白都具有 B3 结构域,而且 21 个 *ARF* 蛋白含都有 3 个结构域,其中 *BoARF2*、*BoARF3*、*BoARF9*、*BoARF21* 和 *BoARF29* 分别含有 2 个 AUX/IAA 结构域,但 *BoARF1*、*BoARF8*、*BoARF10*、*BoARF11*、*BoARF14*、*BoARF18*、*BoARF22* 和 *BoARF27* 只含有 B3 结构域和 Auxin_resp 结构域。拟南芥的 23 个 *AtARF* 蛋白仅有 1 个没有 B3 结构域,其它蛋白质除去 *AtARF3*、*AtARF13* 和 *AtARF17* 外都含有 AUX/IAA 结构域。

WU 等^[19]发现拟南芥中的 *AtARF6* 基因调控种子心皮发育、促进花器官成熟并且能够响应逆境胁迫,通过 Blast 基因序列比对分析,发现甘蓝中的 *BoARF8* 和 *BoARF17* 与 *AtARF6* 基因高度同源,可推测甘蓝中的 *BoARF8* 和 *BoARF17* 基因可能能够调控甘蓝种子的心皮发育,促进甘蓝花器官成熟以及维持甘蓝在逆境胁迫环境中的正常生长。研究表明 *AtARF7* 和 *AtARF19* 作为转录激活子介导了拟南芥侧根形成过程^[20-21],暗示 *BoARF4*、*BoARF15*、*BoARF23* 和 *BoARF28* 基因可能存在相似的功能。OKUSHIMA 等^[22]和 FENG 等^[23]研究发现,拟南芥中 *AtARF11* 基因促进侧根形成,在激素信号途径中起关键作用,而在甘蓝中发现该基因的同源基因有 *BoARF6* 和 *BoARF9*,通过对拟南芥中 *AtARF11* 基因的研究,可推测甘蓝中这 2 个基因可能参与激素信号途径,促进侧根形成。拟南芥中的 *AtARF10*、*AtARF16* 和 *AtARF17* 基因的单突变体没有明显的表型,而 *arf10arf16* 双突变体根冠和根系向地性生长发生异常^[20,24],在甘蓝中共有 4 个 *BoARF* 基因属于这 3 个基因的同源基因,但是具体的功能还有待于进一步的探讨。

WU 等^[25]对番茄的根、茎、叶、芽、子房 5 个组织部位的 *ARF* 基因进行了表达量分析,大多数 *ARF* 基因都是组成型表达,不过每个基因在不同组织部位的表达差异还是比较大的,同一组织部位不同基因的表达量也存在差异,总体的表达模式与甘蓝的相似,并且与烟草^[12]、谷子^[13]、大麦^[14]也是相似。可以推测植物 *ARF* 基因家族的表达模式是一致的,但是每个基因又有它的不同,还有待进一步研究。

ARF 基因家族调控植物的发育和激素代谢,在植物发育过程中起到非常重要的作用。为了更好地了解甘蓝 *ARF* 生长素响应因子的结构和功能,该研究以甘蓝全基因组序列为背景,对甘蓝 *ARF* 基因家族的基因结构、染色体分布、蛋白保守结构域、分子进化及表达谱进行了分析,为将来深入研究该基因家族的表达调控、结构和功能等奠定了坚实的基础,为揭示该基因家族参与植物生长、发育的调控机制提供参考依据。然而,甘蓝 *ARF* 生长素响应因子是如何响应激素信号调控甘蓝的生长发育还需要今后进行深入的功能研究。

参考文献

- [1] TIWARI S B, HAGEN G, GUILFOYLE T. The roles of auxin response factor domains in auxin-responsive transcription[J]. *Plant Cell*, 2003, 15(2): 533-543.
- [2] GUILFOYLE T J, HAGEN G. Auxin response factors[J]. *Current Opinion in Plant Biology*, 2007, 10(5): 453-460.
- [3] ELLIS C M, NAGPAL P, JEFFERY C, et al. Auxin response factor 1 and auxin response factor 2 regulate senescence and floral organ abscission in *Arabidopsis thaliana* [J]. *Development*, 2005, 132(20): 4563-4574.
- [4] CHANDLER J W. Auxin response factors[J]. *Plant, Cell & Environment*, 2016, 39(5): 1014-1028.
- [5] 司马晓娇, 郑炳松. 植物生长素原初响应基因 *Aux/IAA* 研究进展[J]. *浙江农林大学学报*, 2015, 32(2): 313-318.
- [6] LIU S Y, LIU Y M, YANG X H, et al. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes[J]. *Nature Communications*, 2014, 5: 1-10.
- [7] REMINGTON D L, VISION T J, GUILFOYLE T J, et al. Contrasting modes of diversification in the *Aux/IAA* and *ARF* gene families[J]. *Plant Physiol*, 2004, 135(3): 1738-1752.
- [8] WANG D, PEI K, FU Y, et al. Genome-wide analysis of the auxin response factors (ARF) gene family in rice (*Oryza sativa*) [J]. *Gene*, 2007, 394(12): 13-24.
- [9] XING H Y, PUDAKE R N, GUO G G, et al. Genome-wide identification and expression profiling of auxin response factor

- (ARF) gene family in maize[J]. BMC Genomics, 2011, 12(1): 178.
- [10] KUMAR R, TYAGI A K, SHARMA A K. Genome-wide analysis of auxin response factor (ARF) gene family from tomato and analysis of their role in flower and fruit development [J]. Mol Genet Genomics, 2011, 285(3): 245-260.
- [11] KALLURI U C, DIFAZIO S P, BRUNNER A M, et al. Genome-wide analysis of *Aux/IAA* and *ARF* gene families in *Populus trichocarpa* [J]. BMC Plant Biol, 2007, 7(1): 59.
- [12] 孙亭亭, 张磊, 陈乐, 等. 普通烟草 *ARF* 基因家族序列的鉴定与表达分析[J]. 植物遗传资源学报, 2016, 17(1): 162-168.
- [13] 赵艳, 瓮巧云, 马海莲, 等. 谷子 *ARF* 基因家族的鉴定与生物信息学分析[J]. 植物遗传资源学报, 2016, 17(3): 547-554.
- [14] 郭宝健, 李赢, 袁泽宸, 等. 大麦 *ARF* 基因家族的全基因组分析[J]. 麦类作物学报, 2016(11): 1426-1432.
- [15] WU S, LEI J, CHEN G, et al. *De novo* transcriptome assembly of Chinese kale and global expression analysis of genes involved in glucosinolate metabolism in multiple tissues[J/OL]. Frontier in Plant Science, 2017, 2, doi:10.3389/fpls.2017.00092.
- [16] SCHULTZ J, MILPETZ F, BORK P, et al. SMART, a simple modular architecture research tool: Identification of signaling domains[J]. Proc Natl Acad Sci USA, 1998, 95(11): 5857-5864.
- [17] MARCHLER-BAUER A, ANDERSON J B, CHITSAZ F, et al. CDD: Specific functional annotation with the conserved domain database[J]. Nucleic Acids Res, 2009, 37(1): 205-210.
- [18] DENG W K, WANG Y B, LIU Z X, et al. HemI: A toolkit for illustrating heatmaps[J/OL]. PLoS One, 2014, 9: e11198811.
- [19] WU M F, TIAN Q, REED J W. *Arabidopsis* microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction [J]. Development, 2006, 133(21): 4211-4218.
- [20] OKUSHIMA Y, OVERVOORDE P J, ARIMA K, et al. Functional genomic analysis of the auxin response factor gene family members in *Arabidopsis thaliana*: Unique and overlapping functions of ARF7 and ARF19[J]. Plant Cell, 2005, 17(2): 444.
- [21] FINET C, BERNE-DEDIEU A, SCUTT C P, et al. Evolution of the *ARF* gene family in land plants: Old domains, new tricks[J]. Molecular Biology and Evolution, 2013, 30(1): 45-56.
- [22] OKUSHIMA Y, FUKAKI H, ONODA M, et al. ARF7 and ARF19 regulate lateral root formation via direct activation of *LBD/ASL* genes in *Arabidopsis* [J]. Plant Cell, 2007, 19(1): 118-130.
- [23] FENG Z H, ZHU J, DU X, et al. Effects of three auxin-inducible LBD members on lateral root formation in *Arabidopsis thaliana* [J]. Planta, 2012, 236(4): 1227-1237.
- [24] WANG J W, WANG L J, MAO Y B, et al. Control of root cap formation by MicroRNA-targeted auxin response factors in *Arabidopsis* [J]. Plant Cell, 2005, 17(8): 2204-2216.
- [25] WU J, WANG F, CHENG L, et al. Identification, isolation and expression analysis of auxin response factor (ARF) genes in *Solanum lycopersicum* [J]. Plant Cell Reports, 2011, 30(11): 2059-2073.

Genome-wide Identification and Bioinformatics Analysis of *ARF* Gene Family in *Brassica oleracea*

WEI Congcong¹, JIANG Ding², WU Genkun², CHEN Changming²

(1. Yellow River Conservancy Technical Institute, Kaifeng, Henan 475000; 2. Horticulture College, South China Agriculture University, Guangzhou, Guangdong 510642)

Abstract: Auxin response factor (ARF) gene family is the specific transcription factor in plant, and plays an important role in regulation of plant organ development. Based on the published genomic database of *Brassica oleracea*, 29 *ARF* genes were identified by bioinformatics, named *BoARFs*. Their genetic structure, chromosome distribution, conservative domain structure of protein, phylogenetic tree and expression pattern of reference were analyzed for the further research of *Brassica oleracea* *ARF* gene family function to provide the reference basis. The results showed that the gene structure of *BoARF* family was relatively complex and contained 2—14 exons. Except the *BoARF29*, all other genes were located in different chromosomes. All *BoARF* proteins had conserved B3 domain and Auxin_resp domain, while only 21 *BoARF* proteins contained 1 or 2 AUX/IAA domains. Transcriptome data analysis showed that *ARF* gene had different tissue expression patterns, and some genes showed tissue specificity.

Keywords: *Brassica oleracea*; auxin response factor; bioinformatics