

· 研究报告 ·

## 种子植物抗细胞凋亡DAD基因的演化

包颖<sup>†\*</sup>, 梅玉芹<sup>†</sup>

曲阜师范大学生命科学学院, 曲阜 273165

**摘要** 抗细胞凋亡基因(DAD)是一个高度保守的细胞凋亡抑制基因, 在植物生长发育中承担重要功能。为全面了解DAD基因在种子植物中的分布和演化规律, 该文利用31种植物的全基因组数据, 通过生物信息学手段, 深入探讨和分析了不同植物类群中DAD基因的拷贝数目、基因结构和染色体定位, 并综合另外7种裸子植物的转录组数据探讨了其在种子植物中的演化趋势。结果表明, DAD基因属于低拷贝基因, 在不同种子植物中只具有1–3个拷贝; 不同DAD基因编码的氨基酸长度在108–170 aa之间变动。同线性和系统发育分析进一步表明, 种子植物DAD基因的演化具有明显的谱系特异性。随机复制和染色体大片段复制及其随后的基因丢失可能是其维持低拷贝的重要方式。

**关键词** 染色体定位, DAD, 复制类型, 全基因组, 基因结构

包颖, 梅玉芹 (2017). 种子植物抗细胞凋亡DAD基因的演化. *植物学报* 52, 590–597.

程序性细胞死亡(programmed cell death, PCD)是指为了维持机体内部环境的稳定, 由基因控制的细胞自主、有序的死亡方式, 也称作细胞凋亡(apoptosis, APO) (Kerr et al., 1972), 是多细胞生物生长发育过程中普遍存在的现象。PCD由一系列内外因素诱导产生, 在细胞的分化和发育过程中发挥重要作用 (Schwartzman and Cidlowski, 1993; Williams and Smith, 1993; Raff, 1998)。目前对动物PCD的研究比较深入, 其作用机理、关联基因以及发挥作用的信号通路等大部分内容已被揭示。相对而言, 关于植物PCD的研究则较为滞后。尽管在植物发育的许多过程中均已经发现PCD的存在, 如导管的生成、胚柄的退化、糊粉层的消失、植物体的性别决定以及叶片等器官的衰老等 (Nakashima et al., 1993; 贾志蓉等, 2004), 但对其背后的遗传机制却知之甚少。

作为多细胞生物生长发育的必要过程, PCD在细胞内的出现有着精确的时空协调性。其中, 抗细胞凋亡基因(*Defender against Apoptotic Death*, DAD)作为其抑制基因负向调控PCD的发生 (Mondragón-Palomino et al., 2002; Wortman et al., 2003; 杨舒雅等, 2012)。研究表明, DAD是位于内质网上的寡糖基转移酶复合体的必要组分, 能够催化高甘露醇低聚

糖转移到新产生多肽的特定天冬酰胺残基上, 是催化某些N-糖基化的关键酶。DAD的功能或表达量异常会影响寡糖基转移酶功能的正常发挥, 并因此引发细胞凋亡的发生 (Kelleher and Gilmore, 1997; Danon et al., 2004)。目前, 从拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)、棉花(*Gossypium hirsutum*)及番茄(*Solanum lycopersicum*)等多种植物中已成功分离到DAD基因 (Roboti and High, 2012)。序列分析显示, 该基因高度保守, 在物种间同源性极高。例如, 棉花DAD基因与人(*Homo sapiens*)的DAD基因序列一致性达52% (龚文芳等, 2010); 此外, 早期的研究证明, 小鼠(*Mus musculus*)、人、爪蟾(*Xenopus laevis*)、线虫(*Caenorhabditis elegans*)以及水稻的DAD在编码序列上高度同源 (Apte et al., 1995)。Gallois等 (1997) 的实验证实, 拟南芥DAD基因可以对仓鼠(*Mesocricetus auratus*)的同源突变基因进行有效补偿。鉴于DAD在多细胞生物的许多组织和细胞中均有所表达, 因此推测其有可能参与更多生理过程, 在生物有机体的生长发育中起至关重要的作用, 同时该基因也有望在延缓或促进细胞凋亡方面成为基因工程和遗传改良的候选基因。

以往由于缺少必要的信息, 且仅有的研究又

收稿日期: 2016-06-21; 接受日期: 2017-01-10

基金项目: 国家自然科学基金(No.31570218)

† 共同第一作者

\* 通讯作者。E-mail: baoyingus@126.com

过分关注动物类群以及 *DAD* 基因的具体功能, 因此对于此类基因在种子植物各类群中的成员构成、基因结构及染色体分布等基本演化模式缺乏必要的了解, 从而严重限制了对植物 *DAD* 基因的整体把握和有效利用。二代测序以及生物信息处理技术的飞速发展以及大量基因组和转录组等数据的不断累积, 为我们全面梳理和总结种子植物 *DAD* 基因的起源和演化规律奠定了基础。本研究选取具有全基因组数据的31种植物, 比较分析了 *DAD* 基因在植物中, 特别是种子植物中的拷贝数目、基因长度、染色体定位和复制类型等基因特点; 同时, 增加了7种裸子植物的转录组数据, 利用其氨基酸编码序列重建了该基因的系统发育关系, 对 *DAD* 基因在种子植物内的演化历史进行了讨论。

## 1 材料与方法

### 1.1 序列收集

本研究首先选取29种具有全基因组序列的种子植物, 其中包括被子植物中的双子叶植物19种(*Arabidopsis lyrata* L.、*A. thaliana* L.、*Brassica rapa* L.、*Citrullus lanatus* (Thunb.) Matsum. & Nakai、*Cucumis melo* L.、*Capsella rubella* Reut.、*Eucalyptus grandis* W. Hill.、*Fragaria vesca* L.、*Gossypium raimondii* Ulbr.、*Malus domestica* Borkh.、*Manihot esculenta* Crantz.、*Prunus persica* (L.) Batsch.、*Populus trichocarpa* Torr. & A. Gray.、*Ricinus communis* L.、*Solanum lycopersicum* L.、*Solanum tuberosum* L.、*Theobroma cacao* L.、*Thellungiella parvula* (Schrenk) Al-Shehbaz & O'Kane和 *Vitis vinifera* L.), 单子叶植物7种(*Brachypodium distachyon* (L.) P. Beauv.、*Hordeum vulgare* L.、*Musa acuminata* Colla.、*Oryza sativa* ssp. *japonica* L.、*Sorghum bicolor* (L.) Moench.、*Setaria italica* (L.) P. Beauvois和 *Zea mays* L.), 被子植物基部类群1种(*Amborella trichopoda* Baill.), 以及裸子植物2种(*Picea abies* (L.) H. Karst.和 *Pinus taeda* L.). 同时, 为更好地了解 *DAD* 基因在种子植物之前的起源和演化历史, 还选取了1种苔藓植物(*Physcomitrella patens* (Hedw.) Bruch & Schimp.)作为参照。1种藻类植物(*Chlamydomonas reinhardtii* P. A. Dang.)被作为外类群。为了增加裸子植物的代表性, 本研究

也包括了只具有转录组数据的7种裸子植物(*Ginkgo biloba* L.、*Gnetum montanum* Markgr.、*Picea glauca* (Moench) Voss.、*Picea sitchensis* (Bong.) Carr.、*Pinus pinaster* Aiton.、*P. sylvestris* L.和 *Pseudotsuga menziesii* (Mirb.) Franco)。

以GenBank (<http://www.ncbi.nlm.nih.gov/>)数据库中拟南芥2个抗细胞凋亡基因(*AT1G32210*和 *AT2-G35520*)的核苷酸和氨基酸序列分别作为目标序列, 在植物比较基因组PLAZA网站(<http://bioinformatics.psb.ugent.be/plaza/>)的3个子数据库(Dicots PLAZA 3.0、Monocots PLAZA 3.0和Gymno PLAZA 1.0)中, 利用BLASTX程序进行同源查找。BLASTX的阈值设为 $1e^{-5}$ 。将搜寻到的基因组序列与NCBI (<http://www.ncbi.nlm.nih.gov/>)数据库不同植物相应的mRNA表达序列及转录组中的编码信息进行进一步核实与比对, 并结合 *DAD1* 的结构域(pfam02109)进行同步筛查, 去除冗余序列。为明确获得基因在各自基因组中的分布情况, 利用PLAZA数据库(<http://bioinformatics.psb.ugent.be/plaza/>)中的WGMMapping工具以及各基因组的注释信息, 将这些 *DAD* 基因定位到具有全基因组数据植物的染色体或Scaffold上。此外, 对于基因组内包含多个 *DAD* 拷贝的情况, 依据各拷贝所对应的染色体片段间的同线性关系, 判断这些 *DAD* 基因的复制是否可以归入如下类型: 不同拷贝所在染色体片段间存在明显同线性的为染色体大片段复制(Block)类型, 不同拷贝分布在同一染色体内的邻近位置连续出现为串联重复(Tandem)类型, 以及那些同线性比较中没有明显规律的类型, 这里暂定为随机复制(Random)类型。

### 1.2 系统发育关系分析

将完整基因编码的氨基酸序列导入SEAVIEW软件(Gouy et al., 2010), 并采用Clustal W方法进行序列比对。排列好的数据应用PhyML3.0 (Guindon et al., 2010)和MrBayes (Version 3.2) (Ronquist and Huel-senbeck, 2003)软件分别构建最大似然性(maximum likelihood, ML)和贝叶斯(Bayesian)系统发育关系树。其中最大似然性ML树在JJT模型下进行计算, 系统发育树各分支的置信度用100次重复的靴带分析(bootstrap, BS)进行检验。贝叶斯树采用GTR模型, 位点速率模型为gamma, 建立4个马尔科夫链, 运行 $1 \times 10^6$

代, 每100代抽样1次, 重复1次; 10 000代后舍弃老化样本, 剩余样本构建一致树, 各分支的置信度通过后验概率(BBP)来检验。2个树均以藻类的DAD基因作为外类群。

## 2 结果与讨论

### 2.1 种子植物DAD基因的拷贝数和基因特点

利用目标序列对植物基因组和转录组数据进行全面检索及对比筛选后, 我们共找到58个完整的DAD同源基因。它们在植物各类群均以较低的拷贝数出现。例如, 在外类群藻类植物*C. reinhardtii*中只发现1个DAD基因; 在早期陆生植物藓类植物*P. patens*中也仅发现2个DAD基因。在种子植物中, DAD基因的拷贝数仍然持续偏低, 为1–3个不等, 其中以2个居多。例如, 十字花科的3种植物(*A. thaliana*、*A. lyrata*和*C. rubella*)以及裸子植物*P. pinaster*和*P. sitchensis*等均具有2个DAD基因。此外, 我们还查询到4个(MA10-G06540、PAB00057214、PTA00022345和PTA000-22344)不完整的DAD基因以及2个(BR09G26700和PAB00021420)与目标序列有明显分歧的异常基因(见表1及补充材料)。其中位点BR09G26700在其3'端多出了112 aa; 而PAB00021420不但在3'端有增加, 在5'端还有较长的缺失, 与目标基因仅有67 aa能够匹配上; 从基因组的角度进一步分析, 发现位点BR09G26700之所以具有较长的氨基酸长度(201 aa), 是由于其在原来的第4个内含子上多剪切了1个外显子, 同时在原来终止子的位置增加了1个新的内含子和1个新的外显子, 但是这些新增加的序列在其它植物的DAD基因中均没有出现。由于上述不完整或异常的基因序列不能很好地代表植物DAD基因的普遍特点, 因此在进一步的分析中没有包括这6个位点。

分析完整基因的特点, 在长度方面, DAD基因编码的氨基酸长度在108–170 aa之间, 其中以114–119 aa居多(表1)。此外, 基于全基因组数据, 不同植物类群的DAD基因结构趋于稳定, 尽管在外类群*C. reinhardtii*中, DAD基因是由4个外显子和3个内含子组成; 但是从苔藓植物*P. patens*开始一直到所有种子植物, DAD基因倾向由5个外显子和4个内含子组成(表1)。完整基因的成对比较显示, 种子植物种间DAD

基因相似性在70%–100%之间变动, 最大序列分歧存在于*M. acuminata*和*P. sitchensis*的位点MA11G-08020和PSI00016644之间; 完全一致的序列出现在双子叶植物同属的位点ST08G021840和SL08G0-76460, 裸子植物同属的位点PAB00041636和PGL-00020840以及PSI00008404和PGL00010558之间。

染色体定位分析表明, 同一植物的不同DAD基因通常分布在不同的染色体上。如*A. thaliana*中2个DAD基因分别位于1号和2号染色体上。进一步根据DAD基因所在染色体片段在植物类群内的同线性分析, 发现DAD的多拷贝主要是通过染色体大片段复制(19个)或随机复制(14个)形成的, 没有任何串联复制类型存在(表1)。例如, 双子叶植物*A. thaliana*、*G. raimondii*和*P. trichocarpa*等各自的2个拷贝均是染色体大片段复制的产物; 而*M. domestica*和*M. esculenta*等各自的3个拷贝由于没有同线性或串联排列等结构特点, 因此可以被暂时归入随机复制类型。

### 2.2 系统发育关系

为深入了解DAD基因在种子植物中的演化历史, 我们以*C. reinhardtii*的DAD基因作为外类群, 利用全部38种植物的58条DAD基因的氨基酸序列构建了最大似然性(ML)和贝叶斯(BI)系统发育关系树。两树除支持率外, 具有相似的拓扑结构, 因此, 我们仅展示ML树, 并将贝叶斯树的后验概率标注在ML计算的靴带支持率之后(图1)。该系统发育树基本按照植物的大类群进行分级聚类, 苔藓的2个基因逐次位于最外围的基部, 此后种子植物分支中, 被子植物以绝对的支持率(BS=100%, BBP=1.00)构成单系; 裸子植物银杏(*Ginkgo biloba*)和所有松柏类植物聚合成一个单系分支。尽管同样作为裸子植物的买麻藤(*Gnetum montanum*)在系统树上位于被子植物分支的外围基部, 但基本上没有任何支持率(BS<50%, BBP=0.66)。在被子植物分支内部, 无油樟(*Amborella trichopoda*)如预期被置于外围基部, 其它被子植物则基本上按照单子叶植物和双子叶植物分列成不同的分支。

总体上, 在系统发育树的各分支内部, 同一植物或同一科的DAD基因基本上首先聚在一起。例如, 单子叶植物分支中, 芭蕉科*M. acuminata*的2条DAD基因首先聚在一起, 然后和其它的禾本科植物组成的分支构成并系。这种现象在大多数双子叶植物的分支内

表1 38种植物*DAD*同源基因的详细信息Table 1 Detailed information of *DAD* homologous genes of 38 plants

Taxa	Gene ID	Strand	Chromosome	Duplication pattern	Gene structure		Amino acid (aa)
					Intron	Exon	
<b>Angiosperm</b>							
<b>Dicot</b>							
<i>Arabidopsis lyrata</i>	AL1G33450	-	Scaffold_1	Block	4	5	115
	AL4G20890	+	Scaffold_4	Block	4	5	116
<i>A. thaliana</i>	AT1G32210	-	Chr01	Block	4	5	115
	AT2G35520	+	Chr02	Block	4	5	116
<i>Brassica rapa</i>	BR05G20160	+	ChrA05	Block	4	5	115
	BR09G26700 <sup>a</sup>	+	ChrA09	Block	6	7	201 (88) <sup>d</sup>
<i>Capsella rubella</i>	CRU_001G29110	-	Scaffold_1	Block	4	5	115
	CRU_004G16620	+	Scaffold_4	Block	4	5	115
<i>Citrullus lanatus</i>	CL10G00840	+	Chr10		4	5	115
<i>Cucumis melo</i>	CM00021G01290	-	Scaffold00021		4	5	115
<i>Eucalyptus grandis</i>	EG0008G05930	+	Scaffold_8		4	5	115
<i>Fragaria vesca</i>	FV2G07670	+	LG2		4	5	124
<i>Gossypium raimondii</i>	GR03G18540	+	Chr03	Block	4	5	117
	GR08G22400	-	Chr08	Block	4	5	117
<i>Malus domestica</i>	MD05G025840	+	Chr05	Random	4	5	119
	MD10G000120	+	Chr10	Random	4	5	119
<i>Manihot esculenta</i>	ME04430G00010	+	Scaffold04430	Random	4	5	115
	ME07304G00010	+	Scaffold07304	Random	4	5	115
<i>Populus trichocarpa</i>	PT01G13680	-	Chr01	Block	4	5	115
	PT03G09680	+	Chr03	Block	4	5	115
<i>Prunus persica</i>	PPE_004G33460	-	Scaffold_4	Block	4	5	119
	PPE_008G00790	-	Scaffold_8	Block	4	5	119
<i>Ricinus communis</i>	RC29634G00390	+	29634	Random	4	5	115
	RC30068G01440	+	30068	Random	4	5	113
<i>Solanum lycopersicum</i>	SL08G076460	+	Chr08		4	5	116
<i>S. tuberosum</i>	ST08G008690	-	Chr08	Block	4	5	116
	ST08G021840	-	Chr08	Block	4	5	116
<i>Thellungiella parvula</i>	TP1G27890	-	Chr1-1		4	5	115
<i>Theobroma cacao</i>	TC0003G30500	+	Scaffold_3		4	5	116
<i>Vitis vinifera</i>	VV02G01690	+	Chr02		4	5	115
<b>Moncot</b>							
<i>Brachypodium distachyon</i>	BD1G50180	+	Chr01		4	5	114
<i>Hordeum vulgare</i>	HV1571041G00020	-	Contig_1571041	Random	4	5	114
	HV44460G00030	+	Contig_44460	Random	4	5	114
<i>Oryza sativa</i>	OS04G32550	+	Chr04		4	5	114
<i>Setaria italica</i>	SI004G00880	+	Scaffold_4		4	5	114
<i>Sorghum bicolor</i>	SB10G001000	+	Chr10		4	5	114
<i>Zea mays</i>	ZM09G06480	-	Chr09		4	5	114
<i>Musa acuminata</i>	MA07G17850	-	Chr07	Block	4	5	115
	MA10G06540 <sup>b</sup>	-	Chr10	Block	-	1	48 (22) <sup>d</sup>
	MA11G08020	+	Chr11	Block	4	5	170
<b>Basal taxon</b>							
<i>Amborella trichopoda</i>	ATR_00025G00360	+	Scaffold00025		4	5	121
<b>Gymnosperm</b>							

表1 (续) Table 1 (continued)

Taxa	Gene ID	Strand	Chromosome	Duplication pattern	Gene structure		Amino acid (aa)
					Intron	Exon	
<i>Ginkgo biloba</i> <sup>c</sup>	GBI00024938	+	6973		-	-	123
<i>Gnetum montanum</i> <sup>c</sup>	GMO00028447	-	GTHK-0066600		-	-	113
<i>Picea abies</i>	PAB00021420 <sup>a</sup>	-	MA_128105	Random	5	6	162 (67) <sup>d</sup>
	PAB00041636	-	MA_42912	Random	4	5	115
	PAB00057214 <sup>b</sup>	+	MA_8328929	Random	1	2	50 (45) <sup>d</sup>
<i>P. glauca</i> <sup>c</sup>	PGL00020840	+	PUT-39823		-	-	115
			PUT-39823				
	PGL00013662	+	PUT-23726		-	-	115
			PUT-23726				
<i>P. sitchensis</i> <sup>c</sup>	PGL00010558	+	PUT-16509		-	-	115
	PSI00016644	+	PUT-531483		-	-	148
	PSI00008404	+	PUT-21837		-	-	115
<i>Pinus pinaster</i> <sup>c</sup>	PPI00061658	-	Unigene30039		-	-	115
	PPI00006515	-	Cotig25413		-	-	115
<i>Pi. sylvestris</i> <sup>c</sup>	PSY00006071	-	Isotig31028		-	-	115
	PSY00026421	+	Isotig68699		-	-	115
	PSY00004335	+	isotig24244		-	-	115
<i>Pi. taeda</i>	PTA00003657	-	Scaffold464	Random	4	5	115
	PTA00022345 <sup>b</sup>	+	Scaffold214279	Random	2	3	86 (72) <sup>d</sup>
	PTA00022344 <sup>b</sup>	+	Scaffold214279	Random	1	2	44 (44) <sup>d</sup>
<i>Pseudotsuga menziesii</i> <sup>c</sup>	PME00105748	+	Psme_598296871		-	-	146
	PME00105747	+	Psme_598296869		-	-	115
<b>Moss</b>							
<i>Physcomitrella patens</i>	PP00045G01180	+	Scaffold_45		4	5	131
	PP00456G00180	-	Scaffold_456		4	5	114
<b>Alga</b>							
<i>Chlamydomonas reinhardtii</i>	CR02G08990	+	Chr02		3	4	108

<sup>a</sup> 具有差异或 <sup>b</sup> 不完整的基因未被包括在系统发育分析中; <sup>c</sup> 来自转录组的序列资料; <sup>d</sup> 非正常或短位点, 括号内的数字表示 *DAD*-like同源基因的长度; 第6列中的-表示无数据。

<sup>a</sup> Divergent or <sup>b</sup> incomplete genes that not included in phylogeny analysis; <sup>c</sup> Sequence data come from different transcriptomes;

<sup>d</sup> Abnormal or short loci, numbers in parentheses indicate the lengths of *DAD*-like homologous sequences; - in column 6 indicate no data.

也有体现。但是, 在十字花科分支以及裸子植物分支内情况则稍有不同, 在这2个分支内, 基本上每种植物均出现两类 *DAD* 基因并分列入2个平行的小分支, 显示了谱系特异的复制特点(图1)。

### 2.3 讨论

通过对 *DAD* 基因进行分析和比较, 我们进一步证实了该基因在种子植物演化过程中的高度保守性。来自不同种子植物的同源基因, 不仅具有高度相近的编码序列, 而且享有相同的基因结构。例如, 目前已经识别的 *DAD* 同源基因对之间, 其氨基酸序列一致性即使在远缘的 *M. acuminata* 和 *P. sitchensis* 之间也达到了70%; 而在双子叶植物和裸子植物的某些属下类群

可以达到完全一致。29种植物的全基因组数据也显示, 种子植物 *DAD* 基因的正常结构都是由5个外显子和4个内含子组成(表1)。这种演化上的保守性为其功能的高度特化奠定了坚实的基础。以往的研究证明, *DAD* 基因在所有多细胞类群内的功能相对稳定, 它主要作为PCD的负向调控基因, 阻止或延迟PCD的发生(Mondragón-Palomino et al., 2002; Wortman et al., 2003; 杨舒雅等, 2012)。

此外, *DAD* 在种子植物基因组内还具有拷贝数目较少的特点, 其拷贝数目不超过3个。根据早期学者对基因家族类型的划分(Ye et al., 2012), 拷贝数小于2个被归入无家族(non-family, NF)类; 3-10个被归入低拷贝数目家族(low copy-number family, LF)类;

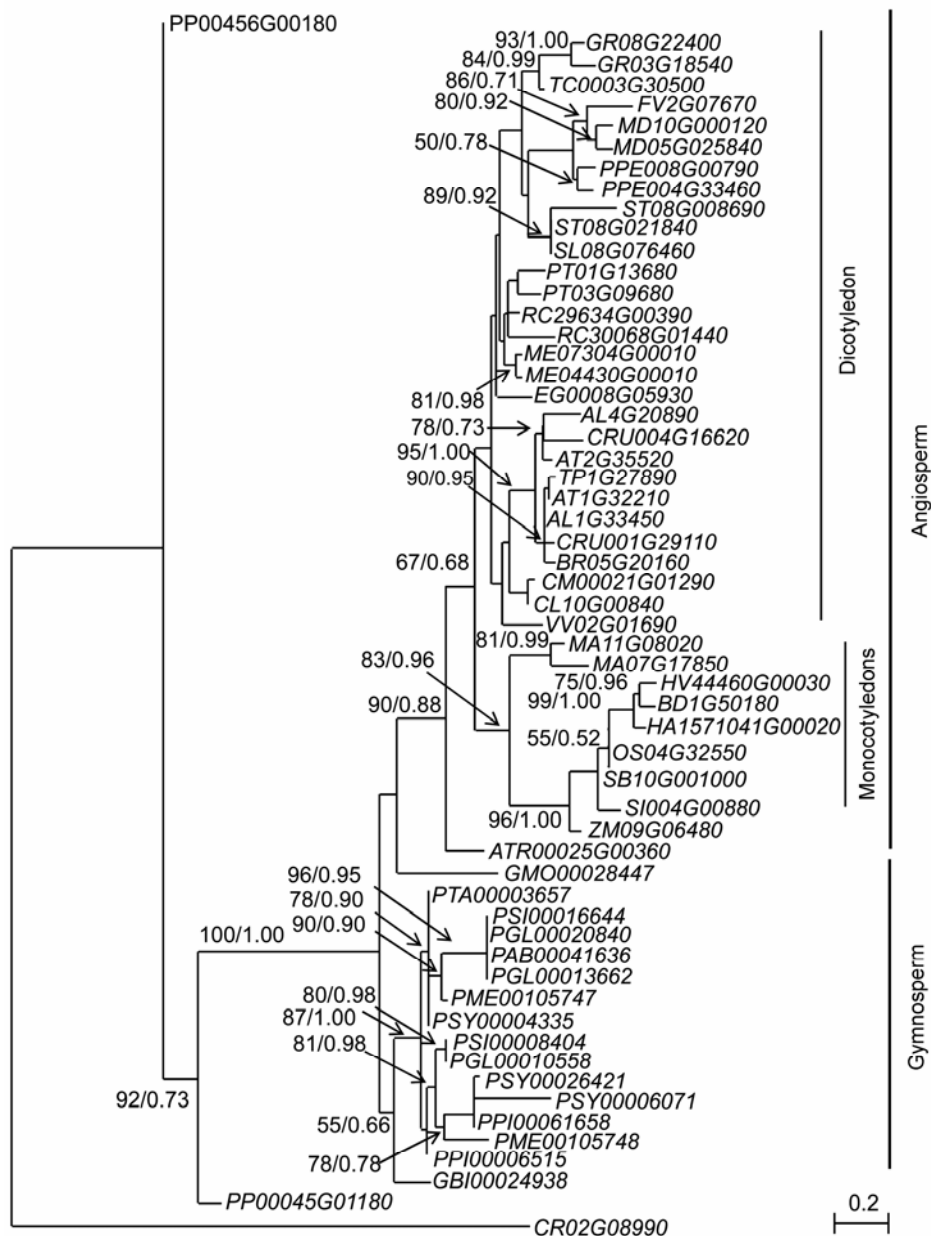


图1 基于38种植物的58条*DAD*基因的编码序列构建的最大似然性系统发育关系树  
分支附近的数值分别代表最大似然性分析中大于50%的自展支持率和贝叶斯分析中的后验率, 各基因位点同表1。

Figure 1 Maximum likelihood tree of *DAD* genes based on 58 amino acid sequences of 38 plants  
Numbers near branches represent bootstrap values (>50%) of the maximum likelihood analysis and posterior rate of the Bayesian analysis. The gene loci in this figure are the same as those listed in Table 1.

大于10个则被归入高拷贝数目家族(high-copy-number family, HF)类。那么*DAD*基因显然应归入NF至LF类型。分析种子植物内拷贝数大于1的*DAD*基因的复制情况, 我们发现基因组内*DAD*拷贝数目的增加主

要源自随机复制和染色体大片的重复。在具有*DAD*复制现象的15个物种中, 6个物种的复制来自随机复制, 而9个物种的复制来自染色体大片重复。进一步的系统发育分析表明, 这些染色体大片重复导致

的复制类型可能与种子植物不同阶层发生的多倍化密切相关。例如, *Arabidopsis*、*Brassica*和*Capsella*是十字花科中隶属于不同族的3个属, 其中*Arabidopsis*与*Capsella*的系统发育关系要比其与*Brassica*的近。在我们构建的系统发育树上(图1), 来自3个属的DAD拷贝并没有完全按照物种进行聚类, 而是分别列入并行的2个小分支, 这种结构暗示着DAD基因在这3个属分开之前曾经历过一次基因复制事件。早期基于全基因组的比较显示, 十字花科植物在*Arabidopsis*和*Brassica*分开之前, 的确经历了一次全基因组加倍事件(Blanc et al., 2003)。

基于BLAST分析, 我们也筛查出4个不完整的DAD基因(表1)。初步推测这些短片段有可能是假基因没有完全丢失的残迹。这种猜测从*M. acuminata*的MA10G06540位点可以得到一定程度的支持。同线性分析表明, 物种内该位点所处的染色体片段和另外2个完整基因各自所在的染色体片段均具有一定的同线性, 且该位点位于同线性片段的中间位置, 这种分布情况说明*M. acuminata*的3个位点是染色体大片段重复产生的。结合系统发育树(图1)上2个完整基因聚合在一起的分支结构, 表明这种大片段的复制应该属于谱系特异扩增的结果。如此, 目前编码序列只有48 aa且仅保留22 aa DAD原始序列的位点MA10G06540不太可能在复制之初就不完整, 而最可能是原来隶属于染色体相应片段中的完整基因, 在后期演化过程中, 为维持其低拷贝和降低冗余而产生的基因沉默和假基因化的结果。Ye等(2012)通过对14种植物基因组的比较分析也证实, 植物基因组内NF和LF在经历低频的染色体大片段复制后, 通常会伴随高频的基因丢失现象。这应该是植物维持低拷贝基因的重要方式。

## 参考文献

- 龚文芳, 喻树迅, 宋美珍, 范术丽, 庞朝友, 肖水平 (2010). 棉花抗细胞凋亡基因*GhDAD1*的克隆、定位及表达分析. 中国农业科学 43, 3713–3723.
- 贾志蓉, 李丹, 吕应堂 (2004). 拟南芥*AtDAD1*超量表达植株对H<sub>2</sub>O<sub>2</sub>抗性的研究. 武汉植物学研究 22, 373–379.
- 杨舒雅, 史娟, 马斌芳, 赵洁, 金晓航 (2012). 抗细胞凋亡因子DAD1的研究进展. 生理科学进展 43, 315–318.
- Apte SS, Mattei MG, Seldin MF, Olsen BR (1995). The highly conserved defender against the death 1 (*DAD1*) gene maps to human chromosome 14q11-q12 and mouse chromosome 14 and has plant and nematode homologs. *FEBS Lett* 363, 304–306.
- Blanc G, Hokamp K, Wolfe KH (2003). A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res* 13, 137–144.
- Danon A, Rotari VI, Gordon A, Mailhac N, Gallois P (2004). Ultraviolet-C overexposure induces programmed cell death in *Arabidopsis*, which is mediated by caspase-like activities and which can be suppressed by caspase inhibitors, p35 and *defender against apoptotic death*. *J Biol Chem* 279, 779–787.
- Gallois P, Makishima T, Hecht V, Despres B, Laudie M, Nishimoto T, Cooke R (1997). An *Arabidopsis thaliana* cDNA complementing a hamster apoptosis suppressor mutant. *Plant J* 11, 1325–1331.
- Gouy M, Guindon S, Gascuel O (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27, 221–224.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59, 307–321.
- Kelleher DJ, Gilmore R (1997). DAD1, the defender against apoptotic cell death, is a subunit of the mammalian oligosaccharyltransferase. *Proc Natl Acad Sci USA* 94, 4994–4999.
- Kerr JFR, Wyllie AH, Currie AR (1972). Apoptosis: a basic biological phenomenon with wide ranging implications in tissue kinetics. *Br J Cancer* 26, 239–257.
- Mondragón-Palomino M, Meyers BC, Michelmore RW, Gaut BS (2002). Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res* 12, 1305–1315.
- Nakashima T, Sekiguchi T, Kuraoka A, Fukushima K, Shibata Y, Komiyama S, Nishimoto T (1993). Molecular cloning of a human cDNA encoding a novel protein, *DAD1*, whose defect causes apoptotic cell death in hamster BHK21 cells. *Mol Cell Biol* 13, 6367–6374.
- Raff M (1998). Cell suicide for beginners. *Nature* 396, 119–122.
- Roboti P, High S (2012). The oligosaccharyltransferase subunits OST48, DAD1 and KCP2 function as ubiquitous and selective modulators of mammalian N-glycosylation. *J Cell Sci* 125, 3474–3484.
- Ronquist F, Huelsenbeck JP (2003). MrBayes 3: Bayesian

phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

**Schwartzman RA, Cidlowski JA** (1993). Apoptosis: the biochemistry and molecular biology of programmed cell death. *Endocr Rev* **14**, 133–151.

**Williams GT, Smith CA** (1993). Molecular regulation of apoptosis: genetic controls on cell death. *Cell* **7**, 777–779.

**Wortman JR, Haas BJ, Hannick LI, Smith RK, Maiti R, Ronning CM, Chan AP, Yu CH, Ayele M, Whitelaw CA, White OR, Town CD** (2003). Annotation of the Arabidopsis genome. *Plant Physiol* **132**, 461–468.

**Ye CY, Li T, Yin H, Weston DJ, Tuskan GA, Tschaplinski TJ, Yang X** (2012). Evolutionary analyses of non-family genes in plants. *Plant J* **73**, 788–797.

## Evolution of *Defender Against Apoptotic Death (DAD)* Genes in Seed Plants

Ying Bao<sup>†\*</sup>, Yuqin Mei<sup>†</sup>

*School of Life Sciences, Qufu Normal University, Qufu 273165, China*

**Abstract** *Defender against apoptotic death (DAD)* is a highly conserved cellular apoptosis gene and plays an important role in seed plant growth and development. To fully understand the evolutionary pattern of the *DAD* gene in seed plants, based on the whole genome data of 31 plants, we analyzed copy number, structure, chromosome location of the *DAD* genes by using bioinformatics. In addition, along with transcriptome data for seven gymnosperms, we discuss the evolutionary trend of the genes in seed plants. The *DAD* gene is a low-copy gene with only 1–3 copies in different seed plants, and the length of different *DAD* protein varies from 108 to 170 aa. Phylogenetic and syntonic analyses further showed that the evolution of the *DAD* gene in these seed species has a lineage-specific characteristic. Random and block duplication and subsequent gene loss were possibly important ways to maintain the low-copy number of *DAD* genes in seed plant genomes.

**Key words** chromosome location, *DAD*, duplication pattern, whole genome, gene structure

**Bao Y, Mei YQ** (2017). Evolution of *Defender against Apoptotic Death (DAD)* genes in seed plants. *Chin Bull Bot* **52**, 590–597.

<sup>†</sup> These authors contributed equally to this paper

\* Author for correspondence. E-mail: baoyingus@126.com

(责任编辑: 朱亚娜)