

白菜防御素基因的预测及结构功能分析*

王瑞雪,沈亮余,赵敬会,李荣冲,郑小敏,张 涛
(重庆师范大学 生命科学学院,重庆 400047)

摘要 采用生物信息学的方法,通过已知防御素基因家族基因氨基酸序列在线 Blast 相似性比对,对白菜(*Brassica rapa*)基因组中防御素基因进行了预测和鉴定,分析了防御素基因在进化过程中各部分结构,包括上游区域、启动子区域、外显子区域、内含子区域以及 UTR 区域的结构变异以及功能变异,对该基因上游序列顺式作用元件和表达模式进行了预测。分析结果表明,白菜防御素基因编码 60~80 个氨基酸短肽,编码氨基酸均具有 8 个保守的半胱氨酸;白菜防御素基因功能结构域保持相对稳定,但部分基因成员前端信号肽序列出现变异;内含子长度出现增长、缩短、消失等 3 种变异模式,该基因上游启动子区域核苷酸变异较其他区域显著,上游序列顺式作用元件大多与光应答、逆境胁迫应答和信号分子应答相关。该基因表达模式预测结果表明白菜防御素基因在进化过程中可能出现分化,主要体现在基因沉默和表达部位的差异。

关键词 白菜;防御素;基因重复;进化

中图分类号:Q37

文献标志码:A

文章编号:1672-6693(2012)06-0080-07

生物体在遭受病原菌入侵时体内会产生天然的免疫应答反应,分泌一种小分子短肽来抵御外源病原菌的传播和入侵,从而达到保护自身的目的,这类短肽就是防御素。防御素是生物抗菌肽的一种,广泛地存在于动物、植物和微生物中,对生物抵御外源病原菌入侵并维持自身生存具有重要意义。目前,防御素的种类主要包括动物防御素(α -防御素、 β -防御素和 θ -防御素)、昆虫防御素和植物防御素^[1]。在天然的免疫应答系统中,只有防御素这类短肽在植物、无脊椎动物和脊椎动物中是相对保守的,它是由富含半胱氨酸的碱性氨基酸通过二硫键的连接形成稳定的三维构象^[2]。国外对动物和昆虫防御素的研究起步较早^[3],而植物防御素直到 20 世纪 90 年代初才第一次从小麦和大麦中分离出来^[4-5]。目前,从不同植物中已经分离得到大约 80 多种植物防御素基因^[2]。植物防御素基因具有 8 个半胱氨酸,通过 4 对二硫键连接,比动物和昆虫防御素基因多两个半胱氨酸和 1 对二硫键。对植物防御素基因的研究,国外报道较多的是拟南芥、萝卜、高粱、油菜等植物^[6],其中对拟南芥防御素基因的研究较为深入。当前,在拟南芥防御素基因家族中已经鉴定了 15 个

防御素类似基因,其中包括 3 个亚家族^[2]。鉴于防御素在植物抗病中的重要作用,因此研究植物防御素基因对于了解植物抗病机理具有重要意义。

白菜作为一个新生的多倍体,在进化过程中先后经历了三倍化和二倍化的过程,先前的 A3 号染色体以及最近的白菜基因组测序结果表明,在白菜基因组中也存在大量的重复基因片段^[7-9]。因此,在白菜基因组序列中存在防御素重复基因的可能性较大。目前为止,对白菜防御素基因的克隆已经有相关报道^[10],但对白菜防御素基因家族的研究相对较少,本文充分利用当前已经报道的植物防御素基因序列,运用生物信息学的方法对白菜基因组序列中的防御素基因进行预测鉴定,分析其进化过程中结构以及功能的变异,并对其表达模式进行预测,进一步确定白菜多倍化事件过程中对防御素基因结构功能以及表达模式的影响,为后续对其基因的克隆以及功能的鉴定等提供理论基础。

1 材料与方法

1.1 基因的鉴定

本研究所用的植物防御素基因种子序列主要来

* 收稿日期 2012-03-23 修回日期 2012-05-13 网络出版时间 2012-11-12 16:42:01

资助项目 国家自然科学基金项目(No. 31171588)

作者简介 王瑞雪,男,硕士研究生,研究方向为生物化学与分子生物学,通讯作者 张涛, E-mail: zht2188@126.com

网络出版地址 http://www.cnki.net/kcms/detail/50.1165.N.20121112.1642.201206.80_018.html

自拟南芥数据库(<http://arabidopsis.org/>)中已经鉴定的拟南芥防御素基因家族序列和植物抗菌肽数据库(<http://phytamp.pfba-lab-tun.org/entrieslist.php?view=GeneralView>)中防御素基因家族(Defensin family)除拟南芥序列以外的植物防御素基因氨基酸序列。所下载的序列均以 fasta 格式保存。运用以上获取的防御素基因氨基酸序列作为查询种子序列,在芸薹属白菜数据库(<http://brassicadb.org/brad/index.php>)运行在线 Blast 程序,运行程序选择 Blastp,数据库选择 *B. rapa*(protein),其他参数默认。在 Blast 比对结果中,为了防止目标序列遗漏,本研究采用序列相似性不低于 40%,E 不大于 -10 的序列,去掉命中的重复序列,将剩余序列作为最终的鉴定序列。为了提高预测的准确性,本文将用于鉴定的目标序列在 SMART 保守结构域预测数据库(<http://smart.embl-heidelberg.de/>)和 Sanger 数据库中基于隐马尔可夫模型的蛋白质保守结构域预测数据库 Pfam(<http://pfam.sanger.ac.uk/>)进行防御素基因保守结构域的预测。由于算法的不同,在 SMART 数据库中预测的防御素基因保守结构域为 Knot1(Knottins),Pfam 数据库中预测的防御素保守结构域为 Gamma thionin,只有 2 个数据库全部预测到目标保守结构域且达到显著性水平,可作为最终白菜基因组防御素基因。

1.2 序列结构及功能分析

基因结构的预测采用 Softberry 在线基因预测分析软件 FGESH,对照物种选择植物(<http://linux1.softberry.com/berry.phtml?topic=fgesh&group=programs&subgroup=gfind>)。启动子区域的预测采用 Softberry 网站 TSSP(prediction of plant promoters)程序(<http://linux1.softberry.com/berry.phtml?topic=tssp&group=programs&subgroup=promoter>)。启动子分析采用 400 bp 长度。基因外显子内含子区域的确定采用在线分析软件 GSDS(<http://gsds.cbi.pku.edu.cn>)。上游序列顺式作用元件的预测运用在线网站 PlantCARE(<http://bioinformatics.psb.ugent.be/webtools/plantcare/html>)。序列保守基序的预测采用在线 MEME(<http://meme.nbcr.net/meme/cgi-bin/meme.cgi>),参数默认。氨基酸理化性质的分析采用在线分析软件(<http://web.expasy.org/protparam/>)。DNA 多态性分析应用 DnaSP Version 5。白菜防御素基因表达模式的预测采用白菜数据库 Blastn 在线程序(<http://brassicadb.org/brad/index.php>),数据

库选择(EST of *B. rapa*(NCBI)),其他参数默认。

2 结果与分析

2.1 基因预测与鉴定

用已知植物防御素基因氨基酸序列在白菜数据库中运行 Blastp 程序,选取相似性在 40% 以上,E 值不大于 -10 的命中序列作为白菜基因组防御素基因鉴定的候选序列。去掉命中的重复序列,本研究中共得到 15 条序列。具体的序列号为 *Bra003740*、*Bra004900*、*Bra008225*、*Bra008226*、*Bra015808*、*Bra015809*、*Bra015811*、*Bra016501*、*Bra017421*、*Bra024110*、*Bra024832*、*Bra025396*、*Bra026615*、*Bra029208* 和 *Bra031312*。用以上初步预测的 15 条白菜基因的氨基酸序列在 SMART 数据库和 Pfam 数据库中进行保守结构域的预测。预测结果显示,基因号为 *Bra015808* 和 *Bra024110* 的两条氨基酸序列在两个数据库中均没有预测到防御素基因相关保守结构域,只在前 30 个氨基酸左右的区域预测到信号肽结构。因此,本研究最终得到 13 条白菜防御素基因家族序列。此外,*Bra015809* 和 *Bra029208* 基因的氨基酸序列中也没有检测到信号肽序列,说明这两个基因在进化过程中信号肽序列发生分化。白菜防御素家族基因主要分布在 A02、A05、A06、A07、A08、A09 等 6 条染色体上。CDS 区间长度在 183 ~ 243 bp 间波动,基因组序列长度在 183 ~ 570 bp 间波动。

2.2 基因结构分析

2.2.1 内含子区域以及外显子区域确定 运用白菜防御素基因 CDS 序列和对应的基因组序列在 GSDS 在线网站确定内含子、外显子数目及区域。从封三彩图 1 可以看出,白菜防御素基因家族在漫长的进化过程中,多数序列保留有两个外显子和 1 个内含子,但 *Bra029208* 基因序列在进化过程中丢失了内含子。剩余序列外显子 1 的长度变化相对保持稳定,均为 64 bp,外显子 2 的长度变化保持相对稳定,在 170 ~ 179 bp 之间变动。内含子长度在进化过程中变化范围相对较大,从 0 ~ 336 bp 之间波动,这说明在白菜防御素基因进化过程中外显子拼接元件在不同拷贝之间的变异较大。

2.2.2 顺式作用元件分析 白菜防御素基因上游序列在 PlantCARE 在线网站进行顺式作用元件分析的结果表明,启动子核心元件、抗逆应答相关元件和信号分子应答元件在白菜防御素基因家族上游调控序列中普遍存在,白菜防御素基因上游调控区存在大量的 TATA-box 和 CAAT-box 启动子核心元件;与环

境相关的调控元件主要包括光应答元件、热激应答元件、防御应答元件、低温诱导元件、激发子诱导元件、厌氧诱导元件、伤害应答元件和干旱诱导元件;与信号分子相关的调控元件主要包括茉莉酸甲酯应答、脱落酸应答、水杨酸应答、赤霉素应答、生长素应答和乙烯应答元件,信号分子应答元件种类和数量在家族成员之间各不相同,只有 *Bra025396* 上游调控区域没有发现信号分子应答元件。在 *Bra003740*

中发现 1 个根部特异表达元件,在 *Bra008226* 和 *Bra016501* 中各发现 1 个种子特异表达元件。此外,除 *Bra025396* 和 *Bra024832* 两个基因外,其他防御素基因均含有 1 个与生理节律控制相关的顺式作用元件(表 1)。由此可见,白菜防御素基因上游调控区域存在大量的与外界环境感知与信号分子应答相关的顺式作用元件,对植物逆境条件下生存具有重要意义。

表 1 白菜防御素基因启动子顺式作用元件个数

	1	2	3	4	5	6	7	8	9	10	11	12	13
TATA-box	59	75	51	54	62	51	73	118	93	35	86	111	46
CAAT-box	34	29	34	24	37	22	32	30	34	10	38	25	29
光应答元件	31	16	24	25	8	24	19	17	24	3	11	27	16
根部特异表达元件	1	-	-	-	-	-	-	-	-	-	-	-	-
低温应答元件	1	-	-	-	1	-	1	-	2	-	-	-	-
生理节律控制	2	1	1	2	3	5	1	1	-	-	2	1	1
伤害应答元件	-	1	-	-	-	-	-	-	-	-	-	-	-
干旱诱导元件	-	1	-	1	-	5	1	2	1	-	-	1	1
厌氧诱导元件	2	2	3	1	1	1	2	2	3	-	2	5	3
真菌激发子	1	-	2	2	2	-	1	-	1	-	-	-	-
顺式调控元件	1	-	3	-	-	-	1	2	1	1	1	1	-
分生组织表达	-	-	-	-	-	-	1	2	1	-	1	-	-
强转录元件	-	-	-	-	-	-	-	-	2	-	1	-	1
种子特异表达	-	-	-	1	-	-	1	-	-	-	-	-	-
Binding site	-	-	-	-	3	2	1	-	1	-	1	1	2
胚乳特异表达	5	2	2	2	7	2	8	-	3	1	2	3	6
热胁迫应答	-	2	-	4	2	-	3	2	-	-	4	4	2
防御胁迫应答	3	1	-	1	2	2	-	1	1	-	1	1	1
脱落酸应答	4	-	-	4	1	1	4	2	2	-	-	2	2
乙烯应答	-	-	-	1	-	1	-	1	1	-	-	2	1
赤霉素应答	2	1	1	-	1	4	-	-	3	-	1	-	5
生长素应答	-	-	1	1	2	2	-	2	-	-	-	-	1
茉莉酸甲酯应答	-	-	-	-	-	4	4	6	4	-	4	-	10
水杨酸应答	2	-	-	-	-	1	1	-	1	-	-	1	1

注:表中第一行 1~13 依次表示基因 *Bra003740*、*Bra004900*、*Bra008225*、*Bra008226*、*Bra015809*、*Bra015811*、*Bra016501*、*Bra017421*、*Bra024832*、*Bra025396*、*Bra026615*、*Bra029208* 和 *Bra031312*

2.2.3 保守基序分析 将白菜防御素基因氨基酸序列运用 MEME 在线网站进行保守基序探测,结果表明在白菜防御素基因家族中主要有 3 种类型的保守基序(Motif)(封三彩图 2),综合 SMART 和 Pfam 数据库预测结果,motif1 和 motif2 是 Knottins 或者 Gamma thionin 保守结构域基序,是白菜防御素基因的核心基序,这与上述两个数据库鉴定的结果是相吻合的,多序列比对结果也可以明确的看到有 8 个非常保守的半胱氨酸(封三彩图 3)。motif3 是信号肽保守基序,*Bra015809*、*Bra016501*、*Bra031312*、*Bra029208* 等 4 条序列并没有探测到保守的信号肽基序,但在 SMART 数据库中的 *Bra016501* 和

Bra031312 序列上检测到了信号肽序列的存在。这可能是由于 MEME 检测算法严格造成的,但可以肯定的是 *Bra015809* 和 *Bra029208* 在进化过程中信号肽序列分化而最终导致丢失而失去分泌蛋白功能,*Bra016501* 和 *Bra031312* 信号肽序列可能在进化过程中也已经发生分化并导致信号肽序列保守性下降。

2.3 基因家族基因多态性与进化分析

表 2 显示,白菜防御素基因在进化过程中,CDS 区域发生插入缺失事件的次数最多,其余由高到低依次为基因组区域、内含子区域、启动子区域、3UTR 和 5UTR。平均插入缺失长度在 CDS 区域最长,其

余由高到低依次为内含子区域、基因组区域、启动子区域、3UTR 和 5UTR。单倍型插入缺失数在启动子区域较多,达 13.00 个,单倍型插入缺失差异同样在启动子区域较大,为 1.00。此外,与其他几个区域相比,启动子区域多样性指数在较高,插入缺失多样性指数为 8.90,位点插入缺失多样性指数为 0.03,序列插入缺失多样性指数为 12.57;并且在该区域,多态性位点数为 232.00 个,突变总数为 536 个,核苷酸多样性指数为 0.59,均高于其他几个区域。由此可见,在进化过程中,启动子区域的变异程度要远高于其他区域。Tajima's D 检测结果表明,白菜防

御素家族基因在进化过程中,均不存在明显的选择作用(显著性统计结果均为不显著)除 CDS 区域有正向选择的趋势外,其他几个区域均呈现负向选择的趋势。进化分析结果显示,白菜防御素基因家族在进化过程中正处于一个不断分化的过程,在进化树中可以看出,白菜防御素基因家族已经分化成两个大的亚家族分支 A 和 B,而亚家族 A 分化为 C、D、E 等 3 个分支,其中 C 分支处于不断分化的状态(图 4,采用 MEGA4.0 邻接法构建,Bootstrap 值采用 1 000 次重复)。

表 2 白菜防御素基因多态性分析

	启动子区域	5UTR	内含子区域	CDS 区域	3UTR	基因组区域
平均插入缺失事件/次	1.90	1.33	2.00	9.42	1.67	2.60
平均插入缺失长度/bp	1.81	1.10	2.30	6.60	1.33	2.28
单倍型插入缺失数/个	13.00	7.00	7.00	5.00	4.00	7.00
单倍型插入缺失差异	1.00	0.87	0.77	0.76	0.72	0.83
插入缺失多样性指数 $k(i)$	8.90	1.49	1.85	1.92	0.97	3.08
位点插入缺失多样性指数 $P(i)$	0.03	0.03	0.02	0.01	0.03	0.02
序列插入缺失多样性指数	12.57	1.93	2.31	2.26	0.97	3.22
多态性位点数 $S/\text{个}$	232.00	37.00	63.00	155.00	28.00	146.00
突变总数 $\theta/\text{个}$	536	70.00	123.00	286.00	62.00	273.00
核苷酸差异平均数 $K/\text{个}$	139.33	18.67	32.26	82.14	15.42	77.28
核苷酸多样性 π	0.59	0.48	0.49	0.46	0.55	0.46
Tajima's D 检验值	-0.89	-0.78	-0.97	1.34	-1.03	-0.56
显著性统计结果	不显著	不显著	不显著	不显著	不显著	不显著

注:显著性统计 p 值均大于 0.1,故不显著

2.4 对应氨基酸序列的基本理化性质分析

白菜防御素基因对应的氨基酸序列组成分析结果见表 3。表 3 显示,该基因家族的氨基酸序列中氨基酸数目在 60~80 个之间,蛋白质分子量为 6 946~9 258.1 Da,等电点值为 5.45~9.82,碱性氨基酸数为 5~15 个,酸性氨基酸数为 3~7 个,蛋白不稳定指数为 27.59~80.54。蛋白指数不稳定的拷贝成员个数为 6 个,有 7 个拷贝成员编码蛋白指数稳定。从氨基酸序列长度来看,白菜防御素基因编码的氨基酸也属于短肽,从氨基酸酸碱性来看,只有 1 个基因 *Bra031312* 编码的氨基酸序列等电点小于 7,这可能是由于序列在进化过程中产生变异的缘故。

2.5 基因表达模式预测

选取白菜防御素基因 CDS 序列在白菜 EST 数据库进行表达模式预测分析,比对结果选取比对长度在 100 bp 以上、E 不大于 -30 以及相似性在 75% 以上的匹配序列。比对结果表明,*Bra029208* 和

Bra016501 没有命中 EST 序列,表明这两个基因在进化过程中可能已经沉默。*Bra008225*、*Bra004900*、*Bra025396* 和 *Bra031312* 序列命中 EST 序列在 5 个以下,其余序列命中的 EST 序列在 21~38 个之间。*Bra004900* 和 *Bra025396* 只命中花蕾 EST 序列,*Bra031312* 只命中 1 条初生叶 EST 序列,这 3 个基因可能呈现组织器官特异性表达。*Bra017421*、*Bra024832* 和 *Bra026615* 序列表达模式相似,表达部位和表达量明显高于其他序列。*Bra003740*、*Bra008226*、*Bra015809*、*Bra015811* 等 4 条序列表达模式相近。由此可见,白菜防御素基因在进化过程中,表达模式已经发生分化,具体表现在功能丢失(基因沉默)、表达量和表达部位的差异(表 4)。综合进化分析和表达模式的预测结果,不难发现白菜防御素基因表达模式的分化和进化分歧是一致的,在 B、C、D 和 E 分支中,每个分支内的基因成员表达模式是相似的,在进化过程中,分支 D 和 E 中的防御素基因家族成员可能已经分化为器官组织特异性

表达,分支 B 和 C 中防御素基因表达模式的差异可能主要体现在表达部位的差异和表达量的差异上。两个可能沉默的防御素基因存在于 B 和 E 两个分支中。

表 3 白菜防御素基因氨基酸序列指数分析

氨基酸序列号	氨基酸数/个	分子量/Da	等电点	碱性氨基酸数/个	酸性氨基酸数/个	不稳定指数	评价
<i>Bra003740</i>	80	8 839.4	8.89	9	4	27.59	稳定
<i>Bra008226</i>	80	8 718.2	8.47	7	4	32.96	稳定
<i>Bra015809</i>	79	8 555.0	8.15	6	4	30.42	稳定
<i>Bra015811</i>	80	8 830.3	8.70	7	3	29.45	稳定
<i>Bra008225</i>	80	8 918.5	9.00	12	6	32.13	稳定
<i>Bra004900</i>	80	9 185.8	9.06	12	6	73.26	不稳定
<i>Bra025396</i>	80	9 258.1	9.41	15	6	42.69	不稳定
<i>Bra016501</i>	78	8 814.2	8.85	9	4	57.46	不稳定
<i>Bra031312</i>	79	8 653.8	5.45	5	7	38.33	稳定
<i>Bra029208</i>	60	6 946.0	8.88	9	4	80.54	不稳定
<i>Bra017421</i>	77	8 691.3	9.82	13	3	49.86	不稳定
<i>Bra024832</i>	77	8 510.0	9.63	12	3	34.44	稳定
<i>Bra026615</i>	77	8 511.0	9.33	11	4	49.43	不稳定

表 4 白菜防御素基因家族表达模式预测

基因序列号	EST 序列命中总个数	不同组织部位 EST 序列命中个数														
		愈伤组织	根	长角果	种子	植株	胚珠	初生叶	成熟叶	老叶	初生花序	花	花蕾	雌蕊	蜜腺	地上部
<i>Bra003740</i>	21	9	-	3	1	1	1	4	1	-	-	-	-	-	-	1
<i>Bra008226</i>	21	9	-	3	1	1	1	4	1	-	-	-	-	-	-	1
<i>Bra015809</i>	21	9	-	3	1	1	1	4	1	-	-	-	-	-	-	1
<i>Bra015811</i>	21	9	-	3	1	1	1	4	1	-	-	-	-	-	-	1
<i>Bra008225</i>	4	1	-	-	-	-	-	1	1	-	-	-	-	-	-	1
<i>Bra004900</i>	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-
<i>Bra025396</i>	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-
<i>Bra016501</i>	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Bra031312</i>	1	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
<i>Bra029208</i>	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Bra017421</i>	38	4	4	6	1	2	-	1	-	3	1	1	11	2	2	-
<i>Bra024832</i>	38	4	4	5	1	2	1	1	-	3	1	1	11	2	2	-
<i>Bra026615</i>	38	4	3	5	1	2	1	2	-	3	1	1	11	2	2	-

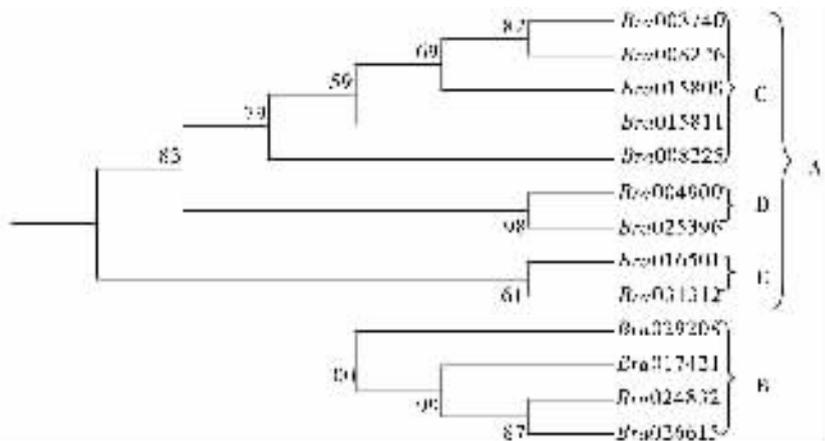


图 4 白菜防御素基因 CDS 序列进化树

3 讨论

通过对白菜防御素基因家族进行预测和鉴定,本研究最终确定了 13 个白菜防御素基因家族成员。该家族成员 CDS 区域在 240 bp 左右,只编码 80 多个氨基酸,这可能与生物体受到外源病原菌侵染时快速翻译、快速应答有关。白菜在进化历史上曾经经历了一个三倍化和随后的二倍化过程,与拥有 15 个基因的拟南芥防御素基因家族相比,白菜

防御素基因家族并没有出现扩张现象,这表明在白菜多倍化进化过程中多数防御素冗余基因可能发生丢失。据报道,基因重复事件对基因组结构和功能的影响是轰动的^[11]。从白菜基因组防御素基因家族结构来看,内含子长度变化在各个成员中非常明显,具体表现为B分支中*Bra029208*内含子丢失,A分支中的成员内含子长度比B分支中的其他3个成员要短,总体来看,内含子在各个成员中长短不一,呈现长度多态性。*Bra015809*和*Bra029208*两个基因在进化过程中丢失了信号肽序列,*Bra016501*和*Bra031312*两个基因在进化过程中信号肽序列已经发生分化,其余拷贝成员之间信号肽序列保守性相对较高。根据先前的研究结果,基因重复事件可以激活基因组中的转座子元件^[12],转座子元件的激活是导致部分基因内含子丢失的一个重要因素^[13-15],内含子的丢失又可能造成重复基因在不同拷贝中基因表达模式出现分化^[16]。此外,孔祥银等^[17]的研究发现,基因发生重复事件后会造体外显子拼接元件在不同拷贝中的迅速变异,非编码区的分化要明显大于编码区。因此,本研究认为内含子的长度变化可能与转座子活性和外显子拼接元件变异有关。理化性质分析结果表明,只有*Bra031312*基因表达的蛋白质呈酸性,其余拷贝成员表达蛋白质均呈碱性,表明基因*Bra031312*表达蛋白的理化环境已经发生改变。此外,各个拷贝成员编码的蛋白质稳定指数评价结果显示,有6个拷贝成员编码的蛋白质不稳定,表明其在进化过程中防御素功能可能已经丧失。从基因表达模式预测来看,各个拷贝成员之间表达模式的分化差异与其进化分歧模式是基本一致的,有两个拷贝成员可能已经发生沉默,有3个拷贝成员表达模式可能呈现组织特异性,其余拷贝成员表达模式主要体现在表达部位差异和表达量差异方面。

先前研究表明,基因重复事件发生后,顺式调控元件区域会发生迅速变异^[18-19]。从白菜防御素基因多态性分析结果来看,虽然插入缺失事件和长度在CDS区域较高,但是插入缺失多样性指数、多态性位点个数、突变总数、核苷酸平均差异数和核苷酸多样性指数要明显低于启动子区域,这表明白菜基因重复事件发生后,上游启动子区域的进化变异速度要远远高于CDS区域。因此,本研究认为上游启动子区域核苷酸多样化可能是造成重复基因功能以及表达模式多样化的又一个重要原因。此外,从白菜防御素基因上游调控区域顺式作用元件分析结果

可以看出,此类基因上游存在大量的光应答元件、逆境伤害应答元件和信号分子应答元件,说明这些基因可以感知外界环境变化,有利于生物体在遭受逆境胁迫时作出迅速应答。本研究发现,白菜防御素基因上游顺式调控元件在不同拷贝之间已经出现分化,不同拷贝之间元件的种类和数目各不相同。据报道,信号分子茉莉酸^[20]、乙烯^[21]、水杨酸^[22]、赤霉素^[23]和脱落酸^[24]等在防御信号调控网络中扮演重要的角色,各个信号分子之间存在一个复杂的信号交谈机制,彼此之间既相互作用,又各自独立,生物体在遭受外源病原菌入侵时会产生多种信号分子从而激活防御素基因的表达,其中有单独依赖水杨酸、乙烯和茉莉酸等信号分子途径,也有共同依赖乙烯和茉莉酸信号分子途径^[25-26]。信号分子之间共同组成一个复杂的防御素信号调控网络,能够在外界环境发生改变时迅速激活防御素应答基因的表达。

参考文献:

- [1] Ganz T. Defensins: antimicrobial peptides of innate immunity [J]. *Nat Rev Immunol*, 2003, 3(9): 710-720.
- [2] Thomma B P, Cammue B P, Thevissen K. Plant defensins [J]. *Planta*, 2002, 216(2): 193-202.
- [3] 付蓝宝,于嘉林,刘伟华. 防御素的生物学特性及其抗病基因工程 [J]. *遗传*, 2011, 33(5): 512-519.
- [4] Colilla F J, Rocher A, Mendez E. Gamma-Purothionins: amino acid sequence of two polypeptides of a new family of thionins from wheat endosperm [J]. *FEBS Lett*, 1990, 270(1/2): 191-194.
- [5] Mendez E, Moreno A, Colilla F. Primary structure and inhibition of protein synthesis in eukaryotic cell-free system of a novel thionin, gamma-hordothionin, from barley endosperm [J]. *Eur J Biochem*, 1990, 194(2): 533-539.
- [6] 王轶,罗勤,吴俊,等. 甘蓝型油菜植物防御素基因的克隆与原核表达 [J]. *四川大学学报:自然科学版*, 2005, 42(4): 831-834.
- [7] Mun J H, Kwon S J, Yang T J, et al. Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication [J]. *Genome Biol*, 2009, 10(10): R111.
- [8] Mun J H, Kwon S J, Seol Y J, et al. Sequence and structure of *Brassica rapa* chromosome A3 [J]. *Genome Biol*, 2011, 11(9): R94.
- [9] Wang X W, Wang H Z, Wang J, et al. The genome of the mesopolyploid crop species *Brassica rapa* [J]. *Nat Genet*, 2011, 43: 1035-1039.
- [10] Park Y S, Jeon M H, Lee S H, et al. Activation of defense

- responses in Chinese cabbage by a nonhost pathogen , *Pseudomonas syringae* pv. tomato [J]. *J Biochem Mol Biol* 2005 ,38(6) :748-754.
- [11] McClintock B. The significance of responses of the genome to challenge [J]. *Science* ,1984 ,226(4676) :792-801.
- [12] Yaakov B ,Kashkush K. Methylation ,transcription and rearrangements of transposable elements in synthetic allopolyploids [J]. *Int J Plant Genomics* 2011(3) :1-7.
- [13] Roy S W ,Hartl D L. Very little intron loss/gain in plasmodium intron loss/gain mutation rates and intron number [J]. *Genome Res* 2006 ,16 :750-756.
- [14] Roy S W ,Penny D. Widespread intron loss suggests retrotransposon activity in ancient apicomplexans [J]. *Mol Biol Evol* 2007 ,24(9) :1926-1933.
- [15] Cohen N E ,Shen R ,Carmel L. The role of reverse- transcriptase in intron gain and loss mechanisms [J]. *Mol Biol Evol* 2012 ,29(1) :179-186.
- [16] Mrinal N ,Nagaraju J. Intron loss is associated with gain of function in the evolution of the gloverin family of antibacterial genes in *Bombyx mori*[J]. *Biol Chem* ,2008 ,283 :23376-23387.
- [17] Zhang Z ,Zhou L ,Wang P et al. Divergence of exonic splicing elements after gene duplication and the impact on gene structures [J]. *Genome Biol* 2009 ,10(11) :R120.
- [18] Kassahn K S ,Dang V T ,Wilkins S J ,et al. Evolution of gene function and regulatory control after whole-genome duplication :comparative analyses in vertebrates [J]. *Genome Res* 2009 ,19(8) :1404-1418.
- [19] Conant G C. Rapid reorganization of the transcriptional regulatory network after genome duplication in yeast [J]. *Proc Bio Sci* 2010 ,277 :869-876.
- [20] Kusnierczyk A ,Tran D H T ,Winge P ,et al. Testing the importance of jasmonate signalling in induction of plant defences upon cabbage aphid (*Brevicoryne brassicae*) attack [J]. *BMC Genomics* 2011 ,12(1) :#23.
- [21] Núñez-Pastrana R ,Arcos-Ortega G ,Souza-Perera R A ,et al. Ethylene but not salicylic acid or methyl jasmonate induces a resistance response against *Phytophthora capsici* in Habanero pepper [J]. *Eur J Plant Pathol* 2011 ,131(4) :669-683.
- [22] Moran P J ,Thompson G A. Molecular responses to aphid feeding in Arabidopsis in relation to plant defense pathways [J]. *Plant Physiol* 2001 ,125(2) :1074-1085.
- [23] Van den Heuvel K J P T ,Hulzink J M R ,Barendse G M W ,et al. The expression of tgas118 ,encoding a defensin in *Lycopersicon esculentum* is regulated by gibberellin [J]. *J Exp Bot* 2001 ,52(360) :1427-1436.
- [24] Chao W S ,Gu Y Q ,Pautot V ,et al. Leucine aminopeptidase RNAs ,proteins ,and activities increase in response to water deficit ,salinity ,and the wound signals systemin ,methyl jasmonate ,and abscisic acid [J]. *Plant Physiol* ,1999 ,120(4) :979-992.
- [25] Liu X Q ,Bai X Q ,Qian Q ,et al. OsWRKY03 ,a rice transcriptional activator that functions in defense signaling pathway upstream of OsNPR1 [J]. *Cell Res* ,2005 ,15(8) :593-603.
- [26] Sun J ,Cardoza V ,Mitchell D M ,et al. Crosstalk between jasmonic acid ,ethylene and nod factor signaling allows integration of diverse inputs for regulation of nodulation [J]. *Plant J* 2006 ,46(6) :961-970.

Prediction and Analysis of Architecture as Well as Function of *Brassica rapa* Defensin Genes

WANG Rui-xue , SHEN Liang-yu , ZHAO Jing-hui , LI Rong-chong , ZHEN Xiao-min , ZHANG Tao

(College of Life Science , Chongqing Normal University , Chongqing 400047 , China)

Abstract :The previous reported plant defensin gene amino acid sequences were performed using blast similar alignment program on *Brassica rapa* database to predict the candidate defensin gene in *Brassica rapa* genome by bioinformatic methods ,the conserved domain of defensin gene has also been confirmed. We analyzed the variation of upstream region , promoter region , exon region , intron region and UTR region during the evolution process. In addition ,the variation of expression pattern as well as cis-regulate element have been analyzed in this paper. The results demonstrated that defensin genes encoding 60-80 amino acid short peptides and each has eight conserved cysteine residues. The major function domain of defensin genes maintain relative stable ,signal peptide sequences have been diverged in partial members. The length of intron has three variation patterns that were increase , contraction and disappear respectively. The nucleotide variations in promoter region were remarkable which compare to other regions. The cis-regulate element mainly involved in light response element , defensin response and signal molecule response element. The prediction of the expression pattern may diverge in the process of evolution such as gene silencing and selective expression differences.

Key words : *Brassica rapa* ; defensin ; duplicate gene ; evolution

(责任编辑 方 兴)