

## 蓖麻 DNA 糖基化酶的序列与基因表达分析

代梦媛<sup>1,3</sup>, 敖涛<sup>1,3</sup>, 徐伟<sup>2,3</sup>, 刘爱忠<sup>2\*</sup>

(1. 中国科学院西双版纳热带植物园热带植物资源开放实验室, 云南 勐腊, 666303;

2. 中国科学院昆明植物研究所, 云南 昆明, 650201; 3. 中国科学院大学, 北京, 100049)

**摘要:**为揭示 DNA 糖基化酶在调节蓖麻生长发育和影响基因印记过程中的分子机理,本研究基于蓖麻全基因组、利用生物信息学的方法结合转录组数据和 RT-PCR 半定量表达分析,鉴别蓖麻 DNA 糖基化酶氨基酸序列的结构特征、系统发生和表达形式。结果显示:三种 DNA 糖基化酶(DME、ROS 和 DML)氨基酸均为不稳定氨基酸,具有典型的与 DNA 结合和碱基切除修复有关的保守结构域;系统进化分析发现三种 DNA 糖基化酶在植物中同源性强,而且是独立演化的结果;在不同组织的表达上,蓖麻的 DME 和 ROS 基因具有相似的表达形式,即在根、茎和叶中不表达,在胚乳中表达最高,而 DML 在各个组织中均未检测到基因的表达。

**关键词:**蓖麻;DNA 糖基化酶;脱甲基化;表观调控**中图分类号:**S565.603 **文献标识码:**A **文章编号:**1007-9084(2014)02-0181-08**Sequence and expression analysis of DNA glycosylase in castor bean (*Ricinus communis* L.)**DAI Meng-yuan<sup>1,3</sup>, AO Tao<sup>1,3</sup>, XU Wei<sup>2,3</sup>, LIU Ai-zhong<sup>2\*</sup>

(1. Key Laboratory of Tropical Plant Resource and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303, China;

2. Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China;

3. Graduate University of the Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:**To reveal the potential function of DNA glycosylases mediating DNA demethylation leading to gene imprinting in the growth and development of castor bean (*Ricinus communis*), the putative genes encoding DNA glycosylases were identified and characterized based on castor bean genome reference data. 3 genes (*RcDME*, *RcROS* and *RcDML*) encoding DNA glycosylases were identified. Their structures, conserved domains, and subcellular localization were characterized. Gene expression was inspected based on the high-throughput transcriptional sequencing data and RT-PCR. Results showed that the amino acids coded by the 3 genes were unstable ones, and the sequences contained typical DNA-binding conserved domains and domains related to base excision repair. Phylogenetically, DME, ROS and DML were homologous and evolved independently in plants. Both *RcDME* and *RcROS* exhibited a similar expression pattern among different tissues with the highest expression in endosperm, while the expression level of *RcDML* was not detected in all tested tissues.

**Key words:***Ricinus communis*; DNA glycosylase; Demethylation; Epigenetic regulation

DNA 分子上的 5-甲基胞嘧啶(5-meC)经常因发生甲基化而引起表观遗传上的 DNA 修饰,广泛存在于真核生物的基因组中。在开花植物中,大约 50% 基因组胞嘧啶被甲基化<sup>[1]</sup>,且发生在 CG、CHG (H 为 A、C 或 T)和 CHH 序列上,大部分甲基化存

在于转座子和其它重复序列中<sup>[2,3]</sup>。DNA 从头甲基化主要是通过 DNA 甲基转移酶来建立和维持<sup>[4]</sup>,在基因组印记、基因和转座子的沉默和基因组的防御等方面具有重要的功能<sup>[5,6]</sup>。然而基因组最终的甲基化水平和模式是由 DNA 甲基化和 DNA 脱甲基化

过程共同决定的。与 DNA 甲基化相比,关于 DNA 脱甲基化作用机制及其关键酶的研究相对较少。在高等植物中,目前认为 DNA 的脱甲基化的机制主要有两种,一种是在 DNA 复制过程中的被动脱甲基化,另一种是 DNA 的主动脱甲基化作用,主要靠 DNA 糖基化酶(DNA glycosylase)来实现<sup>[7,8]</sup>。

DNA 糖基化酶是一类植物特有的大分子蛋白家族,能特异性识别和切除 5 - meC,然后通过碱基切除修复途径(base excision repair, BER),用没有甲基化的胞嘧啶取代原有的甲基胞嘧啶,最终完成 DNA 的脱甲基化途径<sup>[8]</sup>。目前,在拟南芥中发现了三种 DNA 糖基化酶,包括 DEMETER(DME)<sup>[9]</sup>、REPRESSOR OF SILENCING 1(ROS1)<sup>[10]</sup>和 DML 类。其中,DME 特异地雌配子体的中央细胞中表达,导致了胚和胚乳甲基化水平的差异和印记基因的表达,从而控制胚乳和种子的发育<sup>[9,12]</sup>;DML 包括 DML2 和 DML3<sup>[11]</sup>,主要在营养组织中发挥作用,其主要功能是阻止某些特异位点的过度甲基化和维持全基因组甲基化水平的稳定<sup>[11]</sup>;ROS1 的主要功能是抑制 siRNA 引起的 DNA 超甲基化过程和转基因的转录沉默作用<sup>[10]</sup>。近年来,越来越多的研究发现 DNA 糖基化酶在植物的生长发育、生物和非生物胁迫、印迹基因的表达、胚乳和种子的发育方面具有重要的功能<sup>[5~10]</sup>。

蓖麻(*Ricinus communis* L.)属于大戟科(Euphorbiaceae)双子叶一年生或多年生草本植物,是世界上最重要的非食用油料作物之一<sup>[13,14]</sup>。蓖麻胚乳中积累了大量的三酰甘油,其 90% 是由蓖麻油酸(顺式-12-羟基十八碳-9-烯酸)组成<sup>[15]</sup>。蓖麻油酸在高温下不易挥发、低温下不易凝固,而被广泛应用于工业、国防和航空等;另外,蓖麻油酸极性很强,结构稳定,能与酒精完全互溶即可制作成生物柴油。目前,利用蓖麻油作为生物柴油的原料在生物柴油产业化中已经受到了世界各个国家的高度关注和利用<sup>[16]</sup>。促进蓖麻的遗传改良、培育高产优良品种、提高产量是当今蓖麻产业发展的主要任务。理解蓖麻生长发育以及对环境的适应生理机制是蓖麻遗传改良的前提。本研究通过对蓖麻 DNA 糖基化酶序列进行同源比对和基因表达分析,为调查 DNA 脱甲基化作用对蓖麻的生长发育调控方面的影响,以及理解蓖麻生长发育以及对环境适应的表观机制提供一些依据。此外,蓖麻种子是典型的双子叶胚乳型油料种子,深入研究其种子发育与储存物质积累的生理机制对理解植物种子发育与储存物质积累分子基础具有借鉴意义。

## 1 材料与方法

### 1.1 蓖麻 DNA 糖基化酶基因的获取与确定

以拟南芥 DNA 糖基化酶蛋白序列 AtDME(AT5G04560)、AtROS1(AT2G36490)、AtDML3(AT4G34060)和 AtDML2(AT3G10010)(下载自拟南芥基因组数据库:<http://www.arabidopsis.org>)为基序,在蓖麻基因组数据库(<http://castorbean.jcvi.org>)中进行 Blastp 同源序列比对,以获得显著同源候选蛋白( $P < 0.001$ , 相似度  $> 40\%$ ),然后将得到的候选蛋白序列在拟南芥基因组数据库中进行同源比对以及 SMART 分析,进一步检测糖基化酶蛋白的功能结构域的存在。

### 1.2 DNA 糖基化酶的分子进化树构建

利用候选蓖麻糖基化酶蛋白序列在 GenBank 数据库中 Blastp 查找其它物种同源序列;利用 Clustal W<sup>[17]</sup> 软件进行氨基酸多序列比对分析,利用 MEGA 5.0 软件中的 Neighbor - Joining 法构建系统进化树,设置 Bootstrap 的值为 1000 进行校正,生成最终的系统发育树。

### 1.3 蓖麻 DNA 糖基化酶蛋白的理化特性及结构分析

利用 ExPASy Proteomics Server 工具(<http://www.expasy.org/tools>)分析蛋白质理化性质如分子量、等电点、亲水性、稳定性等;利用 SignalP 4.0 程序(<http://www.cbs.dtu.dk/services/SignalP-4.0/>)进行信号肽分析;利用在线软件 Predotar1.03(<http://urgi.versailles.inra.fr/predotar/predotar.html>)和 ProtComp v.9.0(<http://linux1.softberry.com/berry.phtml>)对蛋白质进行亚细胞定位分析;利用 SMART(<http://smart.embl-heidelberg.de/>)在线工具对蛋白结构域进行预测;利用 SOPMA 软件预测蛋白的二级结构如:α - 螺旋(α - helix, H)、β - 转角(β - turn, T)、无规则卷曲(random coil, C)以及延伸链(extended strand, E)等。

### 1.4 蓖麻 DNA 糖基化酶基因的表达分析

实验所用蓖麻品种为淄博 306(ZB306)。蓖麻种子萌发后移植于西双版纳热带植物园温室中。自然条件下生长 60d 后开始取材,分别采集根尖、茎、叶、发育中的雌花和花粉,自花授粉后 35d 的胚和胚乳,将组织样品立即放在液氮中保存。利用 TaKaRa 公司的植物总 RNA 提取试剂盒对各个组织的总 RNA 进行抽提;利用 Nanodrop 仪器测定 RNA 的浓度和质量并用琼脂糖凝胶电泳检测 RNA 的完整性;利用 Transgene 公司的 cDNA Synthesis SuperMi 试剂

盒对分离的 ploy A + RNA 进行反转录。利用 primer3.0 在线软件对 RcDME、RcROS1 和 RcDML 的 CDS 序列进行引物设计(表 1),所有引物由上海捷瑞生物工程有限公司合成。以各个组织反转录的 cDNA 为模板进行 RT-PCR(Semi-quantitative Re-

altime-PCR),PCR 反应条件为 94℃ 预变性 4min; 94℃ 变性 30s,55℃ 退火 30s,72℃ 延伸 30s,35 个循环;72℃ 延伸 7min。对胚乳中得到的 RcDME、RcROS1、RcDML 的 PCR 产物进行纯化并测序验证。

表 1 PCR 引物序列  
Table 1 PCR primer sequences

引物名称 Primer name	引物序列 Primer sequence(5' - 3')	扩增长度 PCR product length/bp	退火温度 Annealing temperature/℃
RcDME - F	AGGATTTCTCTAAACCGACTGGT	432	58.9
RcDME - R	CAAAATGTCTGCATTTCTGCTCT		57.4
RcROS1 - F	GCAACCAGAAAAGCAAAGAGCA	467	62.3
RcROS1 - R	AACTGAAGTGATTCGGGCAGTG		61.6
RcDML - F	AGACAAGAAAAGACAACCAAAGG	475	55.7
RcDML - R	ATGACTGGGTACTCTTCCAAAA		55.9

## 2 结果与分析

### 2.1 蓖麻 DNA 糖基化酶基因的获取确定与氨基酸序列比对

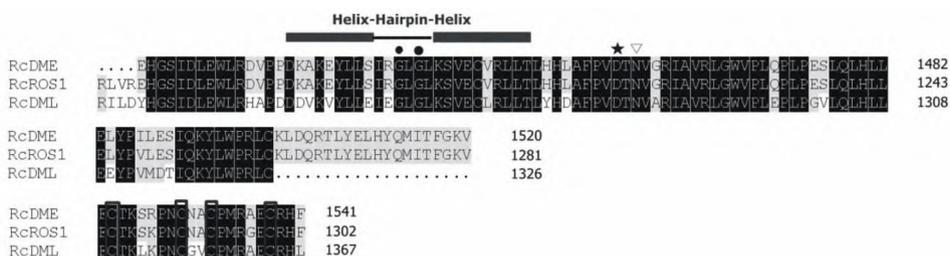
利用拟南芥糖基化酶 AtDME(AT5G04560)、AtROS1(AT2G36490)、AtDML3(AT4G34060)和 AtDML2(AT3G10010)的蛋白序列在蓖麻基因组数据库进行同源序列比对,最终获得了三个相应的蓖麻同源蛋白:RcDME(29428.m000327)、RcROS1(29092.m000452)和 RcDML(29991.m000674)。氨基酸序列同源性分别为 50%、42% 和 48%。同时,cDNA 测序显示出一致的核苷酸组成和序列长度。SMART 和 Pfam 结构功能域分析发现三个酶蛋白都有糖基化酶活性的保守功能域:ENDO3c 和 FES(图

1)。ENDO3c 是具有核酸内切酶 III(endonuclease III)活性的功能域,它能修复 DNA 上受损的嘧啶碱基,同时通过 β 切除机制切除非嘌呤/非嘧啶位点的磷酸二酯键<sup>[18]</sup>,FES 是 endonuclease III 上的铁硫结合域(Fe-S cluster),其中有四个保守的半胱氨酸残基,预测功能与 DNA 结合有关<sup>[19]</sup>。蓖麻 DNA 糖基化酶的蛋白序列比对结果显示,这三个蓖麻糖基化酶蛋白均含有一个 Helix-Hairpin-Helix 的结构元件和其后的一个甘/脯氨酸富集链和一个保守的天冬氨酸(图 2),这些结构元件被认为是 DNA 糖基化酶碱基切除超家族的典型特点<sup>[20-22]</sup>,保守的天冬氨酸在碱基切除反应中作为电子供体参与反应,提供酶催化活性。



图 1 蓖麻 DNA 糖基化酶蛋白结构域

Fig. 1 Predicted protein domains of DNA glycosylase of *Ricinus communis*



注:ENDO3c 结构域:★标示保守天冬氨酸,▽标示单功能糖基化酶保守天冬酰胺,  
●标示 hairpin 上的保守甘氨酸;FES 结构域:方框里是保守半胱氨酸

Note: ENDO3c domain: ★ represents conserved aspartate residues, ▽ represents conserved asparagine residues in monofunctional glycosylases, ● represents conserved glycine residues. FES domain: conserved cysteines in boxes

图 2 蓖麻糖基化酶蛋白两个保守结构域组成

Fig. 2 Two conserved domains of DNA glycosylase of *R. communis*

	Lysine-rich	
RcDME	NYVSLVGNVFNQHNPKDKFLVIVGSAVAISTPFEEFKESCGGSDQVILDLNNTPEOKTPEKRKRRHKFKVIVGEGKPKRTEK	342
JcrDME	DLVSDIDGFNPEKNSNELLHNIGDSFSAVISAIFEEFKDCSKGSDQVILDLNCTPEOKPEKRRKRRHKFKVIVGEGKPKRTEK	335
AthDME	.....MQSLMDSAVNATEATEQNDGSRQDVLEEDLNKTPPEOKPKRRKRRHKFKVIVGEGKPKRTEK	61
AlyDME	QIVTTIGHEITPEKSKDKMSLMDA.SAVNATEVTEQNDGSRQDVLEEDLNKTPPEOKPKRRKRRHKFKVIVGEGKPKRTEK	326
CsaDME	NFVTYFLDEATQKRRHDELLKDLITL.SAISEPT...KVEKGSDDQVILDLNNTPEOKTPEKRKRRHKFKVIVGEGKPKRTEK	290
PtdDME	DLVYVFNELKRNQNSBEELCNLAISVAISTPFEEFKDARDEGQVILDLNCTPEOKPEKRRKRRHKFKVIVGEGKPKRTEK	235
TcaDME	NLITYNANEVVSCHN.CELLONIVLSSVAISTPFEEKRDARSDEGQVILDLNNTPEOKPEKRRKRRHKFKVIVGEGKPKRTEK	342
RcDME	SV.....PKPTVDPNEKAIERKRYVRRKKGQK.....ES	370
JcrDME	AA.....TRKIADPKKATASEKRYVRRKQTOH.....ES	363
AthDME	KPAELPKVVEGKPKRKRPAKAAIQEKVNSKETEKGSARRK...NLKESATKPKPANVGDMSNKSPE...VTLKSCKRA	130
AlyDME	KF.....ATQEKVNSKETEKGSARRKAKAKTNLKEKSAATKPKPAHVGDMSNKSPE...VTLKSCKRA	381
CsaDME	AA.....PKLS...KETSCKRYV...R	288
PtdDME	AA.....TKNTDPEKKEKSTERYVRRKALKEPAIQDTSADTETPPSSAK...RKYGLNES	288
TcaDME	PA.....ITKNINSKENPESGRKRYVRRKGLTESATEQDASITKSDTAAATPAKRYVRRKSLKES	402
RcDME	DQKST...ARRKAGQPAKQKPIITIEEIIYRMEHINNEVKG....DCTAIVPEVKGDCALIFVGDGEFIIKRRKRRFPVLDL	861
JcrDME	GLQSP...ARRKGRPAKQKFPATIEEIIYRMERLSNEKMGK...EONAMVLYKGDCTLVVPEGGEFIIKRRKRRFPVLDL	890
AthDME	GRKS IQDSGKARGPSGELLCDOSIABIETIYRMQNLVYGDKEREQ...EONAMVLYKGDCTLVVPE...SKRRKRRFPVLDI	700
AlyDME	GRKSIQASGARGPSGKLLCDLSDIABIETIYRMQNLVYGDKEREQ...EONAMVLYKGDCTLVVPE...SKRRKRRFPVLDI	963
CsaDME	GRKSIQASGARGPSGKLLCDLSDIABIETIYRMQNLVYGDKEREQ...EONAMVLYKGDCTLVVPE...SKRRKRRFPVLDI	815
PtdDME	HQOQS...AKRRGRPAKQKFSSTIEEIIHMECLSNARSKKIKNKEKONALVPEVGDCTLVVPEGGEFVKKRRKRRFPVLDL	744
TcaDME	YQOSS...KAGPSAKQIYPIPIEETINKFMGLTIDERNNEAKSEVONALVPEVGDCTLVVPEGGEFVKKRRKRRFPVLDL	965
RcDME	LPETEFERWKLMLWRE.GEGEELGTBOEKQWVDEERFVGRGRADSFARMHLVQGDRRSRKRGSVVDSVIGVFLTONVS	940
JcrDME	LPETEFERWKLMLWRE.GEGEELGTBOEKQWVDEERFVGRGRADSFARMHLVQGDRRSRKRGSVVDSVIGVFLTONVS	969
AthDME	LPETEFERWKLMLWRE.GEGEELGTBOEKQWVDEERFVGRGRADSFARMHLVQGDRRSRKRGSVVDSVIGVFLTONVS	780
AlyDME	LPETEFERWKLMLWRE.GEGEELGTBOEKQWVDEERFVGRGRADSFARMHLVQGDRRSRKRGSVVDSVIGVFLTONVS	1042
CsaDME	LPETEFERWKLMLWRE.GEGEELGTBOEKQWVDEERFVGRGRADSFARMHLVQGDRRSRKRGSVVDSVIGVFLTONVS	894
PtdDME	LPETEFERWKLMLWRE.GEGEELGTBOEKQWVDEERFVGRGRADSFARMHLVQGDRRSRKRGSVVDSVIGVFLTONVS	823
TcaDME	LPETEFERWKLMLWRE.GEGEELGTBOEKQWVDEERFVGRGRADSFARMHLVQGDRRSRKRGSVVDSVIGVFLTONVS	1043
Domain A		
RcDME	DHLSSSAFMSLAARFPLKSMRNRCTERDEPRRLIC.EPDIYMLNPNTIKWHEK.LLTFYNGSSMTPHESIEHRRDOET	1018
JcrDME	DHLSSSAFMSLAARFPLKSMRNRNRNRDGTTRILVE.PDVSTSNPNTIKWHENSNCNLYNGISIALYESIEHQRECMT	1048
AthDME	DHLSSSAFMSLAARFPLKSSSRDERNRVSVVVE.DPEGICILMLNEIPSWCEK.VQHSDEMBSVGDSCKEQLRDCSN	858
CsaDME	DHLSSSAFMSLAARFPLKSSSRDERNRVSVVVE.DPEGICILMLNEIPSWCEK.VQHSDEMBSVGDSCKEQLRDCSN	1120
PtdDME	DHLSSSAFMSLAARFPLKSSANLRTQGEVETSIVANESAAVLYPAESIRHWVQLSVRFPMF...QTSINHCNQIAN	970
TcaDME	DHLSSSAFMSLAARFPLKSSNCAAHDSHRKGVILE.EPDMCMQNFNDIKNKSK.FRYELHNGSPMPTIHGSAEPQGES	901
	DHLSSSAFMSLAARFPLKSSNCKRECDGDGVKILIE.EPEFCENPNETIKWHEKILFSLHLDLDRSPSTIYDRRNGEN	1122
RcDME	SCNTRTSIVVAHSHSYSPPEEVLSSQSFSDSSVYVSNVYVSYSGNSLEADPAKGGKHNNHNNTS.NAOKLE...FEFF	1092
JcrDME	SWTRTSIVVAHSHSHPEEVALSSQSFSDSSVYVSNVYVSYSGNSLEADPAKGGKHNNHNNTS.NAOKLE...FEFF	1127
AthDME	SGTRTSIVVAHSHSHPEEVLSSQSFSDSSVYVSNVYVSYSGNSLEADPAKGGKHNNHNNTS.NAOKLE...FEFF	936
AlyDME	SGTRTSIVVAHSHSHPEEVLSSQSFSDSSVYVSNVYVSYSGNSLEADPAKGGKHNNHNNTS.NAOKLE...FEFF	1198
CsaDME	SGTRTSIVVAHSHSHPEEVLSSQSFSDSSVYVSNVYVSYSGNSLEADPAKGGKHNNHNNTS.NAOKLE...FEFF	1049
PtdDME	WCLTRASVMVGAQSHSLSEEFVSSQSFSDSSVYVSNVYVSYSGNSLEADPAKGGKHNNHNNTS.NAOKLE...FEFF	981
TcaDME	FGTRTSIVVAHSHSHPEEVLSSQSFSDSSVYVSNVYVSYSGNSLEADPAKGGKHNNHNNTS.NAOKLE...FEFF	1199
RcDME	THFTNKANLNRNA..SKARKKAESQKDAVDWDLRKRQVLVNDRPKRSESAEMSDLVDAEMSHVNEHEDTKERGMN	1390
JcrDME	VHFTSKVSSINS..LKTRKKAQSKKDAVDWDLRKRQVLANGERKRSQDTMDSLDVBAIKRGSVNVQADAKIKERGMN	1442
AthDME	KEQTAVEYKKTNATILREMKGTLDGKPKTSOWDSLRKQV.EGNEGRCRNNKNNMDSIDYBAIRASISESEAIKERGMN	1239
AlyDME	KEQTAVEYKKTNATILREMKGTLDGKPKTSOWDSLRKQV.EGNEGRCRNNKNNMDSIDYBAIRASISESEAIKERGMN	1505
CsaDME	AHSLSOAHNEGNSIPSKAKRRKVNSEKKGMDWDLRKRQV.EGNEGRCRNNKNNMDSIDYBAIRASISESEAIKERGMN	1358
PtdDME	WHLSDNRELNKSNANGSKARKKVESEKADVFDWDLRKRQV.CTNG.RKERTETMDSLDVBAIKRGSVNVQADAKIKERGMN	1297
TcaDME	IHSSEKENGAYSF..LKSRRKAQESKKNATWDLRKRQVLANGERKRSQDTMDSLDVBAIKRGSVNVQADAKIKERGMN	1507
ENDO3c Helix-Hairpin-Helix		
RcDME	NMLAARIKDFLRLNRLVREHESIDLEWLRDVEPDKARQVYLLSIRGLGLKSVCEVRLLLTHLAFVFDVTVNVR IAVRIGWVPL	1470
JcrDME	NMLAARIQEFLNRLVREHESIDLEWLRDVEPDKARQVYLLSIRGLGLKSVCEVRLLLTHLAFVFDVTVNVR IAVRIGWVPL	1522
AthDME	NMLAARIKDFLRLNRLVREHESIDLEWLRDVEPDKARQVYLLSIRGLGLKSVCEVRLLLTHLAFVFDVTVNVR IAVRIGWVPL	1319
AlyDME	NMLAARIKDFLRLNRLVREHESIDLEWLRDVEPDKARQVYLLSIRGLGLKSVCEVRLLLTHLAFVFDVTVNVR IAVRIGWVPL	1585
CsaDME	NMLAARIKDFLRLNRLVREHESIDLEWLRDVEPDKARQVYLLSIRGLGLKSVCEVRLLLTHLAFVFDVTVNVR IAVRIGWVPL	1438
PtdDME	NMLAARIKDFLRLNRLVREHESIDLEWLRDVEPDKARQVYLLSIRGLGLKSVCEVRLLLTHLAFVFDVTVNVR IAVRIGWVPL	1377
TcaDME	NMLAARIKDFLRLNRLVREHESIDLEWLRDVEPDKARQVYLLSIRGLGLKSVCEVRLLLTHLAFVFDVTVNVR IAVRIGWVPL	1587
Glycosylase Domain FES		
RcDME	QPLPESLOHLLELVYVLESIQRYLWPRCKLDQRTLYVELHYCMITFGKVFCTKSRPNCNACFMRAECRHFASAFASARL	1550
JcrDME	QPLPESLOHLLELVYVLESIQRYLWPRCKLDQRTLYVELHYCMITFGKVFCTKSRPNCNACFMRAECRHFASAFASARL	1602
AthDME	QPLPESLOHLLELVYVLESIQRYLWPRCKLDQRTLYVELHYCMITFGKVFCTKSRPNCNACFMRAECRHFASAFASARL	1399
AlyDME	QPLPESLOHLLELVYVLESIQRYLWPRCKLDQRTLYVELHYCMITFGKVFCTKSRPNCNACFMRAECRHFASAFASARL	1665
CsaDME	QPLPESLOHLLELVYVLESIQRYLWPRCKLDQRTLYVELHYCMITFGKVFCTKSRPNCNACFMRAECRHFASAFASARL	1519
PtdDME	QPLPESLOHLLELVYVLESIQRYLWPRCKLDQRTLYVELHYCMITFGKVFCTKSRPNCNACFMRAECRHFASAFASARL	1457
TcaDME	QPLPESLOHLLELVYVLESIQRYLWPRCKLDQRTLYVELHYCMITFGKVFCTKSRPNCNACFMRAECRHFASAFASARL	1667
RcDME	ALGPEFE..DKSIVTAVLTTTERRSGIVIDPLPIPPAEDN LLTRRGS.DIVSCVPIIIEBATEBQCHHEVLESIEDIEDIFD	1627
JcrDME	ALGPEFE..EKSIVTAVLTTTERRSGIVIDPMPLPPEEN SLKRGEF.DIVSCVPIIIEBATEBQCHHEVLESIEDIEDIFY	1679
AthDME	ALGPEFE..ERLTSATIEVPPESYFVAIPMIELPLELEK SLGASAPNRENEPTIEBATEBQCHHEVLESIEDIEDIYAY	1477
AlyDME	ALGPEFE..ERLTSATIEVPPESYFVAIPMIELPLELEK ALARGAPNRENEPTIEBATEBQCHHEVLESIEDIEDIYAY	1745
CsaDME	ALGPEFE..EKSIVTAVLTTTERRSGIVINLEPLTPEEYSYTAENTS.GPSKCEPIIEBATEBQCHHEVLESIEDIEDIYAY	1595
PtdDME	ALGPEFE..TKDITSSVEMMSRNEVQVNLNMPPLPPEEHN.LLHVGP.NNGSHEPIIIEBATEBQCHHEVLESIEDIEDIFG	1533
TcaDME	ALGPEFE..EKSITSSVEMMSRNEVQVNLNMPPLPPEEHN.LLHVGP.NNGSHEPIIIEBATEBQCHHEVLESIEDIEDIYAY	1743
RcDME	.EDDDEIPTIKNIMEELTNYLNQYMQANMELQEGDMSKALVALNBEAASITPTPKLKNVSLRTEHOVYVLPDSSHLLNLR.	1705
JcrDME	.EDDDEIPTIKNIMEELAVNLYNQYMQANMELQEGDMSKALVALNBEAASITPTPKLKNVSLRTEHOVYVLPDSSHLLKLG.	1757
AthDME	.EDDDEIPTIKNIMEELQFGMTLRHEMERNMELQEGDMSKALVALHITTSITPTPKLKNVSLRTEHOVYVLPDSSHLLDG.	1556
AlyDME	.EDDDEIPTIKNIMEELQFGMTLRHEMERNMELQEGDMSKALVALHITTSITPTPKLKNVSLRTEHOVYVLPDSSHLLDG.	1824
CsaDME	.EDDDEIPTIKNIMEELFTNLRHYIPE.....GMSKALVALNBEAASITPTPKLKNVSLRTEHOVYVLPDSSHLLRE.	1667
PtdDME	.EDDDEIPTIKNIMEELFTNLRHYIPE.....GMSKALVALNBEAASITPTPKLKNVSLRTEHOVYVLPDSSHLLRGN.	1611
TcaDME	.EDDDEIPTIKNIMEELFTNLRHYIPE.....GMSKALVALNBEAASITPTPKLKNVSLRTEHOVYVLPDSSHLLRE.	1821
Domain B		
RcDME	..MDRRPDDPSFYLLAIWTPGPTANSICPEPRHIFQFGQDPLKNEQITFSCNSIRENINOTVRGTLILPCRTAMRGSFF	1783
JcrDME	..MDRRPDDPSFYLLAIWTPGPTANSICPEPRHIFQFGQDPLKNEQITFSCNSIRENINOTVRGTLILPCRTAMRGSFF	1835
AthDME	..MDRRPDDPSFYLLAIWTPGPTANSICPEPRHIFQFGQDPLKNEQITFSCNSIRENINOTVRGTLILPCRTAMRGSFF	1634
AlyDME	..MDRRPDDPSFYLLAIWTPGPTANSICPEPRHIFQFGQDPLKNEQITFSCNSIRENINOTVRGTLILPCRTAMRGSFF	1902
CsaDME	..MDRRPDDPSFYLLAIWTPGPTANSICPEPRHIFQFGQDPLKNEQITFSCNSIRENINOTVRGTLILPCRTAMRGSFF	1745
PtdDME	LQMDRRPDDPSFYLLAIWTPGPTANSICPEPRHIFQFGQDPLKNEQITFSCNSIRENINOTVRGTLILPCRTAMRGSFF	1691
TcaDME	..MDRRPDDPSFYLLAIWTPGPTANSICPEPRHIFQFGQDPLKNEQITFSCNSIRENINOTVRGTLILPCRTAMRGSFF	1899
RcDME	LNGTYFOVNEFADHSSLNPIVPRRAWLNLPRLPVYFGTSVSTIFRGLSTEGLYCFWKGKAYAVEMRTSILEFLSTYR	1863
JcrDME	LNGTYFOVNEFADHSSLNPIVPRRAWLNLPRLPVYFGTSVSTIFRGLSTEGLYCFWKGKAYAVEMRTSILEFLSTYR	1896
AthDME	LNGTYFOVNEFADHSSLNPIVPRDWIIDLPRRTVYFGTSVSTIFRGLSTEGLYCFWKGKAYAVEMRTSILEFLSTYR	1714
AlyDME	LNGTYFOVNEFADHSSLNPIVPRDWIIDLPRRTVYFGTSVSTIFRGLSTEGLYCFWKGKAYAVEMRTSILEFLSTYR	1982
CsaDME	LNGTYFOVNEFADHSSLNPIVPRKWLNLNLPRLPVYFGTSVSTIFRGLSTEGLYCFWKGKAYAVEMRTSILEFLSTYR	1825
PtdDME	LNGTYFOVNEFADHSSLNPIVPRSLTNLPRLPVYFGTSVSTIFRGLSTEGLYCFWKGKAYAVEMRTSILEFLSTYR	1771
TcaDME	LNGTYFOVNEFADHSSLNPIVPRRAWLNLPRLPVYFGTSVSTIFRGLSTEGLYCFWKGKAYAVEMRTSILEFLSTYR	1966
RcDME	LSFYISITONLHYH.....	1876
JcrDME	.....	1896
AthDME	ARLHFFPASKLKNKKT....	1729
AlyDME	ARLHFFPASKLKNKKT....	1997
CsaDME	ARLHFFPASKLKNKQOTE	1844
PtdDME	ARLHFFPASKLKNKQOTE	1789
TcaDME	.....	1966

图3 RcDME 与其他植物 DME 氨基酸序列比对  
 Fig.3 Alignment of amino acid of RcDME and DME homologue of other species

鉴于 DME 糖基化酶在植物繁殖、印迹基因的表达和胚乳发育中的关键作用, 我们进一步解析了 RcDME 蛋白的结构与功能的关系。应用 Phytozome 上的 BlastP 工具进行同源序列查找, 结果显示 RcDME 与小桐子(Jcr4S02159.10 /partial)、拟南芥 AtDME(NP\_196076.2)、琴叶拟南芥(487239)、黄瓜 CsaDME(Cuca.308950.1)、毛果杨 PtrDME(Potri.008G025900.1)、烟草(Thecc1EG043681t1)的 DME 蛋白的同源性分别是: 64%、50.1%、52.9%、52.6%、60.9%和 61%。氨基酸序列比对结果显示(图 3), DME 有三个保守的序列区域, Domain A, Glycosylase Domain 和 Domain B<sup>[23]</sup>。其中 Glycosylase Domain 包含了 DNA 糖基化酶的两个保守功能域, 即 ENDO3c 和 FES 功能域; 除此之外, 在 RcDME 和 RcROS1 蛋白的 N 端还包含一个保守的赖氨酸富集区域, 该区

域被认为与核染色质的相互作用有关<sup>[24]</sup>。

### 2.2 蓖麻 DNA 糖基化酶蛋白的系统发育

为了进一步分析蓖麻 DNA 糖基化酶的进化关系, 我们选择了与蓖麻糖基化酶相似性较高的其它物种的同源蛋白(表 2), 用 MEGA 5.0 软件中 Neighbor-Joining 法构建其系统发育树。如图 4 结果所示, 三类糖基化酶明显地聚类在三个独立的支上, 且在 ROS1 和 DML 分支上, 单子叶和双子叶植物具有明显的分化, 而 DME 仅存在于双子叶植物中, 而在单子叶植物如水稻、玉米和高粱中没有发现其同源物。蓖麻的三种 DNA 糖基化酶在各支上均与其同科植物如小桐子、木薯有近的亲缘关系, 而与模式植物拟南芥的亲缘关系较远, 暗示着蓖麻 DNA 糖基化酶在功能上可能与拟南芥存在着一定的差异。

表 2 进化树分析所用物种的基本信息

Table 2 Basic information of related species in the phylogenetic tree

物种名称 Name	简写名称 Short	分类 Classification	物种名称 Name	简写名称 Short	分类 Classification
蓖麻 Castor	Rc	D	番木瓜 Papaya	Cpa	D
小桐 Jatropha	Jcr	D	大豆 Soybean	Gma	D
木薯 Manioc	Mes	D	葡萄 Grape	Vvi	D
拟南芥 Arabidopsis	Ath	D	棉花 Cotton	Gra	D
琴叶拟南芥 <i>A. lyrata</i>	Aly	D	毛果杨 <i>P. trichocarpa</i>	Ptr	D
芜菁 Turnip	Bra	D	玉米 Maize	Zma	M
黄瓜 Cucumber	Csa	D	水稻 Rice	Osa	M
碧桃 Peach	Ppe	D	高粱 Sorghum	Sbi	M
烟草 Tobacco	Tca	D	二穗短柄草 <i>Brachypodium</i>	Bdi	M

注: D 双子叶植物; M 单子叶植物

Note: D, dicotyledon; M, monocotyledon

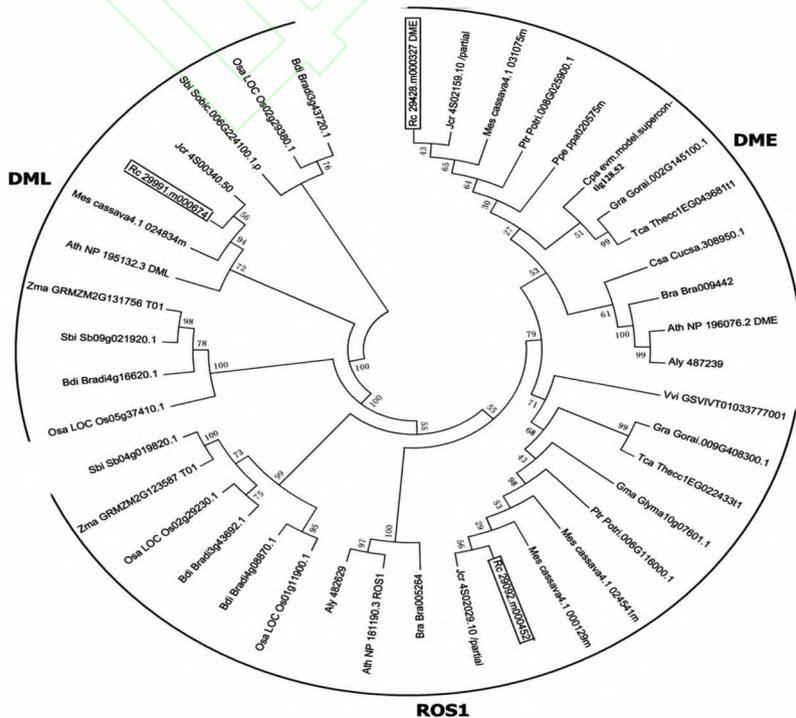


图 4 蓖麻 DNA 糖基化酶蛋白系统发育树

Fig. 4 Phylogenetic tree of DNA glycosylase from *R. communis* and other species

### 2.3 蛋白质理化性质、亚细胞定位预测和二级结构分析

蓖麻 *RcDME*、*RcROS1*、*RcDML* 基因分别编码 1 876、1 634、1 712 个氨基酸残基,相对分子质量分别为 212 279.8、183 163.6 和 191 361.9 kD,均为分子量较大的三种酶蛋白。*RcDME*、*RcDML* 和 *RcROS1* 的酸性氨基酸数目稍微多于碱性氨基酸,其理论等电点分别为 6.90、6.88、6.88,由此可知,蓖麻的三个糖基化酶均属于中性偏酸性蛋白(表 3)。对蛋白质亲/疏水性进行预测,正值越大表明越疏水,负值越大表明越亲水,由此推断这三个蛋白都为亲水性蛋白,其中 *RcDME* 的亲水性最高。三个蛋白的不稳定参数为 49.92、47.11 和 56.83,均为不稳定蛋白。这些理化性质的分析将对蛋白酶的分离与提纯有重要的意义。利用 SignalP 4.0 对蓖

表 3 蓖麻 DNA 糖基化酶基因编码蛋白序列的理化性质

Table 3 Physical and chemical properties of protein coded by DNA glycosylase genes of *R. communis*

基因座位 gene loci	蛋白 protein	编码的氨基酸 Coded amino acids	分子量 MW /kD	碱性氨基酸 Alkaline amino acids	酸性氨基酸 Acidic amino acids	pI	亲水性 Hydrophilicity	稳定性 Stability	信号肽 Signal peptide
29428. m000327	RcDME	1 876	212 279.8	230	236	6.9	-0.784	49.92 不稳定 Unstable	No
29092. m000452	RcROS1	1 634	183 163.6	195	199	6.88	-0.673	47.11 不稳定 Unstable	No
29991. m000674	RcDML	1 712	191 361.9	222	226	6.88	-0.6	56.83 不稳定 Unstable	No

表 4 蓖麻 DNA 糖基化酶基因编码蛋白的亚细胞定位分析

Table 4 Subcellular localization prediction of protein coded by DNA glycosylase genes of *R. communis*

亚细胞定位 Subcellular localization	RcDME	RcROS1	RcDML
细胞核 Nuclear	9.76	9.52	2.43
细胞膜 Plasma membrane			0.10
胞外 Extracellular		0.22	
细胞质 Cytoplasmic	0.02		5.10
线粒体 Mitochondrial	0.23	0.20	
高尔基体 Golgi			2.36
液泡 Vacuolar		0.06	

表 5 蓖麻 DNA 糖基化酶基因编码蛋白的二级结构分析

Table 5 Secondary structure analysis of protein coded by DNA glycosylase genes of *R. communis*

二级结构 Secondary structure	RcDME	RcROS1	RcDML
$\alpha$ -螺旋 Alpha helix (Hh)	527 (28.09%)	441 (26.99%)	514 (30.02%)
$\beta$ -转角 Beta turn (Tt)	61 (3.25%)	52 (3.18%)	98 (5.72%)
延伸链 Extended strand (Ee)	169 (9.01%)	218 (13.34%)	253 (14.78%)
无规则卷曲 Random coil (Cc)	1119 (59.65%)	923 (56.49%)	847 (49.47%)

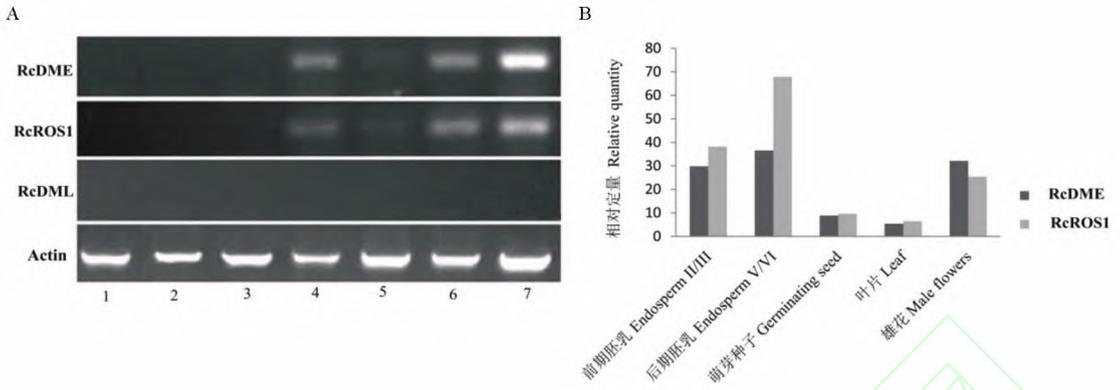
麻糖基化酶的氨基酸进行信号肽预测,结果显示这三种蛋白均不存在信号肽。利用 ProtComp v.9.0 在线软件对蓖麻三种糖基化酶进行亚细胞定位预测分析,如表 4 所示,*RcDME* 蛋白定位于细胞核的得分为 9.76,定位于细胞质的得分为 0.02,定位于线粒体的得分 0.23,而在细胞中其他部分的定位得分均为 0.00,从而推测该蛋白可能最终定位于细胞核,说明 *RcDME* 可能为核靶向蛋白。类似地,*RcROS1* 蛋白也可能定位于细胞核中,但 *RcDML* 蛋白却在细胞质的定位得分最高,预示着 *RcDML* 蛋白可能在细胞质中发挥功能。利用 SOPMA 对蓖麻糖基化酶的二级结构预测表明(表 5),三种蛋白主要是由  $\alpha$ -螺旋和无规则卷曲构成,而延伸链和  $\beta$ -转角散布于整个蛋白质中。

### 2.4 蓖麻 DNA 糖基化酶基因的表达分析

以蓖麻品种 ZB306 为实验材料,检测 *RcDME*、*RcROS1* 和 *RcDML* 在根、茎、叶、雌花和雄花、胚和胚乳组织器官的表达特性。RT-PCR 检测结果表明,*RcDME* 和 *RcROS1* 在雄花、雌花、胚和胚乳中均有表达,并且在雄花中的相对表达量最低,在胚乳中的相对表达量最高;而 *RcDML* 在各个组织均未检测到(如图 5A)。为了进一步检测这三种蛋白在组织间的表达情况,我们从 GenBank 数据库中下载了蓖麻五个组织转录组数据(<http://www.ncbi.nlm.nih.gov/sra/accession/SRX007402> to [SRX007408](http://www.ncbi.nlm.nih.gov/sra/accession/SRX007408))<sup>[25]</sup>。表达谱分析发现,蓖麻 *RcDME* 和 *RcROS1* 在各个组织器官,包括叶片、雄花、萌发的种子和胚乳的两个发育阶段中均有表达,在胚乳组织中表达量最高且在发育过程中表达量有所增高,其次是在雄花中有所表达,后在叶片和萌发的种子有最低的表达。而蓖麻 *RcDML* 基因在检测的所有组织中都未发现相应的转录产物(图 5B),与 RT-PCR 的结果一致。

## 3 结论与讨论

DNA 的主动脱甲基化过程在植物生长发育,响应环境变化和基因组防御过程中发挥着重要的作



注:A:RT-PCR半定量检测蓖麻DNA糖基化酶基因表达。1根,2茎,3叶,4雌花,5雄花,6胚,7胚乳;

B:转录组数据分析蓖麻DNA糖基化酶基因表达<sup>[25]</sup>

Note:A:RT-PCR analysis of DNA glycosylase genes expression level in various tissues.

1:root; 2:stem;3:leaf;4:female flower;5:male flower;6:embryo;7:endosperm.

B:expression analysis of DNA glycosylase genes from Brown's study<sup>[25]</sup>

图5 蓖麻DNA糖基化酶基因组织特异表达

Fig.5 Expression analysis of DNA glycosylase genes from different tissues in *R. communis*

用<sup>[8]</sup>。另外,在植物中DNA主动脱甲基化能够有效阻止siRNA介导的DNA甲基化过程<sup>[8,26]</sup>。最近在植物中发现了一组DNA糖基化酶如DME<sup>[9]</sup>、ROS1<sup>[10]</sup>和DML<sup>[11]</sup>能够通过碱基切除修复机制去除DNA上的甲基化标记。本研究利用拟南芥鉴定的DNA糖基化酶<sup>[11]</sup>通过同源序列比对在蓖麻基因组中鉴定了三种同源的DNA糖基化酶,分别命名为RcDME、RcROS1和RcDML。这三种氨基酸均为不稳定氨基酸,编码较长的氨基酸序列,其中包含了典型的与DNA结合和碱基切除修复有关的保守结构域Domain A, Glycosylase Domain和Domain B<sup>[23]</sup>,但在氨基酸序列的N端存在着高的变异区。然而,在RcDME和RcROS1氨基酸序列的N端存在一个保守的赖氨酸富集区,此区域被认为与核染色质结合有关<sup>[24]</sup>,与亚细胞定位结果一致,而RcDML在N端缺乏该区域,亚细胞定位结果显示其定位在细胞质中。

近年来很多研究发现DNA糖基化酶DME介导的DNA主动脱甲基化作用对印记基因的表达<sup>[27,28]</sup>、胚乳和种子发育有至关重要的作用<sup>[9]</sup>。母源等位基因*dme*突变体种子表现出发育异常甚至败育<sup>[9]</sup>。因此,我们进一步对RcDME糖基化酶的氨基酸序列,系统进化和组织特异性表达进行了详细分析。根据Choi等人的研究,DNA糖基化酶可以分为两种:双功能糖基化酶具有碱基切除和3'磷酸二酯键连接的功能;单功能糖基化酶仅具有DNA糖基化酶活性,DNA缺口的产生还需要AP核酸内切酶的参与。因此,推断RcDME是一种单功能DNA糖基化酶,因为在氨基酸序列的1455位是一个天冬氨酸(双功能糖基化酶为组氨酸)。保守的天冬氨酸能够裂解水分子,然后通过亲核反应替换受损

的或者错配的碱基<sup>[9]</sup>。从系统发育树来看,DME仅出现在双子叶植物中,类似的结果也被Zemach等人水稻基因组甲基化研究中发现,说明DME在双子叶植物中较为保守<sup>[29]</sup>。DME在单子叶植物中的缺失可能是DNA糖基化酶家族的分离进化而引起<sup>[29]</sup>。另外,在其他古老的低等植物中也发现了DME的同源物,显示出DME起源于单个古老的基因<sup>[23]</sup>。而ROS1和DML糖基化酶在单子叶和双子叶植物出现之前就已经产生了分化。蓖麻DNA糖基化酶在进化树上均与拟南芥表现出较远的亲缘关系。RT-PCR和转录组数据分析显示*RcDM*基因在雄花、雌花、种子的胚和胚乳中都表达,不同于拟南芥DME在中央细胞中特异表达,说明RcDME和拟南芥的DME在功能上存在一定的差异,RcDME可能参与不同组织的基因组脱甲基化作用和组织的正常发育。*RcROS1*和*RcDME*基因表现出相似的表达模式,它们是否在功能上存在冗余还需要深入的研究。而蓖麻*RcDML*基因在各个组织中都未检测到。虽然在拟南芥组织中检测到了DML2和DML3的表达,而且其功能的缺失会导致基因组部分区域的低甲基化,但是DML的缺失突变并不影响植物的正常生长发育,因此认为DML可能不是植物生长发育所必需的,而与保护基因免受过度甲基化过程有关<sup>[11]</sup>。

蓖麻作为一种重要的非食用油料作物,从表观遗传水平上研究蓖麻生长发育和抗逆性,特别是胚乳发育的调控机制具有非常重要的现实意义。而且,蓖麻种子具有大的胚乳组织,在整个种子发育阶段持续存在,为研究表观调控因子和印记基因在胚乳发育和种子形成过程中的功能提供了很好的系统。本研究对蓖麻DNA糖基化酶的分离鉴定,理化

性质分析,亚细胞定位预测及其组织表达为深入研究其功能和蓖麻遗传改良提供了理论基础。

### 参考文献:

- [1] Doerfler W. DNA methylation and gene activity[J]. *Ann Rev Biochem*,1983,52:93-124.
- [2] Henderson I R, Jacobsen S E. Epigenetic inheritance in plants[J]. *Nature*,2007,447(7 143):418-424.
- [3] Gehring M, Bubb K L, Henikoff S. Extensive demethylation of repetitive elements during seed development underlies gene imprinting [J]. *Science*, 2009, 324: 1 447 - 1 451.
- [4] Finnegan E J, Kovac K A. Plant DNA methyltransferases [J]. *Plant Mol Biol*,2000,43(2-3):189-201.
- [5] Bird A. DNA methylation patterns and epigenetic memory [J]. *Gene Dev*,2002,16(1):6-21.
- [6] Bender J. DNA methylation and epigenetics [J]. *Annu Rev Plant Biol*,2004,55:41-68.
- [7] 柯跃斌,夏 波. DNA 去甲基化与基因激活过程[J]. *癌变·畸变·突变*,2010,22(2):149-153.
- [8] Zhu J K. Active DNA demethylation mediated by DNA glycosylases [J]. NIH - PA Author Manuscript, 2009, 43:143-166.
- [9] Choi Y, Gehring M, Johnson L, et al. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis* [J]. *Cell*,2002,110:33-42.
- [10] Gong Z, Morales - Ruiz T, Ariza R R, et al. ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase [J]. *Cell*,2002,111:803-814.
- [11] Penterman J, Zilberman D, Huh J H, et al. DNA demethylation in the *Arabidopsis* genome [J]. *PNAS*,2007,104:6 752 - 6757.
- [12] Kinoshita T, Miura A, Choi Y, et al. One - way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation [J]. *Science*,2004,303:521-523.
- [13] 高彩婷,宝力高,刘 涛. 蓖麻研究概况 [J]. *内蒙古民族大学学报(自然科学版)*,2010,25(2):178-181.
- [14] 严兴初,王力军. 蓖麻作为能源开发的现状与前景 [J]. *安徽农业科学*,2007,35(34):11 165 - 11 167.
- [15] Ogunniyi D. Castor oil: A vital industrial raw material [J]. *Bioresource Technology*, 2006, 97(9): 1 086 - 1 091.
- [16] Caupin H J. Products from castor oil: past, present, and future [A]. *Lipid Technologies and Applications [M]*. New York:Marcel Dekker,1997. 787-795.
- [17] Thompson J D, Higgins D G, Gibson T J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position - specific gap penalties and weight matrix choice [J]. *Nucleic Acids Res*,1994,22(22):4 673 - 4 680.
- [18] Tainer J A, Thayer M M, Cunningham R P. DNA repair proteins [J]. *Current Opinion in Structural Biology*, 1995,5:20-26.
- [19] Thayer M M, Ahern H, Xing D X, et al. Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure [J]. *The EMBO Journal*, 1995, 14 (16):4 108 - 4 120.
- [20] Krokan H E, Standal R, Slupphaug G. DNA glycosylases in the base excision repair of DNA [J]. *Biochem*,1997,325:1-16.
- [21] Bruner S D, Norman D P G, Verdine G L. Structural basis for recognition and repair of the endogenous mutagen 8 - oxoguanine in DNA [J]. *Nature*,2000,403(24):859-866.
- [22] Schärer O D, Jiricny J. Recent progress in the biology, chemistry and structural biology of DNA glycosylases [J]. *BioEssays*,2001,23:270-281.
- [23] Mok Y G, Uzawa R, Lee J, et al. Domain structure of the DEMETER 5 - methylcytosine DNA glycosylase [J]. *PNAS*,2010,107(45):19 225 - 19 230.
- [24] Kasinsky H E, Lewis J D, Dacks J B, et al. Origin of H1 linker histones [J]. *The FASEB Journal*,2001,15:34-42.
- [25] Brown A P, Kroon J T M, Swarbreck D, et al. Tissue - specific whole transcriptome sequencing in castor, directed at understanding triacylglycerol lipid biosynthetic pathways [J]. *PLoS One*,2012,7(2):e30100.
- [26] Bauer M J, Fischer R L. Genome demethylation and imprinting in the endosperm [J]. *Curr Opin Plant Biol*, 2011,14(2):162-167.
- [27] Gehring M, Huh J H, Hsieh T F, et al. DEMETER DNA glycosylase establishes MEDEA polycomb gene self - imprinting by allele - specific demethylation [J]. *Cell*, 2006,124(3):495-506.
- [28] Xiao W, Gehring M, Choi Y, et al. Imprinting of the MEA polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase [J]. *Dev Cell*,2003,5(6):891-901.
- [29] Zemach A, Kim M Y, Silva P, et al. Local DNA hypomethylation activates genes in rice endosperm [J]. *PNAS*, 2010,107:18 729 - 18 734.