

# 玉米 RNA 结合蛋白 GRMZM2G052926 基因的生物信息学分析

孙培元<sup>1</sup>, 王俊强<sup>1</sup>, 韩业辉<sup>1</sup>, 于运凯<sup>1</sup>, 许 健<sup>1</sup>, 周 超<sup>1</sup>, 李 昕<sup>2</sup>

(1. 黑龙江省农业科学院 齐齐哈尔分院, 黑龙江 齐齐哈尔 161041; 2. 黑龙江省农业科学院 五常水稻研究所, 黑龙江 五常 150229)

**摘要:** 为了研究玉米 RNA 结合蛋白在玉米生长发育中的调节机制, 以一个玉米 RNA 结合蛋白 GRMZM2G052926 基因为材料, 采用生物信息学的方法, 对其理化性质、跨膜结构域、亚细胞定位、功能域、二级结构等方面进行了预测和分析。结果表明: 基因包含一个 1 335 bp 开放阅读框, 是一种具有亲水能力的不稳定酸性蛋白, 相对分子质量 38.301 1 kDa, 包含两个 RRM 超亲家族(RRM superfamily), 无信号肽序列, 无跨膜结构域, 亚细胞定位在细胞质上, 二级结构由无规则卷曲、 $\alpha$  螺旋、延伸链和  $\beta$  折叠构成; 系统进化树分析表明, GRMZM2G052926 蛋白与水稻 Os05g0437300 蛋白亲缘关系最近, 同源性达到 87%。

**关键词:** 玉米; RNA 结合蛋白; GRMZM2G052926 基因; 生物信息学

**中图分类号:** S513 **文献标识码:** A **文章编号:** 1002-2767(2016)03-0007-05 DOI:10.11942/j.issn1002-2767.2016.03.0007

RNA 结合蛋白是一类含有 RNA 结合结构域(RNA Binding Domain, RBD)的蛋白, 其在 RNA 的剪切、编辑、运输、稳定、翻译等过程中起到至关重要的作用<sup>[1-2]</sup>。已有研究表明, RNA 结合蛋白在细菌、真菌、动物和植物中均有存在, 且广泛参与多种生理生化过程<sup>[1-4]</sup>。2002 年, Crittenden 等人发现, RNA 结合蛋白 FBF-1/-2 在秀丽隐杆线虫的生长过程中通过与 5 个靶标 RNA (fbf-1, fbf-2, gld-3s, gld-1, fem-3) 相结合, 从而调节线虫的性别及维持胚胎干细胞的增殖, 在线虫的种系和早期胚胎发育中起了重要的作用<sup>[5]</sup>。2001 年, Lu 等人通过免疫筛选的方法从人肝癌细胞的 cDNA 表达文库中分离出一种 RNA 结合蛋白 p62, 并通过进一步的研究, 发现 p62 在肝癌细胞中能够对某些生长因子进行调节, 从而非正常表达<sup>[6]</sup>。Richard 等人通过对 PgRNP 基因进行了多种逆境表达研究, 发现该基因在低温、系统伤害、茉莉酸脂(JA)和干旱等逆境环境下, 其表达均会受到影响<sup>[7]</sup>。

玉米(*Zea mays* L.)是我国重要的粮食作物和饲料来源, 其种植面积和单产均位居第一<sup>[8-10]</sup>。随着生物信息学的高速发展以及玉米基因组测序的完成, 玉米生物信息学的研究也越来越深入, 然

而玉米 RNA 结合蛋白的分子机制以及其所参与的生物学过程却鲜见报道。因此, 研究以玉米基因组中一个 RNA 结合蛋白基因 GRMZM2G052926 的序列为材料, 利用相关生物信息学分析工具及数据库对其结构、功能进行系统性分析和预测, 旨在研究该 RNA 结合蛋白在玉米生长发育中的调节机制, 以期为玉米 RNA 结合蛋白的功能研究奠定基础。

## 1 材料与方法

### 1.1 材料

GRMZM2G052926 基因序列来自于 National Center for Biotechnology Information (NCBI) 网站中的 GenBank 数据库 (<http://www.ncbi.nlm.nih.gov/>), 登录号为 AFW81821.1。

### 1.2 方法

试验采用在线软件 ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>), 确定 GRMZM2G052926 基因的 ORF 位置及编码产物大小; 采用 NCBI-CDS 在线分析工具 (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi/>) 对其保守结构域进行分析; 采用在线软件 ProtParam (<http://web.expasy.org/protparam/>) 分析其基本理化性质; 采用在线软件 PSORT (<http://psort.hgc.jp/form.html>) 对其进行亚细胞定位; 采用在线软件 TMHMM Server v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) 寻找其跨膜结构; 采用在线软件 Sig-

收稿日期: 2016-01-13

第一作者简介: 孙培元 (1991-), 男, 黑龙江省绥化市人, 硕士, 研究实习员, 从事玉米分子遗传育种研究。E-mail: 562337271@qq.com。

nalP 4. 1 (<http://www.cbs.dtu.dk/services/SignalP/>)对其信号肽进行预测;采用在线软件 NetPhos 2. 0 Serve(<http://www.cbs.dtu.dk/services/NetPhos/>)对其磷酸化位点进行预测;利用 Coils 在线分析工具([http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html))对其形成卷曲螺旋的倾向性进行预测;采用在线软件 SOPMA([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_sopma.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html))对其二级结构进行分析;采用在线软件 SWISS-MODEL(<http://swissmodel.expasy.org/>)对其

三级结构进行分析;采用 MEGA6. 0 软件将其同其它植物蛋白进行同源比对并构建系统进化树。

2 结果与分析

2.1 GRMZM2G052926 基因编码氨基酸序列分析

在线使用 ORF Finder 确定抗粗缩病候选基因 GRMZM2G052926 的开放阅读框,详见图 1。其中,GRMZM2G052926 基因包含一个完整的开放阅读框,长度为 1 335 bp,起始于 92 bp,终止于 1 426 bp。

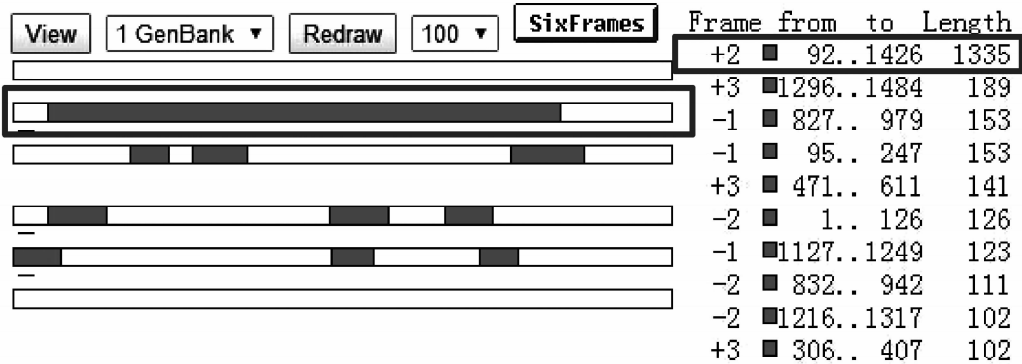


图 1 基因 GRMZM2G052926 的开放阅读框分析  
Fig. 1 ORFs analysis of GRMZM2G052926

2.2 GRMZM2G052926 基因编码蛋白的理化性质预测

利用 Protparam 对 GRMZM2G052926 基因编码蛋白进行理化性质的预测和分析。结果显示,该蛋白分子式为 C<sub>1680</sub> H<sub>2614</sub> N<sub>490</sub> O<sub>509</sub> S<sub>15</sub>,由 5 308个原子构成,相对分子质量 38. 301 1 kDa,理论等电点(pI)为 7. 66,带负电荷的氨基酸残基总数(Asp+Glu)为 31 个,带正电荷的氨基酸残基总数(Arg+Lys)为 32 个,总平均亲水性系数为-0. 450,属于亲水性蛋白,理论推导半衰期是 30 h,不稳定系数为 58. 80,这说明该蛋白是不稳

定蛋白。

2.3 GRMZM2G052926 基因编码蛋白保守结构域分析

结构功能域是指蛋白质三级结构内的独立折叠单元,是介于二级与三级结构之间,且具有一定功能的紧密结构单位<sup>[11]</sup>。利用 NCBI-CDS 在线分析工具预测 GRMZM2G052926 基因编码蛋白中可能存在的保守结构域(见图 2)。结果表明,该基因编码蛋白包含两个 RRM 超亲家族(RRM superfamily)。RRM,即 RNA 识别基序(RNA Recognition Motif),是一种 RNA 结合结构域。

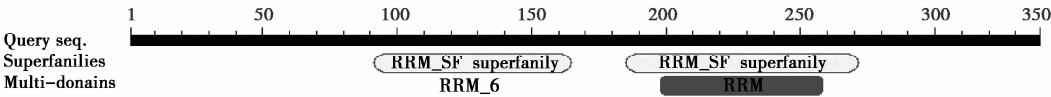


图 2 GRMZM2G052926 基因编码蛋白的保守结构域  
Fig. 2 The conserved domains of GRMZM2G052926 protein

2.4 GRMZM2G052926 基因编码蛋白磷酸化位点预测

在细胞的信号传递过程中,蛋白质磷酸化具有极其重要的作用。利用 NetPhos 2. 0 对

GRMZM2G052926 基因编码蛋白的磷酸化位点进行预测(见图 3)。整个多肽链共存在 16 个氨基酸磷酸化位点(>0. 5),其中丝氨酸(Ser)磷酸化位点有 13 个,在该蛋白中占主导地位;苏氨

酸(Thr)磷酸化位点有 2 个,酪氨酸(Tyr)磷酸化位点有 1 个。因此,推测丝氨酸磷酸化可能在

GRMZM2G052926 基因编码蛋白行使功能中存在重要作用。

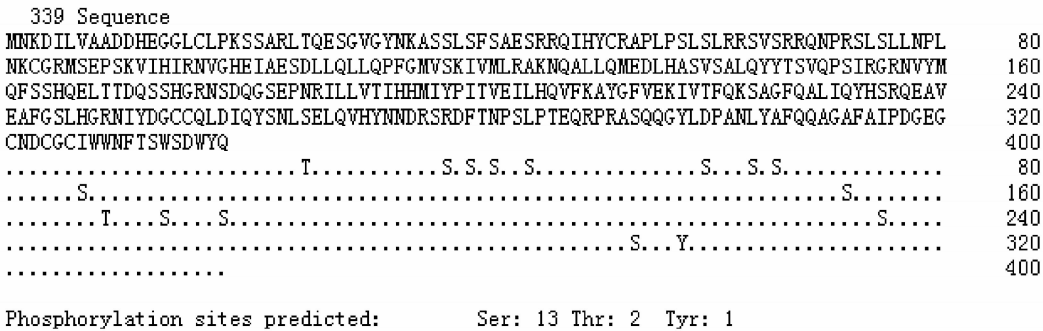


图 3 GRMZM2G052926 基因编码蛋白的磷酸化位点预测  
Fig. 3 Phosphorylation sites prediction of GRMZM2G052926 protein

### 2.5 GRMZM2G052926 基因编码蛋白亚细胞定位

利用在线软件 PSORT 对 GRMZM2G052926 基因编码蛋白进行亚细胞定位(见表 1)。结果表明,该蛋白最可能定位于细胞质,可信度为 0.505;其次定位于微体(过氧化物酶体)中,可信度为 0.300;其在线粒体基质空间和叶绿体类囊膜上也有分布,但可信度较低,均为 0.100。由此推测该蛋白属于细胞质蛋白。

表 1 GRMZM2G052926 基因编码蛋白的亚细胞定位

Table 1 Predicted subcellular localization of GRMZM2G052926

亚细胞结构 Destination of cell compartment	可信度 Certainty
细胞质 Cytoplasm	0.505
微体(过氧化物酶体) Microbody (peroxisome)	0.300
线粒体基质空间 Mitochondrial matrix space	0.100
叶绿体内囊体膜 Chloroplast thylakoid membrane	0.100

### 2.6 GRMZM2G052926 基因编码蛋白跨膜区域及信号肽预测

利用 TMHMM Server v. 2.0 在线分析软件对 GRMZM2G052926 基因编码蛋白进行跨膜结构域分析。结果如图 4 所示,该蛋白不存在跨膜结构域,位于膜外区的可能性最大(可能性 = 1.1),为非跨膜蛋白类。利用网络分析工具 SignalP 4.1 分析 GRMZM2G052926 基因编码蛋白信号肽的结果见图 5。其中,C 值(剪切位点值)为 0.110,在该蛋白中未发现可能的信号肽序列。由此说明,GRMZM2G052926 基因编码蛋白并不

是分泌型蛋白,不可能在细胞中发生迁移。结合之前的亚细胞定位结果,表明该蛋白可能是在细胞质中合成后直接在细胞质中行使功能。

卷曲螺旋(coiled coil,CC)是一类由 2 股或 2 股以上右手  $\alpha$  螺旋相互缠绕而形成的平行或反平行左手超螺旋结构的总称,该结构在蛋白质结构与功能的关系上具有极其重要的作用<sup>[12]</sup>。利用 Coils 分析工具,以 window = 14、21 和 28 为参数,按照几率 > 50% 就可形成螺旋的规则,对 GRMZM2G052926 基因编码蛋白形成卷曲螺旋的倾向性进行预测(见图 6)。结果表明,该蛋白在 windows = 21 时极易形成卷曲螺旋,在 windows = 14 时可能形成卷曲螺旋,而在 windows = 28 时不能形成卷曲螺旋结构。

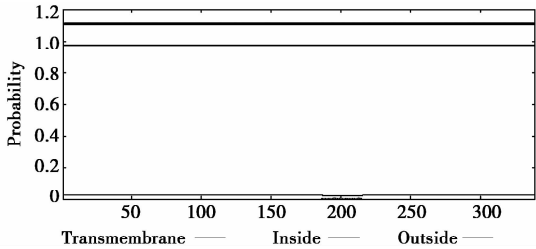


图 4 GRMZM2G052926 基因编码蛋白的跨膜结构预测  
Fig. 4 Transmembrane domain predicted of GRMZM2G052926

### 2.7 GRMZM2G052926 基因编码蛋白二级结构预测

利用 SOPMA 服务器对 GRMZM2G052926 基因编码蛋白的二级结构进行预测分析(见图 7)。结果表明,该蛋白的二级结构主要是由无规则卷曲(41.59%)组成,其次是  $\alpha$ -螺旋(30.38%),再次是由延伸链区(19.47%)和  $\beta$ -折叠构成的  $\beta$ -片层(8.55%),并无其它二级结构。

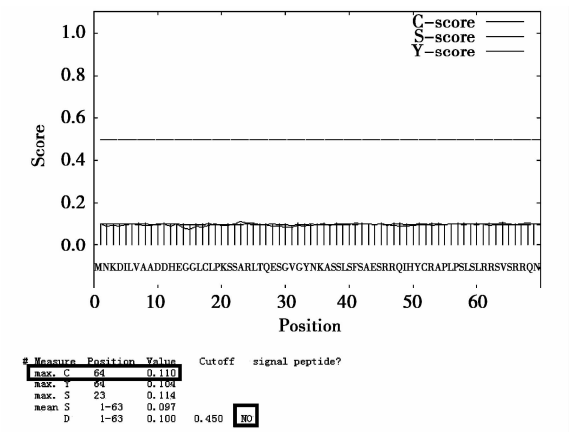


图 5 GRMZM2G052926 基因编码蛋白的信号肽预测  
Fig. 5 The signal peptide and splice site predicted of GRMZM2G052926

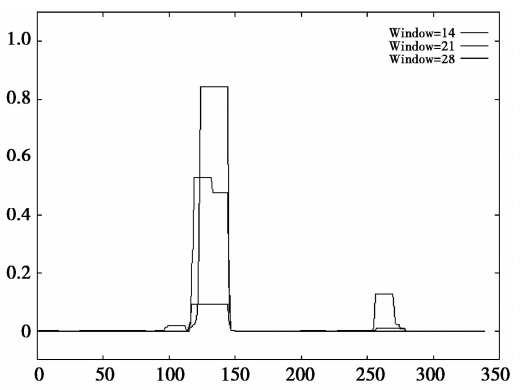


图 6 GRMZM2G052926 基因编码蛋白卷曲螺旋的预测  
Fig. 6 Predicted amino acid sequence of the winded helix of GRMZM2G052926

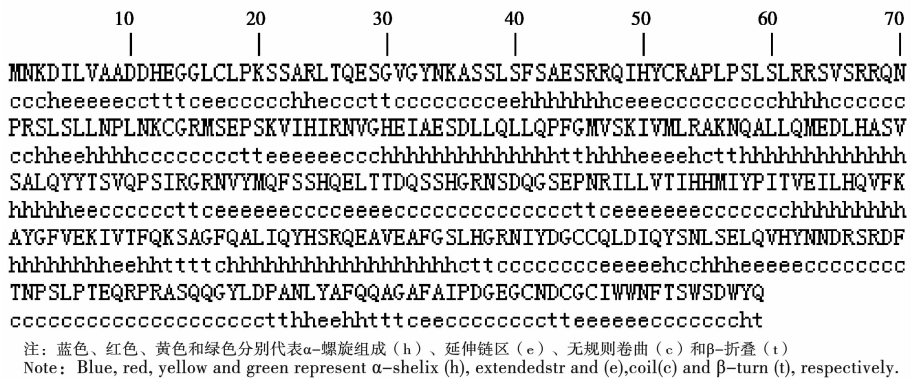


图 7 GRMZM2G052926 基因编码蛋白的二级结构  
Fig. 7 Secondary structure of GRMZM2G052926 protein

2.8 GRMZM2G052926 基因编码蛋白三级结构预测

利用 SWISS-MODEL 在线软件对 GRMZM2G052926 基因编码蛋白的三级结构模型进行预测分析(见图 8)可知,该蛋白由 α 螺旋和 β 折叠盘绕而成,中间则由无规则卷曲连接。这表明该基因能够折叠成特定的空间结构。

2.9 GRMZM2G052926 基因编码蛋白系统进化树的构建

为了研究 GRMZM2G052926 基因编码蛋白的起源和进化关系,对其编码的氨基酸序列在 NCBI 网站中进行 Blast,最终选取 9 条与之相似度较高的序列并下载,通过 MEGA6.0 软件中的 ClustalW 功能进行氨基酸序列比对并采取邻位连接法 (Neighbor-Joining, NJ),构建 GRMZM2G052926 蛋白与其它 9 种蛋白间的系统进化树。如图 9 所示,GRMZM2G052926 蛋白与水稻基因 Os05g0437300 编码蛋白聚为一类,

其同源性达到 87%。RNA 结合蛋白进化上的分组可能揭示其在功能上的异同。



图 8 GRMZM2G052926 基因编码蛋白的三级结构  
Fig. 8 The predicted 3-D structure of GRMZM2G052926

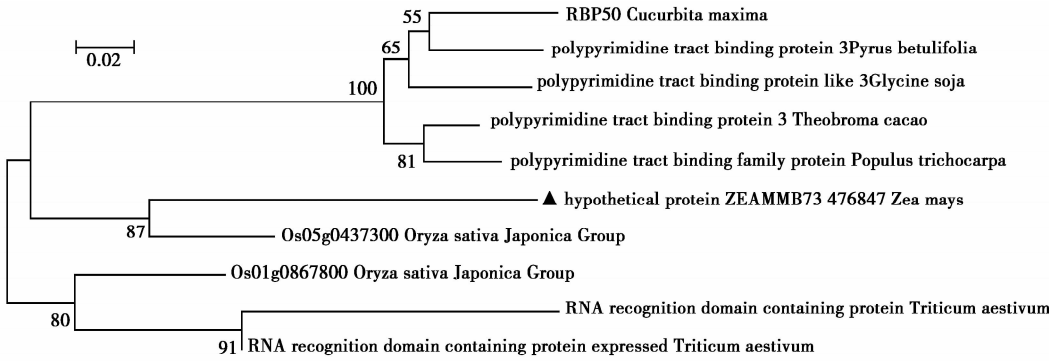


图 9 GRMZM2G052926 基因编码蛋白系统进化树

Fig. 9 Phylogenetic tree of GRMZM2G052926

### 3 结论与讨论

玉米是我国重要的粮食作物之一。根据国家统计局数据,2014 年玉米播种面积共计 3 707.6 万  $\text{hm}^2$ , 单位面积产量 5 817  $\text{kg} \cdot \text{hm}^{-2}$ , 均位居全国第一<sup>[9-10]</sup>。由此可见,确保玉米生产持续稳定地发展对于国家的粮食安全战略至关重要。随着基因组学相关领域研究的深入,分析植物基因的结构特征,探寻植物基因的潜在功能,现已成为当今基因研究的热点<sup>[11,13-14]</sup>。RNA 结合蛋白是一类在生物进化过程中保守的蛋白,其广泛存在于各种植物中。已有研究表明,植物 RNA 结合蛋白在调控叶绿素 mRNA 的稳定和转录、植物抗逆反应、植物开花调控以及植物激素信号调节等过程中均起到重要的作用<sup>[7,15-17]</sup>。因此,RNA 结合蛋白在玉米中的研究便显得尤为重要。

近年来,随着生物信息学的迅速发展,已有大量核苷酸和氨基酸序列被上传至各类分子生物学数据库中,因此,利用生物信息学相关软件进行快速有效地分析已成为当今在基因结构及功能研究领域主流的研究方法。本研究借助相关工具和软件对一个玉米 RNA 结合蛋白 GRMZM2G052926 进行了详尽的生物信息学分析和预测。结果表明,GRMZM2G052926 基因是一个包含 1 335 bp 开放阅读框的胞质基因,为具有一定亲水能力的酸性蛋白,相对分子质量 38.301 1 kDa,属于非跨膜蛋白类,包含两个 RRM 超亲家族(RRM superfamily),其二级结构由无规则卷曲、 $\alpha$  螺旋、延伸链和  $\beta$  折叠构成,所占比例分别为 41.59%、30.38%、19.47% 和 8.55%;系统进化树分析表明,GRMZM2G052926 蛋白与水稻 Os05g0437300 蛋白亲缘关系最近,同源性达到 87%。本研究结果为进一步研究该 RNA 结合蛋白的功能奠定了基础。

#### 参考文献:

[1] 陈璇,李文正,邵岩,等.动植物中 RNA 结合蛋白的研究进

展[J]. 生物技术通报,2007(3):9-15.  
[2] 唐靖.植物 RNA 结合蛋白的研究进展[J]. 安徽农业科学, 2010,38(1):38-41.  
[3] 邢葵益,析骅,曹凯鸣,等.水稻中一种 RNA 结合蛋白 cDNA 的筛选和结构分析[J]. 科学通报,1997,42(4): 442-446.  
[4] 涂雯婷,郑青亮,聂作明,等.家蚕含 RRM 保守结构域蛋白基因(Bmrrm)的克隆及序列分析与表达研究[J]. 蚕业科学,2009,35(3):499-507.  
[5] Crittenden S L, Bernstein D S, Bachorik J L, et al. A conserved RNA-binding protein controls germline stem cells in *Caenorhabditis elegans* [J]. Nature, 2002, 417 (6889): 660-663.  
[6] Lu M, Nakamura R M, Dent D B, et al. Aberrant expression of fetal RNA-binding protein p62 in liver cancer and liver cirrhosis[J]. American Journal of Pathology, 2001, 159(3): 945-953.  
[7] Richard S, Drevet C, Jouanin L, et al. Isolation and characterization of a cDNA clone encoding a putative white spruce glycine-rich RNA binding protein[J]. Gene, 1999, 240(2): 379-388.  
[8] 杨军,仇焕广,董婉璐,等.2013 年国内外玉米供需分析及 2014 年展望[J]. 农业展望,2014,10(2):4-8.  
[9] 习银生,杨丽.2014 年玉米市场形势分析与 2015 年展望[J]. 农业展望,2015,11(1):8-14.  
[10] 杨军,仇焕广,董婉璐,等.2014 年国内外玉米市场分析 及 2015 年展望[J]. 农业展望,2015,11(3):10-14.  
[11] 刘生财,匡华琴,谢礼洋,等.莧菜试管苗 amaNACA 2-like 基因克隆及生物信息学分析[J]. 东北农业大学学报, 2014,45(1):83-89.  
[12] 纪兆林,石刘飞,李健,等.卷曲螺旋结构及其研究进展[J]. 江苏农业科学,2014,42(2):6-12.  
[13] 王安娜,王婢娉,吴蕾,等.大豆 C4H 基因克隆及生物信息学分析[J]. 东北农业大学学报,2010,41(4):12-16.  
[14] 陈红梅.大豆 XHS 基因家族生物信息学分析[J]. 大豆科学,2015,34(3):384-388.  
[15] Hafner M, Landthaler M, Burger L, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP[J]. Cell, 2010, 141(1):129-41.  
[16] Martijn K, Strasser M J, Bijan B, et al. RNA-binding protein Dnd1 inhibits micro RNA access to target mRNA[J]. Cell, 2008, 131(7):1273-86.  
[17] Rouault T A, Stout C D, Kaptain S, et al. Structural relationship between an iron-regulated RNA-binding protein (IRE-BP) and aconitase: functional implications[J]. Cell, 1991, 64(5): 881-883.

# 野生黑果越橘组培快繁技术研究

韩阳花

(新疆农业职业技术学院,新疆 昌吉 831100)

**摘要:**为了促进蓝莓种质资源的保护和开发利用,以黑果越橘茎段为材料,进行了离体培养和植株再生的研究。结果表明:茎段外植体用 0.1% HgCl<sub>2</sub> 消毒 8~10 min 为宜,最佳的丛生芽诱导培养基为改良 WPM+ ZT2.0 mg·L<sup>-1</sup>,丛生芽的诱导率达到 55.17%,且苗健壮;最佳的增殖培养基为改良 WPM+ ZT1.0 mg·L<sup>-1</sup>,增殖倍数为 5.26。采用 IBA 0.5 mg·L<sup>-1</sup>+NAA 0.5 mg·L<sup>-1</sup> 浸泡水苔,包裹试管苗基部进行瓶外生根,生根率可达 91.71%。

**关键词:**黑果越橘;离体快繁;激素

**中图分类号:**S663.9 **文献标识码:**A **文章编号:**1002-2767(2016)03-0012-04 DOI:10.11942/j.issn1002-2767.2016.03.0012

黑果越橘(*Vaccinium myrtillus* L.)是杜鹃花科越橘属的一种野生小灌木,在我国仅新疆有部分分布<sup>[1]</sup>。黑果越橘具有抗寒能力强,果大酸甜适口,可作为北方蓝莓品种改良的重要亲本,属于国内稀缺的野生蓝莓资源。蓝莓果实因富含花青素,具有许多保健功能,近年来越来越受到人们的关注,蓝莓种苗的需求也越来越多。目前,蓝莓

种苗主要依靠植物组织培养技术进行快繁,研究黑果越橘茎段的组织培养不但能够为蓝莓快繁生产提供一定的数据参考,而且在蓝莓种质资源的保护和开发利用方面具有重要的意义。

## 1 材料与方法

### 1.1 材料

试验材料采自喀纳斯景区与哈巴河县牧区交界的山坡,海拔 2 100~2 200 m,分布在浓密的落叶松针叶林下。野外用周围伴生的苔藓包裹根系,实验室内 4℃ 的冰箱下保存 7 d。

改良 WPM 是以 Ca(NO<sub>3</sub>)<sub>2</sub>·4H<sub>2</sub>O 684 mg·L<sup>-1</sup>, KNO<sub>3</sub> 190 mg·L<sup>-1</sup>, C<sub>10</sub>H<sub>13</sub>FeN<sub>2</sub>NaO<sub>8</sub> 73.4 mg·L<sup>-1</sup>

收稿日期:2016-01-28  
基金项目:新疆维吾尔自治区高校科研计划青年教师科研启动基金资助项目(XJEDU2012S049)  
作者简介:韩阳花(1979-),女,四川省西充县人,硕士,讲师,从事资源植物开发利用及植物组织培养研究。E-mail:hhh\_535@163.com。

# Bioinformatics Analysis of the RNA-binding Protein GRMZM2G052926 Gene in Maize

SUN Pei-yuan<sup>1</sup>, WANG Jun-qiang<sup>1</sup>, HAN Ye-hui<sup>1</sup>, YU Yun-kai<sup>1</sup>, XU Jian<sup>1</sup>, ZHOU Chao<sup>1</sup>, LI Xin<sup>2</sup>  
(1. Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar, Heilongjiang 161041; 2. Wuchang Rice Research Institute, Heilongjiang Academy of Agricultural Science, Wuchang, Heilongjiang 150229)

**Abstract:** In order to study the regulatory mechanism of the maize RNA-binding protein in the maize growth and development, taking a RNA-binding protein GRMZM2G052926 gene as materials, its characters, including the physical and chemical properties, transmembrane domain, subcellular localization, functional domain plus the secondary structure were studied by bioinformatics tools. The results showed that this RNA-binding protein was an unstable hydrophilic acidic protein containing an open reading frame of 1 335 bp. The molecular weight of the protein containing two RRM super families was 38. 301 1 kDa. It was no signal peptide, no transmembrane domain, located in the cytoplasm, and contained random coil, alpha-helix, extended strand and beta turn. The phylogenetic results showed that GRMZM2G052926 protein from Zea mays and Os05g0437300 protein from Oryza sativa were the nearest in genetic relationship, reaching to 87%.

**Keywords:** maize; RNA-binding protein; GRMZM2G052926 gene; bioinformatics analysis