

木薯 *Mlo* 基因家族成员的鉴定及其序列特征分析

覃碧¹, 王萌², 刘巧玲³, 郑服丛², 张宇^{2*}

¹中国热带农业科学院橡胶研究所农业部橡胶树生物学与遗传资源利用重点实验室, 海南儋州571737; ²海南大学环境与植物保护学院, 海口570228; ³海南大学应用科学技术学院, 海南儋州571737

摘要: *Mlo* 基因家族是植物重要的抗病基因。本文通过系统分析木薯基因组数据库, 从中共鉴定出21个 *Mlo* 成员, 其中20个具有完整序列, 1个只有部分序列。对其中20个具有完整序列的基因与其他物种的 *Mlo* 基因进行聚类关系分析, 结果显示, 可将木薯 *Mlo* 基因家族分为6类(I~VI), 其中4类都包括有来自拟南芥的 *Mlo* 基因, 第VI类只包括2个木薯 *Mlo* 基因, 可能是木薯中特有的一类 *Mlo*; 6个木薯 *Mlo* 与已知的抗病 *Mlo* 基因分别聚在第IV和第V类, 这6个基因可能是木薯基因组中具有抗病功能的 *Mlo*。对所有的木薯 *Mlo* 蛋白进行结构分析发现, 除了 MeMlo20 外, 其他蛋白均具有6~8个跨膜结构, 其中3个蛋白具有N端信号肽。

关键词: 木薯; 基因家族; *Mlo*; 序列分析

Identification and Sequence Characterization of *Mlo* Gene Family from the Genome of *Manihot esculenta* Crantz

QIN Bi¹, WANG Meng², LIU Qiao-Ling³, ZHENG Fu-Cong², ZHANG Yu^{2*}

¹Key Laboratory of Biology and Genetic Resources of Rubber Tree, Ministry of Agriculture, Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan 571737, China; ²Environment and Plant Protection College, Hainan University, Haikou 570228, China; ³College of Applied Science and Technology, Hainan University, Danzhou, Hainan 571737, China

Abstract: *Mlo* gene family is an important class of plant disease resistance genes. Through systematic analysis of the genome sequence database of *Manihot esculenta*, total 21 *Mlo* genes were identified in this study, of which 20 had full-size and one partial sequences. Phylogenetic analysis was performed to reveal the evolution of the cassava *Mlo* genes with full-size sequences, and *Mlo* genes from other species. The results indicated that the cassava *Mlo* family could be divided into 6 clades (I–VI), of which four contained *Mlo* members from *Arabidopsis*. The VI clade contained only two cassava *Mlo* genes, suggested that it might be a specific clade for cassava genome. Furthermore, 6 cassava *Mlo* genes were grouped into the clades IV and V which contained the *Mlo* genes of known function in disease resistance, suggested that the 6 cassava *Mlo* genes might have their functions in disease resistance. Protein domains analysis revealed that besides MeMlo20, all the other 19 cassava *Mlo* consisted of 6 to 8 membrane-spanning domains, in which 3 proteins contained an N-terminal signal peptide.

Key words: *Manihot esculenta*; gene family; *Mlo*; sequence analysis

Mlo 基因家族是一类植物特有的抗病基因家族, 最早在大麦中发现并得到克隆, 大麦 *Mlo* 基因编码一个膜结合蛋白, 具有至少7个跨膜结构域, 是大麦抗白粉病的负调控因子, 缺乏野生型 *Mlo* 蛋白的大麦突变体 (*mlo*, 隐性纯合体) 表现出对大麦白粉菌具有广谱抗性 (Buschges 等 1997)。随后, *Mlo* 同源基因相继从水稻、拟南芥、亚麻、玉米和小麦等多种植物中分离出来, 构成了植物特有的 *Mlo* 家族 (Devoto 等 2003)。对不同物种的 *Mlo* 基因结构和功能进行分析发现, *Mlo* 是一类比较保守的抗病基因。拟南芥 *AtMlo2*、*AtMlo6* 和 *AtMlo12*

具有类似大麦 *Mlo* 的功能, 它们的缺失突变体也具有抗白粉病的功能 (Consonni 等 2006)。小麦 (*TaMlo1*、*TaMlo2* 和 *TaMlo3*)、水稻 *OsMlo2* 及番茄 *SIMlo1* 也具有抗白粉病的功能 (Bai 等 2008; Elliott 等 2002)。*Mlo* 基因在抗病中的作用使得其在植物抗病育种中占有着十分重要的地位。

收稿 2013-06-03 修定 2013-08-13

资助 农业部热带作物专项(12RZBC-11)、现代农业产业技术体系建设专项资金(CARS-34-GW8)和海南省研究生创新科研课题(Hyb2012-4)。

* 通讯作者(E-mail: yuzhang_rain@163.com; Tel: 0898-66192915)。

木薯是大戟科木薯属植物,广泛种植于热带和亚热带地区,是世界上重要的粮食作物与能源作物。木薯生长过程经常遭受如粉蚧、粉虱、天蛾幼虫、木薯花叶病毒、木薯褐病和木薯萎蔫病菌等的侵袭(Bart等2012;黄强等2013)。过去主要通过化学防治,但化学防治常常出现用药多、成本高、环境污染严重等问题,而培育和推广抗病品种是防治病害最经济有效且环保的措施。随着木薯基因组测序工作的完成,使得鉴定和克隆木薯的抗病虫基因变成可能,从而为进一步通过转基因技术培育木薯抗病虫害新品种提供丰富的基因源。*Mlo*基因家族是植物重要的抗病基因,但有关木薯*Mlo*基因的克隆及其功能研究未见报道。本研究通过系统分析木薯基因组数据库,从中鉴定木薯基因组所有的*Mlo*基因家族成员,通过对木薯与拟南芥基因组*Mlo*成员及其他物种中已知功能的*Mlo*基因进行系统进化关系分析,鉴定木薯基因组中具有抗病功能的*Mlo*候选基因,为进一步研究木薯*Mlo*家族成员的功能及其在抗病分子育种中的应用奠定基础。

材料与方法

采用2种方法系统地搜索和鉴定木薯(*Manihot esculenta* Crantz)基因组中的*Mlo*基因家族成员:一方面以“*Mlo*”作为关键词在木薯基因组数据库(http://www.phytozome.net/search.php?show=blast&method=Org_Mesculenta)中进行搜索,下载其蛋白序列,随后再以候选的*Mlo*序列为检索序列,在NCBI蛋白质数据库中进行Blastp搜索,获得与其他物种中已鉴定的*Mlo*蛋白同源的木薯*Mlo*蛋白;另一方面以已鉴定功能的*Mlo*蛋白(包括大麦Hv*Mlo*,小麦Ta*Mlo*1、Ta*Mlo*2、Ta*Mlo*3,水稻Os*MLO*2,番茄Si*Mlo*1,拟南芥At*Mlo*2、At*Mlo*6、At*Mlo*12)序列为检索序列(表1),分别在木薯基因组数据库(http://www.phytozome.net/search.php?show=blast&method=Org_Mesculenta)中进行Blastp搜索,E值设为 1×10^{-10} ,获得同源的木薯*Mlo*蛋白。通过2种搜索方法得到的木薯*Mlo*蛋白,人工去除重复后,再利用NCBI保守结构域分析工具(<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?>)对所获得的木薯*Mlo*蛋白家族成员进行鉴定,检测是否含有

*Mlo*保守结构域。

采用SMART在线分析软件(<http://smart.embl-heidelberg.de/>)分析蛋白的结构(Letunic等2004);用SignalP 4.1 Server在线分析软件(<http://www.cbs.dtu.dk/services/SignalP/>)分析信号肽结构(Petersen等2011);用TMHMM 2.0 Server在线分析软件(<http://www.cbs.dtu.dk/services/TMHMM/>)分析蛋白跨膜结构域。采用钙调素结合蛋白(CaMBP)数据库(http://calcium.uhnres.utoronto.ca/ctdb/pub_pages/search/index.htm)对可能的抗病相关*Mlo*蛋白进行CaM结合结构域预测。

为了分析木薯*Mlo*基因家族成员与其他物种*Mlo*成员之间的进化关系,进一步将木薯*Mlo*基因家族成员、模式植物拟南芥所有*Mlo*基因家族成员以及大麦、小麦、水稻和番茄中已知功能的*Mlo*基因(表1)构建了系统进化树。利用ClustalX 1.83软件进行多重序列比对(Thompson等2002),采用MEGA 4.1软件通过neighbor-joining方法构建系统进化树,并根据bootstrap重抽样1 000次进行置信度分析(Kumar等2001)。

结果与讨论

1 木薯*Mlo*基因家族成员的鉴定及其序列特征分析

本文采用2种方法对木薯基因组数据库进行系统地搜索,共获得21个木薯*Mlo*基因家族成员,分别命名为Me*Mlo*1~Me*Mlo*21,其中20个成员Me*Mlo*1~Me*Mlo*20具有完整的序列,而Me*Mlo*21只有部分序列(表2)。从表2可以看出,具有完整序列的*Mlo*基因编码蛋白的氨基酸长度、等电点及分子量差异较大,其中Me*Mlo*17是开放阅读框(ORF)最长的一个基因,编码597个氨基酸,蛋白分子量为68.28 kDa;Me*Mlo*20的ORF最短,编码277个氨基酸,蛋白分子量为31.87 kDa;这些*Mlo*的理论等电点从6.51 (Me*Mlo*5)至9.61 (Me*Mlo*17)。进一步对20个具有完整序列的木薯*Mlo*基因的编码蛋白结构进行分析发现,这些*Mlo*蛋白均具有一个*Mlo*保守结构域,包含有多次跨膜结构,跨膜次数为4 (Me*Mlo*20)到8次,其中14个蛋白具有7次跨膜结构,3个蛋白(Me*Mlo*8、Me*Mlo*10和Me*Mlo*18)具有N端信号肽,而且信号肽均与第一次跨膜结构存在重叠区。对拟南芥基因组所有*Mlo*成员及其他物种

表1 拟南芥、大麦、小麦、水稻和番茄Mlo基因及其结构特征

Table 1 The structure characteristics of Mlo genes from *Arabidopsis*, barley, wheat, rice, and tomato

种属	基因	位点/登录号	跨膜结构	信号肽
拟南芥(<i>Arabidopsis thaliana</i>)	<i>AtMlo1</i>	AT4G02600	7	0
	<i>AtMlo2</i>	AT1G11310	7	0
	<i>AtMlo3</i>	AT3G45290	7	0
	<i>AtMlo4</i>	AT1G11000	6	0
	<i>AtMlo5</i>	AT2G33670	7	0
	<i>AtMlo6</i>	AT1G61560	8	0
	<i>AtMlo7</i>	AT2G17430	8	1*
	<i>AtMlo8</i>	AT2G17480	8	1*
	<i>AtMlo9</i>	AT1G42560	6	0
	<i>AtMlo10</i>	AT5G65970	8	1*
	<i>AtMlo11</i>	AT5G53760	7	0
	<i>AtMlo12</i>	AT2G39200	7	0
	<i>AtMlo13</i>	AT4G24250	7	0
	<i>AtMlo14</i>	AT1G26700	7	0
	<i>AtMlo15</i>	AT2G44110	6	0
大麦(<i>Hordeum vulgare</i>)	<i>HvMlo</i>	CAB06083	7	0
小麦(<i>Triticum aestivum</i>)	<i>TaMlo1</i>	CAC25081	7	0
	<i>TaMlo2</i>	AAK94904	7	0
	<i>TaMlo3</i>	CAC25080	7	0
水稻(<i>Oryza sativa</i>)	<i>OsMLO2</i>	AAK94907	7	0
番茄(<i>Solanum lycopersicum</i>)	<i>SlMlo1</i>	AAX77013	6	0

*表示信号肽与第一个跨膜结构之间存在重叠区。

表2 木薯全基因组Mlo基因家族成员及其序列特征

Table 2 The sequence characterizations of Mlo genes from the genome of *M. esculenta*

基因	位点	氨基酸/个	蛋白分子量/kDa	等电点	跨膜结构	信号肽	钙离子结合区
<i>MeMlo1</i>	cassava4.1_033658m.g	564	63.79	9.46	7	0	RWIVAFFRQFFG
<i>MeMlo2</i>	cassava4.1_023541m.g	592	68.07	9.35	7	0	无
<i>MeMlo3</i>	cassava4.1_033120m.g	556	63.56	9.31	7	0	RWIVAFFRQFFA
<i>MeMlo4</i>	cassava4.1_028675m.g	540	61.74	8.18	6	0	无
<i>MeMlo5</i>	cassava4.1_025431m.g	517	59.58	6.51	7	0	无
<i>MeMlo6</i>	cassava4.1_004402m.g	578	66.34	8.78	7	0	VWIVCFRQFFG
<i>MeMlo7</i>	cassava4.1_031151m.g	548	61.63	8.92	7	0	无
<i>MeMlo8</i>	cassava4.1_004507m.g	571	65.31	9.37	8	1*	无
<i>MeMlo9</i>	cassava4.1_029705m.g	514	58.83	7.39	7	0	无
<i>MeMlo10</i>	cassava4.1_031835m.g	453	51.65	9.11	7	1*	无
<i>MeMlo11</i>	cassava4.1_004793m.g	557	63.68	9.02	7	0	无
<i>MeMlo12</i>	cassava4.1_004283m.g	585	67.21	8.82	8	0	VWIVCFRQFFG
<i>MeMlo13</i>	cassava4.1_007579m.g	448	51.77	9.31	6	0	无
<i>MeMlo14</i>	cassava4.1_004823m.g	555	63.59	9.22	7	0	无
<i>MeMlo15</i>	cassava4.1_005640m.g	516	58.15	8.61	7	0	KRKKGLKAAAAAAAA
<i>MeMlo16</i>	cassava4.1_033401m.g	514	58.83	7.69	7	0	无
<i>MeMlo17</i>	cassava4.1_030963m.g	597	68.28	9.61	7	0	无
<i>MeMlo18</i>	cassava4.1_022918m.g	577	65.97	8.60	6	1*	无
<i>MeMlo19</i>	cassava4.1_004964m.g	547	63.09	8.30	7	0	无
<i>MeMlo20</i>	cassava4.1_013598m.g	277	31.87	8.65	4	0	无
<i>MeMlo21**</i>	cassava4.1_029787m.g	191	-	-	-	-	-

*表示信号肽与第一个跨膜结构之间存在重叠区;**表示序列不完整。

中已知功能的Mlo进行结构分析也发现,所有的Mlo蛋白均具有6~8个跨膜结构,拟南芥的3个蛋白AtMlo7、AtMlo8和AtMlo10具有N端信号肽,而且信号肽均与第一次跨膜结构存在重叠区(表1)。

2 木薯Mlo基因家族成员系统进化关系分析

为了揭示木薯Mlo基因家族成员的进化关系及其分类,采用ClustalX 1.83软件对木薯MeMlo1~20、拟南芥基因组所有Mlo成员(AtMlo1~15)、大麦

HvMlo、小麦TaMlo1~3、水稻OsMlo2和番茄SIMlo1共41个基因的编码蛋白进行多重序列比对,并采用MEGA 4.1软件构建系统进化树。结果(图1)显示,可将所有木薯Mlo基因家族成员分为6类(I~VI),除了第IV和第VI类外,其他4类都包括来自拟南芥的Mlo基因成员。第III类是最大的一类,包含有9个成员。第VI类是最小的一类,只包含2个基因木薯(MeMlo18和MeMlo19),因此推测,这可能是木薯中

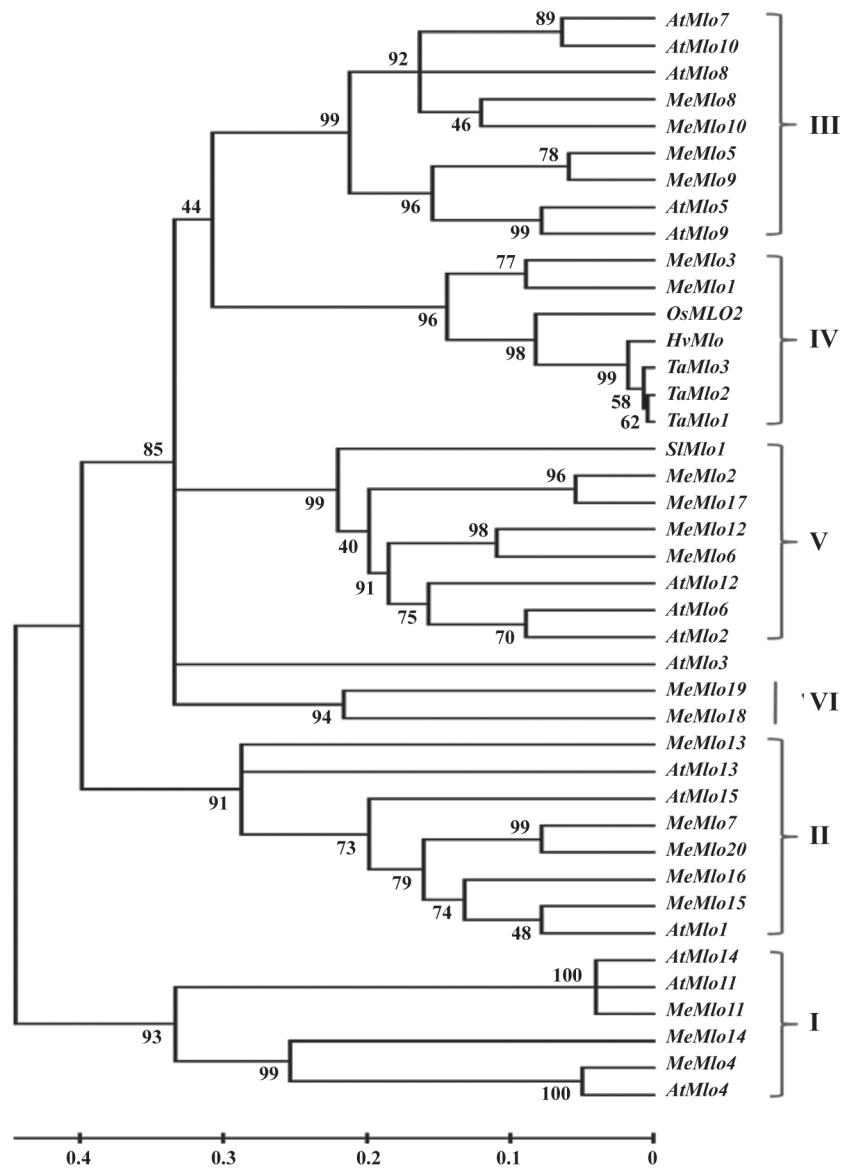


图1 木薯和拟南芥Mlo基因家族成员及其他已知功能Mlo基因的系统进化关系分析

Fig.1 Phylogenetic analysis of Mlo family members from the genomes of *M. esculenta* and *Arabidopsis*, and other Mlo genes of known function

进化树中每个结点上的数值表示bootstrap重抽样1 000次进行置信度分析得到的置信值。

特有的一类Mlo。这一研究结果与前人的研究存在一定差异。Devoto等(2003)对拟南芥、玉米、水稻、大麦、小麦和苔藓的Mlo基因家族成员进行进化关系分析时,将不同物种的Mlo分成了6个亚族和5个类群(I~V),AtMlo3是一个独立的亚族但没有将其进行归类(Devoto等2003)。本文结果的木薯MeMlo18和MeMlo19两个基因可能与AtMlo3类似,应该归为独立的类群,因此,将MeMlo18和MeMlo19划分为第VI类。MeMlo20属于第II类,但该蛋白只有4个跨膜结构域,而且氨基酸序列长度也明显比其他Mlo成员短很多,推测MeMlo20可能是进化过程中终止密码子提前插入,使其突变为一个假基因。第IV类由木薯MeMlo1、MeMlo3及禾本科作物已知的抗病Mlo基因,包括大麦HvMlo、小麦TaMlo1~3和水稻OsMlo2组成(Buschges等1997; Elliott等2002)。木薯MeMlo2、MeMlo6、MeMlo12、MeMlo17与双子叶植物已知的抗病Mlo基因,包括拟南芥(AtMlo2、AtMlo6和AtMlo12)和番茄SIMlo1聚在第V类(Bai等2008; Consonni等2006)。第IV和第V类可能是植物抗病Mlo成员类群,而木薯MeMlo1、MeMlo2、MeMlo3、MeMlo6、MeMlo12和MeMlo17可能是具有抗病功能的

Mlo成员。

3 木薯抗病Mlo候选基因的氨基酸序列特征分析

通过对木薯和拟南芥基因组Mlo基因家族以及大麦、小麦、水稻和番茄的已知抗病Mlo基因进行系统进化关系分析发现,木薯MeMlo1、MeMlo2、MeMlo3、MeMlo6、MeMlo12和MeMlo17基因可能是具有抗病功能的Mlo成员。为了更好地揭示这6个抗病Mlo候选基因的序列特征,采用ClustalX 1.83软件对这6个候选基因的氨基酸序列进行多重序列比对。结果显示,这6个候选基因的编码蛋白在跨膜区及其附近序列具有很高的保守性,而C末端序列相似性比较低(图2)。MeMlo12具有8个跨膜结构,MeMlo1、MeMlo2、MeMlo3、MeMlo6和MeMlo17均具有7个跨膜结构,它们都缺失了MeMlo12的第5个跨膜结构(TM5)(图2,跨膜结构用下划线标出)。这与已知的抗病Mlo蛋白结构类似,AtMlo6和SIMlo1分别具有8个和6个跨膜结构,其他已知的抗病Mlo(HvMlo、TaMlo1-3、OsMlo2、AtMlo2和AtMlo12)均具有7个跨膜结构(表1)。已有的研究表明,Mlo蛋白是一类钙结合蛋白,具有钙离子结合区(CaMBD),而且CaMBD结构域能够提高Mlo蛋白的活性(Kim等

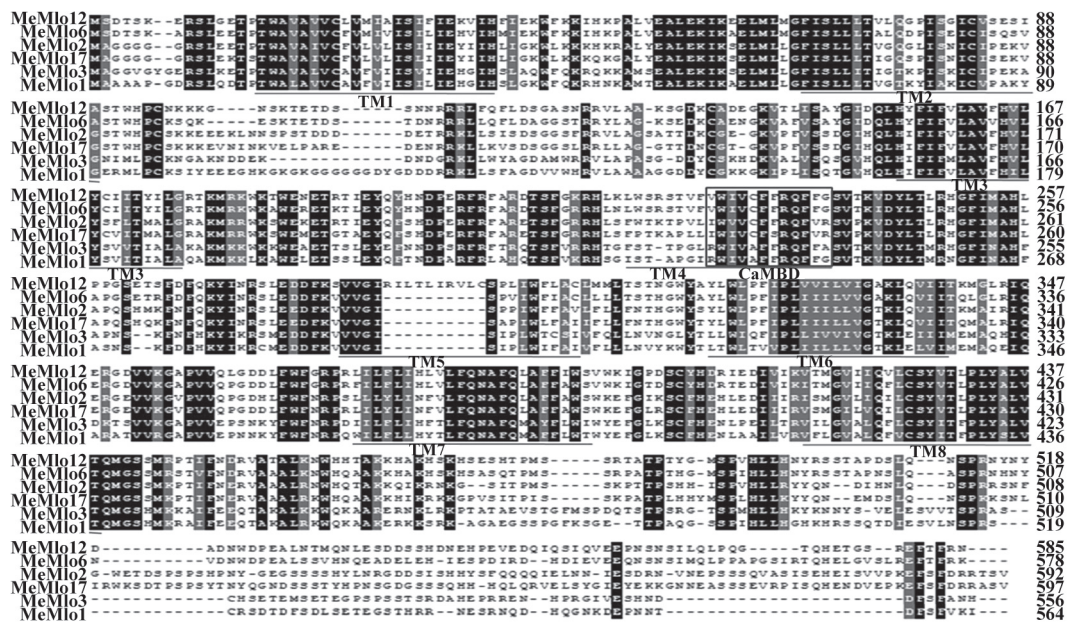


图2 木薯抗病Mlo候选基因的氨基酸多重序列比对

Fig.2 Multiple sequence alignment of cassava Mlo candidate genes for disease resistance in *M. esculenta*
下划线标出的是跨膜结构域(TM1~TM8);方框标出的是钙离子结合区(CaMBD)。

2002), 因此, 采用钙调素结合蛋白(CaMBP)数据库对木薯的20个Mlo蛋白进行CaM结合结构域预测。结果发现, 除MeMlo2和MeMlo17外, 其他4个候选的抗病Mlo即MeMlo1、MeMlo3、MeMlo6和MeMlo12分别在第238、225、226、227个氨基酸可能具有一个CaMBD结构域(表2和图2)。除此之外, MeMlo15虽然没有与已知的抗病Mlo聚在同一类群上, 但它也具有一个CaMBD结构域(位于第461个氨基酸, 表2)。这些抗病Mlo候选基因的鉴定为木薯抗病育种提供了良好的基因资源。

参考文献

- 黄强, 田益农, 黄惠芳, 彭靖茹, 侯学文(2013). 植物分子育种技术在改良木薯种质方面的应用. 广西植物, 33 (2): 148~153
- Bai Y, Pavan S, Zheng Z, Zappel NF, Reinstadler A, Lotti C, De Giovanni C, Ricciardi L, Lindhout P, Visser R et al (2008). Naturally occurring broad-spectrum powdery mildew resistance in a Central American tomato accession is caused by loss of *mlo* function. *Mol Plant Microbe Interact*, 21: 30~39
- Bart R, Cohn M, Kassen A, McCallum EJ, Shybut M, Petriello A, Krasileva K, Dahlbeck D, Medina C, Alicai T et al (2012). High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc Natl Acad Sci USA*, 109: E1972~E1979
- Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J et al (1997). The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell*, 88: 695~705
- Consonni C, Humphry ME, Hartmann HA, Livaja M, Durner J, Westphal L, Vogel J, Lipka V, Kemmerling B, Schulze-Lefert P et al (2006). Conserved requirement for a plant host cell protein in powdery mildew pathogenesis. *Nat Genet*, 38: 716~720
- Devoto A, Hartmann HA, Piffanelli P, Elliott C, Simmons C, Taramino G, Goh CS, Cohen FE, Emerson BC, Schulze-Lefert P et al (2003). Molecular phylogeny and evolution of the plant-specific seven-transmembrane MLO family. *J Mol Evol*, 56: 77~88
- Elliott C, Zhou F, Spielmeier W, Panstruga R, Schulze-Lefert P (2002). Functional conservation of wheat and rice *Mlo* orthologs in defense modulation to the powdery mildew fungus. *Mol Plant Microbe Interact*, 15: 1069~1077
- Kim MC, Lee SH, Kim JK, Chun HJ, Choi MS, Chung WS, Moon BC, Kang CH, Park CY, Yoo JH et al (2002). Mlo, a modulator of plant defense and cell death, is a novel calmodulin-binding protein. Isolation and characterization of a rice Mlo homologue. *J Biol Chem*, 277: 19304~19314
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*, 17: 1244~1245
- Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Res*, 32: D142~D144
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*, 8: 785~786
- Thompson JD, Gibson TJ, Higgins DG (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinform Chap*, 2: 2~3