

DOI: 10.3724/SP.J.1006.2020.94163

甘薯基因组 NBS-LRR 类抗病家族基因挖掘与分析

黄小芳^{1,2,**} 毕楚韵^{1,2,**} 石媛媛² 胡韵卓³ 周丽香⁴ 梁才晓⁴
黄碧芳⁴ 许明^{1,2} 林世强^{1,4,*} 陈选阳^{1,2,5,*}

¹ 福建农林大学作物生物技术福建省高校重点实验室, 福建福州 350002; ² 福建农林大学农学院, 福建福州 350002; ³ 福建农林大学植物保护学院, 福建福州 350002; ⁴ 福建农林大学生命科学学院, 福建福州 350002; ⁵ 福建农林大学教育部作物遗传育种与综合利用重点实验室, 福建福州 350002

摘要: NBS-LRR 类基因家族是植物抗病 *R* 基因(Resistance gene)数量最多的一类, 具有 NBS (Nucleotide-binding site) 和 LRR (Leucine-leucine-repeat) 结构域。甘薯(*Ipomoea batatas*)栽培种基因组已完成测序, 但尚未注释, 本研究对甘薯基因组序列进行外显子预测, 得到甘薯染色体组全基因组蛋白序列, 在此基础上进一步对 NBS-LRR 家族基因鉴定和分析表明, 甘薯基因组中含有 379 个 NBS-LRR 家族基因, 占全基因组基因总数的 0.212%, 其中 N 型亚家族 120 个, NL 型 103 个, CNL 型 133 个, TNL 型 22 个, PN 型 1 个。所有染色体上均有 NBS-LRR 家族基因分布, 但数量明显不同, 其中有 60.9% 的 NBS-LRR 基因序列呈簇状分布。NBS-LRR 基因序列有 15 个保守结构域, 在 N 端较为保守。研究结果为甘薯进一步开展 NBS-LRR 家族基因的功能研究和抗性育种提供了参考。

关键词: 甘薯; NBS-LRR; *R* 基因; 基因家族; 生物信息学

Discovery and analysis of NBS-LRR gene family in sweet potato genome

HUANG Xiao-Fang^{1,2,**}, BI Chu-Yun^{1,2,**}, SHI Yuan-Yuan², HU Yun-Zhuo³, ZHOU Li-Xiang⁴, LIANG Cai-Xiao⁴, HUANG Bi-Fang⁴, XU Ming^{1,2}, LIN Shi-Qiang^{1,4,*}, and CHEN Xuan-Yang^{1,2,5,*}

¹ Key Laboratory of Crop Biotechnology, Fujian Agriculture and Forestry University, Fujian Province Universities, Fuzhou 350002, Fujian, China; ² College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China; ³ College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China; ⁴ College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China; ⁵ Key Laboratory of Genetics, Breeding and Multiple Application of Crops, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China

Abstract: The NBS-LRR gene families possess the most abundant resistance genes in plants. Members of the NBS-LRR gene families contain nucleotide-binding site (NBS) and leucine-leucine repeat (LRR) domains. The genome of sweet potato (*Ipomoea batatas*) cultivar has been sequenced but the genes have not been annotated yet. In this study, we predicted the exons of sweet potato genome and obtained the proteins sequences, which were then used to identify and analyze genes of NBS-LRR family. There were 379 genes within NBS-LRR family, amounting to 0.212% of the total genes of sweet potato. The number of the N type, NL type, CNL type, TNL type and PN type was 120, 103, 133, 22, and 1, respectively. All of the chromosomes had NBS-LRR family genes but varied in number and 60.9% of them were clustered. NBS-LRR genes included 15 conservative domains and the genes were conservative within N terminal domain. The results provide references for further studies on the function of NBS-LRR family genes and resistance breeding of sweet potato.

Keywords: *Ipomoea batatas*; NBS-LRR; *R* gene; gene family; bioinformatics

本研究由福建省科技重大专项子专题(2017NZ0002-2)资助。

This study was supported by the Fujian Provincial Department of Science & Technology (CN)(2017NZ0002-2).

* 通信作者(Corresponding authors): 林世强, E-mail: linshiqiang@fafu.edu.cn, Tel: 059183789367; 陈选阳, E-mail: cxy@fafu.edu.cn, Tel: 059183789483

** 同等贡献(Contributed equally to this work)

第一作者联系方式: 黄小芳, E-mail: 1102718600@qq.com, Tel: 059183789483; 毕楚韵, E-mail: 494028227@qq.com, Tel: 059183789483

Received (收稿日期): 2019-11-01; Accepted (接受日期): 2020-04-15; Published online (网络出版日期): 2020-04-26.

URL: <http://kns.cnki.net/kcms/detail/11.1809.S.20200426.1548.024.html>

在植物育种过程中,通常利用抗病 *R* 基因(resistance gene)来控制植物病害^[1-3]。*R* 基因编码的抗性蛋白通过直接或间接的方式识别对应的由病原微生物无毒基因 *Avr* (Avirulence gene)编码的无毒蛋白后,启动效应因子触发的免疫系统(effector-triggered immunity, ETI),激活植物体内抗病信号途径,在侵染部位产生局部的细胞和组织过敏性坏死(hypersensitive response, HR)反应,最终抵御病原菌的侵染及进一步扩散^[4-5]。

目前克隆的抗病 *R* 基因大多数都属于 NBS-LRR 家族^[6]。这类基因家族根据其 N-末端是否含有 TIR 结构可分为 2 个亚类,一个亚类为 TNL 型,含有 TIR-NBS-LRR 结构,这个亚类 *R* 基因编码的蛋白 N-末端有一个白细胞介素受体(toll/interleukin-1 receptor, TIR)的同源结构域;另一个亚类为非 TIR-NBS-LRR (non-TNL)结构,这类 *R* 基因编码的蛋白 N-末端不含有 TIR 结构,但通常被编码的 CC (Coiled-Coil)替代,称为 CC-NBS-LRR(CNL)型^[7]。

我国是世界上甘薯种植面积最大的国家,甘薯总产量稳居世界首位。据联合国粮食及农业组织 FAO (<http://www.fao.org/faostat/zh/#data>)统计,2018 年中国甘薯栽培面积和总产分别占世界的 30.0%和 58.0%。尽管我国甘薯种植总面积和占薯类比重持续下降,但种植效益仍十分可观^[8]。甘薯在种植过程中受到各种病害、虫害和病毒病的威胁^[9-10],而选育和应用抗病品种是减少病虫害损失的最有效手段。近年来,关于甘薯抗病基因的挖掘逐步得到重视,NBS 类抗病基因的克隆也有了一些有限研究和报道^[11],但从甘薯全基因组中发掘 NBS-LRR 抗病基因的研究还未见报道。

2017 年中国科学院上海辰山植物科学研究中心和植物生理生态研究所,联合德国马克斯普朗克分子遗传研究所和分子植物生理研究所,利用 Illumina 测序技术完成对甘薯全基因组序列测序^[12],为甘薯的基础和应用研究建立了扎实的基础。本研究通过生物信息学方法对甘薯全基因组序列进行了 CDS (Coding Domain Sequence)区预测,检索了 NBS-LRR 家族基因,并对其进行染色体定位、分类、结构分析和系统进化树研究,以为甘薯抗病机制研究及甘薯抗病遗传育种提供科学参考。

1 材料与方法

1.1 甘薯全基因组序列

从 NCBI (<https://www.ncbi.nlm.nih.gov/genome/>)

term= Ipomoea+batatas)下载甘薯全基因组序列,得到 15 条染色体的 DNA 序列。

1.2 snap 基因注释

通过 snap 程序应用拟南芥(*At.hmm*)、线虫(*Ce.hmm*)、水稻(*Os.hmm*)的 HMM 模型对甘薯染色体序列 CDS 进行检索^[13],分别得到 3 个.zff 文件,分别从中随机抽取 500 组合并,作为训练甘薯特异 HMM 模型的序列文件。将甘薯特异的 HMM 模型文件(Ib.hmm)作为检索工具,通过 snap 程序检索预测甘薯染色体 CDS,得到甘薯全基因组蛋白质序列。

1.3 含 NB-ARC 结构域基因预测

从 Pfam 网站^[14-15](<http://pfam.xfam.org/>)下载 NB-ARC(ID:PF00931)家族序列信息,保存为 NB-ARC.hmm 文件。通过 hmmsearch 程序^[16-17]用下载的 NB-ARC.hmm 文件对甘薯全基因组蛋白质序列进行 NB-ARC 结构域预测。从中选取 E-value 值 1E-60 的序列,利用 Samtools 程序^[18-19]新建甘薯的 NB-ARC 蛋白质序列文件(Ib-NBS.fasta)^[20]。利用 Clustal Omega^[21]对 Ib-NBS.fasta 文件进行多重序列比对,得到 Ib-NBS-clustalo.fasta 文件。利用 hmmbuild 程序建立 Ib-NBS-clustalo.fasta 文件的 HMM 模型,获得甘薯特异的 NB-ARC HMM 模型(NB-ARC-specific.hmm)。利用甘薯特异的 HMM 模型再次检索甘薯全基因组蛋白质序列预测 NB-ARC 结构域,选取 E-value<0.01 的蛋白质序列保存为甘薯高特异的 NB-ARC 序列信息(Ib-NBS-specific-aa.fasta)。去除甘薯高特异的 NB-ARC 序列信息中长度小于 200 的蛋白质序列,随后将其导入 NCBI Conserved Domains Tool^[22-25]和 Interproscan^[26](<http://www.ebi.ac.uk/interpro/search/sequence/>)进行筛选,去除结构域缺失严重的序列,最终得到甘薯 NB-ARC 基因家族蛋白质序列数据库。

1.4 NBS 关联的保守结构域分析分类

NBS-LRR 家族基因在其 N-末端含有 CC、TIR^[7]、RPW8^[27]等结构域,在 C-末端通常含有 LRR 结构域^[28]。从 Pfam 数据库^[14-15](<http://pfam.xfam.org/>)中下载 TIR HMMs (PF01582)、RPW8 HMMs (PF05659)和 LRR HMMs (PF00560、PF07723、PF07725 和 PF12799)模型,合并后用 hmmpress 程序建立 TIR/RPW8/LRR HMM 模型(TRL.hmm)。通过 hmmscan 程序用 TRL.hmm 文件对甘薯 NBS-LRR 家族基因信息库进行结构域预测,同时结合 Interproscan 得到结果鉴别出不同类型的 TIR/

RPW8/LRR 结构域, 统计其类型和重复数。通过 Pepcoil 程序^[29]预测甘薯 NBS-LRR 家族基因氨基酸序列信息中的 CC 结构域(coiled-coil domain), 利用 Paircoil2 (P 值临界值为 0.03, 氨基酸长度临界值为 21)^[30]检测 CC 结构域。

1.5 甘薯 NBS-LRR 基因染色体定位

统计甘薯 15 条染色体上 NBS-LRR 家族基因的位置信息, 参考 Jupe 等^[31]的方法绘制 NBS-LRR 家族基因在染色体上的位置。

1.6 甘薯 CNL、TNL、RPW8 亚家族 motif 分析

将甘薯 NBS-LRR 家族基因分类整理后, 从中筛选出 CC-NBS-LRR、TIR-NBS-LRR、RPW8-NBS 三种 NBS 亚家族类型的 90 个氨基酸序列^[32], 通过 MEME 程序^[33]进行 motif 分析和识别(设置搜索功能域数为 15)。将 MEME 程序运行后得到的 .xml 文件利用 TBtools 软件^[34]处理。

1.7 甘薯 CNL、TNL、RPW8 亚家族进化树构建

将筛选得到的 CC-NBS-LRR、TIR-NBS-LRR、RPW8-NBS 三种 NBS-LRR 亚家族类型的 90 个氨基酸序列利用 Clustal Omega 程序^[21]进行多序列比对后, 使用 Gblocks^[35-36]提取保守序列(允许序列保留 50%的缺口数), 进一步利用 Jalview^[37-38]手动矫正提取到的保守序列, 保存为 .fa 序列文件。根据比对结果为基础, 应用 MEGA X^[39]软件, 选择 Maximum Likelihood 法, 设置运行参数模式 WAG with Freqs.(+F)model, 校验参数 Bootstrap=500, 生成甘薯 NBS-LRR 基因家族系统进化树。

2 结果与分析

2.1 甘薯基因组中含有 NB-ARC 结构域的基因鉴别及其分类

采用 snap 程序对甘薯全基因组序列进行 CDS 区识别预测, 得到甘薯 15 条染色体 CDS 序列 178,458 个。利用 hmmsearch 程序筛选甘薯基因组中含有 NB-ARC 结构域的序列, 以 E-value 0.01 为标

准, 获得相关蛋白序列 432 个。从中选取高特异蛋白质序列(E-value 1E-60)构建甘薯特异的 NB-ARC HMM 模型, 利用甘薯特异的 HMM 模型再次预测甘薯全基因组序列的 NB-ARC 保守结构域, 获得甘薯含有 NB-ARC 结构域的特异蛋白序列 735 个(E-value 0.01)。去除氨基酸长度小于 200 以及结构域缺失严重的序列, 最终得到 379 个甘薯 NBS-LRR 家族基因, 占甘薯全基因组基因数目(178,458)的 0.212%。根据 TIR/RPW8/LRR 结构域和 CC 结构域分为 N (NBS)、NL (NBS-LRR)、CNL (CC-NBS-LRR)、TNL (TIL-NBS-LRR)和 PN (RPW8-NBS) 4 个亚家族类型(表 1), 其中 N 型 120 个, NL 型 103 个, CNL 型 133 个, TNL 型 22 个, PN 型 1 个。分析发现, NBS/LRR 结构域重复数和位置存在差异, 在序列结构域读取文件中含有 RX-CC_like 结构域, 此结构域为马铃薯抗 X 病毒和类似蛋白的 Coiled-coil 结构域(Coiled coil domain of the potato virus X resistance protein and similar proteins)^[40], 记为 Cx 结构域, 以区分 Pepcoil 程序得到的 CC 结构域。因此, 根据 NBS/LRR 结构域重复数和位置差异、CC 结构域类型进一步分为 34 个小类(表 1)。

在 NBS-LRR 家族基因的 N 末端区域含有 TIR (Toll/inter-leukin-1 receptor)保守结构域的蛋白称为 TNL 蛋白, 其他的称为非 TNL 蛋白。大部分非 TNL 蛋白的 N 末端有卷曲螺旋结构(Coiled-coil), 称为 CNL 蛋白^[41]。目前已进行 NBS-LRR 家族基因信息分析的植物有木薯^[42]、水稻^[43]、拟南芥^[44]、乌拉尔图小麦^[45]等, 其中 CNL 亚家族数目占其 NBS 家族基因数比例分别为木薯 (35.78%)^[42]、水稻 (32.71%)^[43]、拟南芥 (24.64%)^[44]、乌拉尔图小麦 (19.18%)^[45]。在甘薯 NBS-LRR 家族基因序列结构域分类中, 同时含有 CC、NBS、LRR 结构域的序列有 77 个, 占 NBS-LRR 家族基因总数的 20.32%; 同时含有 TIR、NBS、LRR 结构域的序列有 10 个, 占总基因家族信息数 2.64%。

表 1 甘薯编码 NBS-LRR 蛋白的基因数量及其分类

Table 1 Number of genes encoding NBS-LRR protein in *Ipomoea batatas*

项目 Item	预测结构域 Predicted domain	代码 Code	甘薯 Ib	木薯 ^a Me ^a	水稻 ^b Os ^b	拟南芥 ^c At ^c	乌拉尔图小麦 ^d Tu ^d
N 类型 N type	NB	N	118	14	45	1	270
		NCCC	1	—	—	—	—
	NB NB	NN	1	—	—	—	—
NL 类型 NL type	NB LRR	NL	81	52	301	6	31
		NLN	1	—	—	—	—

(续表 1)

项目 Item	预测结构域 Predicted domain	代码 Code	甘薯 Ib	木薯 ^a Me ^a	水稻 ^b Os ^b	拟南芥 ^c At ^c	乌拉尔图小麦 ^d Tu ^d
CNL 类型 CNL type		NLC	1	—	—	—	—
	NB LRR LRR	NLL	16	—	—	—	19
		NLLC	1	—	—	—	—
	NB LRR LRR LRR	NLLL	1	—	—	—	1
	NB NB LRR	NNL	2	—	3	—	1
	CC NB	CN	22	11	7	5	55
		C _x N	29	—	—	—	—
	CC NB LRR	CNL	15	117	175	51	84
		C _x NL	37	—	—	—	—
		CNCL	1	—	—	—	—
		CNLC	1	—	—	—	—
		C _x NLC	1	—	—	—	—
		LC _x NL	1	—	—	—	—
	CC NB LRR LRR	CNLL	4	—	—	—	8
		C _x NLL	5	—	—	—	—
	CC NB LRR LRR LRR	C _x NLLL	1	—	—	—	1
	CC NB NB LRR LRR	CNNLL	1	—	—	—	—
	CC CC NB	C _x CN	4	—	—	—	—
		C _x C _x N	1	—	—	—	—
	CC CC NB LRR	CCNL	2	—	—	—	—
	C _x CNL	6	—	—	—	—	
CC CC NB LRR LRR	C _x CNLL	1	—	—	—	—	
CC CC CC CC NB LRR LRR	CCCCNLL	1	—	—	—	—	
TNL 类型 TNL type	TIR NB	TN	9	5	3	23	0
	TIR NB NB	TNN	1	—	—	—	—
	TIR NB LRR	TNL	7	29	—	88	0
	TIR NB LRR LRR	TNLL	3	—	—	—	—
	TIR NB LRR LRR LRR	TNLLL	2	—	—	—	—
PN 类型 PN type	RPW8 NB	PN	1	—	—	—	—
其他 Other			—	99	1	33	15
总数 Total			379	327	535	207	485
基因组中基因总数 Total number of genes in the genome			178458	30666	37544	25498	32265
基因组大小 Genome size (Mb)			633.42	~760	389	125	3747.05

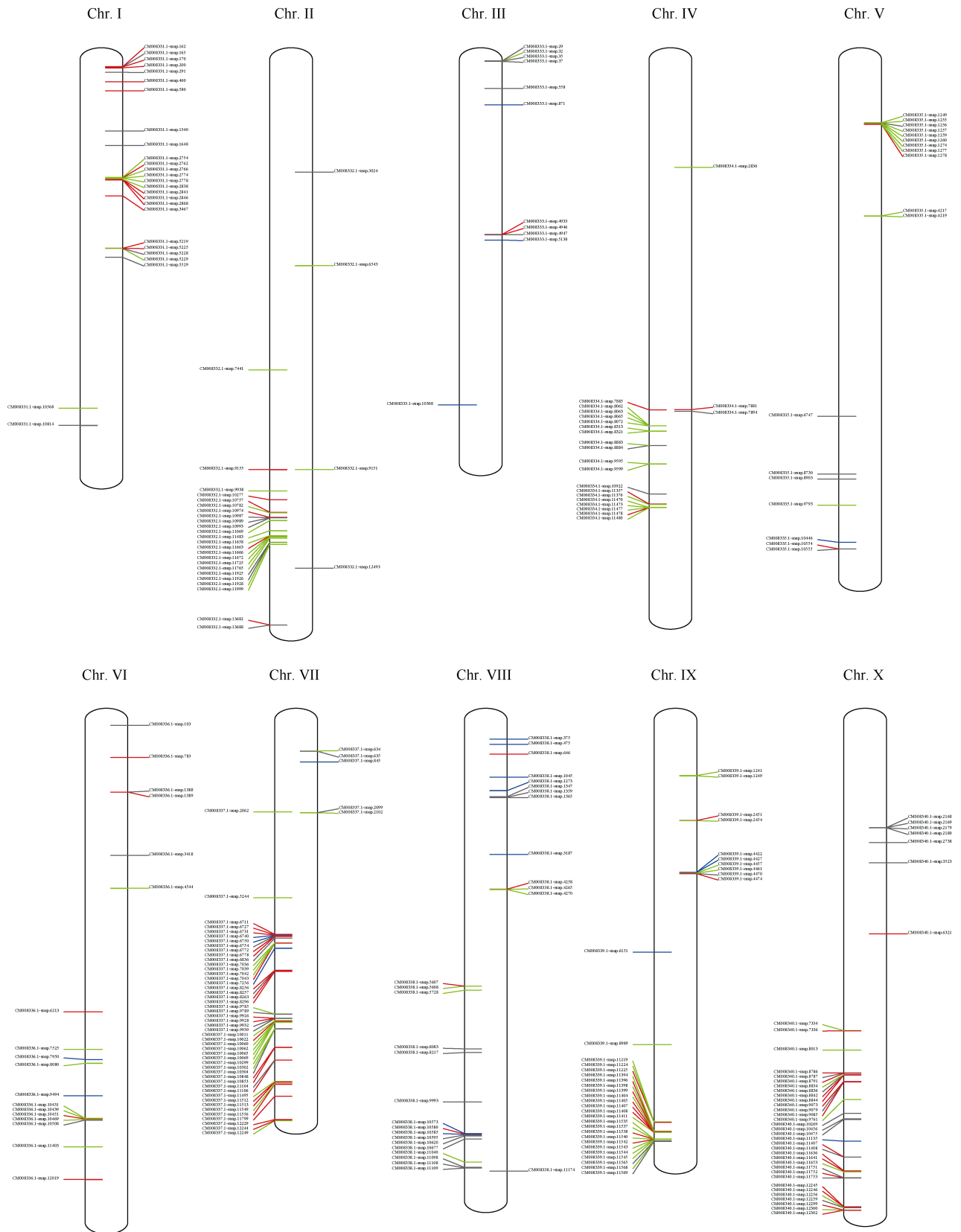
a: 数据来源于[42]; b: 数据来源于[43]; c: 数据来源于[44]; d: 数据来源于[45]。

a: data were from [42]; b: data were from [43]; c: data were from [44]; d: data were from [45]. Ib: *Ipomoea batatas*; Me: *Manihot esculenta*; Os: *Oryza sativa*; At: *Arabidopsis thaliana*; Tu: *Triticum urartu*.

2.2 甘薯 NBS-LRR 家族基因在染色体上的分布

NBS-LRR 家族基因有不同的亚家族, 包括 N 型(黑色)、NL 型(红色)、CNL 型(绿色)、TNL 型(蓝色)、PN 型(青色)(图 1)。基因有正链编码(图 1 中染色体右侧)和负链编码(图 1 中染色体左侧)。NBS-LRR 家

族基因数量在甘薯 15 条染色体上分布并不均匀, 在 11 号染色体上仅有 3 个, 而在 13 号染色体上则有 54 个。在不同染色体上的基因以单基因或基因簇的形式存在。基因簇是指在 200 kb 的核苷酸单位中含有的基因群^[46-47], 甘薯 NBS-LRR 家族基因中共有



(图 1)

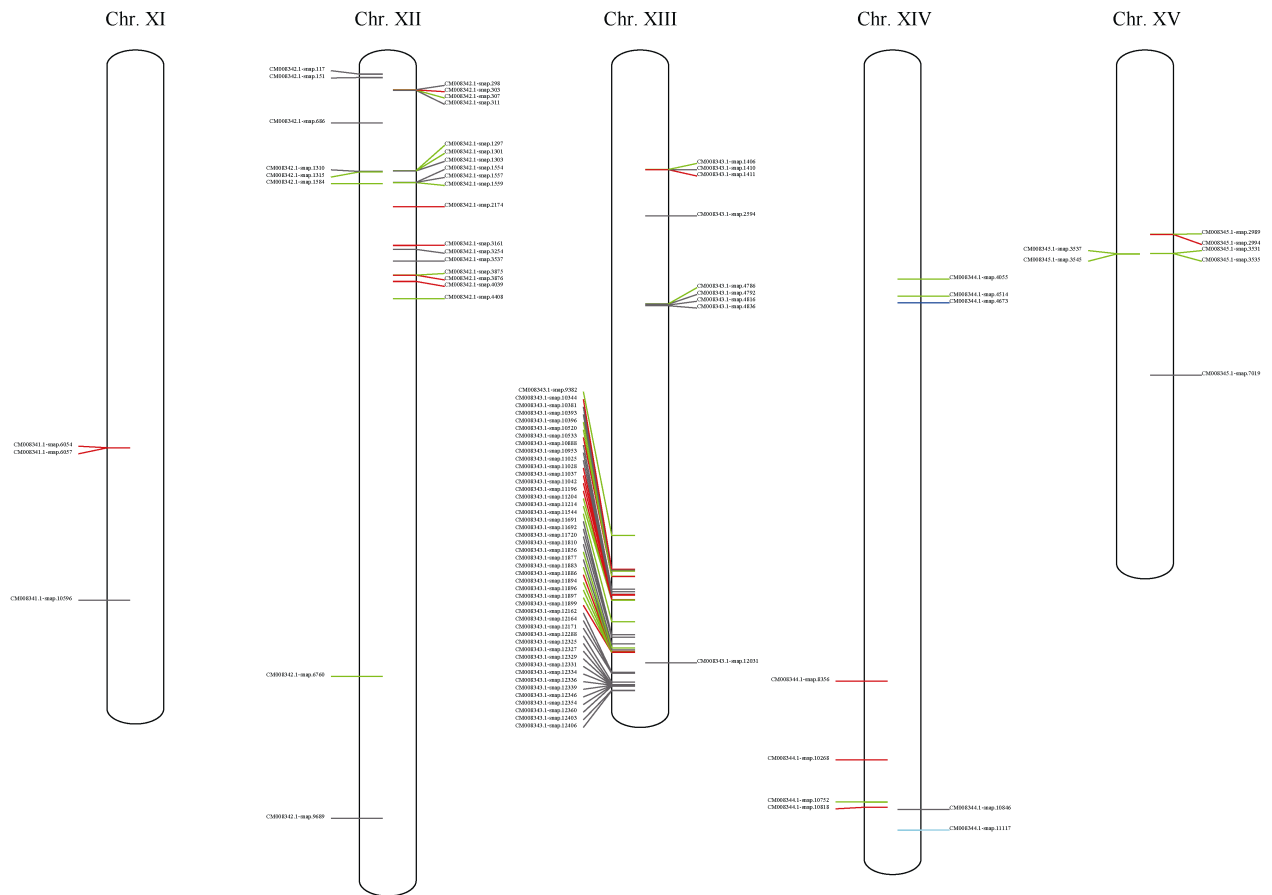


图 1 甘薯 NBS-LRR 家族基因在染色体上的分布

Fig. 1 Distribution of NBS-LRR family genes on *Ipomoea batatas* chromosomes

231 个基因是以基因簇的形式存在。这 231 个基因分布在 81 个基因簇中, 占 NBS-LRR 家族基因总数的 60.9%, 平均每个基因簇含 2.9 个基因。在这些基因簇中含有 2 个基因的基因簇数目最多, 有 46 个, 占总基因簇数的 56.8%。其次是 3~4 个基因的基因簇, 有 18 个, 12 个基因簇含有 4 个基因。含 5 个、6 个和 8 个基因的基因簇均只有 1 个。含有最大基因数的基因簇有 2 个, 分别分布在 9 号和 13 号染色体上, 簇中含有 9 个基因。

甘薯基因组中, 单独分布在染色体上的 NBS-LRR 家族基因有 148 个, 占 NBS-LRR 基因总数 39.1% (图 1 和表 2)。在 14 号染色体上的 9 个 NBS-LRR 家族基因均呈单基因分布, 其中含 RPW8 结构域的 NBS-LRR 家族基因也分布于 14 号染色体。7 号染色体基因簇最多达 16 个, 据此推测 7 号染色体可能发生 NBS-LRR 家族抗病基因的大规模复制, 导致该染色体上 NBS-LRR 家族基因的高密度分布^[46]。

2.3 甘薯 NBS-LRR 家族基因 CNL、TNL 亚家族 motif 分析

分别将甘薯 NBS-LRR 家族基因中 CNL 和 TNL 亚家族序列信息利用 MEME 程序检测, 筛选得到 15 个相似度较高的结构域。77 个 CNL 亚家族基因的结构域分布 (图 2) 存在以下规律: motif 10-motif 11-motif 2(P-loop)-motif 4(RNBS-A)-motif 1(Kinase 2)-motif 13(RNBS-B)-motif 5(GLPL)-motif 15-motif 9(RNBS-D)-motif 6(RNBS-D)-motif 12-motif 3(MHDV)-motif 8-motif 14-motif 7。其中 motif 15 (motif 5 和 motif 9 之间) 仅存在于 21 个序列中, 即位于 GLPL 结构域和 RNBS-D 结构域之间。其中 RNBS-D 保守基序对应于 2 个结构域, 分别为 motif 9 和 motif 6, 共有 70 个序列含有 motif 6, 其中有 17 个序列缺失 motif 9。TNL 亚家族结构域分布 (图 3) 存在以下规律: motif 10(TIR-1)-motif 3-motif 1(TIR-2)-motif 5-(TIR-3)-motif 8(TIR-4)-motif 2(P-loop)-motif 9(Kinase 2)-motif 4(RNBS-B)-motif 13-(RNBS-C)-motif 7(GLPL)-motif 11-motif 15-motif

表 2 甘薯 NBS-LRR 基因家族基因簇统计表

Table 2 Statistics of NBS-LRR gene family clusters of *Ipomoea batatas*

染色体 Chromosome	基因 Gene	基因簇/基因数目 Gene cluster/gene number	最大基因簇 Maximal gene cluster	百分比 Percentage (%)
1	26	4/14	4	53.8
2	28	4/13	4	46.4
3	11	2/7	4	63.6
4	22	5/15	5	68.2
5	18	4/13	6	72.2
6	18	2/4	2	22.2
7	53	16/41	4	77.4
8	29	4/9	3	31.0
9	35	8/30	9	85.7
10	40	12/30	4	75.0
11	3	1/2	2	66.7
12	26	5/14	4	53.8
13	54	11/33	9	61.1
14	9	0/0	0	0
15	7	3/6	2	85.7
总数 Total	379	81/231		60.9

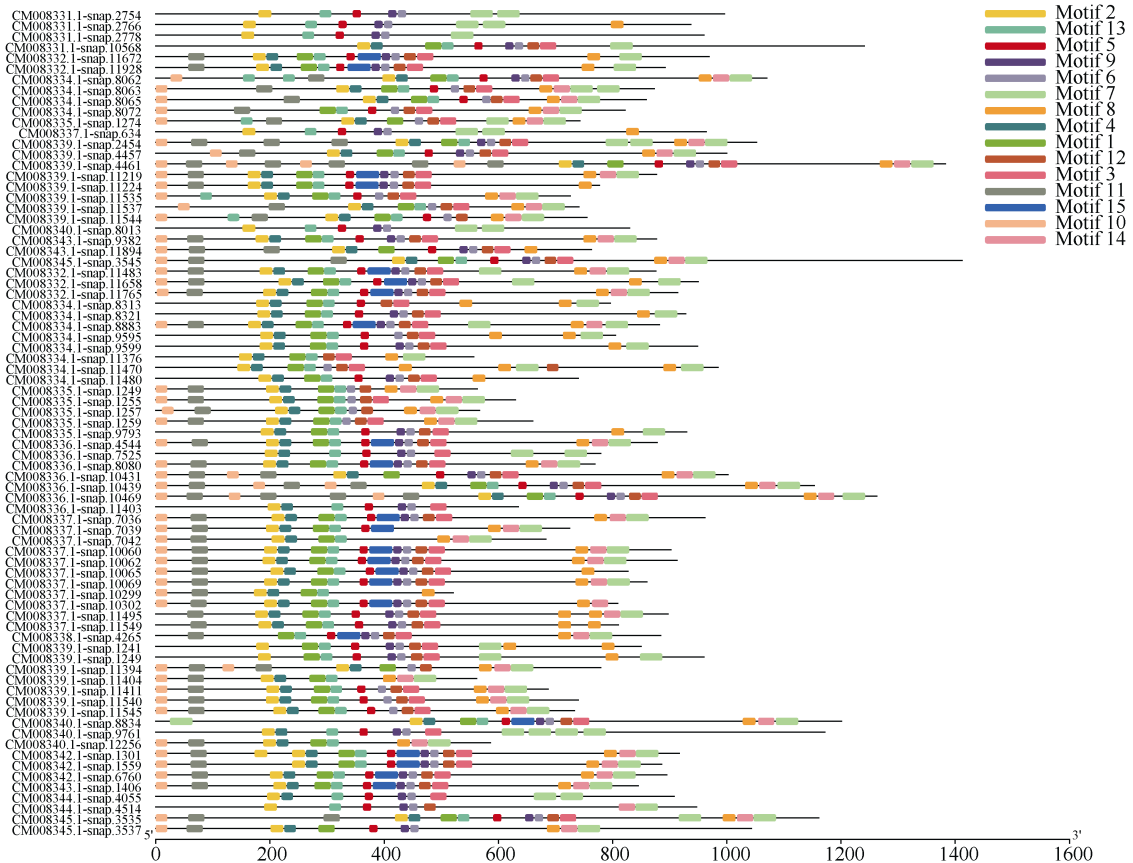


图 2 甘薯 NBS-LRR 家族中 CNL 亚家族保守结构域分布

Fig. 2 Distribution of conservative domains of CNL subfamily of *Ipomoea batatas* NBS-LRR gene family

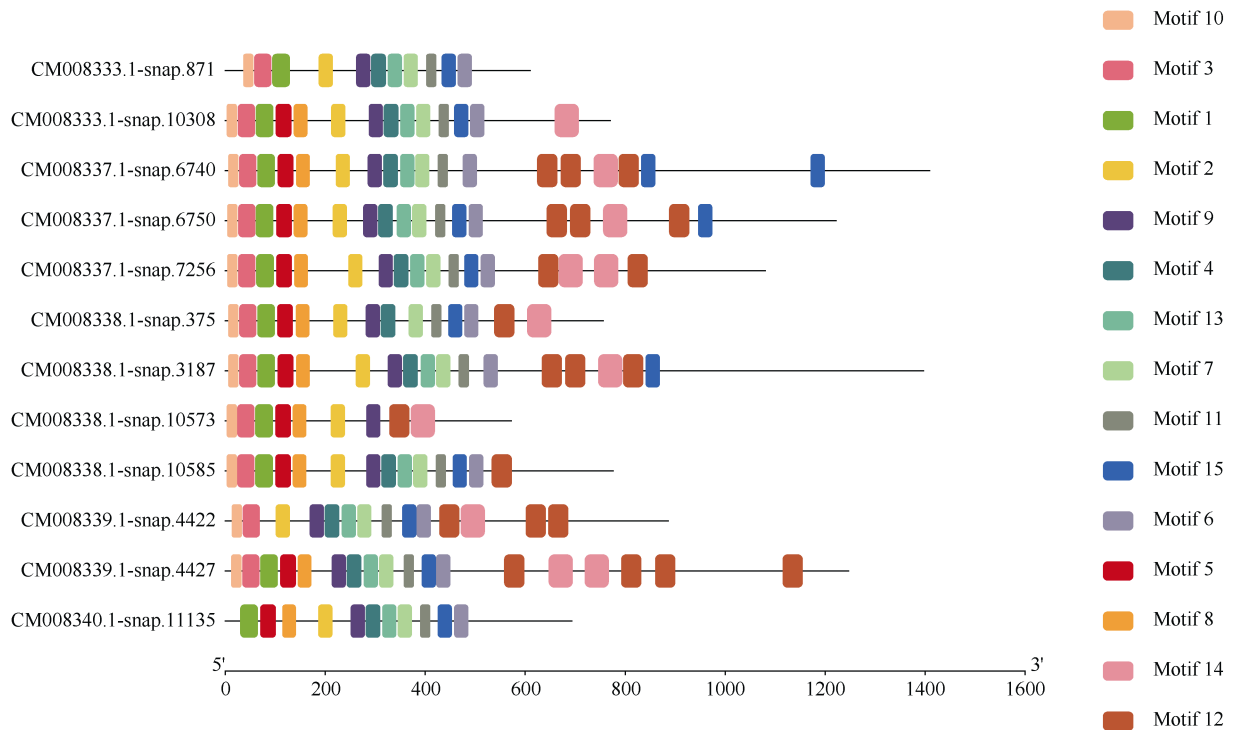


图 3 甘薯 NBS-LRR 家族中 TNL 亚家族保守结构域分布

Fig. 3 Distribution of conservative domain of TNL subfamily of *Ipomoea batatas* NBS-LRR gene family

6-(MHDV)-motif 12-motif 14 或 motif 10(TIR-1)-motif 3-motif 1(TIR-2)-motif 5-(TIR-3)-motif 8(TIR-4)-motif 2(P-loop)-motif 9(Kinase 2)-motif 4(RNBS-B)-motif 13-(RNBS-C)-motif 7(GLPL)-motif 11-motif 6-(MHDV)-motif 12(LRR)-motif 14(LRR)-motif 15。

大部分 motif 15 存在于 motif 11 和 motif 6 之间, 即 MHDV 结构域之前, 仅有 3 个序列 motif 15 位于 motif 14 之后。

2.4 CNL、TNL 亚家族 NB-ARC 结构域保守性分析

植物 NBS-LRR 家族基因中的 NB-ARC 区域相对保守, 通常有 P-loop、RNBS-A、Kinase2、RNBS-B、RNBS-C、GLPL、RNBS-D 和 MHDV 8 个保守基序^[48]。对获得的甘薯 NBS-LRR 家族中 CNL 和 TNL 亚家族序列利用 Clustal Omega 进行多重比对发现, 具有 5 个保守性较高的序列结构域^[44,49-50](图 4), 分别为 P-loop (Kinase 1)、Kinase 2、RNBS-B、GLPL、MHDV。其中 P-loop 最可能氨基酸序列为 SIVGMGGIGKTTLA (画横线部分为氨基酸保守性较强序列), Kinase 2 最可能氨基酸序列为 RYLIVLDDIHW, RNBS-B 最可能的氨基酸序列为 GSRI+LTTRLR, GLPL 最可能的氨基酸序列为 IVSYC+GLPLAIVVVAKRL, MHDV 最可能的氨基

酸序列为 K+TIRMHDLLRDMGR。

根据 MEME 结构域搜索结果, 对 CNL 和 TNL 亚家族基因结构域进行基因定位和识别^[51](图 5), 发现 TNL 亚家族和 CNL 亚家族中分别有 10 个和 7 个保守结构域, 其中 TNL 亚家族 NB-ARC 结构域中保守基序分别为 P-loop、Kinase 2、RNBS-B、RNBS-C、GLPL 和 MHDV, 缺失 RNBS-A 以及 RNBS-D 保守基序; CNL 亚家族 NB-ARC 结构域保守基序分别为 P-loop、RNBS-A、Kinase 2、RNBS-B、GLPL、RNBS-D 和 MHDV, 缺失 RNBS-C 保守基序。TNL 亚家族和 CNL 亚家族中共有的保守基序分别为 P-loop (Kinase 1)、Kinase 2、RNBS-B、GLPL 以及 MHDV, 与多重比对结果相同。此外, 在 TNL 亚家族中 TIR 结构域检测到 4 个保守基序(TIR1-4)^[52]。其中 9 个 NBS-LRR 家族基因含有完整的 TNL 结构域, 占总 TNL 亚家族基因数的 75%, 有 2 个序列同时缺失 TIR 3 和 TIR 4, 有一个序列缺失 TIR 1。

2.5 甘薯 CNL、TNL、RPW8 亚家族系统进化树分析

将甘薯 NBS-LRR 家族基因中 CC/TIR/RPW8/NBS/LRR 结构域的序列提取后, 使用 MEGA X 构建系统进化树。进化树结果中(图 6)有 3 个明显的分支,

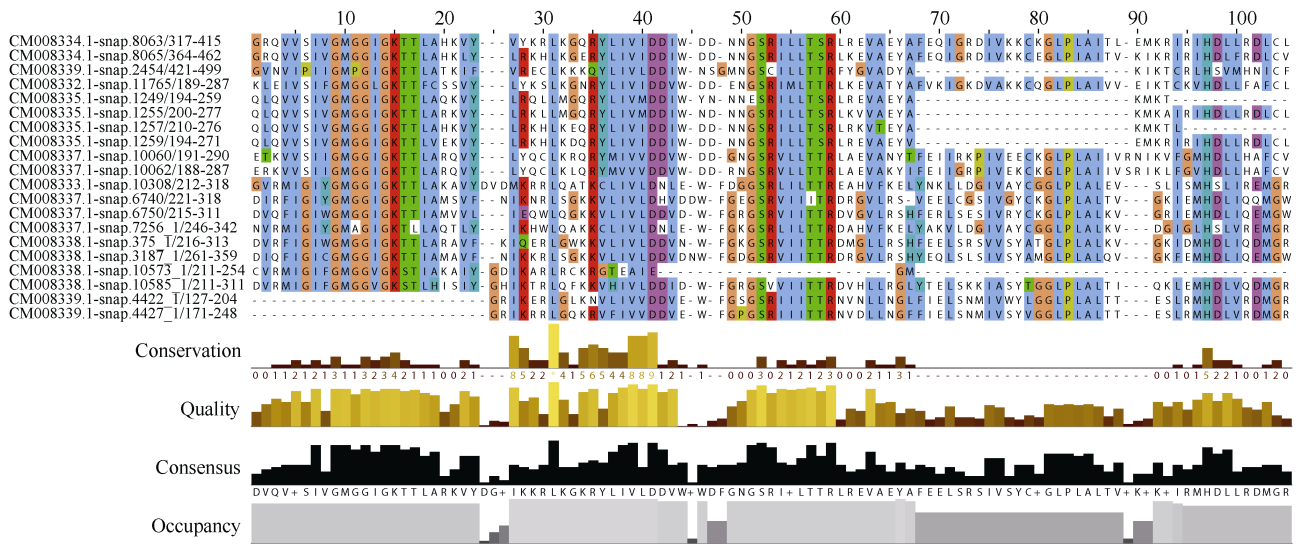


图 4 甘薯 NB-ARC 保守性分析
Fig. 4 Conservation analysis of NB-ARC in *Ipomoea batatas*

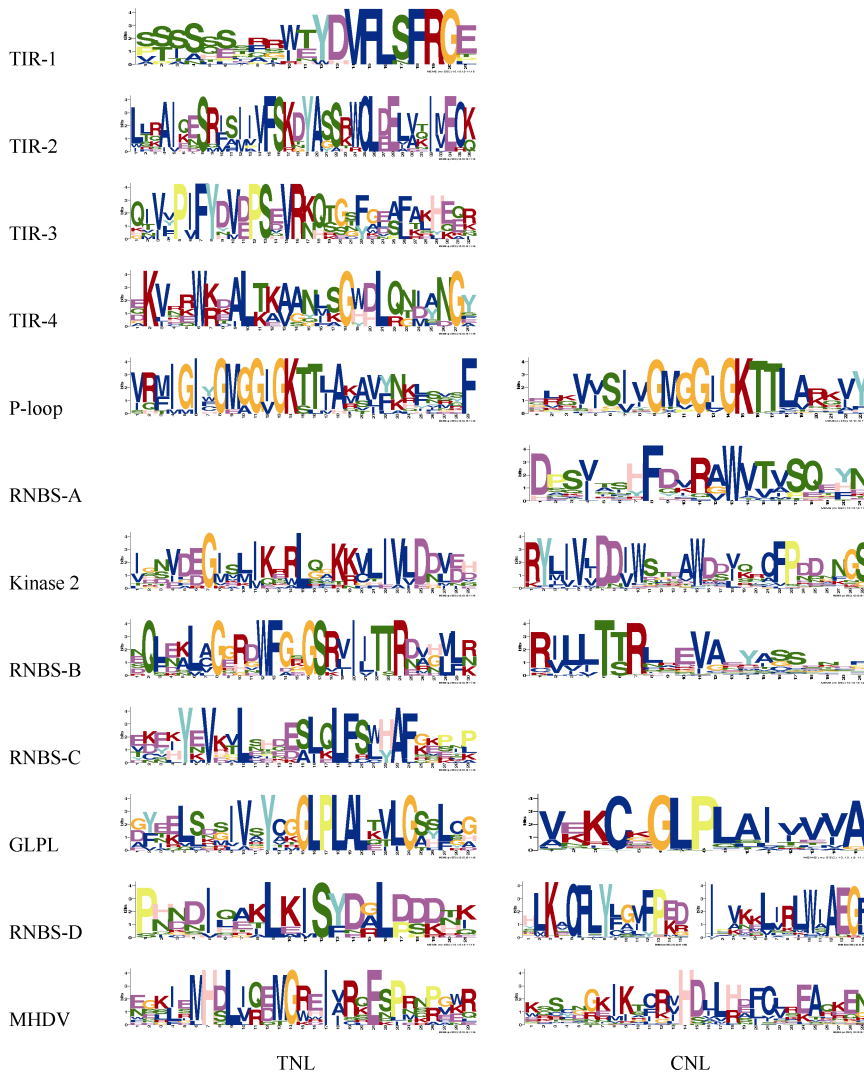


图 5 甘薯 NBS-LRR 家族基因保守结构域及其氨基酸保守性分析
Fig. 5 Conservative domains and amino acids of NBS-LRR family genes in *Ipomoea batatas*

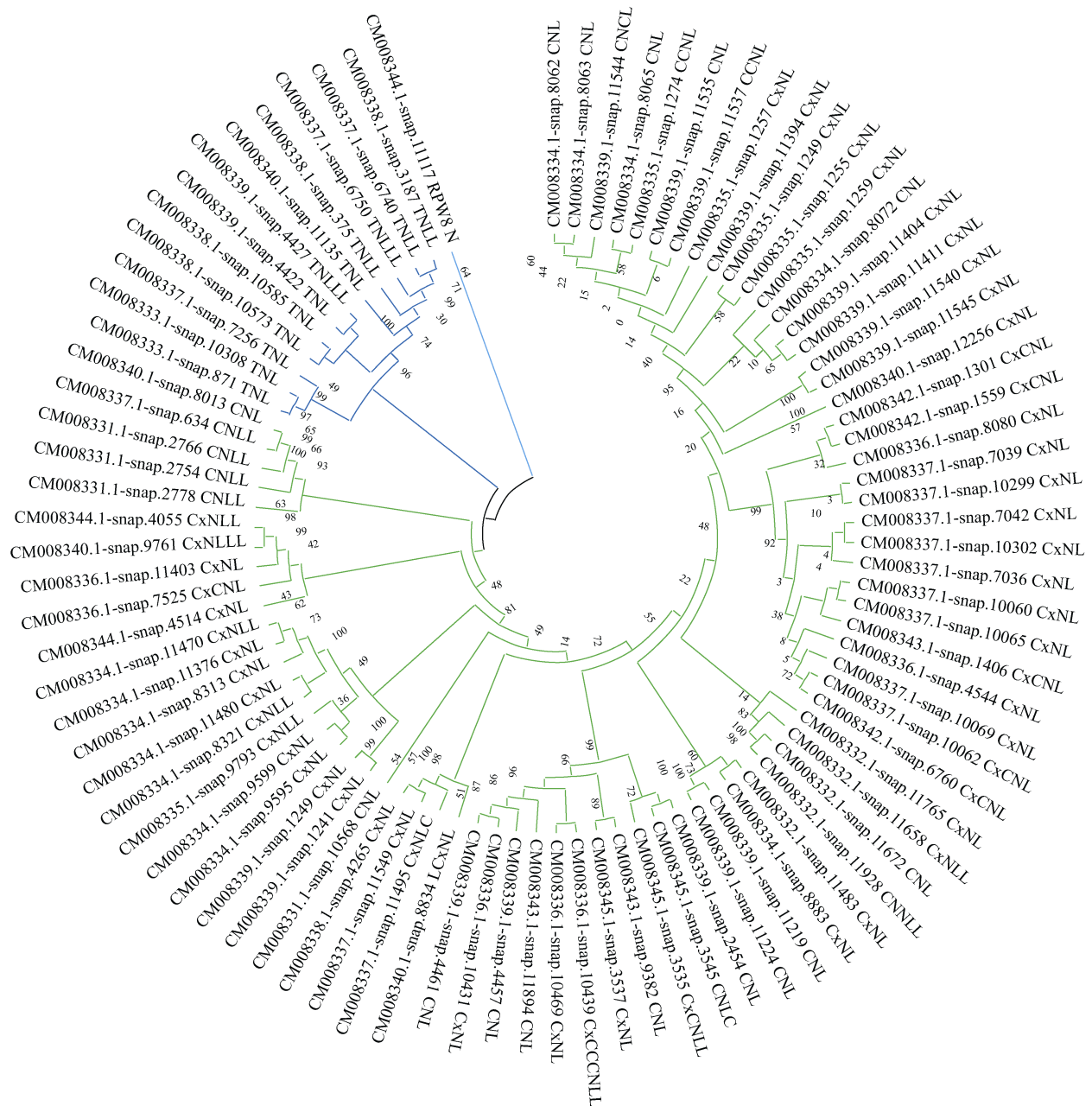


图 6 甘薯 NBS-LRR 家族基因系统进化树

Fig. 6 Phylogenetic tree of NBS-LRR family genes in *Ipomoea batatas*

分别为 CNL 型(绿色)、TNL 型(蓝色)、PN 型(青色)。3 种 NBS-LRR 家族的亚家族存在于不同的进化分支中,但不存在单独的分支,说明这 3 类基因存在相同的进化方式。进化树分支中有些基因分支路径很长,如 CM008340.1-snap12256、CM008331.1-snap10568 等,推测这类基因的“祖先”基因在很早的时候就发生了分化,相应的基因序列也发生了较大的分化。有些基因序列的分支距离较近,如 CM008334.1-snap8062、CM008342.1-snap6760、

CM008334.1-snap11470 等,推测这些基因在短期内发生了复制^[46]。在 CNL 亚家族分支 CC 结构域中,马铃薯 X 病毒类似蛋白^[40]的 CC 结构域(Coiled-coil domain of the potato virus X resistance protein and similar proteins)大致分布于同一小分支,说明马铃薯 X 病毒类似蛋白的 CC 结构域(表 1 中 C_x)与普通 CC 结构域(表 1 中 C)存在进化关系上的差异。

3 讨论

生物信息学方法分析植物抗病基因已被广泛运

用于许多作物, 甘薯是重要的粮食作物, 然而, 甘薯基因组尚未注释, 无法直接查找与分析其基因组的 *R* 基因。本研究应用 snap 软件对甘薯全基因组序列进行了外显子预测, 得到甘薯 CDS 区蛋白质序列, 为甘薯 *R* 基因的搜索及其他相关生物信息学分析提供了序列基础。

本研究利用 HMMER 程序对甘薯全基因组序列进行 NBS-LRR 家族基因检索, 得到 735 个甘薯 NBS-LRR 家族基因。经过筛选、验证、剔除等步骤, 最终得到 379 个序列完整的 NBS-LRR 家族基因, 占甘薯全基因组数目 0.2%, 但与木薯(1.1%)^[42]、水稻(1.4%)^[43]、拟南芥(0.8%)^[44]、乌拉尔图小麦(1.5%)^[45] 相比数量明显偏少。此外, 甘薯 NBS-LRR 家族中 TNL 亚家族基因有 22 个, 与其他作物有较大差别, 如水稻和乌拉尔图小麦中 NBS-LRR 家族基因无 TNL 类型^[43,45], 在其他禾本科和草本的单子叶植物中 TNL 亚家族基因很少^[53-54]。本研究分析了甘薯 NBS-LRR 家族基因的 NB-ARC 结构域保守性, 分别对 CNL 和 TNL 亚家族进行比较, 发现两者在 P-loop、Kinase 2、RNBS-B、GLPL、MHDV 具有较高的相似性。甘薯 P-loop 和 GLPL 基序保守性最高, 其基序分别为 SIVGMGGIGKTTLA 和 IVSYC+GLPLAIVVAKRL, 且 GMGGIGKTT 区域保守性最高。其次为 Kinase 2 和 RNBS-B, Kinase 2 结构域序列为 RYLIVLDDIW, 77 个 CNL 类型中, 有 70 个 Kinase 2 序列末端氨基酸为 W, TNL 类型中末端氨基酸主要为 E 或 D, 未发现 W。研究表明, Kinase 2 序列末端氨基酸 W 是区分 CNL 亚家族和 TNL 亚家族的关键因素^[42,55]。同时, TNL 类型 motif 位置的分布比较保守, 为 TIR(1-4)-P loop-Kinase 2-RNBS B-RNBS C-GLPL-motif 11。CNL 类型 motif 之间位置的排布规律遵循 motif 10-motif 11-P loop-RNBS C-Kinase 2-RNBS B-GLPL 的顺序。

值得注意的是, 本研究所分析的甘薯基因组是基于二代测序组装的单倍体型基因组^[12], 存在数据不完整和一些错误组装的局限性^[56], 本研究结果可能难以完整覆盖甘薯栽培种基因组中 NBS-LRR 家族的所有基因, 在对相关基因进一步深入分析和研究应用时, 还可以结合三浅裂野牵牛(*Ipomoea trifida*)基因组。但是本研究的分析方法和结果, 可为甘薯 NBS-LRR 基因发掘、抗性相关机制研究和抗病育种提供参考。

4 结论

预测了甘薯基因组序列的外显子, 得到甘薯染色体组全基因组蛋白序列, 从中筛选了 379 个 NBS-LRR 家族基因, 并进行了染色体定位, 发现 NBS-LRR 家族基因在不同染色体上的分布数量差异很大, 其中有 60.9% 的 NBS-LRR 基因序列呈簇状分布。NBS-LRR 基因序列有 15 个保守结构域, 在 N 端较为保守, 根据对基因的结构分析构建了系统进化树。研究结果可为甘薯进一步开展 NBS-LRR 家族基因的功能研究和抗性育种提供参考。

References

- [1] Staskawicz B J, Ausubel F M, Baker B J, Ellis J G, Jones J D. Molecular genetics of plant disease resistance. *Science*, 1995, 268: 661-667.
- [2] 李楠洋. 棉花抗黄萎病基因筛选及 NBS-LRR 类抗病基因 *GbaNAI* 功能研究. 中国农业科学院博士学位论文, 北京, 2017.
Li N Y. Screening of Cotton Anti Verticillium Wilt Genes and Functional Study on NBS-LRR Resistance Gene *GbaNAI*. PhD Dissertation of Chinese Academy of Agricultural Sciences, Beijing, China, 2017 (in Chinese with English abstract).
- [3] Changkwan A, Venkatesh J, Lee J H, Han J W, Kwon J K, Siddique M I, Solomon A M, Choi G J, Kim E, Seo Y, Kim Y H, Kang B C. Physical localization of the root-knot nematode (*Meloidogyne incognita*) resistance locus *Me7* in pepper (*Capsicum annuum*). *Front Plant Sci*, 2019, 10: 886.
- [4] 房卫平, 谢德意, 李志芳, 李武, 赵付安, 孙瑶, 段峥峥, 杨晓杰. NBS-LRR 类抗病蛋白介导的植物抗病应答分子机制. 分子植物育种, 2015, 13: 469-474.
Fang W P, Xie D Y, Li Z F, Li W, Zhao F A, Sun Y, Duan Z Z, Yang X J. Molecular mechanism on NBS-LRR proteins-mediated plant disease response. *Mol Plant Breed*, 2015, 13: 469-474 (in Chinese with English abstract).
- [5] Jones J D G, Dangl J L. The plant immune system. *Nature*, 2006, 444: 323-329.
- [6] Hammond-Kosack K E, Parker J E. Deciphering plant-pathogen communication: fresh perspectives for molecular resistance breeding. *Curr Opin Biotechnol*, 2003, 14: 177-193.
- [7] 尹玲, 方辉, 黄羽, 卢江, 曲俊杰. 植物 TIR-NB-LRR 类型抗病基因各结构域的研究进展. 广西植物, 2017, 37: 186-190.
Yin L, Fang H, Huang Y, Lu J, Qu J J. Research progress on domains of plant TIR-NB-LRR resistance genes. *Guihaia*, 2017, 37: 186-190 (in Chinese with English abstract).
- [8] 陆建珍, 汪翔, 秦建军, 戴起伟, 易中懿. 我国甘薯种植业发展状况调查报告(2017 年)——基于国家甘薯产业技术体系产业经济固定观察点数据的分析. 江苏农业科学, 2018, 46(23): 393-398.
Lu J Z, Wang X, Qin J J, Dai Q W, Yi Z Y. Investigation report on the development of sweet potato cultivation in China (2017): Analysis based on the data of industrial economy fixed observation point of the national sweet potato industry technology system.

- Jiangsu Agric Sci*, 2018, 46(23): 393–398 (in Chinese with English abstract).
- [9] 赵永强, 张成玲, 孙厚俊, 徐振, 陈晓宇, 谢逸萍. 甘薯病毒病复合体(SPVD)对甘薯产量的影响. *西南农业学报*, 2012, 25: 909–911.
Zhao Y Q, Zhang C L, Sun H J, Xu Z, Chen X Y, Xie Y P. Effects of viruses (SPVD) on yield of sweet potato. *Southwest China J Agric Sci*, 2012, 25: 909–911 (in Chinese with English abstract).
- [10] 赵冬兰, 唐君, 张安, 周志林, 曹清河, 戴习彬. 甘薯病毒病对不同基因型甘薯产量和品质的影响. *江西农业学报*, 2018, 30(11): 62–65.
Zhao D L, Tang J, Zhang A, Zhou Z L, Cao Q H, Dai X B. Effects of virus diseases on yield and quality of different sweet potato genotypes. *Acta Agric Jiangxi*, 2018, 30(11): 62–65 (in Chinese with English abstract).
- [11] 屈满义, 查向东, 王钰, 杨金环, 蒋琳, 阮龙. 甘薯 NBS-LRR 类抗病基因同源序列的克隆、分析及数目研究. *热带作物学报*, 2008, 29: 610–617.
Qu M Y, Zha X D, Wang Y, Yang J H, Jiang L, Ruan L. Study on cloning, analysis and number of NBS-LRR resistance genes of sweet potato. *Chin J Trop Crops*, 2008, 29: 610–617 (in Chinese with English abstract).
- [12] Yang J, Moeinzadeh M H, Kuhl H, Helmuth J, Xiao P, Haas S, Liu G, Zheng J, Sun Z, Fan W, Deng G, Wang H, Hu F, Zhao S, Fernie A R, Boerno S, Timmermann B, Zhang P, Vingron M. Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nat Plants*, 2017, 3: 696–703.
- [13] Ian K. Gene finding in novel genomes. *BMC Bioinformatics*, 2004, 5: 59.
- [14] Finn R D, Bateman A, Clements J, Coggill P, Eberhardt R Y, Eddy S R, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer E L, Tate J, Punta M. Pfam: the protein families database. *Nucleic Acids Res*, 2014, 42: 222–230.
- [15] Madeira F, Park Y M, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey A R N, Potter S C, Finn R D, Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*, 2019, 47: 636–641.
- [16] Eddy S R. Profile Hidden Markov Models. *Bioinformatics*, 1998, 14: 755–763.
- [17] Pottter S C, Luciani A, Eddy S R, Park Y, Lopez R, Finn R D. Web server issue. *Nucleic Acids Res*, 2018, 46: 200–204.
- [18] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*, 2009, 25: 2078–2079.
- [19] Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011, 27: 2987–2993.
- [20] Brendolise C, Montefiori M, Dinis R, Peeters N, Storey R D, Rikkerink E H. A novel hairpin library-based approach to identify NBS-LRR genes required for effector-triggered hypersensitive response in *Nicotiana benthamiana*. *Plant Methods*, 2017, 13: 32.
- [21] Sievers F, Wilm A, Dineen D, Gibson T J, Karplus K, Li W Z, Lopez R, McWilliam H, Remmert M, Söding J, Thompson J D, Higgins D G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, 2011, 7: 539.
- [22] Marchler-Bauer A, Bryant S H. CD-search: protein domain annotations on the fly. *Nucleic Acids Res*, 2004, 32: 327–331.
- [23] Marchler-Bauer A, Lu S N, Anderson J B, Chitsaz F, Derbyshire M K, De Weese-Scott C, Fong J H, Geer L Y, Geer R C, Gonzales N R, Gwadz M, Hurwitz D I, Jackson J D, Ke Z, Lanczycki C J, Lu F, Marchler G H, Mullokandov M, Omelchenko M V, Robertson C L, Song J S, Thanki N, Yamashita R A, Zhang D C, Zhang N G, Zheng C J, Bryant S H. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res*, 2011, 39: 225–229.
- [24] Marchler-Bauer A, Derbyshire M K, Gonzales N R, Lu S N, Chitsaz F, Geer L Y, Geer R C, He J, Gwadz M, Hurwitz D I, Lanczycki C J, Lu F, Marchler G H, Song J S, Thanki N, Wang Z X, Yamashita R A, Zhang D C, Zheng C J, Bryant S H. CDD: NCBI's conserved domain database. *Nucleic Acids Res*, 2015, 43: 222–226.
- [25] Marchler-Bauer A, Bo Y, Han L Y, He J N, Lanczycki C J, Lu S N, Chitsaz F, Derbyshire M K, Geer R C, Gonzales N R, Gwadz M, Hurwitz D I, Lu F, Marchler G H, Song J S, Thanki N, Wang Z X, Yamashita R A, Zhang D C, Zheng C J, Geer L Y, Bryant S H. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*, 2017, 45: 200–203.
- [26] Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. *Nucleic Acids Res*, 2005, 33: 116–120.
- [27] Barragan C A, Wu R, Kim S T, Xi W Y, Habring A, Hagmann J, Van de Weyer A L, Zaidem M, Ho W W H, Wang G, Bezrukov I, Weigel D, Chae E. RPW8/HR repeats control NLR activation in *Arabidopsis thaliana*. *PLoS Genet*, 2019, 15: e1008313.
- [28] 刘云飞, 万红建, 李志邈, 叶青静, 王荣青, 阮美颖, 姚祝平, 周国治, 韦艳萍, 杨悦俭. 植物 NBS-LRR 抗病基因的结构、功能、进化起源及其应用. *分子植物育种*, 2014, 12: 377–389.
Liu Y F, Wan H J, Li Z M, Ye Q J, Wang R Q, Ruan M Y, Yao Z P, Zhou G Z, Wei Y P, Yang Y J. The structure, function, evolutionary origin and application of plant NBS-LRR resistance genes. *Mol Plant Breed*, 2014, 12: 377–389 (in Chinese with English abstract).
- [29] Lupas A, Dyke M V, Stock J. Predicting coiled coils from protein sequences. *Science*, 1991, 252: 1162–1164.
- [30] McDonnell A V, Jiang T, Keating A E, Berger B. Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics*, 2006, 22: 356–358.
- [31] Jupe F, Pritchard L, Etherington G J, MacKenzie K, Cock P J, Wright F, Sharma S K, Bolser D, Bryan G J, Jones J D, Hein I. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics*, 2012, 13: 75.
- [32] 蒋卉, 张晶, 符真珠, 董晓宇, 王慧娟, 李艳敏, 高杰, 王利民, 张和臣. 蝴蝶兰 NBS-LRR 家族基因挖掘和生物信息学分析. *分子植物育种*, 2018, 16: 2786–2794.
Jiang H, Zhang J, Fu Z Z, Dong X Y, Wang H J, Li Y M, Gao J, Wang L M, Zhang H C. Mining and bioinformatics analysis of NBS-LRR gene family in phalaenopsis. *Mol Plant Breed*, 2018, 16: 2786–2794 (in Chinese with English abstract).
- [33] Bailey T L, Williams N, Misleh C, Li W W. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids*

- Res*, 2006, 34: 369–373.
- [34] Chen C J, Chen H, He Y H, Xia R. TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. *BioRxiv*, 2018. doi: <https://doi.org/10.1101/289660>.
- [35] Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*, 2000, 17: 540–552.
- [36] Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*, 2007, 56: 564–577.
- [37] Waterhouse A M, Procter J B, Martin D M A, Clamp M, Barton G J. Jalview Version 2: a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 2009, 25: 1189–1191.
- [38] Troshin P V, Procter J B, Barton G J. Java bioinformatics analysis web services for multiple sequence alignment—JABAWS: MSA. *Bioinformatics*, 2011, 27: 2001–2002.
- [39] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*, 2018, 35: 1547–1549.
- [40] Hao W, Collier S M, Moffett P, Chai J. Structural basis for the interaction between the potato virus X resistance protein (Rx) and its cofactor Ran GTPase-activating protein 2 (RanGAP2). *J Biol Chem*, 2013, 288: 35868–35876.
- [41] Tarr D E, Alexander H M. TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Res Notes*, 2009, 2: 197.
- [42] Lozano R, Hamblin M T, Prochnik S, Jannink J L. Identification and distribution of the NBS-LRR gene family in the Cassava genome. *BMC Genomics*, 2015, 16: 360.
- [43] Zhou T, Wang Y, Chen J Q, Araki H, Jing Z, Jiang K, Shen J, Tian D. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol Genet Genomics*, 2004, 271: 402–415.
- [44] Meyers B C, Kozik A, Griego A, Kuang H, Michelmore R W. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell*, 2003, 15: 809–834.
- [45] 刘小芳, 袁欣, 聂迎彬, 张晶. 乌拉尔图小麦 NBS-LRR 家族生物信息学分析. 分子植物育种, 2018, 16: 7587–7597. Liu X F, Yuan X, Nie Y B, Zhang J. Bioinformatics Analysis of NBS-LRR gene family in *Triticum urartu*. *Mol Plant Breed*, 2018, 16: 7587–7597 (in Chinese with English abstract).
- [46] 王岩, 李兆阳, 唐心龙, 卢姗, 许鹏, 张静, 方奎, 席景会. 拟南芥基因组 NBS-LRR 类基因家族的生物信息学分析. 中国农学通报, 2009, 25(15): 40–45. Wang Y, Li Z Y, Tang X L, Lu S, Xu P, Zhang J, Fang K, Xi J H. Bioinformatic analysis of the NBS-LRR gene family in *Arabidopsis thaliana*. *Chin Agric Sci Bull*, 2009, 25(15): 40–45 (in Chinese with English abstract).
- [47] Holub E B. The arms race is ancient history in Arabidopsis, the wildflower. *Nat Rev Genet*, 2001, 2: 516–527.
- [48] Liu Z C, Xie J M, Wang H P, Zhong X H, Li H L, Yu J H, Kang J G. Identification and expression profiling analysis of NBS-LRR genes involved in *Fusarium oxysporum* f. sp. *conglutinans* resistance in cabbage. *3 Biotech*, 2019, 9: 202.
- [49] Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers B C, Boerjan W, Martin F. Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Mol Biol*, 2008, 66: 619–636.
- [50] Ahlenstiel G, Lozano R, Ponce O, Ramirez M, Mostajo N, Orjeda G. Genome-wide identification and mapping of NBS-encoding resistance genes in *Solanum tuberosum* group phureja. *PLoS One*, 2012, 7: e34775.
- [51] Li T G, Wang B L, Yin C M, Zhang D D, Wang D, Song J, Zhou L, Kong Z Q, Klosterman S J, Li, J J, Adamu S, Liu T L, Subbarao K V, Chen J Y, Dai X F. The *Gossypium hirsutum* TIR-NBS-LRR gene *GhDSC1* mediates resistance against verticillium wilt. *Mol Plant Pathol*, 2019, 20: 857–876.
- [52] Meyers B C, Morgante M, Michelmore R W. TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. *Plant J*, 2002, 32: 77–92.
- [53] Wu L, Hickson I D. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature*, 2003, 426: 870–874.
- [54] Lakatos L, Csorba T, Pantaleo V, Chapman E J, Carrington J C, Liu Y P, Dolja V V, Calvino L F, López-Moya J J, Burguán J. Small RNA binding is a common strategy to suppress RNA silencing by several viral suppressors. *EMBO J*, 2006, 25: 2768–2780.
- [55] López C E, Zuluaga A P, Cooke R, Delseny M, Tohme J, Verdú V. Isolation of resistance gene candidates (RGCs) and characterization of an RGC cluster in cassava. *Mol Genet Genomics*, 2003, 269: 658–671.
- [56] Wu S, Lau K H, Cao Q H, Hamilton J P, Sun H H, Zhou C X, Eserman L A, Gemenet D C, Olukolu B A, Wang H Y, Crisovan E, Godden G T, Jiao C, Wang X, Kitavi M, Anrique-Carpintero N, Vaillancourt B, Wiegert-Rininger K, Yang X S, Bao K, Schaff J, Wolfgang J K, Gruneberg A K, Ghislain M, Ma D, Jiang J M, Mwangi R O M, Leebens-Mack J, Lachlan J M, Coin G, Yencho G C, Buell C R, Fei Z J. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. *Nat Commun*, 2018, 9: 4580.