

番木瓜赖氨酸基序类受体激酶 LysM-RLK 基因家族鉴定及生物信息学分析

郭盼^{1,2,3}, 孔华^{2,3*}, 王雨^{2,3}, 戴云素⁴, 贾瑞宗^{2,3}, 周瑶^{2,3}, 郭安平^{2,3}, 纪长绵^{2,3*}, 马春花^{1*}

1. 云南农业大学园林园艺学院, 云南昆明 650201; 2. 中国热带农业科学院三亚研究院, 海南三亚 572025; 3. 中国热带农业科学院热带生物技术研究所/海南省南繁生物安全与分子育种重点实验室, 海南海口 571101; 4. 海南大学热带作物学院, 海南海口 570228

摘要: 本研究采用生物信息学方法在番木瓜全基因组范围系统鉴定了 7 个 LysM-RLK 家族基因, 并解析了其染色体分布特征、基因结构特征、蛋白理化性质和亚细胞定位信息。基因家族进化分析表明番木瓜的 7 个 LysM-RLK 基因可划分为 3 个亚组 (Lym-II, Lym-III, Lym-IV), 并分布于 4 条染色体上。基因复制是番木瓜 LysM-RLK 家族进化的主要驱动力, 基因复制分析鉴定得到 7 对复制基因。启动子序列调控元件分析发现其包含多个光响应元件及参与低温应答和逆境相关激素信号应答的元件。大规模比较转录组分析系统地解析了 LysM-RLK 在番木瓜生物胁迫、非生物胁迫条件下及不同组织中的基因表达模式, 揭示其特异性的表达模式及潜在的生物学功能。复制基因对的表达分化、结构域组成差异和蛋白质三级结构的变化, 暗示其在进化过程中的亚功能化现象。生物和非生物胁迫条件下的核心共表达网络分析鉴定了 27 个与 LysM-RLK3 强关联的核心基因, 揭示了 LysM-RLK3 潜在的转录调控模式。相关结果为番木瓜 LysM-RLK 家族起源演化、功能分化及其在逆境响应和生长发育过程中的角色提供重要依据。

关键词: 番木瓜; LysM-RLK 家族; 生物信息分析; 基因表达; 蛋白三维结构

中图分类号: S667.9 文献标识码: A

Identification and Bioinformatics Analysis of Lysine-like Receptor Kinase LysM-RLK Gene Family Based on the Whole Genome of Papaya

GUO Pan^{1,2,3}, KONG Hua^{2,3*}, WANG Yu^{2,3}, DAI Yunsu⁴, JIA Ruizong^{2,3}, ZHOU Yao^{2,3}, GUO Anping^{2,3}, JI Changmian^{2,3*}, MA Chunhua^{1*}

1. College of Landscape Architecture and Horticulture, Yunnan Agricultural University, Kunming, Yunnan 650201, China; 2. Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences, Sanya, Hainan 572025, China; 3. Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences / Hainan Key Laboratory for Biosafety Monitoring and Molecular Breeding in Off-Season Reproduction Regions, Haikou, Hainan 571101, China; 4. College of Tropical Crops, Hainan University, Haikou, Hainan 570228, China

Abstract: In this study, seven LysM-RLK subfamily genes were systematically identified in the whole genome of papaya using bioinformatics methods, and the chromosome distribution, gene structure, protein physical and chemical properties, and subcellular localization information were analyzed. Familial evolutionary analysis showed that papaya

收稿日期 2022-12-16; 修回日期 2023-01-20

基金项目 海南省自然科学基金项目 (No. 322QN418); 海南省重大科技计划项目 (No. ZDKJ202002); 中国热带农业科学院基本科研业务费专项资金 (No. 1630052021017)。

作者简介 郭盼 (1997—), 女, 硕士研究生, 研究方向: 果树学。*通信作者 (Corresponding author): 马春花 (MA Chunhua), E-mail: pony0207@126.com; 纪长绵 (JI Changmian), E-mail: jichangmian@126.com; 孔华 (KONG Hua), E-mail: konghua@itbb.org.cn。

LysM-RLK could be further divided into three subgroups (Lym-II, Lym-III, Lym-IV), distributed on four chromosomes. Recent gene replication is the main driver of the evolution of the papaya LysM-RLK subfamily. The analysis of promoter sequence regulatory elements found that it contained multiple photoresponse elements and elements involved in low temperature response and stress related hormone signal response. Large-scale comparative transcriptome analysis systematically analyzed the gene expression patterns of LysM-RLK under papaya biostress and abiotic stress and different tissues, revealing its specific expression patterns and potential biological functions. The expression differentiation of replicating gene pairs, differences in domain composition, and changes in the tertiary structure of proteins suggest subfunctionalization in the process of evolution. Core co-expression network analysis under biotic and abiotic stress conditions identified 27 core genes strongly associated with LysM-RLK3, revealing the potential transcriptional regulatory patterns of LysM-RLK3. The results would provide important insights into the evolution, functional differentiation and role of the papaya LysM-RLK subfamily in adversity response and growth and development.

Keywords: papaya; LysM-RLK subfamily; bioinformatics analysis; gene expression; protein tertiary structure

DOI: 10.3969/j.issn.1000-2561.2023.05.001

大多数蛋白质编码基因在植物基因组中以多拷贝或基因家族的形式存在^[1-3]。例如,水稻和拟南芥分别有 41 908 和 31 135 个基因,分别包含 13 055 和 10 193 个基因家族^[1]。ZHANG 等^[2]研究发现一个基因家族基因成员的数量不仅在属内物种之间存在差异,而且在一个物种内也存在变化,基因家族成员数量的变化与生物性状的遗传变异显著相关。因此,系统鉴定并解析基因家族进化关系是理解目标基因家族参与植物生物学过程的基础。类受体蛋白激酶家族 (receptor-like kinase, RLK) 是植物基因组中一个大基因家族,前期研究主要集中于基因家族的结构和系统发育分析^[4],较少涉及解析该家族成员的功能和活性。第一个植物 RLK 基因在玉米 (*Zea mays* L.) 中被鉴定^[5]。在双子叶植物模式物种拟南芥 (*Arabidopsis thaliana*) 中,至少鉴定了 610 个 RLK 基因,约占其蛋白质编码基因的 2.5%^[4]。在单子叶植物模式物种水稻 (*Oryza sativa* L.) 中,RLK 家族由 1130 个基因组成,其成员数量是拟南芥的 1.85 倍^[6]。

RLK 基因已被证明参与包括植物生长发育、激素信号感知、抗病和抗逆等在内的生物学过程^[7]。赖氨酸基序受体激酶 (LysM-RLK) 是植物中重要的 RLK 亚家族之一,其成员属于包含赖氨酸基序 (LysM) 的蛋白质,是识别微生物相关分子模式 (microbe-associated molecular patterns, MAMPs) 的模式识别受体 (pattern recognition receptors, PPR)^[8-10]。LysM 首次在芽孢杆菌噬菌体溶菌酶中发现之后,在来自粪肠球菌 (*Enterococcus faecalis*) 的名为肽聚糖水解酶的酶中亦发现了 LysM^[11-12]。植物细胞膜上有一个识别受体,其包含 LysM 结构域,能够与真菌细胞壁上的几丁

质结合并传递信号以激活免疫反应,这一类蛋白根据亚细胞定位的预测和结构域的差异,可以分为 LysM 类受体激酶 (LysM-containing receptor-like kinases, LYKs)、LysM 类受体蛋白 (LysM-containing receptor-like protein, LYPs)、细胞外 LysM 蛋白 (extracellular LysM, LysMe) 及非分泌型胞内 LysM 蛋白 (intracellular non-secretory LysM genes, LysMn),不同物种的单个蛋白含 LysM 结构域的数量有所不同,这是长期进化导致的差异,且与蛋白的功能相关^[13]。随着测序技术的快速发展和相关组学数据的积累,已证实 LysM 结构域蛋白广泛存在于动植物生物体基因组中^[14-15]。

番木瓜 (*Carica papaya* L.) 是热带特色水果,具有很高的营养价值和药用价值,属于热区重要经济作物之一。番木瓜基因组小 (350.3 Mb)、常年开花结果、种子多、生长快、代际短及遗传转化体系成熟,是果树研究重要模式材料^[16]。已有研究报道表明 LysM-RLK 家族参与植物抗病和胁迫响应^[17],然而其在番木瓜全基因组水平基础特征信息、进化关系及基因表达模式研究的不足严重阻碍了其在番木瓜新品种选育中的开发利用。本研究采用生物信息学方法在全基因组水平系统鉴定了番木瓜 LysM 家族基因,并分析了其家族成员的进化关系、染色体分布、基因蛋白和结构特征等。基于已有的 84 份代表性转录组样品数据,系统解析了番木瓜 LysM 家族成员在不同组织、不同发育过程和不同胁迫处理条件下的转录模式及表达分化。相关研究为理解番木瓜 LysM 家族起源演化,复制基因表达分化及其在不同生物学过程中转录表达模式提供重要理论基础,将促进其在番木瓜抗病育种中的研究和应用。

1 材料与方法

1.1 材料

从 Pfam 数据库下载 LysM 种子文件, 用 hmmer search 方法筛选番木瓜和葡萄 (*Vitis vinifera* L.) 基因组中对应的同源基因作为候选基因。

1.2 方法

1.2.1 番木瓜 LysM-RLK 成员鉴定及理化性质与亚细胞定位分析 利用 NCBI Conserved Domain Search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) 数据库, 基于保守结构域、在线分析工具 Pfam (<http://pfam.xfam.org/>) 和 SMART (<http://smart.emblheidelberg.de/>)^[18]预测保守域^[19]。利用 ExpASy (<http://expasy.org/>) 在线软件对从番木瓜基因组中鉴定到的 7 条候选类受体蛋白激酶家族的蛋白序列进行分子量、理论等电点预测。利用 SMART (<http://smart.embl-heidelberg.de/>) 在线软件对所有基因保守结构域的位置进行预测。进一步利用在线工具 WoLFPSORT (<http://wolfpsort.hgc.jp>) 对番木瓜 LysM-RLK 基因进行亚细胞定位预测。

1.2.2 番木瓜 LysM-RLK 家族蛋白的基因结构、结构域分析及蛋白质三级结构预测 使用 TBtools 软件^[20]对基因结构进行可视化。利用 MEME (<http://meme-suite.org/>) 在线软件基于相似度较高的基序 (motif), 分析番木瓜 LysM-RLK 基因家族成员蛋白序列保守特征及特征展示。使用 SMART (<https://smart.embl.de/>) 在线软件进行 LysM-RLK 基因结构域预测。使用人工智能网络算法工具 AlphaFold2^[21]预测番木瓜 LysM-RLK 家族蛋白三维结构, 获得蛋白三维结构 PDB 文件。将获得的 PDB 文件输入到 PyMOL V2.3^[22]软件包中, 对蛋白三维结构进行可视化展示。

1.2.3 番木瓜 LysM-RLK 蛋白系统发育分析及 LysM-RLK 基因染色体定位 为了阐明番木瓜 (*Carica papaya*)、拟南芥 (*Arabidopsis thaliana*) 和葡萄 (*Vitis vinifera* L.) 中的 LysM-RLK 家族成员之间的系统发育关系, 构建 LysM-RLK 家族进化树。采用 muscle v3.8.31 软件对来自番木瓜 (7 个 LysM-RLKs)、拟南芥 (4 个 LysM-RLKs) 和葡萄 (11 个 LysM-RLKs) 的 22 个 LysM-RLK 蛋白质序列进行多重序列比对。使用 BMGE1.12 软件, 选用参数为默认参数, 在多重序列比对中选择适合系统发育推断的高质量比对区域。最后, 利用 IQ-TREE (V2.1.2) 软件^[23]构建 LysM-RLK

亚基因家族的最大似然 (maximum-likelihood, ML) 系统发育树。

根据番木瓜基因组注释文件, 提取番木瓜 LysM-RLK 基因在染色体上的位置信息, 利用在线程序 (<http://mg2c.iask.in/mg2c v2.1/>) 展示其在番木瓜基因组上位置分布特征。

1.2.4 复制基因鉴定及选择压力 (Ka/Ks) 分析及复制基因共线性鉴定 使用 Dup Gen finder^[24]流程鉴定番木瓜 LysM-RLK 基因家族的复制基因对及复制类型。进一步计算复制基因对的 Ka/Ks 值, 评估其在进化过程中面临的选择压力。首先利用 Muscle (<https://www.ebi.ac.uk/Tools/msa/muscle/>) 在线软件对每一个复制基因对蛋白序列进行全局比对。然后, 利用 ParaAT (V2.0) 软件^[25]将比对序列反向翻译成对应的 cds 序列。最后, 利用 KaKs calculator 软件^[26]计算非同义替换率 (Ka)、同义替换率 (Ks) 和 Ka/Ks^[24]值, 使用多重序列共线性搜索工具 MCscan 搜索番木瓜全基因组的同源共线性片段。

1.2.5 番木瓜 LysM-RLK 家族启动子分析 根据番木瓜基因组注释 gff 文件, 提取得到番木瓜 LysM 基因翻译起始位点上游 2000 bp 的序列作为候选启动子区域。通过在线软件 PlantCARE (<http://bioinformatic.psb.ugent.be/webtools/plantcare/html/>) 分析 LysM-RLK 家族启动子顺式作用元件。

1.2.6 LysM-RLK 基因在生物与非生物胁迫条件下的表达模式分析 番木瓜转录组数据收集。从 NCBI 数据库中下载 84 份与番木瓜相关转录组数据, 其中非生物胁迫相关的转录组测序数据 29 条、不同组织转录组测序数据 15 条, 生物胁迫相关转录组数据 40 条, 共组成 43 组转录组数据, 具体信息见附表 S1。

基因表达定量。转录原始测序数据经过 Fastp 软件包质控以后, 用 HISAT2 软件^[27]将高质量的转录组测序数据比对到番木瓜参考基因组得到序列比对文件。然后采用 StringTie 软件^[28]组装比对上的 reads, 构建转录本序列, 并采用 FPKM 算法对 read count 做归一化处理计算基因和转录本的表达量。

差异表达及表达分化基因鉴定。最后利用 DESeq2 软件^[29]鉴定差异表达基因: 差异倍数 FC \geq 2, 差异显著性 FDR > 0.05。利用 R 语言的 heatmap 函数绘制 LysM-RLK 基因表达量热图。同一组样

品中复制基因对之间表达量的比值 ≥ 2 , 定义为表达显著分化的复制基因对。

1.2.7 番木瓜 LysM-RLK 家族蛋白的基因结构、结构域分析及蛋白质三级结构预测 使用 TBtools^[20]软件对基因结构进行可视化。利用 MEME (<http://meme-suite.org/>) 在线软件基于相似度较高的基序, 即 motif, 分析番木瓜 LysM-RLK 基因家族成员蛋白序列保守特征及特征展示。使用在线软件 SMART (<https://smart.embl.de/>) 进行 LysM-RLK 基因结构域预测。使用人工智能网络算法工具 AlphaFold2^[21]预测番木瓜 LysM-RLK 家族蛋白三维结构, 获得蛋白三维结构 PDB 文件。将获得的 PDB 文件输入到 PyMOL V2.3^[22]软件包中, 对蛋白三维结构进行可视化展示。

1.2.8 番木瓜 LysM-RLK 家族加权基因共表达网络分析 通过炭疽病、番木瓜畸形花叶病毒 (*Papaya leaf-distortion mosaic virus*, PLDMV)、禾生素 (CTS)、干旱等处理的转录组数据筛选差异表达基因 (DEGs)。处理组与对照组差异基因鉴定参数为 $|\log_2 \text{FC}(\text{FPKM})| \geq 1$, $\text{FDR} \leq 0.05$ 。共筛选出 10 490 个差异基因。采用 R 语言的 WGCNA 包构建 LysM-RLK 家族基因的加权共表达网络。分析鉴定共表达模块, 在模块中鉴定基因共表达差异表达基因作为核心基因 (hub-genes), 并利用 Cytoscape 软件对核心基因的共表达网络 (hub-network) 可视化。

2 结果与分析

2.1 LysM-RLK 家族基因鉴定及特征分析

利用 5 个拟南芥的 LysM-RLK 家族基因作为参考^[4], 将其比对到番木瓜和葡萄的蛋白序列, 根据序列相似性筛选到 LysM-RLK 候选同源基因 (identify $\geq 45\%$)。如果其编码区同时存在 N 末端 LysM 基序、1 个跨膜结构域和 C 末端 Pkinase 保守结构域的经典的 LysM-RLK 结构域模式, 则认为其属于 LysM-RLK 基因家族。本研究在番木瓜、拟南芥和葡萄基因组中分别鉴定到 7 个、4 个和 11 个 LysM-RLK 家族基因。为获得番木瓜 LysM-RLK 家族蛋白的理化特征信息, 利用生物信息学方法解析了番木瓜的 7 个 LysM-RLK 基因家族成员的蛋白长度、分子量、理论等电点和亚细胞定位。结果表明其氨基酸序列大小为 445~692, 分子量为 49.85~76.10 kDa, 等电点为 5.12~8.69 (表 1)。利用在线软件 WoLFPSORT 预测番木瓜 7 个 LysM-RLK 基因的亚细胞定位信息, 结果发现这 7 个家族成员定位于不同亚细胞区域。其中 1 个基因 (*CpLYK5*) 位于细胞质膜, 1 个基因 (*CpLYK5-like*) 位于胞质, 1 个基因位于 (*CpLYK2*) 胞外区域, 其余 4 个基因 (*CpLYK4-1*、*CpLYK3*、*CpLysM-RLK3*、*CpLYK4*) 定位于叶绿体 (表 1)。该分析结果暗示不同的 LysM-RLK 基产物可能被转运到番木瓜不同的亚细胞区域, 进而参与不同的生物学过程。

表 1 番木瓜 RLK 基因家族 LysM-RLK 成员理化性质

Tab. 1 Physicochemical properties of LysM-RLK family genes in papaya

基因 ID Gene ID	基因名称 Gene name	基因长度 Gene length/bp	氨基酸数量 Amino acids	分子量 Molecular mass/Da	等电点 pI	亚细胞定位 Subcellular localization
Cpa02g024430	<i>CpLYK5-like</i>	2109	445	49 853.81	8.40	胞质
Cpa04g004680	<i>CpLYK2</i>	6865	686	76 096.56	6.86	胞外
Cpa04g020240	<i>CpLYK5</i>	2079	692	76 089.45	6.22	质膜
Cpa04g023090	<i>CpLYK4-1</i>	1995	447	49 470.89	8.69	叶绿体
Cpa07g009890	<i>CpLYK3</i>	53 476	604	66 869.80	7.55	叶绿体
Cpa07g011660	<i>CpLYK4</i>	2465	626	69 005.18	5.78	叶绿体
Cpa09g008160	<i>CpLysM-RLK3</i>	9312	605	67 371.71	5.12	叶绿体

2.2 LysM-RLK 家族的基因结构特征及保守结构域分析

外显子的获得或丢失可能引起基因对应的蛋白质结构域结构和蛋白质功能的变化^[30]。因而, 本研究比较了 LysM-RLK 家族成员之间的基因结构特征。结果发现 LysM-RLK 基因主要由 1~11 个外显

子构成, 其中 *CpLYK5* 属于单外显子基因, *CpLYK3* 基因长度达到 53 476 kb, 包含 11 个较小的外显子。*CpLYK3* 存在多个大内含子插入的现象, 导致其基因长度明显大于该家族的其他基因 (图 1A)。外显子数量的增加可以增加 mRNA 剪接模式多样性, 并导致包括 DNA 结合亲和力^[31], 蛋白质溶解度^[32]

和蛋白质功能^[30]等改变，内含子的插入能够增强目标基因的表达并影响转录本（mRNA）的衰变或定位^[30, 33]。以上证据暗示内含子插入和外显子数量差异可能会引起番木瓜 LysM-RLK 家族不同成员之间的结构差异和功能分化。

基因家族中共有的基序（motif）可能是该基因家族保守结构域序列，暗示其是基因家族基础

功能核心序列。利用 MEME（<http://meme-suite.org/>）在线工具对番木瓜的 LysM-RLK 基因家族成员进行保守结构域鉴定（图 1B，表 2）。结果表明番木瓜 LysM-RLK 蛋白的 motif 长度最小为 15，最大为 50。在 MEME 鉴定得到的 10 个 motif 中，motif2 的保守性得分最高，并且按照一致的顺序排布（图 1B）。

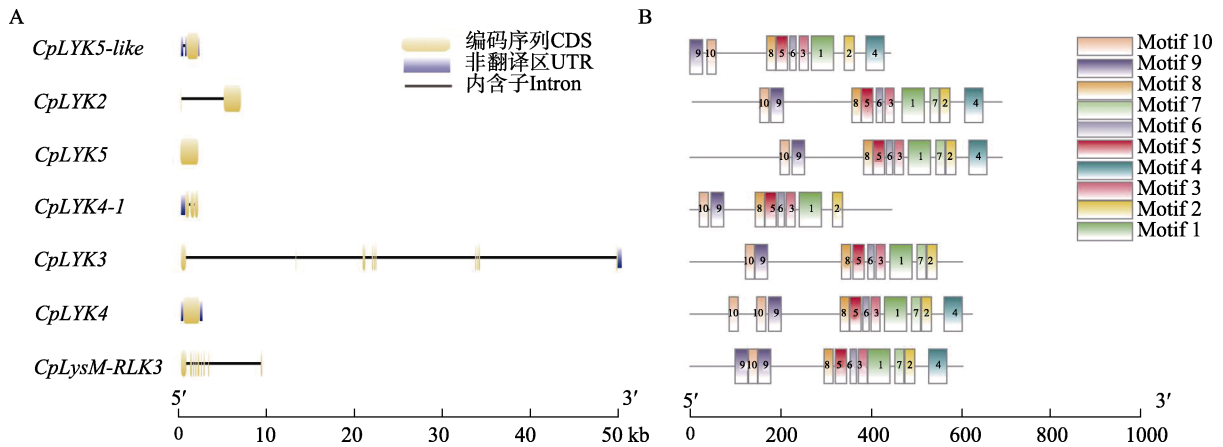


图 1 番木瓜 LysM-RLK 基因家族的基因结构 (A) 和蛋白保守结构域分析 (B)

Fig. 1 Gene structure (A) and motif analysis (B) of LysM-RLK gene family on papaya genome

表 2 番木瓜 LysM-RLK 蛋白的预测 motif

Tab. 2 Sequences of motif consensus of LysM-RLK protein

序号 No.	名称 Name	序列 Sequence	E 值 E-value
1	Motif 1	WKQRVQIALDAARGLEYJHEHTKPPYVHRBIKSSNILLDSBFRAKIABFG	4.60E-599
2	Motif 2	GLVTPKLDVYAFGVVLELJSGK	6.60E-215
3	Motif 3	GSLYLVEYEAZNGSLSDWLHG	1.60E-184
4	Motif 4	EKLREFIDPCLGDNYPLDLAYKMAQLAKSCVEEDPNLRPSM	1.60E-243
5	Motif 5	GSVYYGLJRGDEAAIKMKGEKSKEF	5.40E-182
6	Motif 6	TKIHHSNLVRLSGYC	6.20E-129
7	Motif 7	DLALTRHVVGTFGYLAPEYLE	3.10E-128
8	Motif 8	VYSYEELRKATDNFSES NKIG	2.70E-121
9	Motif 9	VKYLVTYPVRPGDSVSSJAERFGVSTESI	1.10E-127
10	Motif 10	RNLRVGARLLVPLRCACPTGK	5.60E-73

2.3 LysM-RLK 基因家族进化关系及染色体位置分布特征

为了揭示 LysM-RLK 基因家族成员的进化关系，本研究利用极大似然法构建了葡萄、番木瓜和拟南芥的 LysM-RLK 基因家族系统发育树（图 2）。结果表明番木瓜的 LysM-RLK 基因家族成员分成 3 个独立分支（LysM- II~LysM-IV），其中分支 LysM- II 中含有 2 个番木瓜 LysM-RLK 基因，分支 LysM- III 中含有 2 个番木瓜 LysM-RLK 基因，分支 LysM-IV 中含有 3 个番木瓜 LysM-RLK

基因，暗示其在进化过程中产生了功能分化。

解析 LysM-RLK 在番木瓜基因组上分布特征，结果表明番木瓜的 7 个 LysM-RLK 家族基因主要分布在 chr02、Chr04、chr07 和 chr09 染色体上（图 3）。CpLYK5-like 分布在 2 号染色体的远端粒区（chr02: 37 470 447~37 472 556 bp）。CpLYK2（chr04: 4 182 588~4 189 453 bp）、CpLYK5（chr04: 32 118 333~32 120 412 bp）、CpLYK4-1（chr04: 34 635 116~34 637 111 bp）分别位于 4 号染色体远端粒区域。CpLYK3（chr07: 9 452 026~9 505 502 bp）、

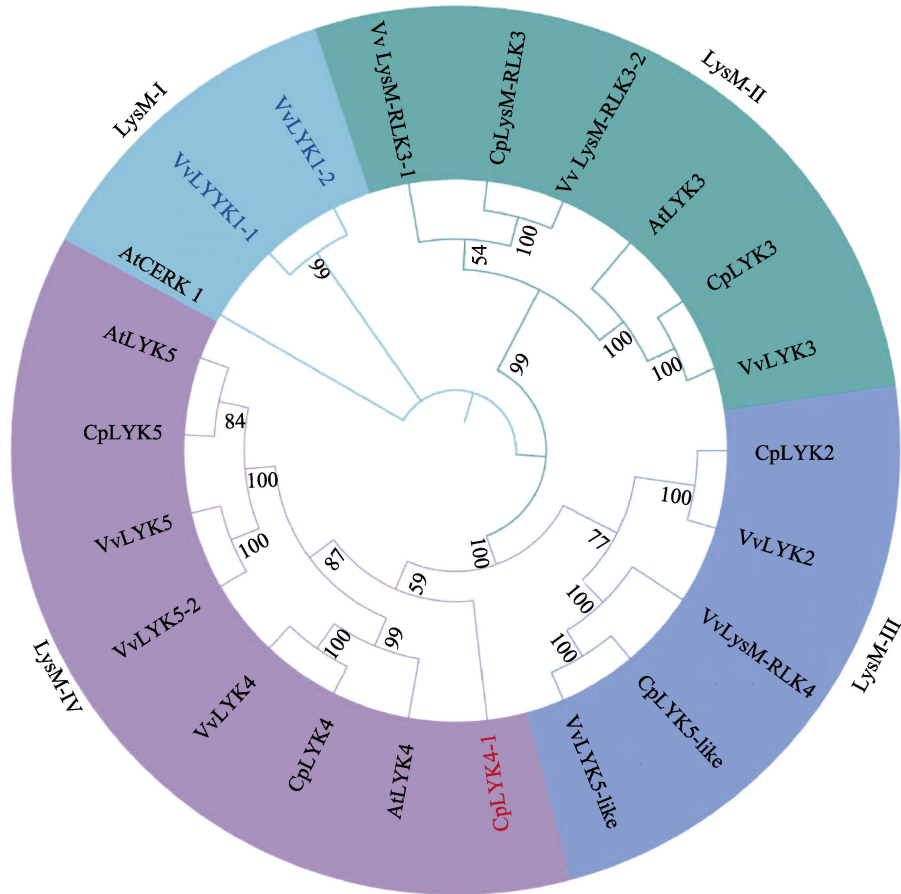


图 2 番木瓜、拟南芥和葡萄的 LysM-RLK 基因家族进化关系

Fig. 2 Phylogenetic relationship of LysM-RLK subfamilies in *C. papaya* L., *A. thaliana* and *V. vinifera* L.

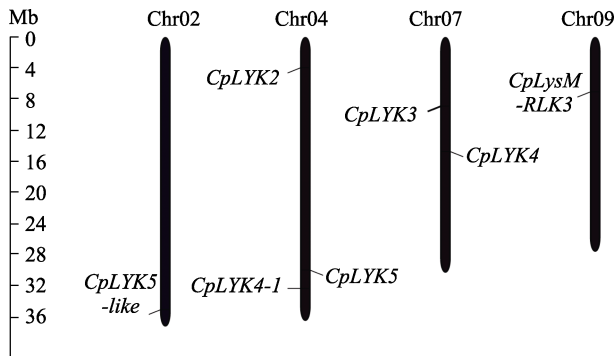


图 3 番木瓜 LysM-RLK 基因染色体分布特征

Fig. 3 Chromosomal distribution of LysM-RLK genes in *C. papaya*

CpLYK4 (chr07: 15 723 626~15 726 091 bp) 分布在 7 号染色体的远端粒区域。*CpLysM-RLK3* (chr09: 7 478 070~7 487 382 bp) 分布于 9 号染色体的远端粒区域。

2.4 早期基因复制事件驱动番木瓜 LysM-RLK 家族的形成

已有研究表明基因复制是导致植物基因家族扩张的重要进化驱动力，促进基因家族内部同源

基因之间功能分化，为植物的遗传多样性提供重要支撑^[33]。基因复制分析发现番木瓜 LysM-RLK 家族中 5 个基因均来源于 dispersed 类型复制 (表 3)，表明基因复制对其家族成员形成具有重要贡献。番木瓜与葡萄、拟南芥一样经历了大约 1.2 亿年前的一次真双子叶植物共有的祖先全基因组三倍化事件 (γ -WGT)^[34-35]。然而，本研究未发现由全基因组复制事件 (whole genome duplication, WGD) 对番木瓜 LysM-RLK 基因家族的贡献 (表 3)，暗示 LysM-RLK 基因家族的基因复制可能发生在 γ -WGT 事件之后。因而，本研究进一步评估了基因复制发生的时间 (表 3)。结果发现番木瓜 LysM-RLK 家族的基因发生复制的时间大约在 1.83 百万年~3.00 百万年之间，远远晚于真双子叶 γ -WGT 事件发生时间。推测祖先基因组三倍化事件 (γ -WGT) 产生的 LysM-RLK 复制基因 (wgd-type duplicated genes) 可能在番木瓜基因组进化过程中被丢失^[33, 35]，而番木瓜基因组中新的基因复制促进了番木瓜 LysM-RLK 家族的形成。

表 3 番木瓜 LysM-RLK 基因家族基因对复制类型
Tab. 3 Duplicated gene pairs of LysM-RLK family on papaya genome

序号 No.	复制基因对 Duplicated gene pairs	基因复制类型 Duplicated type	复制时间 (百万年) Duplicated time (Ma)
1	CpLYK5-like/CpLYK2	dispersed	1.83
2	CpLYK5-like/CpLYK5	dispersed	3.00
3	CpLYK5-like/CpLYK4-1	dispersed	2.87
4	CpLYK2/CpLYK5	dispersed	3.17
5	CpLYK5/CpLYK4-1	dispersed	3.01
6	CpLYK4-1/CpLYK2	dispersed	2.83
7	CpLysM-RLK3/CpLYK5	dispersed	2.83

为了明确重复基因的位点关系，在番木瓜基因组上用红线将 LysM-RLK 复制基因对相连（图 4）。结果发现 4 号染色体上含有 3 个基因参与复制，2 号染色体和 9 号染色体上都含有 1 条基因参与复制。在番木瓜的 7 个 LysM-RLK 基因中有 5 个基因被鉴定为基因复制来源基因。LysM-RLK 基因家族中，位于 2 号染色体的基因 *CpLYK5-like* 是 LysM-RLK 基因家族中最早出现基因复制的基因，并分化出基因 *CpLYK2* 和基因 *CpLYK4-1* 两

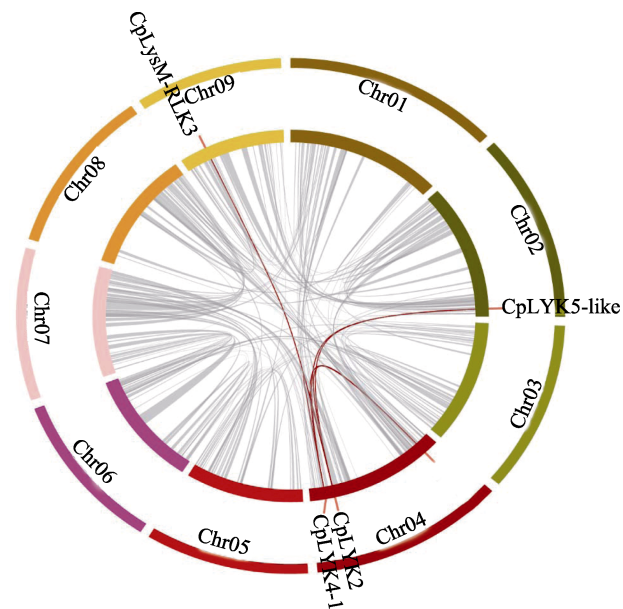


图 4 番木瓜 LysM-RLK 家族成员复制起源
Fig. 4 Putative origin of LysM-RLK family genes on papaya genome

个基因，推测 4 号染色体的 *CpLYK2*、*CpLYK4-1* 可能来自 2 号染色体 *CpLYK5-like* 内部的基因复制事件。*CpLysM-RLK3* 由 *CpLYK5* 复制得到，推测 9 号染色体的 *CpLysM-RLK3* 可能来自 4 号染色体 *CpLYK5* 的基因复制事件。

2.5 番木瓜 LysM-RLK 家族基因启动子区域的顺式作用元件预测

本研究将番木瓜 LysM-RLK 基因 5'UTR 上游 2000 bp 的序列作为启动子序列。通过 PlantCARE 软件分析启动子区域的顺式作用元件。发现番木瓜 LysM-RLK 基因启动子区域的包括组织特异性元件、非生物胁迫相关元件、激素响应元件和功能相关元件 4 类顺式作用元件（图 5）。其中 *CpLYK4* 和 *CpLysM-RLK3* 含有参与低温响应顺式元件。*CpLYK5*、*CpLYK4-1* 和 *CpLYK4* 含有参与水杨酸反应的顺式作用元件。部分基因的启动子序列中包含包括多种激素响应、低温应答及逆境响应等生物学过程相关的顺式作用元件。上述结果暗示 LysM-RLK 基因可能参与调控番木瓜的抗逆防御和生长发育等重要生物学过程。

2.6 LysM-RLK 基因在番木瓜不同组织和生物学过程中表达模式

为了研究番木瓜中 LysM-RLK 家族潜在的生物学功能，本研究系统地解析其在番木瓜生物胁迫、非生物胁迫条件下及不同组织中的基因表达模式（图 6A）。通过对 84 个转录组数据进行表达定量分析，发现番木瓜 LysM-RLK 基因家族成员明显地分成高表达和低表达 2 个模式。*CpLYK4-1* 和 *CpLYK5-like* 基因整体表现出低表达的趋势，而 *CpLYK2*、*CpLysM-RLK3*、*CpLYK4*、*CpLYK5* 和 *CpLYK3* 倾向于高表达。值得注意的是 *CpLYK4-1* 和 *CpLYK4* 在不同病毒和炭疽病侵染条件下（ANT, AIN1, PAPMV, SIN2, G20, Y61）显著差异表达，并且相对于对照组其表达量明显上调，暗示它们在番木瓜抗病响应中具有重要生物学功能。相对于其他组织，*CpLYK5-like* 在正常的根组织中特异性高表达，并且在干旱胁迫条件下表达量明显下调，表明其在根组织中负调控干旱胁迫。转录组数据分析表明，干旱胁迫过程（0~20 d）中，*CpLYK4* 在叶片组织表达量逐渐上调，在根组织表达量逐渐下调。同时，*CpLYK4* 在番木瓜炭疽病侵染过程中高表达量，并且随着侵染时间延长（0、24、48 h）表达量逐渐上调。以上结果表明 *CpLYK4* 同时参与响应叶片干旱胁迫和炭疽病

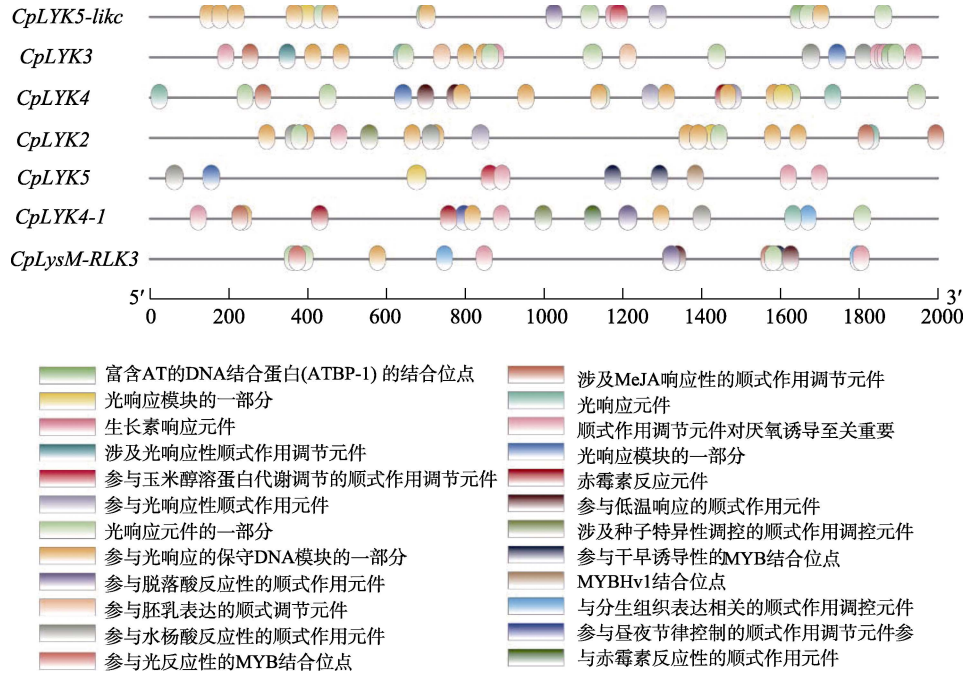
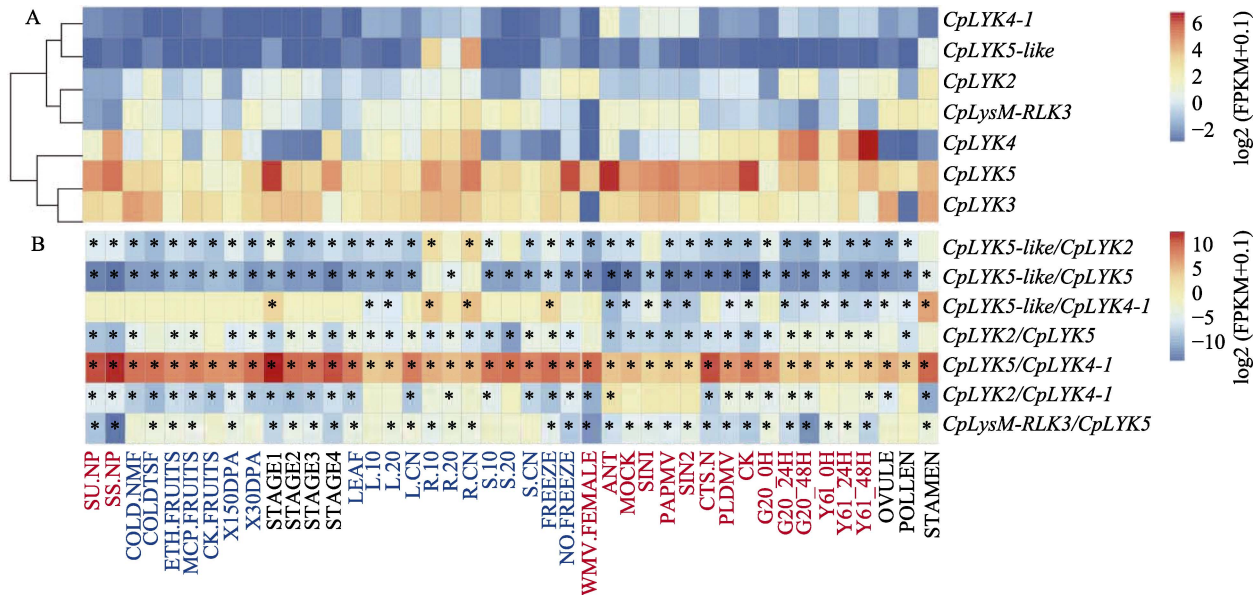


图 5 番木瓜 LysM-RLK 基因家族启动子分析
Fig. 5 Papaya LysM-RLK gene family promoter analysis



A: LysM-RLK 基因在不同组织、不同发育过程、生物胁迫和非生物胁迫条件下的表达量热图; B: LysM-RLK 复制基因对在不同组织、不同发育过程、生物胁迫和非生物胁迫条件下基因表达分化现象。红色样品名称表示生物胁迫, 蓝色样品名称表示非生物胁迫, 黑色样品名称表示不同组织, 详细信息见附表 S1。

A: Expression profile of LysM-RLK gene family among different tissues, under biotic- and abiotic-stress; B: Expression divergence of duplicated gene pairs of LysM-RLK among different tissues, under biotic- and abiotic- stress. Red sample names indicate biotic stresses, blue sample names indicate abiotic stresses, black sample names indicate different tissues, Detailed information is given in Tab S1.

图 6 番木瓜 LysM-RLK 基因潜在生物学功能

Fig. 6 Exploration of functional roles for LysM-RLK genes

侵染过程。*CpLYK5* 和 *CpLYK3* 两个基因在不同组织、不同发育过程、生物胁迫和非生物胁迫条件下整体呈现高表达现象, 暗示它们的功能对番木瓜生长发育和胁迫响应至关重要。*CpLYK3* 在

WMV 病毒侵染和花粉组织中不表达, 表明其不参与 WMV 病毒侵染过程和花粉发育。

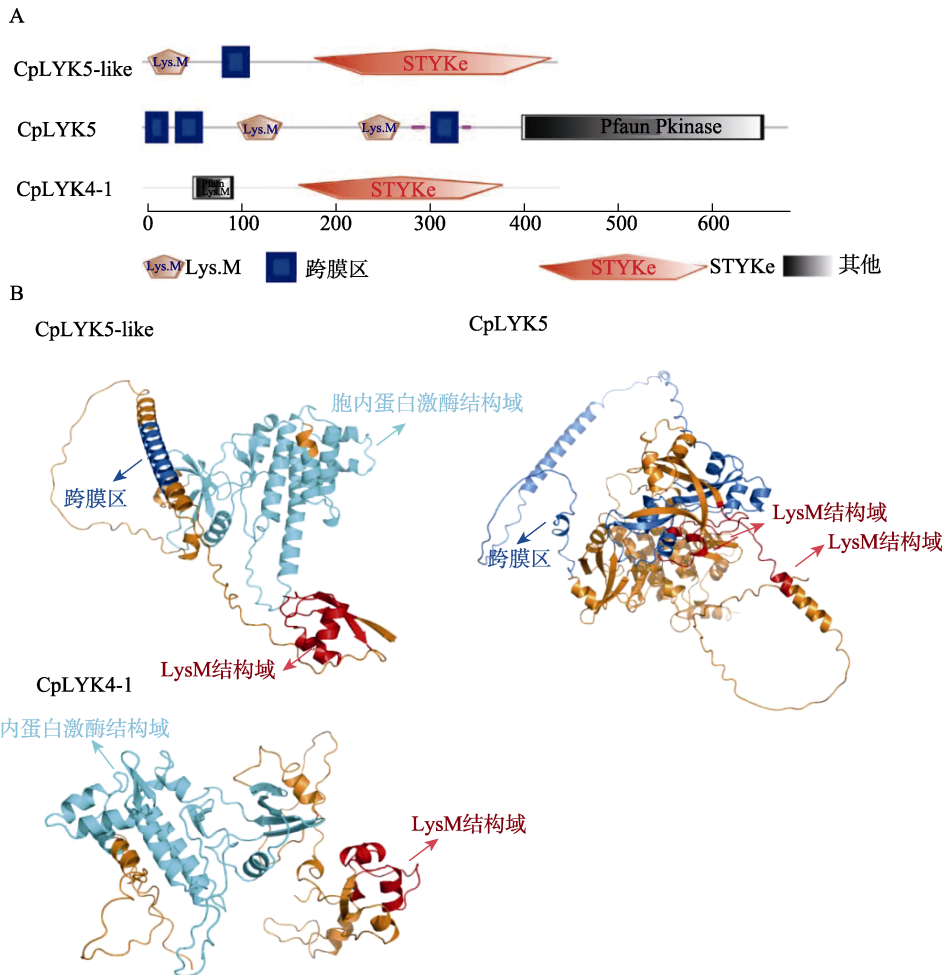
为了解析复制基因之间功能差异, 本研究重点关注鉴定得到的 7 对复制基因的表达情况 (图

6B)。发现 *CpLYK5-like/CpLYK2*、*CpLYK5-like/CpLYK4-1*、*CpLYK2/CpLYK5*、*CpLYK4-1/CpLYK2* 和 *CpLysM-RLK3/CpLYK5* 五对复制基因在不同组织、生物胁迫和非生物胁迫条件下表现出明显的基因表达量分化现象。其中约 3 百万年前产生的复制基因 *CpLYK5-like*、*CpLYK5* 和 *CpLYK4-1* 之间出现明显的表达量分化。*CpLYK5* 在不同组织、生物胁迫和非生物胁迫条件下显著高表达，而其复制基因 *CpLYK5-like* 和 *CpLYK4-1* 则表现出低表达量，暗示复制基因在进化过程中出现了功能分化现象。

2.7 LysM-RLK 家族的功能结构域和蛋白序列三级结构解析

根据表达分析结果，推测复制基因对 *CpLY*

K5-like/CpLYK5 和 *CpLYK5-like/CpLYK4-1* 之间表达分化可能引起功能变化。因此对 *CpLYK5-like*、*CpLYK5* 和 *CpLYK4-1* 三个基因的蛋白功能结构域和蛋白质三维结构进行了比较分析。结构域分析结果发现 3 个 LysM-RLK 基因都含有典型胞外 LysM 结构域，*CpLYK5-like* 和 *CpLYK5* 分别包含 1 个跨膜结构域和 1 个胞内蛋白激酶结构域，属于典型的 LysM 型受体激酶，而 *CpLYK4-1* 跨膜结构域缺失（图 7A）。为了解析这 3 个复制基因之间蛋白空间结构变化，本研究利用人工智能网络算法 AlphaFold2 进行了蛋白三维结构预测（图 7B）。结果发现 LysM 结构域和跨膜区均存在 α -螺旋结构且基因 *CpLYK5-like* 的 LysM 结构域包含 β -折叠，然而它们的蛋白三维空间结构已经表现



A: LysM-RLK 保守结构域图；B: AlphaFold2 预测的蛋白三维结构信息；LysM 结构域序列标记为红色，跨膜区记为蓝色，胞内蛋白激酶结构域标记为浅蓝色。

A: Conserved domain diagram; B: Predicted protein three-dimensional structures of them; The LysM domain is marked in red, the transmembrane region is marked in blue, and the intracellular protein kinase domain is marked in light blue.

图 7 LysM-RLK1、LysM-RLK3 和 LysM-RLK4 蛋白 LysM-RLK 保守结构域与三维结构信息

Fig. 7 Conserved domains and three-dimensional structural information among LysM-RLK1, LysM-RLK3, and LysM-RLK4 proteins

出明显差异。结合其在不同转录组中的基因表达分化, 本研究认为番木瓜复制基因 *CpLYK5-like*、*CpLYK5* 和 *CpLYK4-1* 在长期进化过程中已经表现明显的基因功能分化。

2.8 番木瓜基因 *LysM-RLK3* 加权基因共表达网络分析 (WGCNA)

为了进一步揭示 3 个复制基因(*CpLYK5-like*、*CpLYK5* 和 *CpLYK4-1*) 的转录调控网络, 本研究对基于 4 个不同胁迫条件下[炭疽病、番木瓜畸形花叶病毒 (PLDMV)、禾生素 (CTS)、干旱等]的转录组数据构建的差异表达基因的共表达网络发现, *CpLYK5* (Cpa04g020240) 位于 *plum1* 共表达模块。选取模块内连接度最高的前 27 个基因作为该模块的核心基因, 构建了核心共表达网络。基因表达热图 (图 8A) 表明, 这些核心基因在不同胁迫条件下表达模式出现明显的表达差异, 尤其是在炭疽病侵染过程中表达量显著下调, 而在 PLDMV 病毒侵染过程中表达量明显上调, 暗示这些核心基因协同参与不同生物和非生物胁迫响应。Cpa01g017940、Cpa02g003800 和 Cpa04g021650 分别为编码膜结合蛋白苏氨酸激酶 (*CpP HOT2*)、糖转运蛋白 (*CpSWEET1*)、3-酮基-CoA 合成酶 6 (*CpKCS6*) 的基因, 其在炭疽病侵染的果实和干旱胁迫的茎组织中表达量下调, 而在番木瓜畸形花叶病毒 (PLDMV) 和禾生素 (CTS) 处理叶片组织中表达量上调。*CpLysM-RLK3* 是拟南芥 *AtLYK5* 的同源基因, 属于主要的几丁质受体, 可与几丁质激发子受体激酶 (*AtCERK1*) 结合形成几丁质介导的复合物, 诱导植物免疫。在核心共表达网络中 (图 8B), *CpLysM-RLK3* 与膜结合蛋白苏氨酸激酶 (*CpPHOT2*)、糖转运蛋白 (*CpSWEET1*)、3-酮基-CoA 合成酶 6 (*CpKCS6*) 等核心基因表现出强共表达关系, 暗示这些核心基因通过相互之间的转录调控协同响应植物不同逆境胁迫。*CpLYK5-like*、*CpLYK5* 和 *CpLYK4-1* 不在同一个核心共表达网络, 为复制基因之间的功能分化提供新的证据, 其具体功能差异需要后续进行更加深入的研究。

3 讨论

LysM 型类受体蛋白激酶基因在植物生长发育及逆境胁迫中发挥重要功能, 已成为当今植物中基因功能研究的热点之一^[36]。本研究采用生物信息学方法对番木瓜基因组中的 *LysM-RLK* 家族

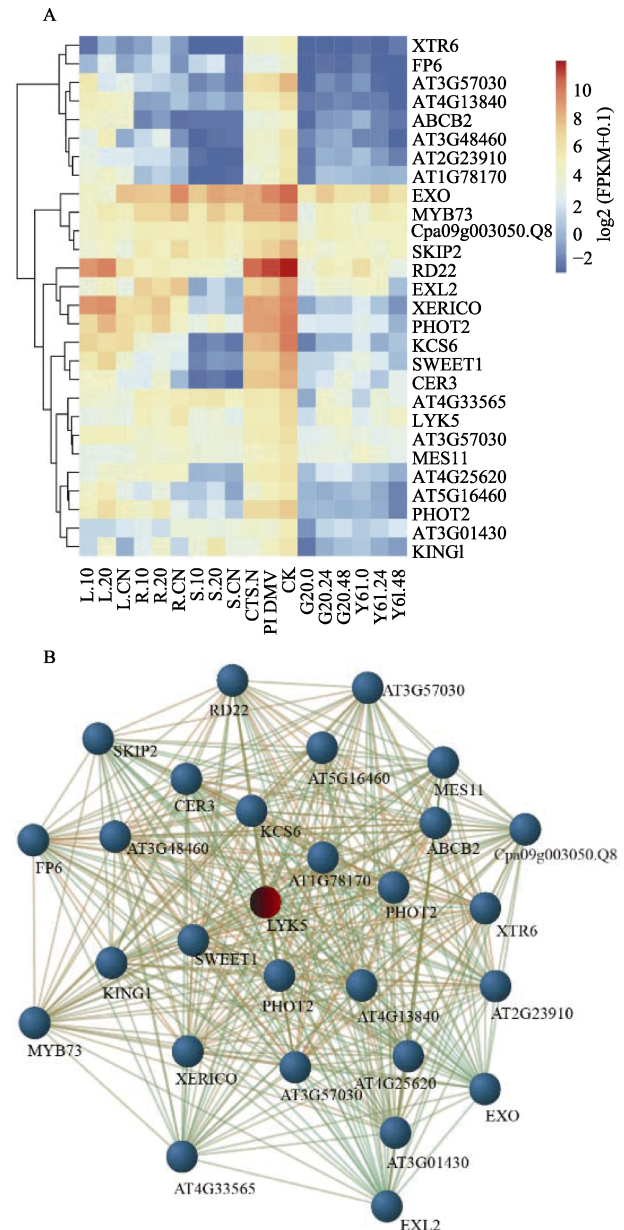


图 8 WGCNA 模块 *plum1* 核心基因表达量热图和 *plum1* 模块核心共表达网络

Fig. 8 Expression heatmap of hub-genes in *plum1* module of WGCNA and hub-network of hub-genes in *plum1* module

基因进行系统鉴定、分类及理化特征评估。同时比较基因组学角度来阐明番木瓜 *LysM-RLK* 家族形成的演化历史, 明确近期的基因复制对其家族形成重要贡献。

本研究鉴定了番木瓜 7 个 *LysM-RLK* 家族基因, *LysM-RLK* 家族成员数多于拟南芥 (4 个), 小于葡萄 (11 个)。高梅等^[36]研究的小立碗藓 (*Physcomitrium patens*) 是早期登陆的植物代表, 其 *LysM-RLK* 家族成员数量多于拟南芥和水稻, 推

测该家族成员在小立碗藓中可能存在基因功能的冗余。随着不断进化发展,一些高等植物中 LysM-RLK 家族成员数量得到扩展,意味着该家族新功能的引入。

通过系统进化分析,将番木瓜 LysM-RLK 基因分为 3 个亚组(LysM-II、LysM-III 和 LysM-IV),它们均与拟南芥和葡萄的部分成员聚为一类,而高等植物中 LysM-RLK 家族分为 4 个亚组,推测番木瓜中 LysM-RLK 家族成员的功能相对单一^[37]。

顺式作用调控元件作为一个重要的分子开关,如王东东等^[37]发现大多数 LysM-RLK 基因的启动子序列中富含多种激素相关的作用元件,如水杨酸(salicylic acid, SA)响应元件 TCA-element (S)、茉莉酸甲酯(methyl jasmonate, MeJA)响应元件 CGTCA-motif(M)等,但不同基因之间的元件含量和种类均不相同,表明 LysM-RLK 家族不同成员能对不同激素发生响应且响应程度不同^[38]。本文预测的番木瓜顺式作用调控元件表明 LysM-RLK 参与各种环境因素(干旱、低温等)、光响应和植物激素[生长素(auxin, IAA)、脱落酸(abscisic acid, ABA)、水杨酸(salicylic acid, SA)、茉莉酸甲酯、赤霉素(gibberellins, GAs)等],与前人研究一致。

植物胁迫反应机制的一个潜在创新来源是基因复制,ZOU 等^[38]对复制基因参与应激调节的启动子顺式作用 DNA 调节元件的分析表明,祖先应激反应的不对称分配可能部分归因于 DNA 调控元件的差异;失去大部分顺式作用元件的重复基因不参与应激反应,但可能参与其他反应。因此,失去顺式作用元件的复制基因可能形成新功能而保留下来,拟南芥重复基因之间的顺式调节元件部分解释了为什么重复基因对胁迫有不同的应答反应。

左存武等^[39]发现 LysM-RLK 在苹果腐烂病发生前后,发生差异表达的基因较多推测部分苹果 LysM-RLK 基因可能在部分苹果真菌病害发生的过程中起重要的感受和调控,秦春晓^[40]研究表明星星草(*Borago officinalis*) Put LysM 基因可能参与种子萌发与根的盐应答调控过程,表明 LysM-RLK 与抗病抗逆相关,并参与组织发育。番木瓜基因表达谱分析表明 LysM-RLK 广泛参与生物胁迫(肯尼亚不知名病毒、PapMV 和 PRSV 协同处理、PRSV 和 PapMV 协同处理、PapMV、

PLDMV、炭疽病)、非生物胁迫(低温胁迫、乙烯处理、1-甲基环丙烯、干旱胁迫、冻融处理)和不同组织发育(胚珠、花粉、雄蕾、叶、根、茎、雌花、雄花和不同发育时期的果实)。

本研究在转录组水平上挖掘番木瓜抗病和抗逆相关的 LysM-RLK 基因,揭示 LysM-RLK 基因参与番木瓜抗病和抗逆反应。为番木瓜的遗传改良和育种提供了必要的 LysM-RLK 基因资源。

参考文献

- [1] SCHNABLE P S, WARE D, FULTON R S, STEIN J C, WEI F, PASTERNAK S, LIANG C, ZHANG J, FULTON L, GRAVES T A, MINX P, REILY A D, COURTNEY L, KRUCHOWSKI S S, TOMLINSON C, STRONG C, DELEHAUNTY K, FRONICK C, COURTNEY B, ROCK S M, BELTER E, DU F, KIM K, ABBOTT R M, COTTON M, LEVY A, MARCHETTO P, OCHOA K, JACKSON S M, GILLAM B, CHEN W, YAN L, HIGGINBOTHAM J, CARDENAS M, WALIGORSKI J, APPLEBAUM E, PHELPS L, FALCONE J, KANCHI K, THANE T, SCIMONE A, THANE N, HENKE J, WANG T, RUPPERT J, SHAH N, ROTTER K, HODGES J, INGENTHON E, CORDES M, KOHLBERG S, SGROJ, DELGADO B, MEAD K, CHINWALLA A, LEONARD S, CROUSE K, COLLURA K, KUDRNA D, CURRIE J, HE R, ANGELOVA A, RAJASEKAR S, MUELLER T, LOMELI R, SCARA G, KO A, DELANEY K, WISSOTSKI M, LOPEZ G, CAMPOS D, BRAIDOTTI M, ASHLEY E, GOLSER W, KIM H, LEE S, LIN J, DUJMIC Z, KIM W, TALAG J, ZUCCOLO A, FAN C, SEBASTIAN A, KRAMER M, SPIEGEL L, NASCIMENTO L, ZUTAVERN T, MILLER B, AMBROISE C, MULLER S, SPOONER W, NARECHANIA A, REN L, WEI S, KUMARI S, FAGA B, LEVY M J, MCMAHAN L, VAN BUREN P, VAUGHN M W, YING K, YE H C T, EMRICH S J, JIA Y, KALYANARAMAN A, HSIA A P, BARBAZUK W B, BAUCOM R S, BRUTNELL T P, CARPITA N C, CHAPARRO C, CHIA J M, DERAGON J M, ESTILL J C, FU Y, JEDDELOH J A, HAN Y, LEE H, LI P, LISCH D R, LIU S, LIU Z, NAGEL D H, MCCANN M C, SANMIGUEL P, MYERS A M, NETTLETON D, NGUYEN J, PENNING B W, PONNALA L, SCHNEIDER K L, SCHWARTZ D C, SHARMA A, SODERLUND C, SPRINGER N M, SUN Q, WANG H, WATERMAN M, WESTERMAN R, WOLFGRUBER T K, YANG L, YU Y, ZHANG L, ZHOU S, ZHU Q, BENNETZEN J L, DAWE R K, JIANG J, JIANG N, PRESTING G G, WESSLER S R, ALURU S, MARTIENSSEN R A, CLIFTON S W, MCCOMBIE W R, WING R A, WILSON R K. The B73 maize genome: complexity, diversity, and dynamics[J]. Sci-

- ence, 2009, 326(5956): 1112-1115
- [2] ZHANG M, WU Y H, LEE M K, LIU Y H, RONG Y, SANTOS T S, WU C, XIE F, NELSON R L, ZHANG H B. Numbers of genes in the NBS and RLK families vary by more than four-fold within a plant species and are regulated by multiple factors[J]. *Nucleic Acids Research*, 2010, 38(19): 6513-6525.
- [3] KARLGREN A, GYLLENSTRAND N, KÄLLMAN T, SUNDSTRÖM J F, MOORE D, LASCOUX M, LAGERCRANTZ U. Evolution of the PEBP gene family in plants: functional diversification in seed plant evolution[J]. *Plant Physiology*, 2011, 156(4): 1967-1977.
- [4] SHIU S H, KARLOWSKI W M, PAN R, TZENG Y H, MAYER K F, LI W H. Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice[J]. *The Plant Cell*, 2004, 16(5): 1220-1234.
- [5] WALKER J C, ZHANG R. Relationship of a putative receptor protein kinase from maize to the S-locus glycoproteins of *Brassica*[J]. *Nature*, 1990, 345(6277): 743-746.
- [6] ZHANG X C, CANNON S B, STACEY G. Evolutionary genomics of LysM genes in land plants[J]. *BMC Evolutionary Biology*, 2009, 9: 183.
- [7] 汪青. 水稻类受体蛋白激酶 OsRPK1 激酶域的纯化及体外激酶活性验证[D]. 南京: 南京农业大学, 2010.
- WANG Q. Expression purification and *in vitro* kinase activity validation of OsRPK1 kinase domain of rice receptor-like protein kinase[D]. Nanjing: Nanjing Agricultural University, 2010. (in Chinese)
- [8] CHISHOLM S T, COAKER G, DAY B, STASKAWICZ B J. Host-microbe interactions: shaping the evolution of the plant immune response[J]. *Cell*, 2006, 124(4): 803-814.
- [9] LIU S, WANG Z, ZHU R, WANG F, CHENG Y, LIU Y. Three differential expression analysis methods for rRNA sequencing: limma, EdgeR, DESeq2[J]. *Journal of Visualized Experiments*, 2021(175): e62528.
- [10] TAN X, MEYERS B C, KOZIK A, WEST M A, MORGANTE M, ST CLAIR D A, BENT A F, MICHELMORE R W. Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in *Arabidopsis*[J]. *BMC Plant Biology*, 2007, 7(1): 56.
- [11] BÉLIVEAU C, POTVIN C, TRUDEL J, ASSELIN A, BELLEMARE G. Cloning, sequencing, and expression in *Escherichia coli* of a *Streptococcus faecalis* autolysin[J]. *Journal of Bacteriology*, 1991, 173(18): 5619-5623.
- [12] YIN J, LI F, ZHOU Y, MOU M, LU Y, CHEN K, XUE J, LUO Y, FU J, HE X, GAO J, ZENG S, YU L, ZHU F. INTEDE: interactome of drug-metabolizing enzymes[J]. *Nucleic Acids Research*, 2021, 49(D1): D1233-D1243.
- [13] 江聪, 黄敏仁, 徐立安. LysM 结构域及其与植物-真菌相互关系[J]. *植物学报*, 2014, 49(2): 221-228.
- JIANG C, HUANG M R, XU L A. LysM structural domain and its relationship with plant-fungal interactions[J]. *Journal of Botany*, 2014, 49(2): 221-228. (in Chinese)
- [14] BATEMAN A, BYCROFT M. The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD)[J]. *Journal of Molecular Biology*, 2000, 299(4): 1113-1119.
- [15] LI B, TANG J, YANG Q, LI S, CUI X, LI Y, CHEN Y, XUE W, LI X, ZHU F. NOREVA: normalization and evaluation of MS-based metabolomics data[J]. *Nucleic Acids Research*, 2017, 45(W1): W162-W170.
- [16] TUO D, YAN P, ZHAO G, CUI H, ZHU G, LIU Y, YANG X, WANG H, LI X, SHEN W, ZHOU P. An efficient papaya leaf distortion mosaic potyvirus vector for virus-induced gene silencing in papaya[J]. *Horticulture Research*, 2021, 8(1): 144.
- [17] LIANG X, ZHOU J M. Receptor-like cytoplasmic kinases: central players in plant receptor kinase-mediated signaling[J]. *Annual Review of Plant Biology*, 2018, 69: 267-299.
- [18] SCHULTZ J, MILPETZ F, BORK P, PONTING C P. SMART, a simple modular architecture research tool: identification of signaling domains[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1998, 95(11): 5857-5864.
- [19] QIN B, XIAO T, DING C, DENG Y, LV Z, SU J. Genome-wide identification and expression analysis of potential antiviral tripartite motif proteins (TRIMs) in grass carp (*Ctenopharyngodon idella*)[J]. *Biology*, 2021, 10(12): 1252.
- [20] CHEN C, CHEN H, ZHANG Y, THOMAS H R, FRANK MH, HE Y, XIA R. TBtools: an integrative toolkit developed for interactive analyses of big biological data[J]. *Molecular Plant*, 2020, 13(8): 1194-1202.
- [21] JUMPER J, EVANS R, PRITZEL A, GREEN T, FIGURNOV M, RONNEBERGER O, TUNYASUVUNAKOOL K, BATES R, ŽÍDEK A, POTAPENKO A, BRIDGLAND A, MEYER C, KOHL S A A, BALLARD A J, COWIE A, ROMERA-PAREDES B, NIKOLOV S, JAIN R, ADLER J, BACK T, PETERSEN S, REIMAN D, CLANCY E, ZIELINSKI M, STEINEGGER M, PACHOLSKA M, BERGHAMMER T, BODENSTEIN S, SILVER D, VINYALS O, SENIOR A W, KAVUKCUOGLU K, KOHLI P, HASSABIS D. Highly accurate protein structure prediction with AlphaFold[J]. *Nature*, 2021, 596(7873): 583-589.
- [22] VAN DONGEN M J P, KADAM R U, JURASZEK J, LAWSON E, BRANDENBURG B, SCHMITZ F, SCHEPENS W B G, STOOPS B, VAN DIEPEN H A, JONGENELEN M, TANG C, VERMOND J, VAN EIJGEN-OBREGOSO REAL A, BLOKLAND S, GARG D,

- YU W, GOUTIER W, LANCKACKER E, KLAP J M, PEETERS D C G, WU J, BUYCK C, JONCKERS T H M, ROYMANS D, ROEVENS P, VOGELS R, KOUDSTAAL W, FRIESEN R H E, RABOISSON P, DHANAK D, GOUDSMIT J, WILSON I A. A small-molecule fusion inhibitor of influenza virus is orally active in mice[J]. *Science*, 2019, 363(6431): 18-24.
- [23] MINH B Q, SCHMIDT H A, CHERNOMOR O, SCHREMPF D, WOODHAMS M D, VON HAESELER A, LANFEAR R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era[J]. *Molecular Biology and Evolution*, 2020, 37(5): 1530-1534.
- [24] QIAO X, LI Q, YIN H, QI K, LI L, WANG R, ZHANG S, PATERSON A H. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants[J]. *Genome Biology*, 2019, 20(1): 38-39.
- [25] ZHANG Z, XIAO J, WU J, ZHANG H, LIU G, WANG X, DAI L. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments[J]. *Biochemical and Biophysical Research Communications*, 2012, 419(4): 779-781.
- [26] ZHANG Z, LI J, ZHAO X Q, WANG J, WONG G K, YU J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging[J]. *Genomics, Proteomics & Bioinformatics*, 2006, 4(4): 259-263.
- [27] KIM D, PAGGI J M, PARK C, BENNETT C, SALZBERG S L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype[J]. *Nature Biotechnology*, 2019, 37(8): 907-915.
- [28] PERTEA M, PERTEA G M, ANTONESCU C M, CHANG T C, MENDELL J T, SALZBERG S L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads[J]. *Nature Biotechnology*, 2015, 33(3): 290-295.
- [29] LIU B, GAO X, ZHANG H. BioSeq-Analysis2.0: an updated platform for analyzing DNA, RNA and protein sequences at sequence level and residue level based on machine learning approaches[J]. *Nucleic Acids Research*, 2019, 47(20): e127.
- [30] MERKIN J J, CHEN P, ALEXIS M S, HAUTANIEMI S K, BURGE C B. Origins and impacts of new mammalian exons[J]. *Cell Reports*, 2015, 10(12): 1992-2005.
- [31] GABUT M, SAMAVARCHI-TEHRANI P, WANG X, SLOBODENIUC V, O'HANLON D, SUNG H K, ALVARIZ M, TALUKDER S, PAN Q, MAZZONI E O, NEDELEC S, WICHTERLE H, WOLTJEN K, HUGHES T R, ZANDSTRA P W, NAGY A, WRANA J L, BLENCOWE B J. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming[J]. *Cell*, 2011, 147(1): 132-146.
- [32] IZQUIERDO J M, MAJÓS N, BONNAL S, MARTÍNEZ C, CASTELO R, GUIGÓ R, BILBAO D, VALCÁRCCEL J. Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition[J]. *Molecular Cell*, 2005, 19(4): 475-484.
- [33] SHAUL O. How introns enhance gene expression[J]. *The International Journal of Biochemistry & Cell Biology*, 2017, 91(B): 145-155.
- [34] MURAT F, ARMERO A, PONT C, KLOPP C, SALSE J. Reconstructing the genome of the most recent common ancestor of flowering plants[J]. *Nature Genetics*, 2017, 49(4): 490-496.
- [35] JAILLON O, AURY J M, NOEL B, POLICRITI A, CLEPET C, CASAGRANDE A, CHOISNE N, AUBOURG S, VITULO N, JUBIN C, VEZZI A, LEGEAI F, HUGUENEY P, DASILVA C, HORNER D, MICA E, JUBLOT D, POULAIN J, BRUYÈRE C, BILLAULT A, SEGURENS B, GOUYVENOUX M, UGARTE E, CATTONARO F, ANTHOUARD V, VICO V, DEL FABBRO C, ALAUX M, DI GASPERO G, DUMAS V, FELICE N, PAILLARD S, JUMAN I, MOROLDO M, SCALABRIN S, CANAGUIER A, LE CLAINCHE I, MALACRIDA G, DURAND E, PESOLE G, LAUCOU V, CHATELET P, MERDINOGLU D, DELLEDONNE M, PEZZOTTI M, LECHARNY A, SCARPELLI C, ARTIGUENAVE F, PÈ ME, VALLE G, MORGANTE M, CABOCHE M, ADAM-BLONDON A F, WEISSENBACH J, QUÉTIER F, WINCKER P. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla[J]. *Nature*, 2007, 449(7161): 463-467.
- [36] 高梅, 辛健康, 姜山. 小立碗藓 LysM 型类受体激酶基因家族生物信息学分析[J]. *广西植物*, 2021, 41(6): 979-988.
- GAO M, XIN J K, JIANG S. Bioinformatics analysis of the LysM-type receptor-like kinase gene family in small standing bowl moss[J]. *Guangxi Plants*, 2021, 41(6): 979-988. (in Chinese)
- [37] 王东东, 温有福, 赵丹, 朵虎, 王超, 左存武. 苹果和梨 LysM-RLK 响应腐烂病信号成员鉴定[J]. *植物生理学报*, 2022, 58(2): 381-392.
- WANG D D, WEN Y F, ZHAO D, DUO H, WANG C, ZUO C W. Identification of LysM-RLK response to rot disease signaling members in apple and pear[J]. *Journal of Plant Physiology*, 2022, 58(2): 381-392. (in Chinese)
- [38] ZOU C, LEHTI-SHIU M D, THOMASHOW M, SHIU S H. Evolution of stress-regulated gene expression in duplicate genes of *Arabidopsis thaliana*[J]. *PLoS Genetics*, 2009, 5(7): e1000581.
- [39] 左存武, 张卫娜, 毛娟, 姜雪峰, 马宗桓, 苏静, 陈佰鸿. 苹果 LysM 类受体激酶基因家族鉴定与表达分析[J]. *园艺学报*, 2017, 44(4): 733-742.
- ZUO C W, ZHANG W N, MAO J, JIANG X F, MA Z H, SU J, CHEN B H. Identification and expression analysis of LysM-like receptor kinase gene family in apple[J]. *Journal of Horticulture*, 2017, 44(4): 733-742. (in Chinese)
- [40] 秦春晓. 星星草 (*Puccinellia tenuiflora*) LysM 基因家族鉴定与分析[D]. 哈尔滨: 东北林业大学, 2022.
- QIN C X. Identification and analysis of the LysM gene family of starry grass (*Puccinellia tenuiflora*)[D]. Harbin: Northeast Forestry University, 2022. (in Chinese)