



辣椒R2R3-MYB转录因子家族的全基因组鉴定与比较进化分析

康凯程^{1,2}, 牛西强², 黄先忠^{2*}, 胡能兵², 隋益虎², 张开京², 艾昊²

¹石河子大学生命科学学院, 植物基因组学实验室, 石河子 832003; ²安徽科技学院农学院, 凤阳 233100

摘要 MYB转录因子作为植物中最大的转录因子家族之一, 参与植物的生长、代谢、抵御生物和非生物胁迫等多种生理生化过程。R2R3-MYB是MYB转录因子家族的主要存在形式。辣椒是具有重要经济价值的蔬菜作物, 其R2R3-MYB转录因子缺乏系统的研究。从一年生辣椒(*Capsicum annuum*)、浆果状辣椒(*C. baccatum*)和中国辣椒(*C. chinense*)基因组中分别鉴定出94、92和94个R2R3-MYB基因, 基于系统发育关系将其分为28个亚族。共线性分析表明, 3种辣椒间存在73组直系同源R2R3-MYB基因, 一年生辣椒、浆果状辣椒和中国辣椒分别存在5、4和2个特有的R2R3-MYB基因。鉴定出12对重复基因, 其中8对是串联重复基因, 它们在3种辣椒分化前就已经存在。比较基因组学分析表明, 在辣椒进化过程中同源R2R3-MYB转录因子发生了功能分化。组织表达分析表明, 辣椒R2R3-MYB基因主要有3种表达特征: 在根、叶、茎和花中均高表达, 如CaMYB13/CbMYB12/CcMYB13; 仅在花中高表达, 如CaMYB93/CbMYB86/CcMYB12; 仅在根中高表达, 如CaMYB48/CbMYB47/CcMYB51。研究结果为深入揭示R2R3-MYB转录因子在辣椒生长发育中的生物学功能奠定了基础。

关键词 一年生辣椒, 浆果状辣椒, 中国辣椒, R2R3-MYB转录因子, 直系同源基因

康凯程, 牛西强, 黄先忠, 胡能兵, 隋益虎, 张开京, 艾昊 (2021). 辣椒R2R3-MYB转录因子家族的全基因组鉴定与比较进化分析. 植物学报 56, 315–329.

转录因子在基因转录调节中起重要作用, 通过激活或抑制靶基因的转录, 调控植物的生长发育(Pabo and Sauer, 1992; Kasuga et al., 1999; Riechmann et al., 2000; Valliyodan and Nguyen, 2006)。根据转录因子与其靶基因结合位点的相似性, 可将其分成不同的家族, 如AP2/EREBP、bHLH、HB、MYB和WRKY (Riechmann et al., 2000)。MYB转录因子具有与DNA结合的特征性MYB结构域, 该结构域在植物中高度保守。MYB转录因子包含1–4个不完全重复的MYB结构域, 分别命名为R1 (Repeat1)、R2、R3和R4, 每个MYB结构域约有52个氨基酸残基, 形成3个 α -螺旋, 其中第2和第3螺旋形成螺旋-转角-螺旋(helix-turn-helix)结构并与DNA大沟槽结合(Ogata et al., 1996; Jia et al., 2004; Dubos et al., 2010)。根据包含的MYB结构域数目, MYB转录因子可分为4类, 即

1R-MYB、2R-MYB (R2R3-MYB)、3R-MYB (R1R2R3-MYB)和4R-MYB (Dubos et al., 2010)。1R-MYB转录因子又称MYB-related转录因子, 包含1个完整或部分MYB结构域(Stracke et al., 2001; Chen et al., 2006)。3R-MYB转录因子包含3个连续的MYB结构域(R1、R2和R3), 在大多数真核生物基因组中都发现了3R-MYB转录因子, 其在调控细胞周期中发挥作用(Haga et al., 2007)。4R-MYB转录因子是最小的一类, 每个基因都含有4个R1/R2 MYB结构域, 它们在植物中发挥的作用尚不明确。包含2个MYB结构域(R2和R3)的R2R3-MYB转录因子是高等植物中MYB转录因子的主要存在形式, 该家族基因可能是因3R-MYB转录因子中R1 MYB结构域缺失进化而来(Rosinski and Atchley, 1998), 也有观点认为3R-MYB转录因子是从R2R3-MYB转录因子进化来的, R1 MYB结构

收稿日期: 2020-08-10; 接受日期: 2021-02-01

基金项目: 国家自然科学基金(No.31860393)、安徽省科技攻关(No.1704g07020111)、安徽省高校优秀青年人才支持计划(No.gxyqZD-2017070)和安徽科技学院学科带头人引进人才启动经费(No.NXYJ202001)

* 通讯作者。E-mail: huangxz@ahstu.edu.cn

域从R2R3-MYB转录因子获得(Jiang et al., 2004)。通常在MYB转录因子的N末端检测到高度保守的MYB结构域, C末端MYB结构域是高度分化的激活结构域, 这也决定了MYB转录因子具有多种调节作用(Dias et al., 2003; Matus et al., 2008)。

MYB转录因子广泛分布于植物基因组中, 是植物中最大的转录因子家族之一(Riechmann et al., 2000; Dubos et al., 2010)。首个植物MYB转录因子是从玉米(*Zea mays*)中鉴定出的COLORED1(Paz-Ares et al., 1987)。目前对MYB转录因子研究不断深入, R2R3-MYB转录因子已成为研究重点之一。据报道, R2R3-MYB转录因子参与调控许多重要的生理生化过程, 包括调节植物的初级和次生代谢、控制生长发育以及参与响应各种生物和非生物胁迫(Stracke et al., 2007)。例如, 马铃薯(*Solanum tuberosum*) *StMYB44*在高温下负调控花青素的生物合成(Liu et al., 2019); 金银花(*Lonicera japonica*) *LjMYB12*在拟南芥(*Arabidopsis thaliana*)中异位表达可以提高类黄酮含量, 其对类黄酮的生物合成的上下游途径具有不同的调控模式(Qi et al., 2019); 拟南芥*AtMYB20*、*AtMYB42*、*AtMYB43*和*AtMYB85*能够特异性地抑制类黄酮的生物合成(Geng et al., 2020); 棉花(*Gossypium hirsutum*) *GhMYB108-like*在干旱和盐胁迫下发挥重要作用(Ullah et al., 2020); 苹果(*Malus domestica*) *MdMYB24-like*参与茉莉酸甲酯诱导的花青素苷生物合成(Wang et al., 2019), *MdMYB30*在苹果角质层蜡质的积累中起重要作用, 并能增强苹果的抗病性(Zhang et al., 2019); 芝麻(*Sesamum indicum*) *SiMYB75*在拟南芥中的异位过表达显著促进根系生长, 提高植株对干旱、盐胁迫和渗透胁迫的耐受性(Dossa et al., 2020); 盐、干旱及脱落酸胁迫促进甘薯(*Ipomoea batatas*) *IbMYB3*的表达, 暗示*IbMYB3*在甘薯非生物胁迫应答中起重要作用(李格等, 2020)。

随着测序技术的发展, R2R3-MYB转录因子家族先后在许多高等植物的全基因组水平得到系统鉴定和分析, 包括拟南芥(Stracke et al., 2001; Chen et al., 2006)、甘蓝型油菜(*Brassica napus*) (Hajeb-rahimi et al., 2017)、大豆(*Glycine max*) (Du et al., 2012)、番茄(*Solanum lycopersicum*) (Zhao et al., 2014)和马铃薯(Li et al., 2019; Sun et al., 2019)。基

因组测序技术促进了园艺植物基因组学和遗传学研究(唐嘉璈等, 2020)。辣椒是茄科(*Solanaceae*)辣椒属(*Capsicum*)一年生或多年生草本植物, 是具有重要经济价值的蔬菜作物, 在世界范围内广泛种植。辣椒属包含27种, 被人类驯化长期栽培的辣椒有5种, 分别为一年生辣椒(*Capsicum annuum*)、浆果状辣椒(*C. baccatum*)、中国辣椒(*C. chinense*)、灌木状辣椒(*C. frutescens*)和绒毛辣椒(*C. pubescens*) (Ibiza et al., 2012)。我国主要栽培品种是一年生辣椒, 其它4个栽培种具有许多优良性状, 是对一年生辣椒进行品种改良的重要种质资源(隋益虎和陈劲枫, 2009; 魏家香等, 2018)。本研究以一年生辣椒、浆果状辣椒和中国辣椒为材料, 对3种辣椒R2R3-MYB转录因子家族在全基因组范围内进行鉴定, 在此基础上分析了辣椒R2R3-MYB基因的序列、motif分布、外显子-内含子结构、染色体定位、比较进化和不同组织(根、茎、叶和花)中的表达情况。通过共线性分析, 发现3种辣椒都存在特有的R2R3-MYB基因, 一年生辣椒和浆果状辣椒中的R2R3-MYB转录因子家族发生了基因组重复事件。研究结果为揭示辣椒R2R3-MYB转录因子功能及辣椒分子遗传育种奠定了理论基础。

1 材料与方法

1.1 辣椒R2R3-MYB转录因子家族的鉴定

一年生辣椒(*Capsicum annuum* L.) (ECW)、浆果状辣椒(*Capsicum baccatum* L.) (PBC81)和中国辣椒(*Capsicum chinense* Jacq.) (PI159236)的基因组数据从辣椒基因组平台(<http://peppergenome.snu.ac.kr>)中获得(Kim et al., 2017)。从拟南芥(*Arabidopsis thaliana* L.)全基因组数据库(<https://www.arabidopsis.org/>)中下载拟南芥R2R3-MYB转录因子家族的氨基酸序列作为query序列, 分别与3种辣椒的蛋白序列进行BLASTP比对, 参数设置*E-value*≤1e-10, 每条query保留5条匹配结果, 其余参数为默认值。将得到的候选序列使用hmmer (<http://www.hmmer.org/>)比对到PFAM32.0数据库中, 基于Pfam的MYB模型(PF00249)进行进一步比对筛选, 参数为默认值, 从而确定辣椒的R2R3-MYB转录因子基因成员。使用WebLogo (Crooks et al., 2004)绘制R2R3-MYB结构域seqlogo图。

使用基于BioPerl (<https://bioperl.org/>)编写的脚本获得辣椒R2R3-MYB转录因子的氨基酸长度、等电点和分子量等理化性质。利用ProtComp v. 9.0 (<http://linux1.softberry.com/berry.phtml>)进行R2R3-MYB转录因子的亚细胞定位预测。

1.2 系统进化树的构建、motif分布和基因结构分析

使用MEGA 7.0 (Kumar et al., 2016)软件构建系统进化树,采用邻接法(neighbor-joining, NJ),自检次数设为1 000次。使用MEME (<http://meme-suite.org/tools/meme>)分析CaMYB、CbMYB和CcMYB蛋白的保守motif, motif最大发现数目为10个,其它参数设为默认值。使用GSDS 2.0 (Hu et al., 2015) (<http://gsds.cbi.pku.edu.cn/>)分析CaMYB、CbMYB和CcMYB转录因子基因的外显子和内含子分布模式。

1.3 基因比较进化分析

共线性分析使用MCScanX (Wang et al., 2012)预测物种间和物种内存在的同源基因,参数设为默认值。将3种辣椒的基因组进行两两比较,设置 $E\text{-value}\leq 1e-10$,提取R2R3-MYB基因家族中的基因重复对信息,可筛选出3种辣椒间的直系同源基因;分别将3种辣椒的基因组与自身基因组进行比较,设置 $E\text{-value}\leq 1e-10$,根据结果筛选出3种辣椒内的旁系同源基因,即重复基因;如果1对重复基因位于染色体上的邻近位置(<100 kb),就可以认定为串联重复基因。使用InteractiVenn (Heberle et al., 2015)绘制Venn图。使用Circos (Krzywinski et al., 2009)显示一年生辣椒、浆果状辣椒和中国辣椒同源基因的共线性关系。使用ParaAT2.0 (Zhang et al., 2012)对同源基因对的CDS序列和ORF序列进行比对,参数设为默认值;使用KaKs_Calculator2.0 (Wang et al., 2010)读取比对结果,采用 γ -MYN法(Wang et al., 2009)计算同源基因对的非同义替换率(K_a)和同义替换率(K_s),通过 K_a/K_s 值来评估同源基因对中同源基因在进化过程中受到的选择压力。 $K_a/K_s>1$,基因受到正选择; $K_a/K_s<1$,基因受到纯化选择; $K_a/K_s=1$,基因中性进化(Yadav et al., 2015)。

1.4 组织表达分析

一年生辣椒、浆果状辣椒和中国辣椒的转录组测序数

据从BioProject数据库(<https://www.ncbi.nlm.nih.gov/bioproject>)中获得,登录号分别为PRJNA223222、PRJNA308879和PRJNA331024。测序样品为生长6周植株的根、茎和叶组织和生长4个月植株盛开的花(Kim et al., 2017)。使用fastp (Chen et al., 2018)对测序数据进行过滤,参数设为默认值。使用RSEM (Li and Dewey, 2011)将过滤后的数据比对到对应物种的基因组序列上,比对方式为bowtie2 (Langmead and Salzberg, 2012),参数设为默认值。将比对结果进行标准化处理,得到基因的每百万个转录片段中每千碱基片段的映射读数(fragments per kilobase of exon model per million mapped reads, FPKM)。FPKM值经 $\log_{10}(\text{FPKM}+1)$ 转换后,使用R包pheatmap (<https://CRAN.R-project.org/package=pheatmap>)绘制热图。

采用实时荧光定量PCR (qRT-PCR)检测辣椒MYB基因的组织表达,根据一年生辣椒MYB基因CDS序列设计qRT-PCR引物(表1)。实验材料是由安徽科技学院辣椒课题组提供的一年生辣椒自交系(编

表1 本研究使用的引物序列

Table 1 Primer sequences used in this study

Gene name	Primer sequences (5'→3')	Use
CaMYB62	ATCGCAGCGTATGAACACAC TTGTTACCGTGTTTGGCCTG	qRT-PCR
CaMYB35	GGCGGTGAATGTGATGATCC GGAAACCCCGTAGAAGCAAC	qRT-PCR
CaMYB92	TGATCCAACAACACACAGGC TTGTTCTCGAAATGGGCTGC	qRT-PCR
CaMYB39	TGGGGCAATCGTTGGTCTAA TGGATGTAAGTGGTGGTGGG	qRT-PCR
CaMYB25	AACCACCTACCTCGGCATTT TCACTTTTGTGAGCCTGCA	qRT-PCR
CaMYB2	CATCATCACCAGCAGTCACA GATCGACTTGCCAGCAGAAC	qRT-PCR
CaMYB8	CAGATCCTAGGGCTCTCGTG GGCCCGACTAATCCTGAGAT	qRT-PCR
CaMYB80	TCTTGGCAATAGGTGGTCTGA TGCTGCTCTCGTGGATAATCT	qRT-PCR
CaMYB19	GGACACAACATGGACAGCAG CTGAATCAGATGGGGAAGGA	qRT-PCR
CaUBI-3	TGTCCATCTGCTCTCTGTTG CACCCCAAGCACAATAAGAC	Internal reference gene

个内含子和3个外显子组成,少数*R2R3-MYB*基因由1个内含子和2个外显子构成。S1、S4、S7、S8、S18、S24、S25、S26和S28亚家族成员有2种不同的基因结构; *CaMYB9*、*CaMYB12*、*CaMYB66*、*CbMYB8*、*CbMYB9*、*CbMYB17*、*CcMYB10*、*CcMYB11*和*CcMYB65*具有3个内含子和4个外显子,它们分别属于S1、S7和S18亚家族; S25中的*CaMYB79*和*CbMYB76*具有11个内含子和12个外显子,但在中国辣椒中未发现具有11个内含子和12个外显子结构的

基因。这些结果表明,3种辣椒中*R2R3-MYB*基因的结构在物种长期进化过程中发生了明显的改变,外显子/内含子结构的演化可能是*R2R3-MYB*转录因子家族成员增多的原因之一。

通过对辣椒*R2R3-MYB*转录因子进行motif分析,共预测出10个保守的motifs,聚集在同一亚家族的基因具有相同的motif分布(图3C)。其中motif2、motif4和motif5是所有*R2R3-MYB*转录因子都具有的,这3个motif与*R2R3-MYB*转录因子所包含的R2和R3

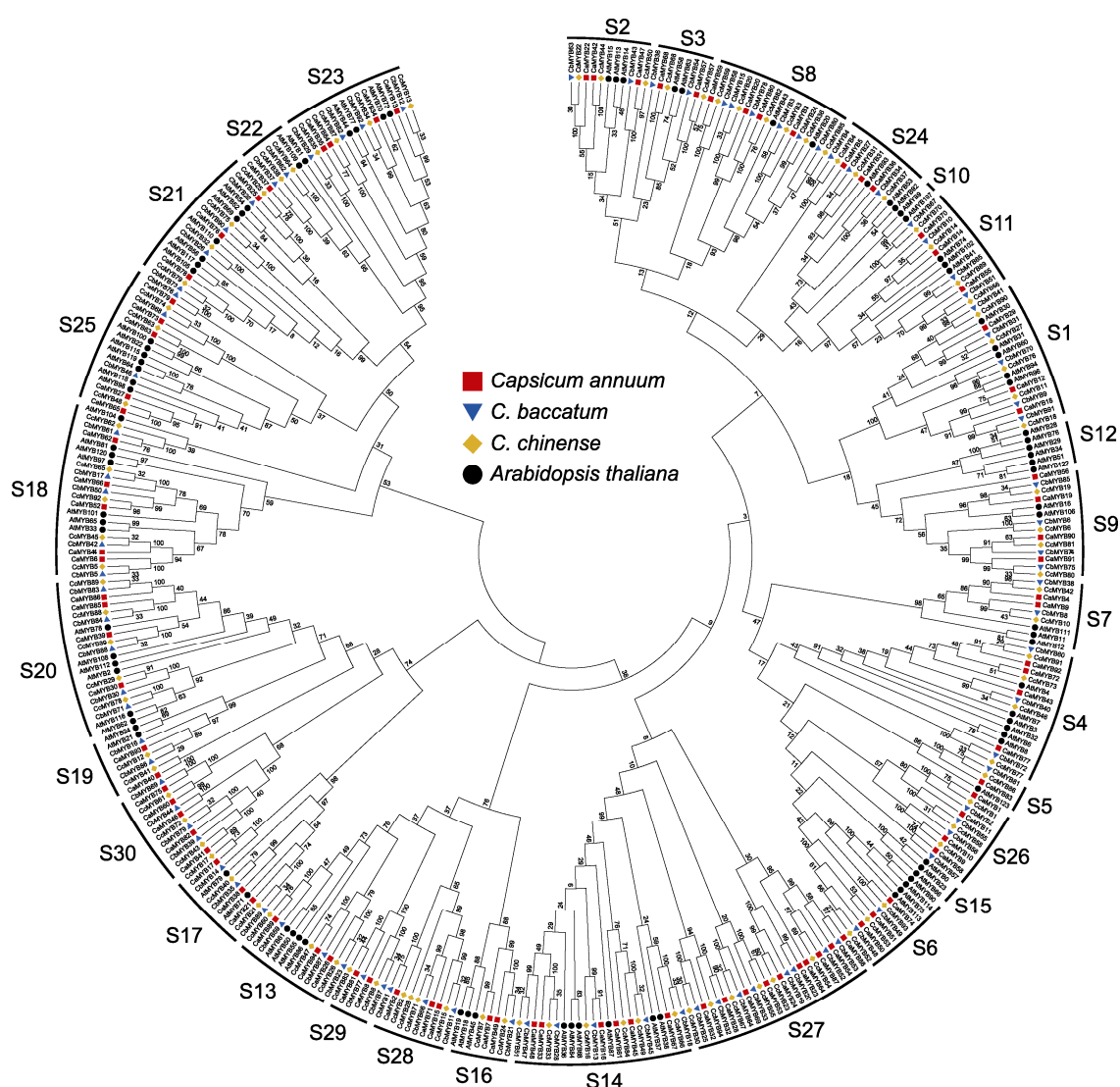


图2 3种辣椒和拟南芥*R2R3-MYB*转录因子家族的系统进化树
S1-S30分别代表30个亚组。

Figure 2 Phylogenetic tree of the *R2R3-MYB* transcription factor family from three *Capsicum* species and *Arabidopsis thaliana*
S1-S30 represent 30 subgroups.

MYB结构域相对应。

2.3 3种辣椒R2R3-MYB转录因子的系统进化关系

在植物进化过程中经常会出现基因重复，这也是植物基因家族发生扩张的重要原因之一(Cannon et al., 2004)。前人研究表明，大约在170万年前浆果状辣椒和一年生辣椒、中国辣椒的祖先发生分化，在110万年前一年生辣椒和中国辣椒也发生了分化(Kim et al., 2017)。为了确定3种辣椒间R2R3-MYB同源基因以及每种辣椒内部可能出现的基因重复现象，本研究鉴定了3种辣椒中R2R3-MYB基因的进化关系。最终在3种辣椒中鉴定出73组直系同源R2R3-MYB基因(图4)，

分别占一年生辣椒R2R3-MYB基因的77.66% (73/94)、浆果状辣椒R2R3-MYB基因的79.35% (73/92)和中国辣椒R2R3-MYB基因的77.66% (73/94)。这说明大多数R2R3-MYB基因在它们的祖先物种中存在，并且在辣椒的进化过程中高度保守。

除了3种辣椒中都存在的直系同源R2R3-MYB基因，我们检测发现25组R2R3-MYB基因只在2种辣椒中鉴定到。其中有9组(CbMYB6/CcMYB6、CbMYB-24/CcMYB36、CbMYB26/CcMYB32、CbMYB41/CcMYB90、CbMYB62/CcMYB64、CbMYB65/CcMYB69、CbMYB70/CcMYB76、CbMYB71/CcMYB78和CbMYB80/CcMYB85) R2R3-MYB

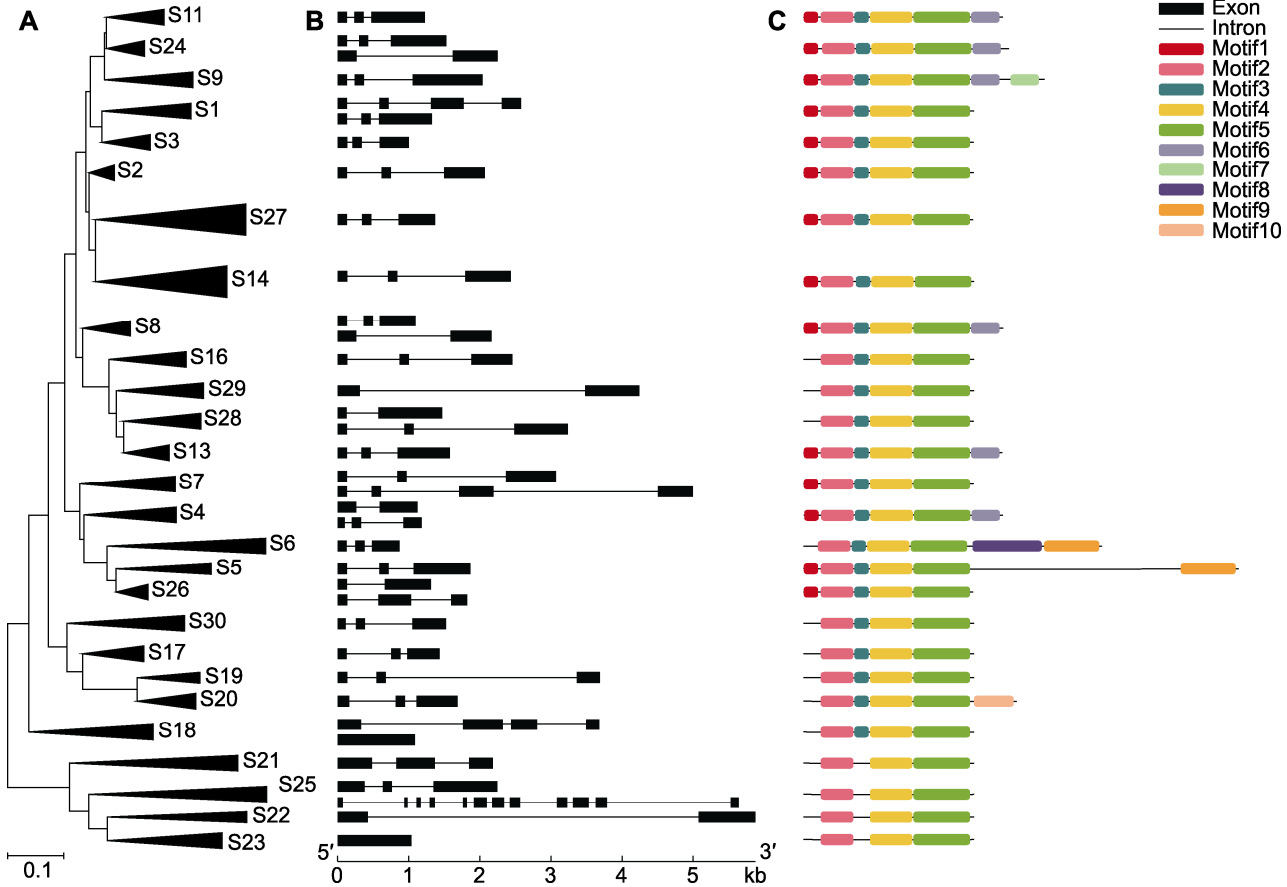


图3 3种辣椒R2R3-MYB转录因子家族的系统进化、基因外显子-内含子结构和motif分析
(A) 27个亚家族的系统进化树; (B) 27个亚家族的外显子-内含子结构(黑色方框表示外显子，黑色线条表示内含子); (C) 27个亚家族的motif分布(不同颜色的方框表示不同的motif)

Figure 3 Phylogenetic tree, gene exon-intron structure and motif analysis of R2R3-MYB transcription factor family in three *Capsicum* species

(A) Phylogenetic tree of the 27 subfamilies; (B) Exon-intron structure of the 27 subfamilies (Black boxes indicate exons and black lines indicate introns); (C) The motif distribution of the 27 subfamilies (The boxes with different colors indicate different motifs)

基因在一年生辣椒中未鉴定到直系同源基因; 有10组(*CaMYB7/CcMYB7*、*CaMYB40/CcMYB41*、*CaMYB42/CcMYB44*、*CaMYB60/CcMYB61*、*CaMYB63/CcMYB63*、*CaMYB65/CcMYB48*、*CaMYB68/CcMYB68*、*CaMYB72/CcMYB73*、*CaMYB74/CcMYB93*和*CaMYB82/CcMYB72*) 在浆果状辣椒中未鉴定到直系同源基因; 有6组(*CaMYB23/CbMYB20*、*CaMYB46/CbMYB44*、*CaMYB58/CbMYB57*、*CaMYB75/CbMYB69*、*CaMYB79/CbMYB76*和*CaMYB88/CbMYB53*) 在中国辣椒中未鉴定到直系同源基因(图4)。一年生辣椒还存在5个特有的R2R3-MYB基因(*CaMYB27*、*CaMYB54*、*CaMYB56*、*CaMYB61*和*CaMYB64*)，浆果状辣椒和中国辣椒分别存在4个(*CbMYB27*、*CbMYB54*、*CbMYB61*和*CbMYB64*)和2个(*CcMYB28*和*CcMYB84*)特有的R2R3-MYB基因(图4)。上述结果表明，在3种辣椒分化之后，它们可能丢失了一些基因，并获得了一些新基因。

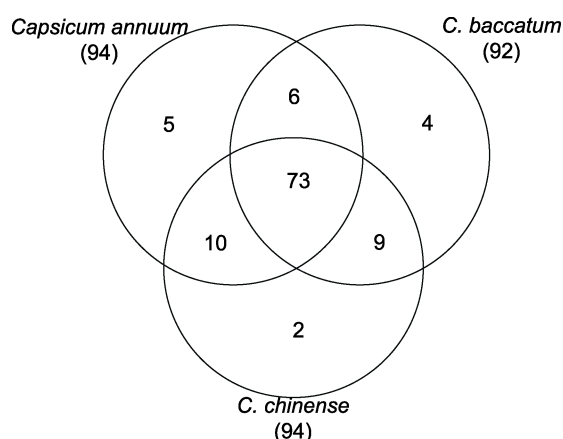


图4 一年生辣椒、浆果状辣椒和中国辣椒中R2R3-MYB转录因子直系同源基因的数量

Figure 4 The number of R2R3-MYB transcription factor orthologous genes in *Capsicum annum*, *C. baccatum*, and *C. chinense*

辣椒R2R3-MYB基因中有12对重复基因，其中8对是串联重复基因。在这些串联重复基因中，*CaMYB50/CbMYB48/CcMYB52*和*CaMYB51/CbMYB49/CcMYB53*属于同一组直系同源基因；*CaMYB87/CbMYB52*和*CaMYB88/CbMYB53*属于同一组直系同源基因；*CaMYB90/CbMYB74/CcMYB81*和*CaMYB91/CbMYB75/CcMYB80*属于同一组直系同源基因。但*CaMYB87/CbMYB52*在中国辣椒中的直系同

源基因*CcMYB54*未鉴定到串联重复现象，与其串联重复的*CaMYB88/CbMYB53*在中国辣椒中未找到直系同源基因。

为进一步探明3种辣椒R2R3-MYB转录因子家族的进化机制，我们基于直系同源基因和串联重复基因构建了3种辣椒的Circos图(图5)。结果表明，R2R3-MYB基因在所有染色体上均有分布，主要位于染色体两臂的末端。3种辣椒间直系同源基因在染色体上的位置十分接近，锚定在高度保守的共线性区块。值得注意的是，一些直系同源基因，如*CaMYB10/CbMYB56*位于不同的染色体上，这可能是由于染色体结构变异所致(Kim et al., 2017)。目前发布的基因组信息显示，一年生辣椒和浆果状辣椒中有8个R2R3-MYB基因没有定位到染色体上，中国辣椒中有5个R2R3-MYB基因没有定位到染色体上，但与它们对应的直系同源基因却定位到染色体上。如在*CaMYB87/CbMYB52/CcMYB54*中，*CaMYB87*没有定位到染色体上，但*CbMYB52*和*CcMYB54*分别位于浆果状辣椒和中国辣椒的7号染色体上，造成这一现象的原因可能是在染色体组装时产生的误差。

通过计算非同义替换率(K_a)和同义替换率(K_s)可以评估3种辣椒中同源基因的进化模式。 $K_a/K_s > 1$ 表示受到正选择， $K_a/K_s < 1$ 表示受到纯化选择， $K_a/K_s = 1$ 表示基因中性进化(Yadav et al., 2015)。本研究表明，所有直系同源基因中有17对直系同源基因(*CaMYB7/CcMYB7*、*CaMYB9/CbMYB8*、*CaMYB11/CbMYB55*、*CaMYB18/CbMYB91*、*CaMYB29/CcMYB27*、*CaMYB30/CbMYB30*、*CaMYB37/CbMYB37*、*CaMYB37/CcMYB38*、*CaMYB58/CbMYB57*、*CaMYB63/CcMYB63*、*CaMYB67/CcMYB66*、*CaMYB90/CcMYB81*、*CaMYB94/CbMYB87*、*CbMYB8/CcMYB10*、*CbMYB30/CcMYB29*、*CbMYB43/CcMYB50*和*CbMYB74/CcMYB81*)的 $K_a/K_s > 1$ (图6)，说明这些基因对受到正选择，暗示这17对直系同源基因可能存在功能上的分化。有1对同源基因(*CbMYB4/CcMYB4*) $K_a/K_s = 1$ ，可能是中性进化。其余同源基因对和12对重复基因的 K_a/K_s 均小于1，说明这些基因很可能受到纯化选择。

2.4 3种辣椒R2R3-MYB基因组织表达特征

为了解R2R3-MYB基因在辣椒不同组织中的表达模式，我们利用3种辣椒的根、叶、茎和花组织的转录

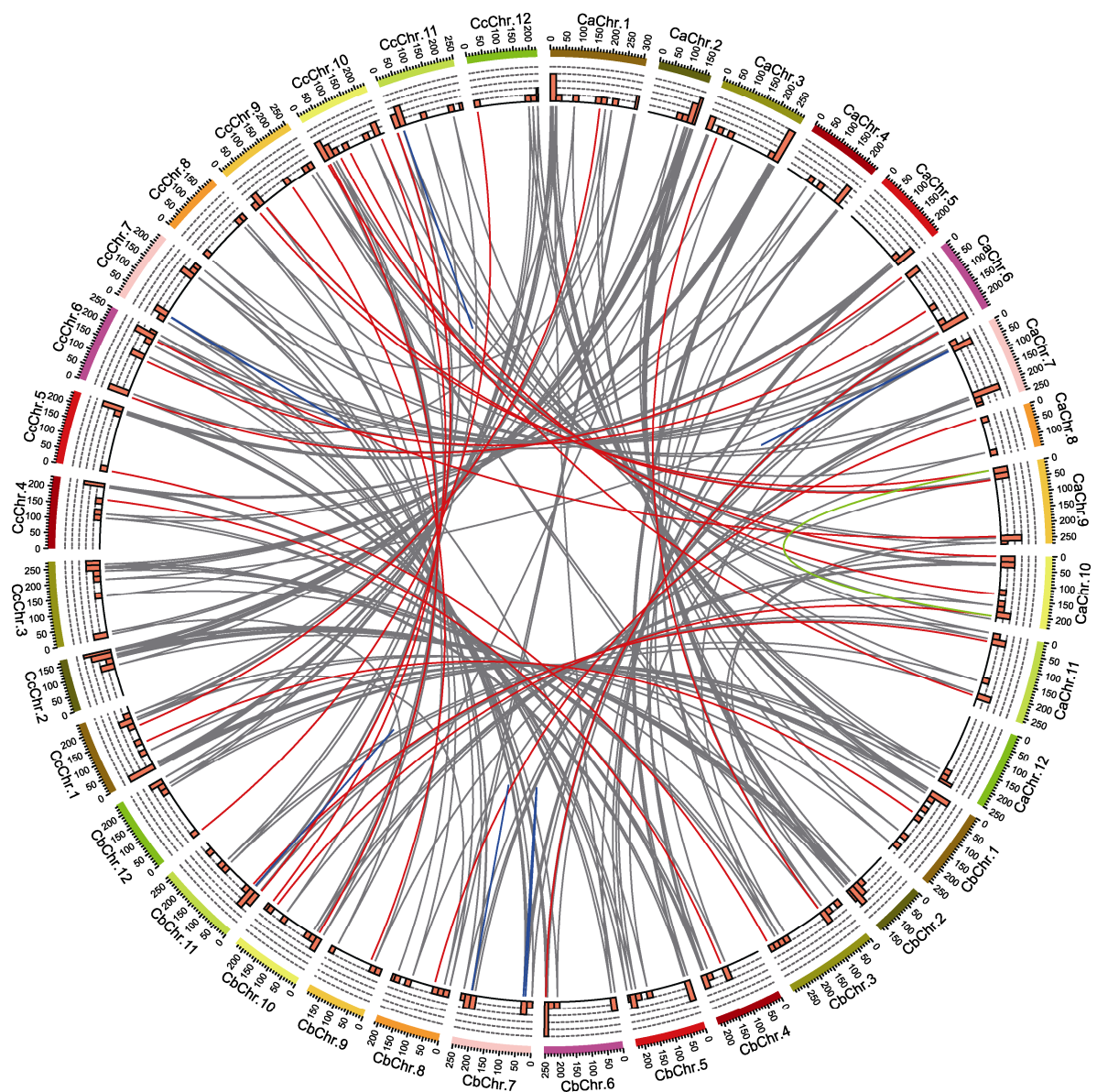


图5 一年生辣椒(Ca)、浆果状辣椒(Cb)和中国辣椒(Cc)中R2R3-MYB转录因子的同源基因对
从外向内依次是3种辣椒的染色体数目、20 Mb窗口统计R2R3-MYB基因的数量(0–5)和3种辣椒中R2R3-MYB同源基因对。灰色线连接3种辣椒中都存在的同源基因对；红色线连接只在两种辣椒中存在的同源基因对；绿色线连接重复基因；蓝色线连接串联重复基因。

Figure 5 Homologous R2R3-MYB transcription factor gene pairs in *Capsicum annuum* (Ca), *C. baccatum* (Cb), and *C. chinense* (Cc)
Tracks from outside to inside are the number of chromosomes, number of R2R3-MYB genes (0–5) (20 Mb windows), and homologous gene pairs among three *Capsicum* species. The gray lines connect the homologous gene pairs that exist in three *Capsicum* species; the red lines connect the homologous gene pairs in two species; the green lines connect the replicated genes; and the blue lines connect the tandem repeated genes.

组数据分析转录丰度。结果表明，在一年生辣椒、浆果状辣椒和中国辣椒中分别有14、6和10个R2R3-MYB基因未检测到其转录产物(FPKM<0.5)，表明这

些基因在其它组织中或在这些组织中具有不同的时空表达特征。

基于73组直系同源基因的转录组数据，我们发

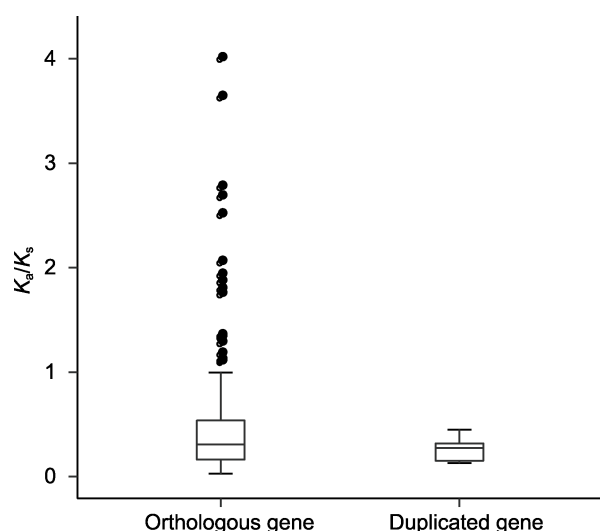


图6 一年生辣椒、浆果状辣椒和中国辣椒中R2R3-MYB转录因子直系同源和重复基因对的 K_a/K_s 比率

黑色点表示非同义替换率(K_a)/同义替换率(K_s)>1的基因对。

Figure 6 The K_a/K_s ratios for the orthologous and duplicated gene pairs of R2R3-MYB transcription factor in *Capsicum annuum*, *C. baccatum*, and *C. chinense*. The black dots represent gene pairs with non-synonymous substitution rate (K_a)/synonymous substitution rate (K_s)>1.

现不同辣椒的R2R3-MYB基因在4种组织中的表达具有明显差异(图7)。综合分析发现主要有3种表达模式:(1)在根、叶、茎和花4种组织中均高表达,但表达量不尽相同,如CaMYB13/CbMYB12/CcMYB13、CaMYB34/CbMYB92/CcMYB34和CaMYB35/CbMYB29/CcMYB35,表明这些基因可能在调控辣椒生长发育过程中起重要作用;(2)在花中特异或高表达的基因,如CaMYB93/CbMYB86/CcMYB12;(3)在根中特异或高表达的基因,如CaMYB48/CbMYB47/CcMYB51和CaMYB80/CbMYB78/CcMYB82,这些基因可能参与相应组织发育的某些生物学过程。此外,大多数直系同源基因在3种辣椒中的表达特征类似,但还存在一些直系同源基因在同一组织中表达模式不同。例如,CaMYB31在所有组织中均不表达,而它的直系同源基因CbMYB27和CcMYB31在根中表达,表明这些直系同源基因可能发生了功能分化。

利用qRT-PCR分析了9个基因在根、叶、茎和花中的表达,结果与RNA-seq分析结果虽不完全一致,但趋势非常吻合,进一步验证辣椒R2R3-MYB基因的组织表达特征(图8)。其中,CaMYB62在4个组织中

均高表达,但在根中的表达明显较低;CaMYB35在根和花中的表达量明显高于叶和茎;CaMYB92在花中高表达;而CaMYB39在根、叶和茎中几乎不表达,但在花中高表达。CaMYB25在根中表达最高,其次是茎中;CaMYB2主要在花中表达。CaMYB8在花中的表达最低;而CaMYB80在根中的表达量明显高于其它组织;CaMYB19除了在根中表达量较低以外,在其它组织中均高表达。

2.5 讨论

随着基因测序技术的不断发展,已在基因组水平上对许多植物的R2R3-MYB转录因子家族进行了鉴定,如拟南芥(125个)(Stracke et al., 2001; Chen et al., 2006)、甘蓝型油菜(249个)(Hajiebrahimi et al., 2017)、大豆(244个)(Du et al., 2012)、番茄(121个)(Zhao et al., 2014)和马铃薯(112个)(Li et al., 2019; Sun et al., 2019),它们的基因组大小分别为125 Mb (Zapata et al., 2016)、912 Mb (Chalhoub et al., 2014)、1.03 Gb (Schmutz et al., 2010)、828 Mb (The Tomato Genome Consortium, 2012)和663 Mb (The Potato Genome Sequencing Consortium, 2011)。这表明物种中R2R3-MYB基因数目与物种基因组的大小并不成正比。一年生辣椒、浆果状辣椒和中国辣椒的基因组大小分别为2.85、2.99和2.80 Gb (Kim et al., 2017)。本研究从中分别鉴定出94、92和94个R2R3-MYB基因,也支持这一观点。与茄科植物番茄(Zhao et al., 2014)和马铃薯(Li et al., 2019; Sun et al., 2019)相比,3种辣椒中R2R3-MYB基因数目明显较少,说明R2R3-MYB转录因子家族在辣椒中发生了明显收缩。

基于3种辣椒和拟南芥R2R3-MYB基因序列的相似性,可将其划分为30个不同的亚家族(图2)。与马铃薯(Li et al., 2019; Sun et al., 2019)类似,辣椒R2R3-MYB转录因子家族中存在部分亚家族成员丢失,也获得了一些新的亚家族成员,表明辣椒R2R3-MYB转录因子家族也有扩张。一年生辣椒、浆果状辣椒和中国辣椒中R2R3-MYB基因的结构分析(图3)显示,它们主要由2个内含子、3个外显子和1个内含子、2个外显子这两类基因结构组成,相同亚家族成员也普遍具有相同的基因结构。辣椒R2R3-MYB转录因子中包含2个MYB结构域与所有转录因子共有的motif

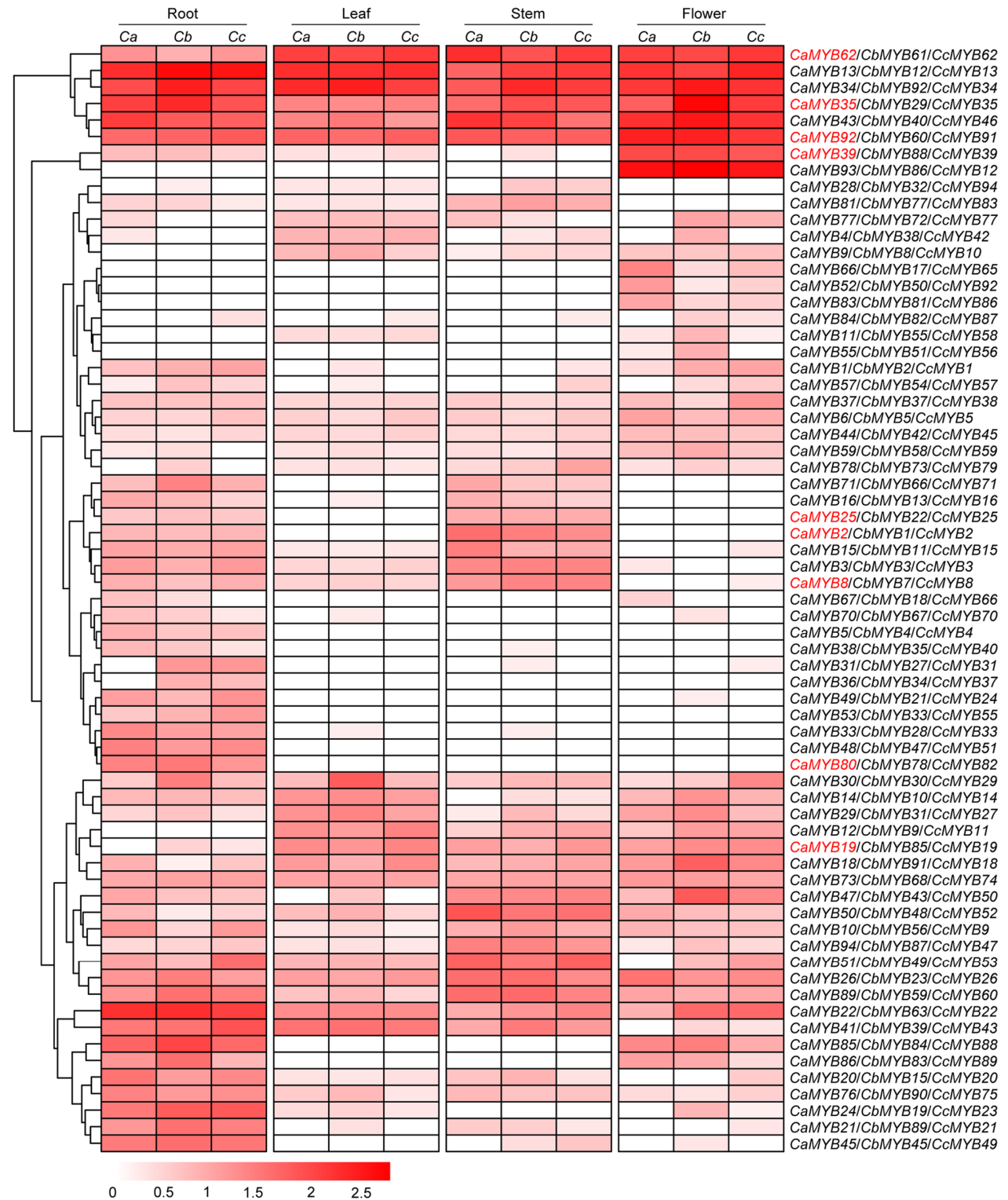


图7 一年生辣椒(Ca)、浆果状辣椒(Cb)和中国辣椒(Cc) *R2R3-MYB*基因的组织表达特征
颜色条表示*R2R3-MYB*基因在不同组织中 $\log_{10}(\text{FPKM}+1)$ 值的变化范围。红色字体的基因进行了qRT-PCR验证。

Figure 7 Expression profile of *R2R3-MYB* genes in different tissues from *Capsicum annuum* (Ca), *C. baccatum* (Cb), and *C. chinense* (Cc)
The color bars indicate the variation range of $\log_{10}(\text{FPKM}+1)$ values of *R2R3-MYB* genes in different tissues. The expression of the genes in red was verified by qRT-PCR.

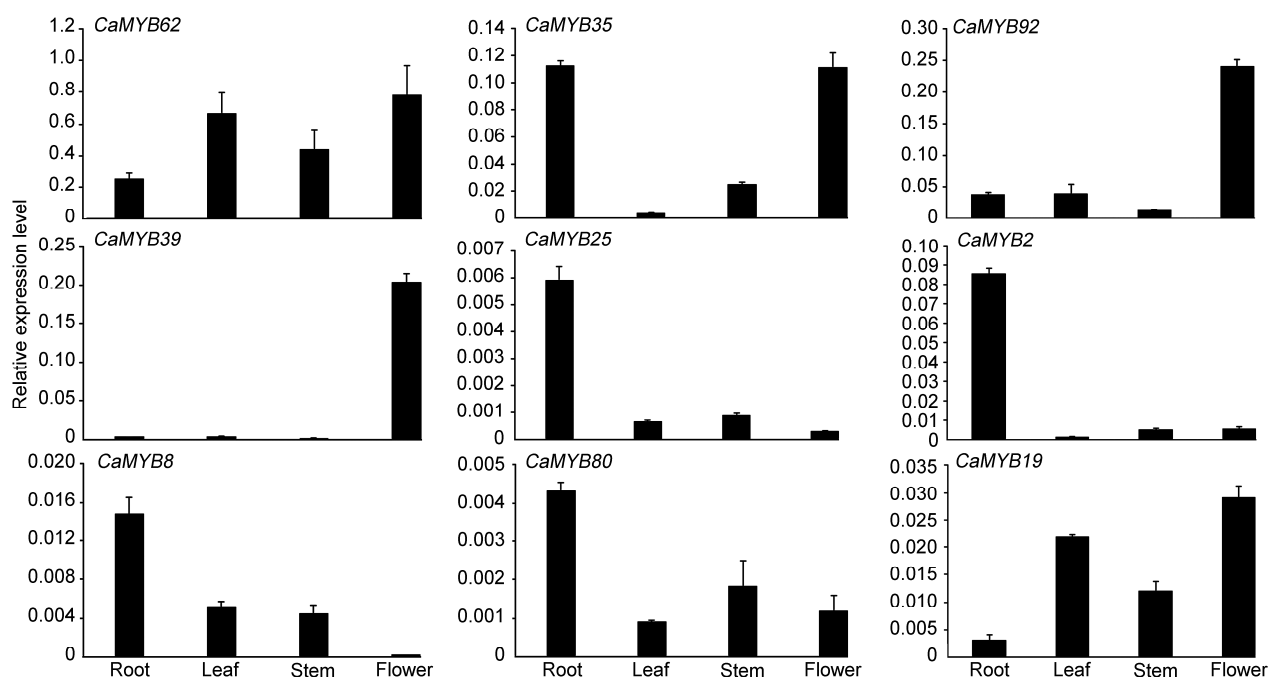


图8 实时荧光定量PCR分析9个*R2R3-MYB*基因的组织表达特征

Figure 8 Tissue expression profile of nine *R2R3-MYB* genes by qRT-PCR

相对应,位于同一亚家族的*R2R3-MYB*基因motif也都相同,这说明*R2R3-MYB*基因的亚家族分类主要由C端序列决定。

本研究发现3种辣椒基因组中存在直系同源基因丢失。通过鉴定只找到73组在3种辣椒中均存在的直系同源*R2R3-MYB*基因,其它直系同源*R2R3-MYB*基因存在不同程度的丢失(图4)。此外,本研究还鉴定出12对重复基因,其中8对是串联重复基因。这些串联重复基因分属3组直系同源基因,但*CaMYB87/CaMYB88*和*CbMYB53/CbMYB53*在中国辣椒中只找到1个直系同源基因*CcMYB54*。这进一步证明在辣椒进化的过程中,*R2R3-MYB*转录因子家族中不同的亚家族成员既有明显的收缩也有扩张趋势。

*R2R3-MYB*基因在辣椒12条染色体上的数目不尽相同,但它们主要位于每条染色体的长臂或短臂末端(附表1;图5)。3种辣椒直系同源基因在染色体上的位置很接近,锚定在高度保守的共线性区块上。研究发现,浆果状辣椒中3号染色体长臂与5号染色体短臂发生了易位,3号染色体短臂和9号染色体长臂发生了易位(Kim et al., 2017),位于这些位点的*R2R3-MYB*基因也发生了变化。例如,*CbMYB24*位于浆果状

辣椒3号染色体上,它的同源基因*CcMYB36*位于中国辣椒5号染色体上。此外,还有少量直系同源*R2R3-MYB*基因位于不同的染色体上。例如,*CbMYB38/CcMYB42*位于6号染色体,其直系同源基因*CaMYB4*位于1号染色体;*CaMYB10/CcMYB9*位于1号染色体,其直系同源基因*CbMYB56*位于8号染色体,这可能是由染色体结构变异所致。基于已发布的基因组信息,3种辣椒中有21个*R2R3-MYB*基因尚未定位到染色体上,但和它们对应的直系同源基因却已定位到染色体上。导致这一现象的原因不排除基因组组装错误,后续可通过进一步精细测序以及构建高质量辣椒基因组图谱来消除。直系同源基因和串联重复基因的 K_a/K_s 分析结果(图6)表明,有17对直系同源基因可能受到正选择,暗示着它们的功能可能发生了变化。其中,*CaMYB37/CbMYB37*、*CaMYB37/CcMYB38*的 K_a/K_s 均大于1,但*CbMYB37/CcMYB38*的 K_a/K_s 却小于1,表明*CaMYB37*和它的直系同源基因*CbMYB37/CcMYB38*之间可能出现了功能分化。*CbMYB8*和*CbMYB30*也有类似的进化模式。

*R2R3-MYB*转录因子参与辣椒不同组织的发育,并且有部分转录因子表现出组织特异性表达(图7)。

例如, *CaMYB39/CbMYB88/CcMYB39*和*CaMYB93/CbMYB86/CcMYB12*主要在花中特异表达, 且*CaMYB93*与拟南芥*AtMYB112*序列相似性较高, 而*AtMYB112*在花青苷积累中起重要的调节作用(Lotkowska et al., 2015), 表明*CaMYB93/CbMYB86/CcMYB12*可能是辣椒花青素积累的重要调节因子。一些在根或花中特异性表达的*R2R3-MYB*基因也是进行功能分析的理想候选基因。此外, 还有一些直系同源基因表现出不同的表达模式。例如, *CaMYB30/CbMYB30/CcMYB29*中*CaMYB30*在所有组织中的表达量较低, 而*CbMYB30*在叶片和根中高表达, 但*CcMYB29*在花中高表达。表明这些直系同源基因虽然存在一定的功能保守性, 但也存在功能分化。12对重复基因中*CaMYB54/88*、*CaMYB60/75*、*CaMYB-87/88*、*CbMYB52/53*和*CcMYB52/53*表达模式出现差异, 说明虽然这些重复基因之间的序列相似性较高, 但也存在一定程度的功能分化。此外, 受到正选择的直系同源基因在不同组织中的表达模式也出现差异(如*CaMYB18/CbMYB91*); 受到纯化选择的一些直系同源基因对也表现出组织表达差异(图7)。这些结果表明, 除了受环境影响之外, 还有其它原因导致同源基因出现功能分化, 值得进一步研究。

近期, 居利香等(2020)在辣椒中开展了MYB基因家族的鉴定, 但未进行MYB基因的组织表达特征及在3种辣椒之间的比较基因组分析。浆果状辣椒和中国辣椒具有很多优异的性状, 是对一年生辣椒进行遗传改良的重要材料, 但目前缺少分子水平的研究。越来越多的研究表明, MYB转录因子参与植物许多重要的生理生化过程。R2R3-MYB转录因子作为MYB转录因子中最大的一类, 对该家族成员的准确鉴定和功能分析, 为今后深入揭示辣椒R2R3-MYB转录因子的功能奠定了基础, 并对辣椒的遗传改良具有参考价值。

参考文献

居利香, 雷欣, 赵成志, 舒黄英, 汪志伟, 成善汉 (2020). 辣椒MYB基因家族的鉴定及与辣味关系分析. 园艺学报 47, 875–892.

李格, 孟小庆, 李宗芸, 朱明库 (2020). 甘薯盐胁迫响应基因*lbMYB3*的表达特征及生物信息学分析. 植物学报 55, 38–48.

隋益虎, 陈劲枫 (2009). 辣椒属种间远缘杂交育种研究进展. 热带作物学报 30, 557–563.

唐嘉璐, 邱杰, 黄学辉 (2020). 基因组学技术大发展助力园艺植物研究取得新进展. 植物学报 55, 1–4.

魏家香, 俞佳虹, 程远, 叶青静, 王荣青, 阮美颖, 李志邈, 姚祝平, 周国治, 杨悦俭, 万红建 (2018). 辣椒种间杂交的现状及其研究进展. 分子植物育种 16, 5474–5482.

Blanc G, Wolfe KH (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16, 1667–1678.

Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol* 4, 10.

Chalhoub B, Denoeud F, Liu SY, Parkin IAP, Tang HB, Wang XY, Chiquet J, Belcram H, Tong CB, Samans B, Corr  a M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao MX, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnell N, Le Paslier MC, Fan GY, Renault V, Bayer PE, Golicz AA, Manoli S, Lee TH, Thi VHD, Chalabi S, Hu Q, Fan CC, Tollenaere R, Lu YH, Battail C, Shen JX, Sidebottom CHD, Wang XF, Canaguier A, Chauveau A, B  rard A, Deniot G, Guan M, Liu ZS, Sun FM, Lim YP, Lyons E, Town CD, Bancroft I, Wang XW, Meng JL, Ma JX, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury JM, Adams KL, Batley J, Snowden RJ, Tost J, Edwards D, Zhou YM, Hua W, Sharpe AG, Paterson AH, Guan CY, Wincker P (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953.

Chen SF, Zhou YQ, Chen YR, Gu J (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.

Chen YH, Yang XY, He K, Liu MH, Li JG, Gao ZF, Lin ZQ, Zhang YF, Wang XX, Qiu XM, Shen YP, Zhang L, Deng XH, Luo JC, Deng XW, Chen ZL, Gu HY, Qu LJ (2006). The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol Biol* 60, 107–124.

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004). WebLogo: a sequence logo generator. *Genome Res* 14, 1188–1190.

Dias AP, Braun EL, McMullen MD, Grotewold E (2003). Recently duplicated maize *R2R3 Myb* genes provide evidence for distinct mechanisms of evolutionary divergence

- after duplication. *Plant Physiol* **131**, 610–620.
- Dossa K, Mmadi MA, Zhou R, Liu AL, Yang YX, Diouf D, You J, Zhang XR (2020). Ectopic expression of the sesame MYB transcription factor *SiMYB305* promotes root growth and modulates ABA-mediated tolerance to drought and salt stresses in *Arabidopsis*. *AoB Plants* **12**, plz081.
- Du H, Yang SS, Liang Z, Feng BR, Liu L, Huang YB, Tang YX (2012). Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol* **12**, 106.
- Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010). MYB transcription factors in *Arabidopsis*. *Trends Plant Sci* **15**, 573–581.
- Geng P, Zhang S, Liu JY, Zhao CH, Wu J, Cao YP, Fu CX, Han X, He H, Zhao Q (2020). MYB20, MYB42, MYB43, and MYB85 regulate phenylalanine and lignin biosynthesis during secondary cell wall formation. *Plant Physiol* **182**, 1272–1283.
- Haga N, Kato K, Murase M, Araki S, Kubo M, Demura T, Suzuki K, Muller I, Voß U, Jürgens G, Ito M (2007). R1R2R3-Myb proteins positively regulate cytokinesis through activation of *KNOLLE* transcription in *Arabidopsis thaliana*. *Development* **134**, 1101–1110.
- Hajiebrahimi A, Owji H, Hemmati S (2017). Genome-wide identification, functional prediction, and evolutionary analysis of the R2R3-MYB superfamily in *Brassica napus*. *Genome* **60**, 797–814.
- Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R (2015). InteractVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169.
- Hu B, Jin JP, Guo AY, Zhang H, Luo JC, Gao G (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**, 1296–1297.
- Ibiza VP, Blanca J, Cañizares J, Nuez F (2012). Taxonomy and genetic diversity of domesticated *Capsicum* species in the Andean region. *Genet Resour Crop Evol* **59**, 1077–1088.
- Jia L, Clegg MT, Jiang T (2004). Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies *indica* and *japonica* genomes. *Plant Physiol* **134**, 575–585.
- Jiang CZ, Gu JY, Chopra S, Gu X, Peterson T (2004). Ordered origin of the typical two- and three-repeat MYB genes. *Gene* **326**, 13–22.
- Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1999). Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat Biotechnol* **17**, 287–291.
- Kim S, Park J, Yeom SI, Kim YM, Seo E, Kim KT, Kim MS, Lee JM, Cheong K, Shin HS, Kim SB, Han K, Lee J, Park M, Lee HA, Lee HY, Lee Y, Oh S, Lee JH, Choi E, Choi E, Lee SE, Jeon J, Kim H, Choi G, Song H, Lee J, Lee SC, Kwon JK, Lee HY, Koo N, Hong YJ, Kim RW, Kang WH, Huh JH, Kang BC, Yang TJ, Lee YH, Bennetzen JL, Choi D (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol* **18**, 210.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009). Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639–1645.
- Kumar S, Stecher G, Tamura K (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**, 1870–1874.
- Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359.
- Li B, Dewey CN (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.
- Li XX, Guo C, Ahmad S, Wang Q, Yu J, Liu C, Guo YF (2019). Systematic analysis of MYB family genes in potato and their multiple roles in development and stress responses. *Biomolecules* **9**, 317.
- Liu YH, Kui LW, Espley RV, Wang L, Li YM, Liu Z, Zhou P, Zeng LH, Zhang XJ, Zhang JL, Allan AC (2019). StMYB44 negatively regulates anthocyanin biosynthesis at high temperatures in tuber flesh of potato. *J Exp Bot* **70**, 3809–3824.
- Livak KJ, Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408.
- Lotkowska ME, Tohge T, Fernie AR, Xue GP, Balazadeh S, Mueller-Roeber B (2015). The *Arabidopsis* transcription factor MYB112 promotes anthocyanin formation during salinity and under high light stress. *Plant Physiol* **169**, 1862–1880.
- Matus JT, Aquea F, Arce-Johnson P (2008). Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes. *BMC Plant Biol* **8**, 83.
- Ogata K, Kanei-Ishii C, Sasaki M, Hatanaka H, Nagadoi A, Enari M, Nakamura H, Nishimura Y, Ishii S, Sarai A

- (1996). The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and *trans*-activation. *Nat Struct Biol* **3**, 178–187.
- Pabo CO, Sauer RT** (1992). Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* **61**, 1053–1095.
- Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H** (1987). The regulatory *c1* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J* **6**, 3553–3558.
- Qi XW, Fang HL, Chen ZQ, Liu ZQ, Yu X, Liang CY** (2019). Ectopic expression of a R2R3-MYB transcription factor gene *LjaMYB12* from *Lonicera japonica* increases flavonoid accumulation in *Arabidopsis thaliana*. *Int J Mol Sci* **20**, 4494.
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu GL** (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* **290**, 2105–2110.
- Rosinski JA, Atchley WR** (1998). Molecular evolution of the MYB family of transcription factors: evidence for polyphyletic origin. *J Mol Evol* **46**, 74–83.
- Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu SQ, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du JC, Tian ZX, Zhu LC, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA** (2010). Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183.
- Stracke R, Ishihara H, Huep G, Barsch A, Mehrrens F, Niehaus K, Weisshaar B** (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J* **50**, 660–677.
- Stracke R, Werber M, Weisshaar B** (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr Opin Plant Biol* **4**, 447–456.
- Sun WJ, Ma ZT, Chen H, Liu MY** (2019). MYB gene family in potato (*Solanum tuberosum* L.): genome-wide identification of hormone-responsive reveals their potential functions in growth and development. *Int J Mol Sci* **20**, 4847.
- The Potato Genome Sequencing Consortium** (2011). Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189–195.
- The Tomato Genome Consortium** (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641.
- Ullah A, UI Qamar MT, Nisar M, Hazrat A, Rahim G, Khan AH, Hayat K, Ahmed S, Ali W, Khan A, Yang XY** (2020). Characterization of a novel cotton MYB gene, *GhMYB108-like* responsive to abiotic stresses. *Mol Biol Rep* **47**, 1573–1581.
- Valliyodan B, Nguyen HT** (2006). Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Curr Opin Plant Biol* **9**, 189–195.
- Wang DP, Wan HL, Zhang S, Yu J** (2009). γ -MYN: a new algorithm for estimating K_a and K_s with consideration of variable substitution rates. *Biol Direct* **4**, 20.
- Wang DP, Zhang YB, Zhang Z, Zhu J, Yu J** (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom Proteom Bioinf* **8**, 77–80.
- Wang YC, Liu WJ, Jiang HY, Mao ZL, Wang N, Jiang SH, Xu HF, Yang GX, Zhang ZY, Chen XS** (2019). The R2R3-MYB transcription factor MdMYB24-like is involved in methyl jasmonate-induced anthocyanin biosynthesis in apple. *Plant Physiol Biochem* **139**, 273–282.
- Wang YP, Tang HB, Debarry JD, Tan X, Li JP, Wang XY, Lee TH, Jin HZ, Marler B, Guo H, Kissinger JC, Patterson AH** (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49.
- Yadav CB, Bonthala VS, Muthamilarasan M, Pandey G, Khan Y, Prasad M** (2015). Genome-wide development of transposable elements-based markers in foxtail millet and construction of an integrated database. *DNA Res* **22**, 79–90.
- Zapata L, Ding J, Willing EM, Hartwig B, Bezdan D, Jiao WB, Patel V, Velikkakam James G, Koornneef M, Ossowski S, Schneeberger K** (2016). Chromosome-level assembly of *Arabidopsis thaliana* Ler reveals the extent of translocation and inversion polymorphisms. *Proc Natl Acad Sci USA* **113**, E4052–E4060.
- Zhang YL, Zhang CL, Wang GL, Wang YX, Qi CH, Zhao Q, You CX, Li YY, Hao YJ** (2019). The R2R3 MYB transcription factor MdMYB30 modulates plant resistance against pathogens by regulating cuticular wax biosynthesis.

sis. *BMC Plant Biol* **19**, 362.

Zhang Z, Xiao JF, Wu JY, Zhang HY, Liu GM, Wang XM, Dai L (2012). ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem Biophys Res Commun* **419**, 779–781.

Zhao PP, Li Q, Li J, Wang LN, Ren ZH (2014). Genome-wide identification and characterization of *R2R3MYB* family in *Solanum lycopersicum*. *Mol Genet Genomics* **289**, 1183–1207.

Genome-wide Identification and Comparative Evolutionary Analysis of the R2R3-MYB Transcription Factor Gene Family in Pepper

Kaicheng Kang^{1,2}, Xiqiang Niu², Xianzhong Huang^{2*}, Nengbing Hu²
Yihu Sui², Kaijing Zhang², Hao Ai²

¹Plant Genomics Laboratory, College of Life Sciences, Shihezi University, Shihezi 832003, China

²College of Agriculture, Anhui Science and Technology University, Fengyang 233100, China

Abstract As one of the largest transcription factor (TF) families in plants, MYB TFs are involved in various physiological and biochemical processes, such as plant growth, metabolism, and response to various biotic and abiotic stresses. R2R3-MYB is the main form of MYB TFs in higher plants. Pepper is a vegetable crop with important economic value, but the R2R3-MYB TF family has not been systematically studied in pepper. In this study, 94 *CaMYB*, 92 *CbMYB*, and 94 *CcMYB* TFs genes were identified with comparative genomic analysis in *Capsicum annuum*, *C. baccatum*, and *C. chinense*, respectively. These genes were categorized into 28 subfamilies. Collinearity analysis indicated that there were 73 groups of orthologous *R2R3-MYB* genes among three pepper species. There were five, four, and two unique *R2R3-MYB* genes in *C. annuum*, *C. baccatum*, and *C. chinense*, respectively. In addition, we identified 12 pairs of duplicated genes, and eight of which are tandemly repeated genes, which already existed before the divergence of three pepper species. Comparative genomics analysis suggested that the homologous *R2R3-MYB* TFs underwent functionally divergence during the evolution of pepper. Analysis on the expression profile showed that *R2R3-MYB* genes were expressed in three major patterns: high expression in roots, leaves, stems, and flowers, such as *CaMYB13/CbMYB12/CcMYB13*; high expression in flowers, such as *CaMYB93/CbMYB86/CcMYB12*; high expression in roots, such as *CaMYB48/CbMYB47/CcMYB51*. These results lay a foundation for further study on the biological functions of *R2R3-MYB* TFs in the growth and development of pepper.

Key words *Capsicum annuum*, *C. baccatum*, *C. chinense*, R2R3-MYB transcription factor, orthologous gene

Kang KC, Niu XQ, Huang XZ, Hu NB, Sui YH, Zhang KJ, Ai H (2021). Genome-wide identification and comparative evolutionary analysis of the R2R3-MYB transcription factor gene family in pepper. *Chin Bull Bot* **56**, 315–329.

* Author for correspondence. E-mail: huangxz@ahstu.edu.cn

(责任编辑: 白羽红)

附表 1 一年生辣椒 R2R3-MYB 蛋白基本信息

Appendix table 1 Basic information of the R2R3-MYB proteins from *Capsicum annuum*

<http://www.chinbullbotany.com/fileup/1674-3466/PDF/t20-143.pdf>

附表 1 一年生辣椒 R2R3-MYB 蛋白基本信息

Appendix table 1 Basic information of the R2R3-MYB proteins from *Capsicum annuum*

Gene ID	GenBank ID	Subfamilies	Chr.	Protein length (aa)	Molecular weight (kDa)	Isoelectric point	Predict the sub-cellular localization
CaMYB1	PHT92173.1	S26	CaChr01	260	29.81	8.95	Nuclear
CaMYB2	PHT92354.1	S28	CaChr01	338	38.58	6.28	Nuclear
CaMYB3	PHT92630.1	S8	CaChr01	278	31.70	5.09	Nuclear
CaMYB4	PHT92940.1	S7	CaChr01	321	36.81	4.61	Nuclear
CaMYB5	PHT93006.1	S24	CaChr01	299	34.10	4.85	Nuclear
CaMYB6	PHT93927.1	S18	CaChr01	509	55.30	5.56	Nuclear
CaMYB7	PHT94384.1	S16	CaChr01	280	31.87	8.37	Nuclear
CaMYB8	PHT94622.1	S29	CaChr01	367	40.76	6.37	Nuclear
CaMYB9	PHT94937.1	S7	CaChr01	410	45.86	4.96	Extracellular
CaMYB10	PHT95917.1	S26	CaChr01	272	31.55	9.71	Nuclear
CaMYB11	PHT96064.1	S26	CaChr01	277	32.29	10.26	Nuclear
CaMYB12	PHT89671.1	S1	CaChr02	330	36.93	5.29	Nuclear
CaMYB13	PHT90336.1	S23	CaChr02	345	38.25	8.5	Nuclear
CaMYB14	PHT90680.1	S11	CaChr02	381	43.31	6.09	Nuclear
CaMYB15	PHT90879.1	S28	CaChr02	341	38.72	7.17	Nuclear
CaMYB16	PHT91447.1	S14	CaChr02	330	37.44	7.23	Nuclear
CaMYB17	PHT91635.1	S17	CaChr02	257	29.49	8.94	Nuclear
CaMYB18	PHT91875.1	S1	CaChr02	304	34.33	7.34	Nuclear
CaMYB19	PHT91893.1	S9	CaChr02	415	45.20	5.83	Extracellular
CaMYB20	PHT91999.1	S8	CaChr02	255	29.41	6.31	Nuclear
CaMYB21	PHT85511.1	S17	CaChr03	261	30.70	9.29	Nuclear
CaMYB22	PHT85687.1	S2	CaChr03	278	31.50	4.98	Nuclear
CaMYB23	PHT86364.1	S27	CaChr03	233	27.01	9.78	Nuclear
CaMYB24	PHT86610.1	S27	CaChr03	250	28.83	9.95	Nuclear
CaMYB25	PHT87835.1	S21	CaChr03	245	29.10	8.71	Nuclear
CaMYB26	PHT88024.1	S13	CaChr03	332	37.96	5.63	Nuclear
CaMYB27	PHT88073.1	S25	CaChr03	372	42.88	5.89	Nuclear
CaMYB28	PHT88110.1	S27	CaChr03	355	39.75	6.98	Extracellular
CaMYB29	PHT88362.1	S1	CaChr03	334	36.72	6.68	Nuclear
CaMYB30	PHT88689.1	S20	CaChr03	293	33.72	7.22	Nuclear
CaMYB31	PHT83486.1	S24	CaChr04	347	39.27	5.97	Nuclear
CaMYB32	PHT83629.1	S27	CaChr04	332	37.57	7.1	Nuclear
CaMYB33	PHT84462.1	S14	CaChr04	330	37.35	8.34	Nuclear
CaMYB34	PHT84596.1	S23	CaChr04	326	35.52	9.71	Nuclear
CaMYB35	PHT84665.1	S23	CaChr04	262	29.32	8.45	Nuclear
CaMYB36	PHT82022.1	S24	CaChr05	312	35.75	9.4	Nuclear
CaMYB37	PHT82135.1	S22	CaChr05	406	44.04	4.94	Extracellular
CaMYB38	PHT82223.1	S17	CaChr05	237	27.91	9.91	Nuclear
CaMYB39	PHT82246.1	S20	CaChr05	317	36.80	6.79	Nuclear
CaMYB40	PHT77586.1	S19	CaChr06	231	27.07	6.51	Nuclear
CaMYB41	PHT78012.1	S30	CaChr06	238	28.08	8.04	Nuclear
CaMYB42	PHT78581.1	S2	CaChr06	280	32.26	7.24	Nuclear
CaMYB43	PHT79168.1	S4	CaChr06	288	32.11	9.41	Nuclear
CaMYB44	PHT79962.1	S18	CaChr06	495	54.20	5.59	Nuclear
CaMYB45	PHT80075.1	S14	CaChr06	301	34.26	7.73	Nuclear
CaMYB46	PHT80171.1	S30	CaChr06	225	26.35	6.37	Nuclear
CaMYB47	PHT80426.1	S2	CaChr06	237	26.99	5.42	Nuclear
CaMYB48	PHT75195.1	S14	CaChr07	342	38.73	6.16	Nuclear
CaMYB49	PHT75521.1	S16	CaChr07	279	31.76	6.29	Nuclear
CaMYB50	PHT75561.1	S6	CaChr07	214	24.80	9.56	Nuclear
CaMYB51	PHT75572.1	S6	CaChr07	214	24.72	9	Nuclear
CaMYB52	PHT76439.1	S18	CaChr07	485	53.83	6.85	Extracellular
CaMYB53	PHT76531.1	S27	CaChr07	229	26.57	6.4	Nuclear
CaMYB54	PHT76534.1	S27	CaChr07	249	29.00	8.73	Nuclear
CaMYB55	PHT76539.1	S11	CaChr07	337	38.47	6.16	Nuclear
CaMYB56	PHT76677.1	S9	CaChr07	275	31.98	9.32	Nuclear
CaMYB57	PHT77523.1	S3	CaChr07	298	33.79	6.73	Nuclear
CaMYB58	PHT73830.1	S26	CaChr08	293	32.95	7.71	Nuclear
CaMYB59	PHT74496.1	S8	CaChr08	310	34.60	6.93	Nuclear

CaMYB60	PHT71769.1	S30	CaChr09	118	13.96	10.3	Nuclear
CaMYB61	PHT71785.1	S14	CaChr09	311	34.98	6.77	Nuclear
CaMYB62	PHT72159.1	S18	CaChr09	361	41.33	9.75	Nuclear
CaMYB63	PHT72198.1	S25	CaChr09	451	50.14	6.9	Nuclear
CaMYB64	PHT73137.1	S27	CaChr09	172	19.87	8.94	Nuclear
CaMYB65	PHT73144.1	S25	CaChr09	521	59.64	6.79	Extracellular
CaMYB66	PHT73196.1	S18	CaChr09	479	53.59	6.24	Extracellular
CaMYB67	PHT73553.1	S14	CaChr09	268	30.47	6.12	Nuclear
CaMYB68	PHT69515.1	S3	CaChr10	225	25.92	8.87	Nuclear
CaMYB69	PHT69534.1	S27	CaChr10	318	35.62	8.17	Nuclear
CaMYB70	PHT69931.1	S11	CaChr10	321	36.46	7.74	Nuclear
CaMYB71	PHT69965.1	S28	CaChr10	331	38.04	6.23	Nuclear
CaMYB72	PHT70378.1	S4	CaChr10	263	29.90	7.95	Nuclear
CaMYB73	PHT70552.1	S25	CaChr10	539	60.08	8.94	Extracellular
CaMYB74	PHT70593.1	S6	CaChr10	258	29.63	8.49	Nuclear
CaMYB75	PHT70989.1	S30	CaChr10	203	23.65	6.36	Nuclear
CaMYB76	PHT71270.1	S21	CaChr10	343	39.72	7.35	Extracellular
CaMYB77	PHT67233.1	S4	CaChr11	162	18.70	9.85	Nuclear
CaMYB78	PHT67771.1	S21	CaChr11	452	49.67	7.5	Extracellular
CaMYB79	PHT67883.1	S25	CaChr11	476	53.28	6.51	Extracellular
CaMYB80	PHT68519.1	S8	CaChr11	224	25.99	6.71	Nuclear
CaMYB81	PHT68695.1	S29	CaChr11	357	39.95	6.51	Nuclear
CaMYB82	PHT68777.1	S30	CaChr11	220	25.89	7.6	Nuclear
CaMYB83	PHT66282.1	S5	CaChr12	352	39.60	6.51	Nuclear
CaMYB84	PHT66425.1	S23	CaChr12	242	27.22	8.33	Nuclear
CaMYB85	PHT66981.1	S20	CaChr12	290	34.18	7.07	Nuclear
CaMYB86	PHT66982.1	S20	CaChr12	269	31.54	7.11	Nuclear
CaMYB87	PHT62995.1	S27	CaSca.1085	247	28.42	6.62	Nuclear
CaMYB88	PHT62996.1	S27	CaSca.1085	258	29.65	8.9	Nuclear
CaMYB89	PHT62538.1	S13	CaSca.1248	420	47.23	7.26	Nuclear
CaMYB90	PHT62351.1	S9	CaSca.1335	342	39.13	8.87	Nuclear
CaMYB91	PHT62353.1	S9	CaSca.1335	346	39.45	8.54	Nuclear
CaMYB92	PHT64037.1	S4	CaSca.588	239	27.61	9.64	Nuclear
CaMYB93	PHT63325.1	S19	CaSca.973	196	22.56	6.8	Nuclear
CaMYB94	PHT63293.1	S13	CaSca.993	319	36.98	6.85	Nuclear