

簸箕柳 R2R3-MYB 转录因子家族全基因组分析

王岚春, 沈方圆, 欧阳丹, 李 校

(四川大学生命科学学院 生物资源与生态环境教育部重点实验室, 成都 610064)

摘要: 为研究簸箕柳中 R2R3-MYB 转录因子家族成员的分类和进化, 以及在抗重力刺激中可能具有的功能, 本研究使用全基因组鉴定策略和其它生物信息学手段对簸箕柳相关家族成员进行详细的分析. 结果表明, 在簸箕柳中共鉴定出 158 个 R2R3-MYB 蛋白, 这些成员进一步被分为 23 个亚组. 基因结构和基序组成分析表明, 在同一亚组中的基因通常具有相似的内含子-外显子结构和相似的基序组成, 进一步支持了系统发育分析的结果. 通过共线性分析了 SsMYB 家族的进化, 85 对 SsMYB 基因被预测来自串联或者片段复制事件, 这在 SsMYB 家族的扩展中起着重要作用. 此外, 本研究利用 RNA-seq 数据分析重力刺激下 SsMYB 基因的表达情况, 筛选出 20 个潜在的 MYB 抗重力刺激蛋白.

关键词: 簸箕柳; R2R3-MYB; 转录因子; 全基因组分析

中图分类号: Q811.4 **文献标识码:** A **DOI:** 10.19907/j.0490-6756.2022.036001

Genome-wide analysis of the R2R3-MYB transcription factor family in *Salix suchowensis*

WANG Lan-Chun, SHEN Fang-Yuan, OUYANG Dan, LI Xiao

(Key Laboratory of Bio-resources and Eco-environment of Ministry of Education,
College of Life Sciences, Sichuan University, Chengdu 610064, China)

Abstract: In order to study the taxonomy and evolution of the R2R3-MYB transcription factor family members and their possible functions under gravistimulation, this study used genome-wide identification strategy and other bioinformatics methods to analyze the related family members of *Salix suchowensis* in detail. The results showed that a total of 158 R2R3-MYB proteins were identified in *Salix suchowensis*. These members were further divided into 23 subgroups. The analysis of gene structure and motif composition showed that genes in the same subgroup usually had similar exon/intron structure and similar motif composition, which further supported the results of phylogenetic analysis. Syntenic analysis was performed to explore the evolution of the SsMYB family. 85 pairs of SsMYB genes were predicted to arise from tandem or segment duplication events, which played an important role in the expansion of the SsMYB family. In addition, RNA-seq data was used to analyze the expression pattern of SsMYB genes under gravistimulation, and 20 potential SsMYB genes were screened out which involved in negative gravitropism of the stem.

Keywords: *Salix suchowensis*; R2R3-MYB; Transcription factors; Genome-wide analysis

收稿日期: 2021-08-15

基金项目: 国家自然科学基金(61001149)

作者简介: 王岚春(1997-), 女, 河北人, 硕士研究生, 研究方向为生物信息学. E-mail: wanglc@stu.scu.edu.cn

通讯作者: 李校. E-mail: lix@scu.edu.cn

1 引言

生物体会通过基因的表达和沉默来调控自身的生长发育,其调控过程是相互独立而又相互依存的.这种调控可能发生在基因表达过程中的任意一个阶段,如 DNA 的转录过程,mRNA 的加工过程,以及 mRNA 的翻译过程等,且调控的发生需要各种酶和调节蛋白的相互配合^[1].其中转录因子参与 DNA 转录起始过程来影响基因的表达.转录因子通常被定义为具有序列特异性的 DNA 结合并能够激活或抑制转录的蛋白质^[2].基因表达在转录水平上的调控影响或控制着细胞或有机体中的许多生物学过程,如细胞周期的进展、代谢和生理平衡以及对环境的反应.

漫长的自然选择与进化过程当中,植物形成了其特有的基因表达和调控机制,以适应不断改变的生存环境.转录因子在这一过程中起到了很大的作用.MYB 转录因子广泛分布于所有真核生物中,是植物界最大的转录因子家族之一^[1].大多数 MYB 蛋白起着转录因子的作用,具有不同数量的 MYB 结构域重复序列,赋予它们结合 DNA 的能力^[1].MYB 蛋白是控制发育、代谢和对生物和非生物胁迫反应的调控网络中的关键因子^[3].

MYB 转录因子 (TFs) 是一组由 1 到 4 个 MYB 重复序列定义的全真核转录因子.50 个氨基酸组成三个 α 螺旋,每个重复序列的第二个和第三个螺旋形成一个螺旋-旋转-螺旋 (HTH) 结构,其中有三个间距规则的色氨酸(或疏水)残基,形成疏水核心.每个重复序列的第三个螺旋是与 DNA 直接接触的 DNA 识别螺旋^[4].HTH 与启动子中的调节元件相互作用,而 C-末端区域负责与真核转录机制的其他成分建立蛋白质-蛋白质相互作用^[5].在 DNA 接触过程中,两个 MYB 重复序列紧密地堆积在主沟中,使两个识别螺旋协同结合到特定的 DNA 识别序列基序上.MYB-DNA 结合域的长度为 100~160 个残基,这取决于 N-末端区域不完全重复(称为 R)的数量.根据 MYB 重复数和 MYB 重复序列的特性,MYB 蛋白质一般分为 MYB-related、R2R3-MYB、R1R2R3-MYB 和 4R-MYB 蛋白质^[1].在这四种类型中,R2R3MYB 是高等植物特有的,在大多数植物中数量占优势,其特点是存在保守的 MYB 结构域和高度可变的 C-末端区域^[6, 7].MYB TFs 是植物中最大的 TF 类之一,该类 TF 的规模主要归因于 R2R3-MYB TF 家

族的迅速扩张^[8].先前的研究表明,R2R3-MYB TFs 的数量随着绿色植物进化过程而增加^[9].

越来越多的证据表明,R2R3MYB 转录因子参与植物许多生理生化过程,如叶毛状体分化^[10]、次生壁形成^[11]、花药和花粉发育^[12, 13]、腋生分生组织形成^[14],二次代谢的调节包括类黄酮^[15]、花青素^[16]和木质素^[17].除此之外,R2R3MYB 家族成员还参与植物对各种非生物和生物胁迫的防御和响应^[18-20],并在调节植物对包括吲哚乙酸^[21]、脱落酸^[21, 22]、赤霉素^[23, 24]、乙烯、水杨酸和茉莉酸^[24]等植物激素线索的反应中发挥了作用,以及一些环境信号响应,如水可用性^[25],光^[26]和营养元素^[27].

随着多种植物的全基因组测序完成,实验技术及手段的进步,使得科研人员可以对 R2R3-MYB 基因家族在内的基因资源进行更为广泛和深入的研究.对于 R2R3-MYB 转录因子的生物功能研究也有了很大的进展,越来越多的 R2R3-MYB 转录因子的功能被揭示.虽然在一些物种中已经对该家族进行了全基因组分析,但对簸箕柳(*Salix suchowensis*)的 R2R3-MYB 基因知之甚少.簸箕柳属于杨柳科中的柳属,筐柳组,因其枝条强韧,经常被用来编制筐篮等农具,是一种有发展前途的经济植物.除此之外,簸箕柳也可用作固砂树种、河堤边的防浪林树种,具有一定的生态价值.对簸箕柳进行 R2R3-MYB 基因的全基因组鉴定,且对其进行该家族的多样性、进化和其功能研究,有助于了解该蛋白的扩展过程和机制以及随后该蛋白的亚功能化和新功能化机制.进一步探索 R2R3-MYB 参与簸箕柳生命生长过程的途径、方式、调控机制还可以为改良作物遗传及抗逆提供理论依据及支持,也可以直接作为基因资源加以利用.更为详细的关于系统发育分析、辅助基序和 DNA 结合特异性的发现也会为深入了解植物 R2R3-MYB 转录因子的进化史提供线索.除此之外,因簸箕柳枝条强劲、十分耐涝,是河堤边防浪林树种的良好选择,其具有耐性和韧性的枝条在水波中起到了很好的缓冲作用,但洪水或者大风天过后往往造成树木的倒伏,研究簸箕柳的抗重力刺激响应为后期提升簸箕柳的抗倒伏性能提供了一定的理论基础.

2 材料与方法

2.1 材料

从 PopGenIE^[28] 数据库 (<http://popgenie.org/>) 下载簸箕柳的基因组、蛋白质序列和注释文

件;从转录因子数据库 PlantTFDB^[29] v5. 0 (<http://planttfdb.gao-lab.org/>) 下载拟南芥 MYB 转录因子蛋白序列。

2.2 方法

2.2.1 簸箕柳 R2R3-MYB 转录因子家族的鉴定及其保守结构域分析 根据 Pfam 数据库^[30] (<http://pfam.xfam.org/>) 中 MYB 转录因子家族保守结构域的 HMM 文件 (PF00249), 用 HMMER 3. 0^[31] 软件对簸箕柳蛋白质序列进行本地搜索. 将拟南芥 R2R3-MYB 转录因子蛋白序列作为查询序列, 对簸箕柳蛋白质数据库进行 BLASTP 搜索. 将簸箕柳蛋白质序列传入 iTAK 数据库^[32] (<http://itak.feilab.net/cgi-bin/itak/index.cgi>) 进行线上转录因子家族鉴定. 将上述方法得到的蛋白序列整合、去除冗余后, 批量搜索 Pfam 数据库^[30] 和 NCBI-CDD 数据库^[33] (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) 去除结构域不完整的蛋白序列. 将鉴定出的蛋白进行重新命名: SsMYB001-SsMYB158.

将上一步鉴定出的簸箕柳 R2R3-MYB 转录因子家族的蛋白序列导入 ClustalW^[34] 进行多序列比对, 将得到的结果进行统计. 利用网络服务器 MEME^[35] (<http://meme-suite.org/tools/meme>) 绘制保守结构域 sequence logo 图形.

2.2.2 系统发育、基序组成、基因结构分析 将上一步鉴定出来的簸箕柳 R2R3-MYB 转录因子家族成员的序列导入 MAGA-X^[36] 软件, 采用 MAGA-X 自带的 ClustalW 进行多序列比对, 邻接法 (NJ), 检验参数 (Bootstrap) 设置为 1000, 进行系统发育分析, 将得到的结果进行亚组分类. 使用网络服务器 MEME^[35] 寻找簸箕柳 R2R3-MYB 转录因子家族成员的保守基序, 预期搜索数量设置为 20, 允许最小宽度为 6, 允许最大宽度为 50.

使用在线软件 iTOL^[37] (<https://itol.embl.de/>) 进行系统进化分析和基序组成分析可视化. 根据簸箕柳注释文件, 使用 TBtools^[38] 进行基因结构分析可视化.

2.2.3 亚细胞定位及理化性质分析 使用在线软件 CELLO^[39] (<http://cello.life.nctu.edu.tw/>) 对簸箕柳 R2R3-MYB 转录因子家族成员进行亚细胞定位预测. 采用本地 Perl 脚本对簸箕柳 R2R3-MYB 转录因子家族成员的基本理化性质进行分析, 如理论等电点 (pI)、分子量等.

2.2.4 染色体定位和基因复制分析 采用多重共

线性工具包 MCScanX^[40] 对簸箕柳 R2R3-MYB 转录因子家族成员进行基因复制事件分析, 利用 TBtools^[38] 和 Circos^[41] 工具绘制共线性图形. 采用 kaks_Calculator2. 0^[42] 软件计算基因对的 ka/ks 值. 使用 RepeatMasker^[43] 对簸箕柳基因组进行重复序列鉴定, 并统计重复序列和基因密度, 基于相对于染色体其他区域的高重复序列丰度和低基因含量, 预测了每条染色体的着丝粒位置.

2.2.5 GO 注释与基因表达分析 将 SsMYBs 与 NCBI 的 Nr 数据库进行 blast 比对, 结果导入 Blast2GO^[44], 最后使用 R 语言的 GGplot2 将注释结果可视化.

从 NCBI_SRA 数据库 (<https://www.ncbi.nlm.nih.gov/sra/>) 下载簸箕柳重力刺激相关 RNA-Seq 数据 (SRA 号: SRR9849616、SRR9849619-SRR9849623). 运用 HISAT2 + Stringtie^[45, 46] 流程计算各基因表达量, 再使用 DEseq2^[47] 对两组数据进行差异表达分析, 选择 logFC 绝对值大于 1 且 FDR 小于 0. 01 的基因为差异表达基因.

3 结果分析

3.1 簸箕柳 R2R3-MYB 转录因子家族鉴定及保守结构域分析

在簸箕柳中, 经 Pfam 数据库和 NCBI-CDD 数据库验证, 去除相关结构域不完整的蛋白序列, 一共鉴定出 158 个簸箕柳 R2R3-MYB 转录因子家族成员. 将鉴定出的蛋白进行重新命名: SsMYB001-SsMYB158.

在植物界, R2R3-MYB 是 MYB 蛋白中最大的一个亚群, 它含有由两个相邻 MYB 重复序列组成的高度保守的 DNA 结合域^[48]. 为了探索 SsMYBs (R2R3-MYB) 成员结构域的特征, 提取了 SsMYBs 成员的 MYB 结构域, 并对其氨基酸序列进行了多重序列比对, 结果可以用来检测 SsMYBs 蛋白的 R2 和 R3 重复序列中每个残基位置的保守性. 通过数据统计和 MEME 可视化得到结构域的一致性序列 (图 1) 和相关 sequence logo (图 2). 结果显示, SsMYBs 结构域平均约含有 105 个氨基酸残基 (图 2), 两个重复序列中插入或缺失频率很低.

根据先前的报道, R2 和 R3 重复序列具有特征性氨基酸, 包括一系列分布均匀且高度保守的 Trp(W) 色氨酸 (或疏水氨基酸) 残基^[7]. 在 158 个

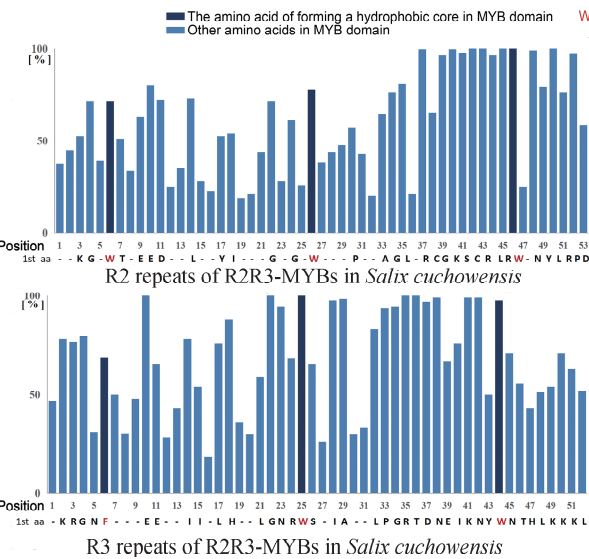


图 1 簸箕柳 R2R3-MYB 转录因子家族的保守结构域
横坐标为重复序列中氨基酸出现的相对位点,纵坐标为该位点占比最大氨基酸的百分比
Fig. 1 The conserve domain of *S. suchowensis* R2R3-MYB proteins

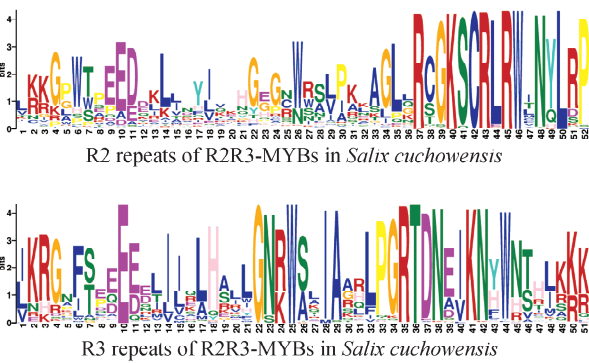


图 2 *SsMYBs* 保守结构域 sequence logo 图
Fig. 2 The sequence logo of conserve domain of *SsMYBs*

SsMYBs 蛋白中, R2 重复序列含有 3 个色氨酸残基, 其分别位于第 6、26 和 46 位, 这些色氨酸残基形成疏水核心, 是植物 MYB 结构域的典型标志。其中, 其中第一个色氨酸残基有部分被组氨酸(H)丝氨酸等残基替代, 第二个色氨酸残基有部分被天冬酰胺(N)替代。在 R3 重复序列中, 大多数成员的第一个色氨酸残基(位于 6)被苯丙氨酸(F)取代, 第二个色氨酸残基(位于 25)和第三个色氨酸残基(位于 44)在几乎所有 *SsMYBs* 中都很好地保存, 尤其是存在于所有成员中的第二个色氨酸残基。除了保守的 W 外, R2 重复序列末端还存在十二个高度保守的氨基酸残基: R-37、G-39、K-40、S-41、C-42 等, R3 重复序列中的 E-10、G-22、A-29 等也高度保守(图 2)。如图 1 所示, *SsMYBs* 结合域中的保守区域主要位于两个 R 重复序列的第二个和第

三个色氨酸残基(每个重复序列中 HTH 结构域的第三螺旋)之间。 *SsMYBs* 结合域中每个 R 重复序列的第一个和第二个色氨酸残基(第一螺旋)之间的氨基酸序列相对不保守。除此之外, 通过 sequence logo(图 2)可以更直观的看出 R3 重复片段结构域比 R2 要更为保守。相比之下, *SsMYBs* 结构域以外的区域在长度和氨基酸组成方面的保守性较差。

3.2 系统发育、基序组成、基因结构分析

通过多重序列比对和 Bootstrap 分析, 研究了 *SsMYBs* 转录因子家族成员的系统发育分析(图 3)。除 *SsMYB101* 未能很好地成簇外, 其余 157 个家族成员一共被分为了 23 个亚组, 不同的颜色代表了不同的亚组(图 3)。基因结构也可以提供进化信息^[49]。为了深入了解 *SsMYBs* 家族的进化, 对 158 个 *SsMYBs* 基因的内含子外显子的数目和排列进行了分析。结果表明, 同一亚组中的基因通常具有相似的内-外显子结构(图 3)。 *SsMYBs* 成员的基因结构存在显著差异, 包括外显子和内含子的数目和相对位置。其中, 内含子的数目从 0 到 12 不等, 大多数基因有两个内含子(65%)或一个内含子(23%)。在四个 *SsMYBs* 基因中发现了三个内含子: *SsMYB069*, *SsMYB027*, *SsMYB033*, *SsMYB053*。而在 *SsMYB059*、*SsMYB034*、*SsMYB058* 和 *SsMYB046* 中分别发现了 7、10、11 和 12 个内含子。此外, 10 个 *SsMYBs* 基因被发现是无内含子的。同一亚群中同源性最高的成员通常具有相同或平行的外显子/内含子模式, 表现出相似的数量、位置和外显子长度。例如, C2 中的七个 *SsMYBs* 拥有两个外显子和一个内含子、C4 中的两个 *SsMYBs* 没有内含子、C11 中的十个 *SsMYBs* 包括三个外显子和两个内含子。但是, 在 C12 和 C18 中各存在一个例外, *SsMYB034*、*SsMYB059* 的外显子和内含子的位置以及长度与同组其他成员存在显著差异, 成员间的遗传相似性较低。值得注意的是, 在每个亚组的末端节点中发现一对或多对高度同源的 *SsMYBs*, 这表明这些蛋白质具有相似的功能。

利用 MEME 程序预测了 *SsMYBs* 蛋白的保守基序, 共鉴定出 20 个不同的基序。这些基序被命名为基序 1~20, 同一亚组的 *SsMYB* 成员具有相似的基序结构。值得注意的是, motif14 只在 C7 亚组中存在, motif8、motif10、motif13 只在 C2 亚组中存在且十分保守, motif12 只在 C6 亚组中存在, motif17 只在 C10 亚组中存在, motif17 只在 C17

亚组中存在, motif15 只在 C18 亚组中存在. 这些亚组中特有的 motif 可能有助于其功能分化. 在同一亚组的成员, 通常具有相似的基序组成和相似的基因结构(图 3).

3.3 亚细胞定位及理化性质分析

SsMYBs 蛋白的长度在 164 到 1041 个氨基酸之间, 长度普遍位于 200~400 aa, 平均长度为 332 aa. 其中, SsMYB099 是最长的蛋白质, 含有 1041 个氨基

酸; SsMYB152 和 SsMYB106 是最短的蛋白质, 各含有 164 个氨基酸. SsMYBs 的分子量在 18.3 到 117.3 kD 之间, 平均分子量是 37.2 kD. SsMYBs 的理论等电点在 4.61 到 10.14 之间. 158 个 SsMYBs 蛋白呈现不规则的理论等电点特征, 75 个酸性蛋白(47.5%)和 83 个碱性蛋白(52.5%). 大部分 SsMYBs 蛋白(87%)的亚细胞定位结果显示其位于细胞核内的.

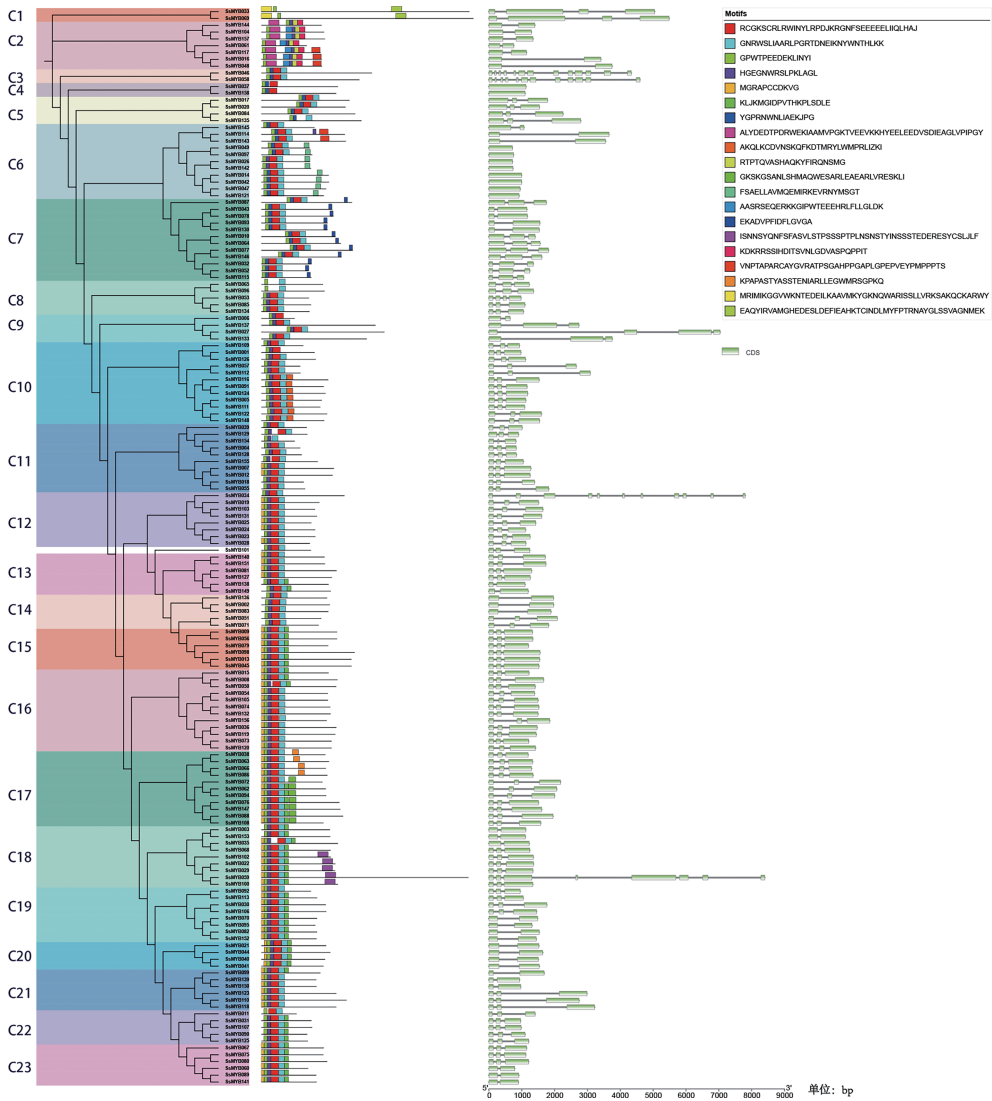


图 3 簸箕柳 R2R3-MYB 转录因子家族的保守结构域的系统发育树, motif 组成及基因结构
Fig. 3 Phylogenetic analysis, gene structure and motif composition of SsMYBs from *S. suchowensis*

3.4 染色体定位和基因复制分析

片段重复和串联复制是导致植物基因家族扩张的两个主要原因, 由同一基因座扩张而来的基因往往具有相似度较高的序列. 根据簸箕柳注释信息, 对 SsMYB 基因进行染色体定位分析, 有 7 个基因无法定位到簸箕柳 19 条染色体上, 而其余 151 个基因则不均匀、无规律地分布于 19 条染色体上(图 4). 其中 1 号染色体上该家族基因数目最多, 有

15 个; 11 号染色体上的最少, 只有一条. SsMYB 基因在 1、2、3 号染色体有较高的密度. 相反, 一些大的染色体区域缺乏 SsMYB 基因, 如 11 号染色体的中央和底部部分、14 号染色体的底部.

在 SsMYB 基因中存在两组串联复制, 涉及 5 个 SsMYB 基因, 分别位于 13 号染色体和 19 号染色体. 此外, 使用 MCScanX 对 SsMYB 基因进行片段重复或全基因组重复分析(图 5). 共鉴定出 83 个

片段重复对,涉及 101 个 *SsMYB* 基因. 在这 101 个 *SsMYB* 基因中,58 个基因只出现一次记录,43 个基因存在于不止一次的片段重复事件中. 这些结果表明,约(104)65%的 *SsMYB* 基因可能是由重复事件产生的,在簸箕柳 MYB 基因家族的扩展中起主要作用.

Ka/Ks 比值用于估计中性突变、纯化选择和有益突变之间的平衡. 计算了串联和片段复制基因对的 Ka/Ks 比值. 结果表明,所有 Ka/Ks 比值均小于 1,表明簸箕柳 MYB 复制成员在进化过程中可能经历了纯化选择压力.

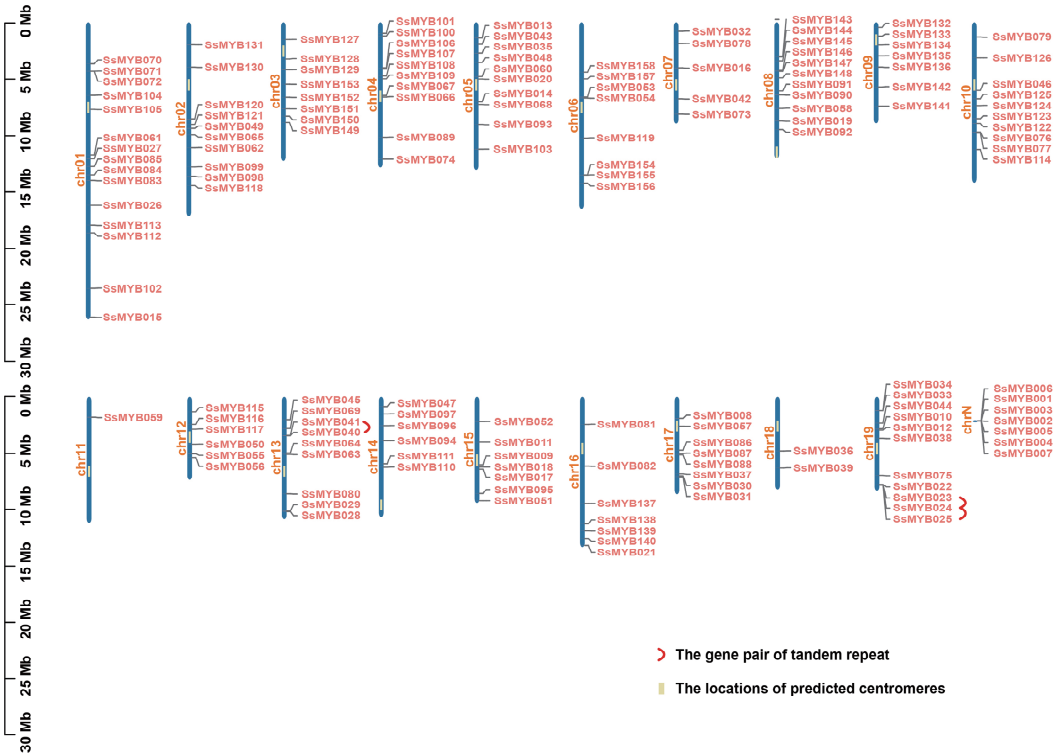


图 4 簸箕柳 R2R3-MYB 转录因子家族在染色体上的分布和串联复制事件
Fig. 4 Chromosome distribution and tandem duplication events for R2R3-MYB gene from *S. suchowensis*

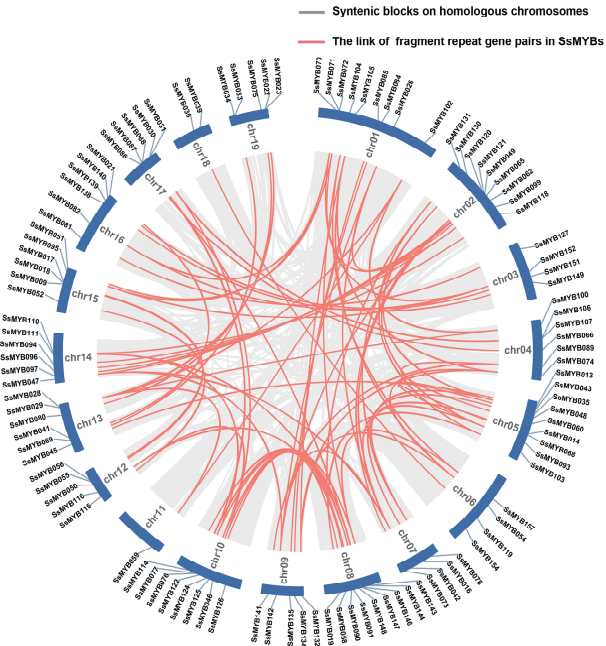


图 5 簸箕柳 R2R3-MYB 转录因子家族的片段复制事件
Fig. 5 Fragment duplication events for R2R3-MYB gene from *S. suchowensis*

3.5 GO 注释

相似的基因在不同物种中,其功能往往保守的. 为了探索 *SsMYBs* 的功能,对这 158 个成员进行了 GO 注释. GO 注释可以预测一个功能未知基因可能执行的分子功能(Molecular Function)、可能处于的细胞组分(Cellular Component)以及可能参与的生物学过程(Biological Process)^[44].

簸箕柳 MYB 成员的 GO 注释中分子功能结果显示(图 6),大部分成员被预测为起转录调控(46)、细胞分化(23)、气孔运动调节(7)等功能. 对于细胞组分,约 96%(151)的成员预测为处于细胞核内. 对于生物学过程,大部分成员被预测为参与 DNA 结合、转录调控区域特异性 DNA 结合、DNA 结合转录因子活性、转录协同调节活性、序列特异性 DNA 结合等生物学过程.

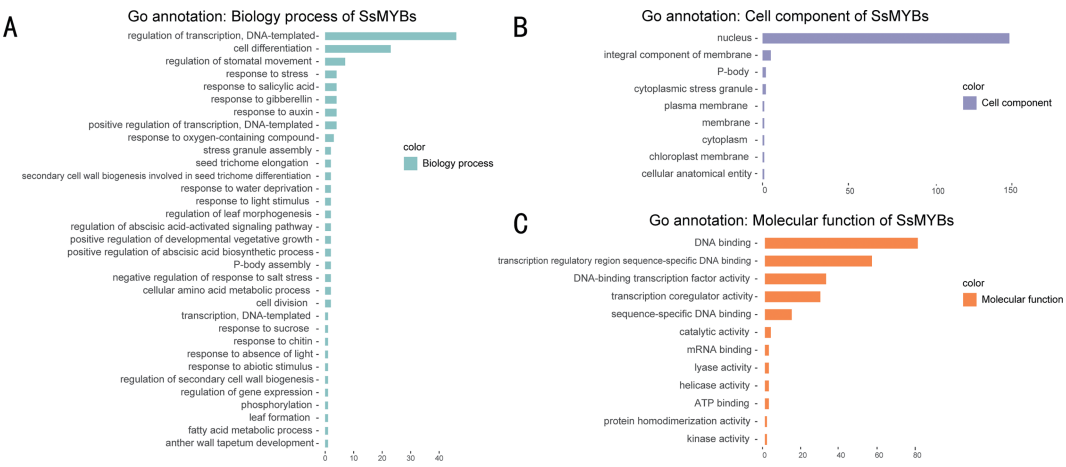


图 6 SsMYB 蛋白的 GO 注释
Fig. 6 Gene ontology annotation of SsMYB proteins

3.6 SsMYBs 在簸箕柳重力刺激下的差异表达分析

R2R3-MYB 家族成员参与植物对各种非生物和生物胁迫的防御和响应^[18-20],并在调节植物对包括吲哚乙酸^[21]、脱落酸^[21, 22]、生长素^[50]等植物激素线索的反应中发挥重要作用. 重力是调节植物生长和发育的一种普遍输入,各种植物谱系和器官已经进化不同的机制来调节相对于重力的生长方向^[51]. 在被子植物中,重力刺激的反应木被称为张力木,形成于茎的上侧,且产生张力,将茎向上拉. 张力木是通过维管形成层中细胞分裂速率的增加而产生的,其特征是导水导管元件数量减少,并且含有凝胶细胞壁层(G 层)的特殊张力木纤维被认为是张力产生的核心^[52].

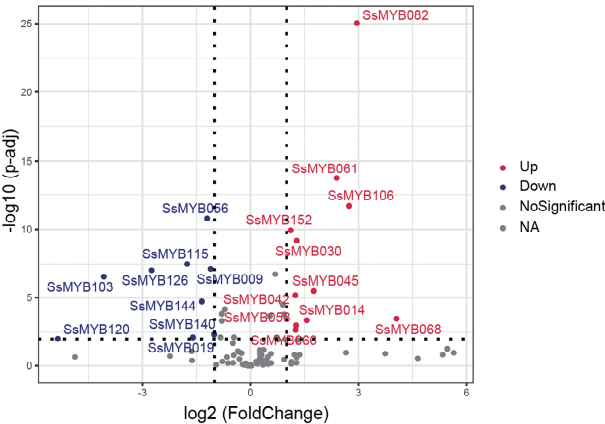


图 7 SsMYB 基因在簸箕柳重力刺激下的差异表达分析
Fig. 7 Expression analysis of SsMYB gene under gravistimulation of *S. suchowensis*

为了探究 SsMYBs 在簸箕柳应对重力刺激时可能发挥的作用,对正常茎和重力刺激下茎的张力

木进行差异表达分析,选取 $\log_{2}FC>1$ 或 <-1 和 $p_{adj}<0.01$ 的 MYB 基因为差异表达基因. 共筛选出 20 个差异表达 R2R3-MYB 基因(图 7),上调基因有 11 个,下调基因 9 个. 其中, SsMYB103 为显著下调基因,其在拟南芥中的同源基因 (*AtMYB15*)为拟南芥激活木质素生物合成基因所必需的调节因子^[53]. G 层的特征是低木质素、高纤维素^[52]. 推测 SsMYB103 参与了 G 层的形成进而推动了簸箕柳茎的背地性响应.

4 讨论

MYB 家族是植物最大的转录因子家族之一,参与了植物的多种重要生物学过程,如初级和次级代谢、发育过程、生物和非生物胁迫反应、细胞和器官形态发生以及细胞周期控制^[3, 54]. MYB 转录因子的进化和功能一直是研究的热点. MYB 转录因子已在一些植物物种中被鉴定与分析,如拟南芥、水稻、马铃薯和茶树^[55-58],但对于簸箕柳 MYB 家族的了解却很少.

本研究在簸箕柳中一共鉴定出 158 个 R2R3-MYB 转录因子家族成员. 对这 158 个成员先后进行了保守结构域、系统进化、基因结构、基序组成、亚细胞定位、理化性质、染色体定位、基因复制事件以及 GO 注释等分析.

多序列比对后,发现 SsMYBs 结构域平均约含有 105 个氨基酸残基,其中, R2 重复序列含有 3 个色氨酸残基来形成疏水核心, R3 重复序列中,其疏水核心中第一个色氨酸残基被苯丙氨酸取代,第二个和第三个色氨酸残基保守性很强. 相比之下, SsMYBs 结构域以外的区域在长度和氨基酸组成

方面的保守性较差,与之前其他植物该家族结构域的研究一致。随后进行系统进化分析,除 SsMYB101 未能很好地成簇外,其余 157 个家族成员一共被分为了 23 个亚组。在同一亚组中的基因通常具有相似的内含子-外显子结构和相似的基序组成。值得注意的是,某些 motif 只存在于特定的亚组,这些亚组特异性 motif 可能有助于其功能分化。簸箕柳染色体组装还不足够完善,这可能是导致 7 个 SsMYBs 无法定位到染色体上的原因,其余 151 个基因则在 19 条染色体上不均匀、无规律地分布。片段重复和串联复制是导致植物基因家族扩展的两个主要原因。在 SsMYBs 中存在两组串联复制事件,和 83 个片段重复对,这些结果表明,片段重复和串联复制在簸箕柳 MYB 基因家族的扩展中起主要作用。Ka/Ks 比值用于估计中性突变、纯化选择和有益突变之间的平衡。计算了串联和片段复制基因对的 Ka/Ks 比值。结果显示,所有 Ka/Ks 比值均小于 1,表明簸箕柳 MYB 复制成员在进化过程中可能经历了纯化选择压力。采用 Blast2GO^[44]对 SsMYBs 进行 GO 注释。大多数 SsMYBs 被预测位于细胞核内,约占 96%。SsMYBs 被预测参与簸箕柳很多的生物学过程,其中 46 位成员被预测为参与转录调控过程,23 位成员被预测为参与细胞分化过程等。大多数 SsMYBs 被预测起 DNA 绑定分子功能。除此之外,本研究对簸箕柳茎抗重力刺激应答相关的 SsMYB 进行了分析。共鉴定出 20 个差异表达基因。其中, SsMYB103 为显著下调基因,为拟南芥 *AtMYB15* 的同源基因,可能通过参与了 G 层的形成进而推动了簸箕柳茎的背地性响应。

综上所述,本研究对簸箕柳 R2R3-MYB 转录因子家族进行了全基因组水平的鉴定与分析,为簸箕柳 R2R3-MYB 基因功能研究和家族特性提供了全面系统的信息,并为改善簸箕柳的生长调节和抗倒伏性能提供了有价值的信息。

参考文献:

- [1] 牛义岭,姜秀明,许向阳. 植物转录因子 MYB 基因家族的研究进展[J]. 分子植物育种, 2016, 14: 2050.
- [2] Riechmann J L, Heard J, Martin G, *et al.* *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes [J]. Science, 2000, 290: 2105.
- [3] Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in *Arabidopsis thaliana* [J]. Curr Opin Plant Biol, 2001, 4: 447.
- [4] Ogata K, Kanei-Ishii C, Sasaki M, *et al.* The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation [J]. Nat Struct Biol, 1996, 3: 178.
- [5] Matus J T, Aquea F, Arce-Johnson P. Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes [J]. BMC Plant Biol, 2008, 8: 83.
- [6] Du H, Yang S S, Liang Z, *et al.* Genome-wide analysis of the MYB transcription factor superfamily in soybean [J]. BMC Plant Biol, 2012, 12: 106.
- [7] Du H, Feng B R, Yang S S, *et al.* The R2R3-MYB transcription factor gene family in maize [J]. PLoS One, 2012, 7: e37463.
- [8] Dubos C, Stracke R, Grotewold E, *et al.* MYB transcription factors in *Arabidopsis* [J]. Trends Plant Sci, 2010, 15: 573.
- [9] Feller A, Machemer K, Braun E L, *et al.* Evolutionary and comparative analysis of MYB and bHLH plant transcription factors [J]. Plant J, 2011, 66: 94.
- [10] Oppenheimer D G, Herman P L, Sivakumaran S, *et al.* A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules [J]. Cell, 1991, 67: 483.
- [11] Chai G, Wang Z, Tang X, *et al.* R2R3-MYB gene pairs in *Populus*: evolution and contribution to secondary wall formation and flowering time [J]. J Exp Bot, 2014, 65: 4255.
- [12] Brownfield L, Hafidh S, Borg M, *et al.* A plant germline-specific integrator of sperm specification and cell cycle progression [J]. PLoS Genet, 2009, 5: e1000430.
- [13] Millar A A, Gubler F. The *Arabidopsis* GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development [J]. Plant Cell, 2005, 17: 705.
- [14] Keller T, Abbott J, Moritz T, *et al.* *Arabidopsis* REGULATOR OF AXILLARY MERISTEMS1 controls a leaf axil stem cell niche and modulates vegetative development [J]. Plant Cell, 2006, 18: 598.
- [15] Mehrkens F, Kranz H, Bednarek P, *et al.* The *Arabidopsis* transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis

- [J]. *Plant Physiol*, 2005, 138: 1083.
- [16] Teng S, Keurentjes J, Bentsink L, *et al.* Sucrose-specific induction of anthocyanin biosynthesis in *Arabidopsis* requires the MYB75/PAP1 gene [J]. *Plant Physiol*, 2005, 139: 1840.
- [17] Bedon F, Grima-Pettenati J, Mackay J. Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*) [J]. *BMC Plant Biol*, 2007, 7: 17.
- [18] Oh J E, Kwon Y, Kim J H, *et al.* A dual role for MYB60 in stomatal regulation and root growth of *Arabidopsis thaliana* under drought stress [J]. *Plant Mol Biol*, 2011, 77: 91.
- [19] Cominelli E, Galbiati M, Vavasseur A, *et al.* A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance [J]. *Curr Biol*, 2005, 15: 1196.
- [20] He Y, Li W, Lv J, *et al.* Ectopic expression of a wheat MYB transcription factor gene, TaMYB73, improves salinity stress tolerance in *Arabidopsis thaliana* [J]. *J Exp Bot*, 2012, 63: 1511.
- [21] Seo P J, Xiang F, Qiao M, *et al.* The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in *Arabidopsis* [J]. *Plant Physiol*, 2009, 151: 275.
- [22] Abe H, Urao T, Ito T, *et al.* *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling [J]. *Plant Cell*, 2003, 15: 638.
- [23] Gocal G F, Sheldon C C, Gubler F, *et al.* GAMYB-like genes, flowering, and gibberellin signaling in *Arabidopsis* [J]. *Plant Physiol*, 2001, 127: 1682.
- [24] Li J, Li X, Guo L, *et al.* A subgroup of MYB transcription factor genes undergoes highly conserved alternative splicing in *Arabidopsis* and rice [J]. *J Exp Bot*, 2006, 57: 1263.
- [25] Jung C, Seo J S, Han S W, *et al.* Overexpression of AtMYB44 enhances stomatal closure to confer abiotic stress tolerance in transgenic *Arabidopsis* [J]. *Plant Physiol*, 2008, 146: 623.
- [26] Hartmann U, Sagasser M, Mehrtens F, *et al.* Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes [J]. *Plant Mol Biol*, 2005, 57: 155.
- [27] Devaiah B N, Madhuvanthi R, Karthikeyan A S, *et al.* Phosphate starvation responses and gibberellic acid biosynthesis are regulated by the MYB62 transcription factor in *Arabidopsis* [J]. *Mol Plant*, 2009, 2: 43.
- [28] Sjödin A, Street N R, Sandberg G, *et al.* The Populus genome integrative explorer (PopGenIE): a new resource for exploring the Populus genome [J]. *New Phytol*, 2009, 182: 1013.
- [29] Tian F, Yang D C, Meng Y Q, *et al.* PlantReg-Map: charting functional regulatory maps in plants [J]. *Nucleic Acids Res*, 2020, 48: D1104.
- [30] Mistry J, Chuguransky S, Williams L, *et al.* Pfam: The protein families database in 2021 [J]. *Nucleic Acids Res*, 2021, 49: D412.
- [31] Mistry J, Finn R D, Eddy S R, *et al.* Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions [J]. *Nucleic Acids Res*, 2013, 41: e121.
- [32] Zheng Y, Jiao C, Sun H, *et al.* iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases [J]. *Mol Plant*, 2016, 9: 1667.
- [33] Lu S, Wang J, Chitsaz F, *et al.* CDD/SPARCLE: the conserved domain database in 2020 [J]. *Nucleic Acids Res*, 2020, 48: D265.
- [34] Thompson J D, Higgins D G, Gibson T J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice [J]. *Nucleic Acids Res*, 1994, 22: 4673.
- [35] Bailey T L, Boden M, Buske F A, *et al.* MEME SUITE: tools for motif discovery and searching [J]. *Nucleic Acids Res*, 2009, 37: W202.
- [36] Kumar S, Stecher G, Li M, *et al.* MEGA X: molecular evolutionary genetics analysis across computing platforms [J]. *Mol Biol Evol*, 2018, 35: 1547.
- [37] Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments [J]. *Nucleic Acids Res*. 2019, 47: W256.
- [38] Chen C, Chen H, Zhang Y, *et al.* TBtools: an integrative toolkit developed for interactive analyses of big biological data [J]. *Mol Plant*, 2020, 13: 1194.
- [39] Yu C S, Lin C J, Hwang J K. Predicting subcellular localization of proteins for Gram-negative bacte-

- ria by support vector machines based on n-peptide compositions[J]. *Protein Sci*, 2004, 13: 1402.
- [40] Wang Y, Tang H, Debarry J D, *et al.* MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity[J]. *Nucleic Acids Res*, 2012, 40: e49.
- [41] Krzywinski M, Schein J, Birol I, *et al.* Circos: an information aesthetic for comparative genomics [J]. *Genome Res*, 2009, 19: 1639.
- [42] Wang D P, Wan H L, Zhang S, *et al.* Gamma-MYN: a new algorithm for estimating Ka and Ks with consideration of variable substitution rates[J]. *Biol Direct*, 2009, 4: 20.
- [43] Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences [J]. *Curr Protoc Bioinformatics*, 2009 (S25): 4.10.1.
- [44] Conesa A, Götz S, García-Gómez J M, *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research [J]. *Bioinformatics*, 2005, 21: 3674.
- [45] Kim D, Paggi J M, Park C, *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype [J]. *Nat Biotechnol*, 2019, 37: 907.
- [46] Kovaka S, Zimin A V, Pertea G M, *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2 [J]. *Genome Biol*, 2019, 20: 278.
- [47] Love M I, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 [J]. *Genome Biol*, 2014, 15: 550.
- [48] Li Q, Zhang C, Li J, *et al.* Genome-wide identification and characterization of R2R3MYB family in *Cucumis sativus* [J]. *PLoS One*, 2012, 7: e47576.
- [49] Li X, Guo C, Ahmad S, *et al.* Systematic analysis of MYB family genes in potato and their multiple roles in development and stress responses [J]. *Bio-molecules*, 2019, 9: 317.
- [50] Gibbs D J, Voß U, Harding S A, *et al.* AtMYB93 is a novel negative regulator of lateral root development in *Arabidopsis* [J]. *New Phytol*, 2014, 203: 1194.
- [51] Gerttula S, Zinkgraf M, Muday G K, *et al.* Transcriptional and hormonal regulation of gravitropism of woody stems in populus[J]. *Plant Cell*, 2015, 27: 2800.
- [52] Mellerowicz E J, Gorshkova T A. Tensional stress generation in gelatinous fibres: a review and possible mechanism based on cell-wall structure and composition [J]. *J Exp Bot*, 2012, 63: 551.
- [53] Kim S H, Lam P Y, Lee M H, *et al.* The *Arabidopsis* R2R3 MYB transcription factor MYB15 is a key regulator of lignin biosynthesis in effector-triggered immunity [J]. *Front Plant Sci*, 2020, 11: 583153.
- [54] Ambawat S, Sharma P, Yadav N R, *et al.* MYB transcription factor genes as regulators for plant responses: an overview [J]. *Physiol Mol Biol Plants*, 2013, 19: 307.
- [55] Chen Y H, Yang X Y, He K, *et al.* The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family [J]. *Plant Mol Biol*, 2006, 60: 107.
- [56] Smita S, Katiyar A, Chinnusamy V, *et al.* Transcriptional regulatory network analysis of MYB transcription factor family genes in rice [J]. *Front Plant Sci*, 2015, 6: 1157.
- [57] Li Y, Lin-Wang K, Liu Z, *et al.* Genome-wide analysis and expression profiles of the StR2R3-MYB transcription factor superfamily in potato (*Solanum tuberosum* L.) [J]. *Int J Biol Macromol*, 2020, 148: 817.
- [58] Arce-Rodríguez M L, Martínez O, Ochoa-Alejo N. Genome-Wide Identification and Analysis of the MYB Transcription Factor Gene Family in Chili Pepper (*Capsicum* spp.) [J]. *Int J Mol Sci*, 2021, 22: 2229.

引用本文格式:

中文: 王岚春, 沈方圆, 欧阳丹, 等. 簸箕柳 R2R3-MYB 转录因子家族全基因组分析[J]. 四川大学学报: 自然科学版, 2022, 59: 036001.

英文: Wang L C, Shen F Y, Ouyang D, *et al.* Genome-wide analysis of the R2R3-MYB transcription factor family in *Salix suchowensis* [J]. *J Sichuan Univ: Nat Sci Ed*, 2022, 59: 036001.