

DOI: 10.3724/SP.J.1005.2013.01307

杨树 *MIR169* 基因家族分子进化分析

刘志祥^{1,2,3}, 曾超珍¹, 谭晓风^{2,3}

1. 中南林业科技大学生命科学与技术学院, 长沙 410004;
2. 中南林业科技大学林学院, 长沙 410004;
3. 经济林育种与栽培国家林业局重点实验室, 长沙 410004

摘要: MicroRNA(miRNA)是真核生物中普遍存在的一类参与基因表达调控的小分子 RNA。*ptc-MIR169* 基因家族有 33 个成员, 是杨树中规模最大的 miRNA 基因家族。研究 *MIR169* 基因家族的进化对揭示杨树 miRNA 基因的进化机制具有重要的意义。文章对毛果杨 *MIR169* 基因家族的分子系统发育、基因倍增模式、表达分化和靶基因进行了分析。结果表明, 染色体大片段重复和串联重复在毛果杨 *MIR169* 基因家族扩张中均具有重要作用; *MIR169* 基因家族在表达方式上已经出现了较大的分化; *MIR169* 基因家族可能在杨树中已经形成了复杂的调控网络, 对杨树的生长发育和适应性等具有重要的调控作用。文章为杨树及杨柳科相关植物中 miRNA 基因家族的分子进化研究提供了参考。

关键词: 毛果杨; *MIR169* 基因家族; 分子进化

Molecular evolution of the poplar *MIR169* gene family

LIU Zhi-Xiang^{1,2,3}, ZENG Chao-Zhen¹, TAN Xiao-Feng^{2,3}

1. College of Life Science and Technology, Central South University of Forestry and Technology, Changsha 410004, China;
2. College of Forestry, Central South University of Forestry and Technology, Changsha 410004, China;
3. Key Lab of Non-Wood Forest Products of State Forestry Administration, Changsha 410004, China

Abstract: MicroRNAs (miRNAs) are a class of short non-coding RNAs found widely in eukaryotic organisms. The *ptc-MIR169* gene family, which consists of 33 members, is the largest miRNA gene family in poplar (*Populus trichocarpa*). It is significant to analyze the evolution of the *ptc-MIR169* gene family in order to understand the evolutionary mechanisms of miRNAs in poplar. In the present study, we investigated the molecular phylogeny, duplication, expression and target genes of the *MIR169* gene family in poplar. Both tandem duplications and chromosome segmental duplications contributed to the expanding of *ptc-MIR169* gene family, and the expression patterns diversified obviously among the gene family. These findings suggest that the *ptc-MIR169* gene family is involved in complex regulatory networks, and plays significant roles in development and stress response in poplar. This paper provides a reference for the evolutionary study of miRNAs in poplar and related species in Salicaceae.

Keywords: *Populus trichocarpa*; *MIR169* gene family; molecular evolution

收稿日期: 2013-05-21; 修回日期: 2013-07-03

基金项目: 湖南省自然科学基金项目(编号: 13JJ3095)和中南林业科技大学青年基金项目(编号: QJ2011042B)资助

作者简介: 刘志祥, 博士研究生, 讲师, 研究方向: 植物分子遗传学。Tel: 0731-85623486; E-mail: liuzx@csuft.edu.cn

通讯作者: 谭晓风, 博士, 教授, 研究方向: 林业生物技术。E-mail: nwfp@foxmail.com

网络出版时间: 2013-8-6 18:48:28

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20130806.1848.001.html>

MicroRNA(miRNA)是一类广泛存在于动植物的内源调控小分子RNA,在植物中miRNA主要通过序列互补介导靶mRNA的降解从而抑制靶基因的表达^[1]。miRNA广泛参与植物发育调控^[2]与逆境胁迫响应^[3,4]。对miRNA的研究有助于揭示很多生命现象的分子机理,并可能应用于作物和林木遗传改良^[3]。

杨树是重要的用材树种,随着毛果杨(*Populus trichocarpa*)基因组测序的完成^[5]及相关遗传学和生物信息学研究的开展^[6],目前杨树已成为公认的本模式植物。对于许多生物学问题,例如木材形成、多年生生长、季节性生长等,杨树作为研究材料是拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)等无法替代的,在开花调控、生物进化、逆境胁迫响应等方面,杨树也具有重要的研究价值^[7,8]。根据miRBase(www.mirbase.org, release 19)的数据,在毛果杨中已经发现了323个miRNA基因^[9-11]。

在miRBase数据库中,拟南芥、水稻、毛果杨中miRNA基因家族成员数量平均分别为1.54、1.78、3.11;这3种植物共有的保守而古老的miRNA基因家族(从MIR156至MIR408^[12])成员平均数量分别为4.05、6.90、8.29,从而可以看出,毛果杨miRNA基因家族规模普遍大于拟南芥和水稻。植物miRNA基因家族成员在基因扩张后经历了明显的重排,导致在基因组中多呈分散分布^[13]。这说明杨树miRNA基因家族的进化过程更加复杂,具有重要的研究价值。

MIR169基因家族是植物界迄今所发现的较大的miRNA基因家族。miRBase数据库中部分植物MIR169基因家族成员数分别为:拟南芥14个,玉米(*Zea mays*)18个,蒺藜苜蓿(*Medicago truncatula*)18个,葡萄(*Vitis vinifera*)25个,毛果杨33个,均为这些物种中最大的miRNA基因家族^[14]。为什么在多种植物中MIR169基因家族都是最大的miRNA基因家族?这与其功能有无关联?对这些问题目前还没有相关的文献报道。本文通过分析杨树MIR169基因家族分子进化过程和表达分化情况,以期研究植物MIR169基因家族的进化和功能提供基础。

1 材料和方法

1.1 数据库搜索

毛果杨基因组数据从Phytozome数据库([http://](http://www.phytozome.net/)

www.phytozome.net/, Phytozome v9.0, Poplar v3.0)查询和下载, MIR169基因家族信息从miRBase查询和下载。

1.2 基因倍增模式分析

基因倍增模式分析参考Maher等^[15]的方法并略加修改。具体方法如下:利用Phytozome毛果杨基因组数据库的信息将MIR169基因家族成员定位到染色体上。同一家族的不同miRNA基因位于同一或相邻的基因间区域(Intergenic region),则这些miRNA基因是串联重复的产物。取miRNA基因上下游各10个蛋白质编码基因,将每一个蛋白质编码基因与毛果杨基因组数据库中73 013个蛋白质序列进行BLASTP比对,获取与之最佳非自我匹配基因(The best no-self matches),设置E值范围为小于0.001。然后统计MIR169基因家族成员之间最佳非自我匹配基因对的数量,若两个MIR169基因家族成员的侧翼蛋白质编码基因中有1对以上最佳非自我匹配基因对,则这些miRNA基因为染色体片段重复的产物。

1.3 基因倍增时间估算

由于miRNA基因是非编码基因,所以只能采用倍增块(Duplication block)中的保守蛋白质编码基因对MIR169基因家族倍增时间进行估算^[15]。首先对倍增块中保守的蛋白质编码基因采用Clustal X进行氨基酸序列联配(Alignment),然后以氨基酸序列联配结果为指导,进行编码序列联配,联配结果保存为AXT文件格式,并使用KaKs_Calculator软件^[16]采用YN算法^[17]计算序列间同义替换率(Ks)。当Ks>2.0时存在饱和现象,故大于2.0的Ks值均舍弃^[18]。每一倍增块中保守蛋白质基因的平均Ks值用于计算倍增时间,其计算公式为: $D=Ks/2E$ 。

1.4 多序列比对与系统发育分析

MIR169基因多序列比对采用ClustalW 2.1进行,比对采用空位开放罚分(Gap opening penalty)为5、空位延伸罚分(Gap extension penalty)为2,采用较低的空位罚分的目的是增强对miRNA前体序列中非保守的环区序列的比对。比对结果输入MEGA5,分别采用邻接法、最大似然法构建系统进化树,采用自展法(Bootstrap method)检验进化树,重复值(Replicates)设定为1 000。

1.5 基于 EST 的表达分析

将杨树 *MIR169* 基因家族前体序列提交NCBI与杨树EST数据进行比对, 参数为默认值, 搜索到的结果如果与miRNA匹配覆盖率为 100%、相似度 >90%、 E 值 < 10^{-10} 、匹配方式为正链-正链匹配, 则作为该miRNA的表达序列, 并按照组织来源进行统计。

1.6 基于小 RNA 高通量测序的表达分析

从毛果杨 4 种不同来源的组织样品的SOLiD Total RNA-Seq高通量测序结果^[11]中提取 *MIR169* 基因家族成员原始测序读数, 并按照如下公式进行均一化处理: 表达量 = 原始读数 / 与基因组匹配的总读数 $\times 10^8$, 以比较各基因在不同组织样品中的表达量。

1.7 靶基因预测

杨树 *MIR169* 基因家族靶基因预测采用 psRNA Target(<http://plantgrn.noble.org/psRNATarget/>), 将 miRNA 成熟序列提交 psRNATarget, 靶基因搜索范围设置为 *Populus trichocarpa* transcript(phytozome v8.0, genome V3.0, internal number 210), 参数为默认值。

2 结果与分析

2.1 毛果杨 *MIR169* 基因家族部分成员成簇分布

从 miRBase 查询到毛果杨 *MIR169* 基因家族成员序列, 提交毛果杨基因组数据库进行比对, 从而获得基因家族成员在基因组上的位置。杨树 *MIR169* 基因家族 32 个成员分布于 12 条染色体, 另有 1 个

成员 *ptc-MIR169g* 定位于 scaffold_127, 其中 20 个成员成簇分布(图 1)。染色体 Chr.01 上有 2 个基因簇, 染色体 Chr.05、Chr.06、Chr.07、Chr.15、Chr.17、Chr.18 上各有 1 个基因簇。在染色体 Chr.18 上 *ptc-MIR169aa*、*ptc-MIR169n*、*ptc-MIR169z* 和 *ptc-MIR169p* 聚为最大的一簇。

2.2 *MIR169* 基因家族的倍增模式

2.2.1 串联重复

在Phytozome数据库中通过Gbrowse查看成簇分布的 *MIR169* 基因家族成员与侧翼蛋白质编码基因的位置关系, 同一家族的不同成员如果位于同一个或相邻的基因间区域, 则这些成员为串联重复关系。按照此标准, 杨树 *MIR169* 基因家族有 20 个成员形成了 8 个串联重复组(表 1)。

在 8 组串联重复基因中, 5 组由 2 个基因组成, 2 组由 3 个基因组成, 1 组由 4 个成员构成。仅第三组存在反向串联重复, 其余 7 组均为正向串联重复。染色体 Chr.05 上的 *ptc-MIR169y* 与 *ptc-MIR169ac* 间隔长度仅 98 bp, 另外, *ptc-MIR169q* 与 *ptc-MIR169w*、*ptc-MIR169i* 与 *ptc-MIR169j*、*ptc-MIR169y* 与 *ptc-MIR169ac*、*ptc-MIR169l* 与 *ptc-MIR169m* 其间距

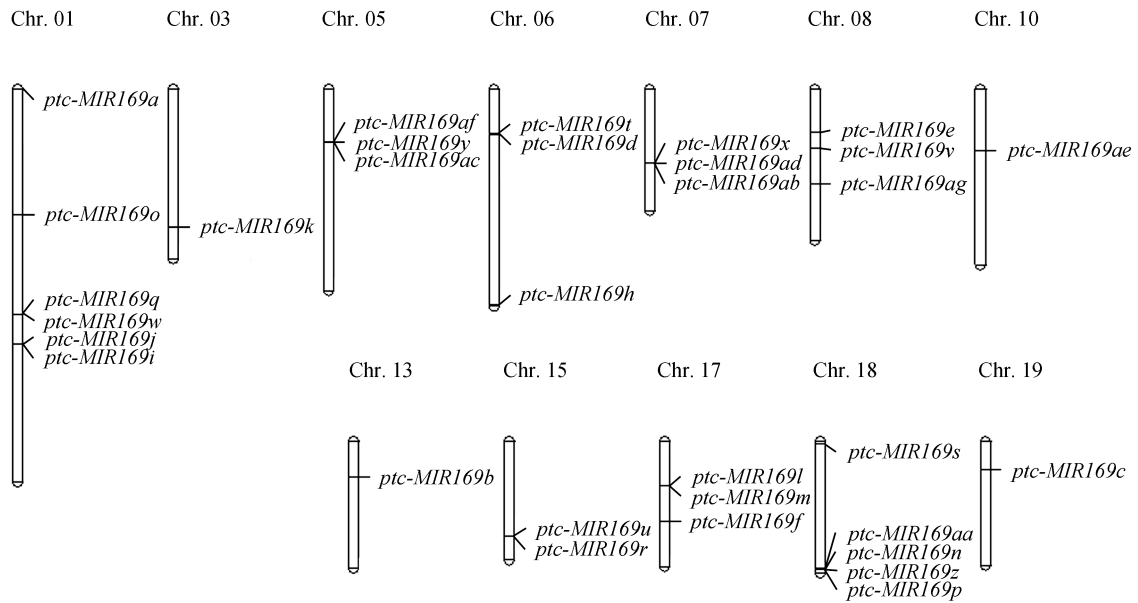


图 1 *MIR169* 基因家族成员在毛果杨基因组中的分布

表 1 *MIR169* 基因家族串联重复

串联重复组	基因名称	染色体	起始位置	终止位置	链	与前一基因的间隔长度(bp)
1	<i>ptc-MIR169q</i>	Chr.01	28934136	28934250	+	-
	<i>ptc-MIR169w</i>	Chr.01	28934380	28934488	+	129
2	<i>ptc-MIR169j</i>	Chr.01	32753148	32753284	-	-
	<i>ptc-MIR169i</i>	Chr.01	32753432	32753554	-	147
3	<i>ptc-MIR169af</i>	Chr.05	6806058	6806164	-	-
	<i>ptc-MIR169y</i>	Chr.05	6809763	6809885	+	3 598
	<i>ptc-MIR169ac</i>	Chr.05	6809984	6810089	+	98
4	<i>ptc-MIR169t</i>	Chr.06	5781772	5781879	-	-
	<i>ptc-MIR169d</i>	Chr.06	5794392	5794502	-	12 512
5	<i>ptc-MIR169x</i>	Chr.07	9527690	9527807	+	-
	<i>ptc-MIR169ad</i>	Chr.07	9531522	9531622	+	3 714
	<i>ptc-MIR169ab</i>	Chr.07	9532261	9532367	+	638
6	<i>ptc-MIR169u</i>	Chr.15	12187400	12187521	+	-
	<i>ptc-MIR169r</i>	Chr.15	12195358	12195476	+	7 836
7	<i>ptc-MIR169l</i>	Chr.17	5766969	5767111	-	-
	<i>ptc-MIR169m</i>	Chr.17	5767273	5767391	-	161
	<i>ptc-MIR169aa</i>	Chr.18	16370855	16370963	-	-
8	<i>ptc-MIR169n</i>	Chr.18	16375673	16375779	-	4 709
	<i>ptc-MIR169z</i>	Chr.18	16381172	16381289	-	5 392
	<i>ptc-MIR169p</i>	Chr.18	16384470	16384633	-	3 180

非常小(98 bp~161 bp), 而其间隔区序列分析表明, 没有启动子元件特征, 所以它们应该是作为一个转录单位进行转录的。绝大部分基因家族成员位于基因间区域, 另有 6 个成员(*MIR169ae/af/c/h/n/y*)位于

蛋白质基因内部(包括 5'UTR、3'UTR)。

2.2.2 染色体片段重复

从Phytozome数据库分别获取杨树*MIR169* 基因家族各成员侧翼 20 个蛋白质编码基因, 通过

BLASTP 比对分析, 寻找它们在杨树中的最佳匹配基因, 若两个 *MIR169* 基因家族成员的侧翼蛋白质编码基因中有 1 对以上在杨树中为最佳匹配关系, 则这两个 *MIR169* 基因家族成员所定位的染色体区段是基因组进化过程中染色体大片段重复的产物, 即这些 miRNA 基因位于染色体倍增块 (Duplicated block) 上。分析表明, 杨树 *MIR169* 基因家族 33 个成员中有 30 个分布于 13 对倍增块上 (表 2), 如果将两侧保守蛋白质基因数大于或等于 4 定义为严谨标准 (Strict definition) 倍增块, 将两侧保守蛋白质基因数大于或等于 1 定义为宽松标准 (Loose definition) 倍增块^[15], 则 13 对倍增块中有 7 对为严谨标准倍增块, 24 个基因家族成员定位于严谨标准倍增块。

在 13 对倍增块中存在交叉匹配的情况, 即一个染色体片段同时与 2 个或 3 个染色体片段构成倍增块, 例如: *MIR169l/m-MIR169i/j-MIR169k* 倍增块 (图 2A)、*MIR169g-MIR169aa/n/p/z-MIR169d/t-MIR169s-MIR169h* 倍增块 (图 2B)、*MIR169a-MIR169c-MIR169b-MIR169f-MIR169e* 倍增块 (图 2C)。说明这些倍增块可能是多次染色体片段重复的产物。

2.3 基因倍增时间估算

在蛋白质编码基因的进化过程中, 不导致氨基酸序列改变的核苷酸变化称为同义突变 (Synonymous mutation), 核苷酸的同义替换突变被认为是不受选择压力的影响, 因而同义替换是随时间而累积的^[19]。对于侧翼保守蛋白质数大于 4 的重复片段, 分别计算保守蛋白质基因的同义替换率 (K_s), 计算平均 K_s 值, 然后以 $D=K_s/2E$ 计算分化时间。目前杨树分子进化速率 (E) 目前还没有明确。已有的分子进化分析结合化石证据表明, 杨树分子进化速率较拟南

表 2 大片段重复及其时间估算

倍增基因对	侧翼保守蛋白质基因对数量	平均 K_s	K_s 标准差	倍增时间 (MYA)
<i>MIR169i/j</i> 与 <i>MIR169l/m</i>	9	0.242693571	0.047215946	48.54
<i>MIR169ac/afly</i> 与 <i>MIR169ab/ad/x</i>	9	0.415415833	0.175370557	83.08
<i>MIR169q/w</i> 与 <i>MIR169v</i>	7	0.336759600	0.232438806	67.35
<i>MIR169d/t</i> 与 <i>MIR169aa/n/p/z</i>	7	0.254755600	0.039950905	50.95
<i>MIR169h</i> 与 <i>MIR169s</i>	6	0.286655000	0.047744179	57.33
<i>MIR169b</i> 与 <i>MIR169c</i>	6	0.232426500	0.027394104	46.49
<i>MIR169aa/n/p/z</i> 与 <i>MIR169g</i>	6	0.083147575	0.024711877	16.63
<i>MIR169a</i> 与 <i>MIR169b</i>	1	-	-	-
<i>MIR169o</i> 与 <i>MIR169ag</i>	1	-	-	-
<i>MIR169b</i> 与 <i>MIR169f</i>	1	-	-	-
<i>MIR169d/t</i> 与 <i>MIR169s</i>	1	-	-	-
<i>MIR169e</i> 与 <i>MIR169f</i>	1	-	-	-
<i>MIR169i/j</i> 与 <i>MIR169k</i>	1	-	-	-

芥慢, 约为拟南芥的 $1/6$ ^[5], 而拟南芥的 E 值为 1.5×10^{-8} 替换/同义替换位点/年^[20], 故本文中杨树 E 值采用拟南芥 E 值的 $1/6$, 即 2.5×10^{-9} 替换/同义替换位点/年。大片段重复时间估算结果如表 2 所示。

杨树与拟南芥在 100~120 百万年前 (Million years ago, MYA) 分歧之后, 经历了一次全基因组重复事件, 即杨柳科重复事件 (Salicoid duplication event), 发生的时间约为 60~65 MYA^[5]。对杨树 *MIR169* 基因家族所经历的大片段重复时间估算结

果显示, *MIR169aa/n/p/z* 与 *MIR169g* 所在重复片段重复时间较晚, 约为 16.6 MYA; *MIR169ac/afly* 与 *MIR169ab/ad/x* 所在重复片段重复时间较早, 约为 83.1 MYA; 而其余的大片段重复时间与杨柳科重复事件发生时间较接近。基因组定位分析表明, *MIR169i/j* 与 *MIR169l/m*、*MIR169ac/afly* 与 *MIR169ab/ad/x*、*MIR169d/t* 与 *MIR169aa/n/p/z*、*MIR169h* 与 *MIR169s* 定位于杨柳科重复所产生的染色体重复片段上^[5], 说明这部分成员可能也是这一

次全基因组重复的产物。

2.4 分子系统进化树的构建

采用最大似然法和邻接法分别构建毛果杨 *MIR169* 基因家族的分子系统进化树, 两种方法构建的进化树拓扑结构基本一致, 邻接法构建的进化树如图 3 所示。将分子系统进化树与基因家族倍增模式分析结果进行比较分析, 可以发现: 串联重复基因 *MIR169u* 与 *MIR169r* 在进化树上分歧较晚, 这说明 *MIR169u* 与 *MIR169r* 之间的串联重复发生的时间较晚; 而倍增基因对 *MIR169i/j* 与 *MIR169l/m* 中, *MIR169i* 与 *MIR169m* 分歧较晚, *MIR169j* 与 *MIR169l* 分歧较晚, 这说明在这一对倍增块中串联重复是先发生的, 染色体大片段重复是后发生的; 在倍增基因对 *MIR169q/w* 与 *MIR169v* 中, *MIR169w* 与 *MIR169v* 分歧较晚, 而 *MIR169q* 与前两者分歧较早, 说明可能 *MIR169q* 与 *MIR169w* 之间的串联重复早于 *MIR169q/w* 与 *MIR169v* 之间的大片段重复, 而 *MIR169q* 的另一个大片段重复的基因拷贝在进化过程中丢失了。

从进化树还可以看出, 杨树 *MIR169* 基因明显分为 2 个亚族 A 与 B, 其成员分别与 miRBase 数据库中 *MIR169_1* 和 *MIR169_2* 两个家族基本对应。例外的是 *ptc-MIR169e* 在 miRBase 中归类于 *MIR169_1*, 而在进化树中归于亚族 B。这可能是由于基于不同的算法和模型分析导致的差异。

在基因家族扩增模式分析中所发现的串联重复和染色体大片段重复全部发生在 *MIR169_1* 和 *MIR169_2* 两个家族内部。这说明 *MIR169_1* 和 *MIR169_2* 两个家族的分化在时间上早于上述的串联重复和染色体大片段重复事件。在其它双子叶植物, 例如拟南芥、蒺藜苜蓿 (*M. truncatula*)、大豆 (*Glycine max*)、葡萄 (*V. vinifera*), 以及单子叶植物, 例如水稻、高粱 (*Sorghum bicolor*)、玉米 (*Z. mays*) 中

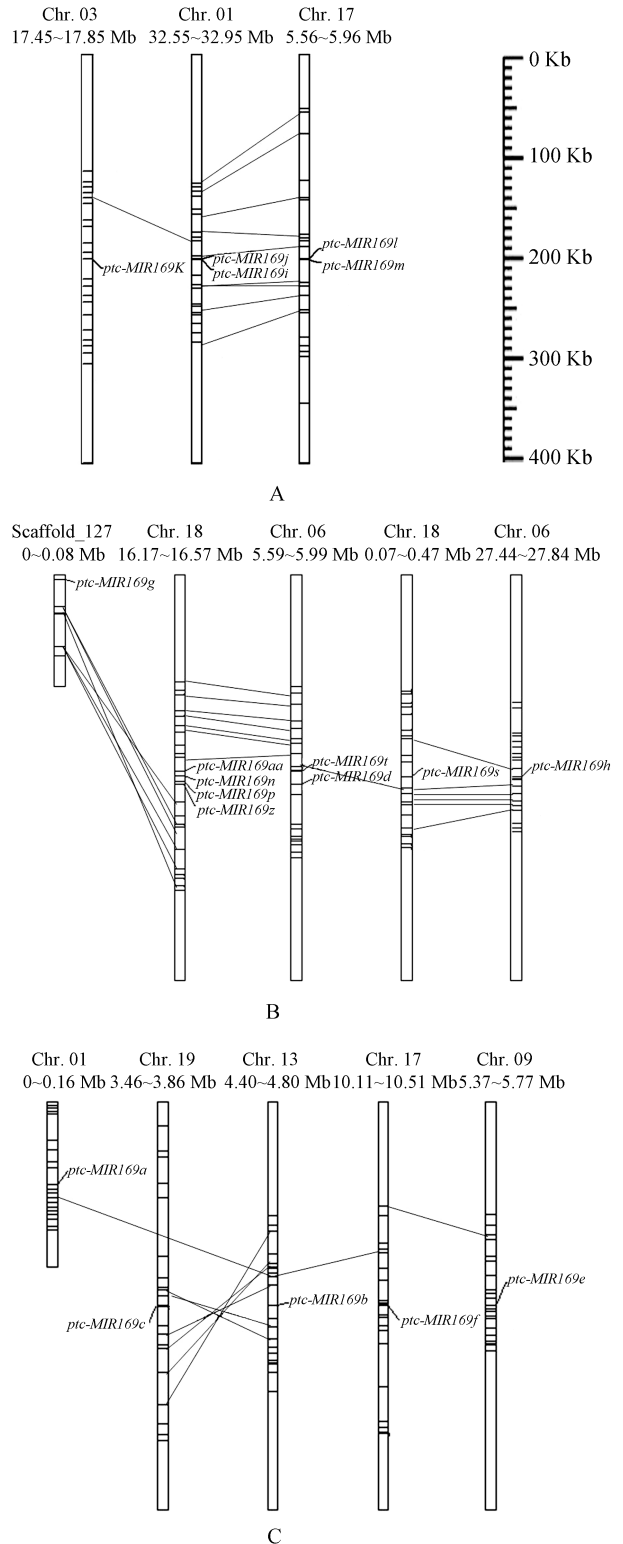


图 2 毛果杨 *MIR169* 基因家族侧翼保守蛋白质基因
A : *MIR169k-MIR169i/j-MIR169l/m* 倍增块; B : *MIR169g-MIR169aa/n/p/z-MIR169d/t-MIR169s-MIR169h* 倍增块; C : *MIR169a-MIR169c-MIR169b-MIR169f-MIR169e* 倍增块。

都已经出现了 *MIR169_1* 家族和 *MIR169_2* 家族的分化, 所以它们之间的分化早于单子叶和双子叶植物的分化。

在 miRBase 中还有成员较少的 2 个家族, 即 *MIR169_3* 家族(包括 *ptc-MIR169aa* 和 *ptc-MIR169t*) 和 *MIR169_4* 家族(包括 *ptc-MIR169o*), *MIR169_3* 家族的 2 个成员在进化树中均归于亚族 B, 而 *MIR169_4* 家族的成员 *ptc-MIR169o* 在进化树中归于亚族 A。 *MIR169_3* 和 *MIR169_4* 家族目前仅在毛果杨中发现, 而在其它植物中还没有发现。说明 *MIR169_3* 和 *MIR169_4* 家族分别从 *MIR169_2* 和 *MIR169_1* 家族中分化出来, 而且发生的时间较晚, 可能发生在杨树与拟南芥分歧(100~120 MYA)之后。

2.5 杨树 *MIR169* 基因家族的表达情况

将 *MIR169* 基因家族成员提交 NCBI, 通过 BLAST 比对搜索与之匹配的 EST 序列, 搜索范围包括所有杨属物种以及杂交种, 结果共搜索到 13 条杨树 *MIR169* 基因家族表达序列, 分别对应于 *MIR169* 基因家族中的 5 个成员(表 3)。

利用已有的毛果杨 sRNA (Small RNA) 高通量测序数据^[11], 从中提取 *MIR169* 基因家族表达数据, 并进行均一化处理。结果表明杨树 *MIR169* 基因家族中有 13 个成员在高通量测序中检测到表达, 且其表达方式已经表现出较大的分化(图 4)。 *MIR169a/c* 在混合样品中均具有最高的表达量; *MIR169d/g/h/s* 在混

合样品中具有较高的表达量, 在机械胁迫处理的木质部(MTX)中也有明显的表达; *MIR169n* 在木质部中具有较高的表达量; *MIR169u* 在 3 个样品(叶片、木质部、MTX)中均具有最高的表达活性。

综合上述 2 种表达分析的结果, 杨树 *MIR169* 基因家族 33 个成员中共有 17 个成员在不同组织样品中检测到了表达, 且其表达方式已经呈现明显差异。基因芯片和小 RNA 深度测序也表明, 胡杨 (*Populus euphratica*) *MIR169* 基因家族成员在干旱处理条件下也呈现不一样的表达水平变化^[21]。这些结果表明杨树 *MIR169* 基因家族已经表现出较高分化的表达方式分化。

2.6 杨树 *MIR169* 靶基因预测结果

将 *MIR169* 成熟序列提交 psRNA Target 进行靶基

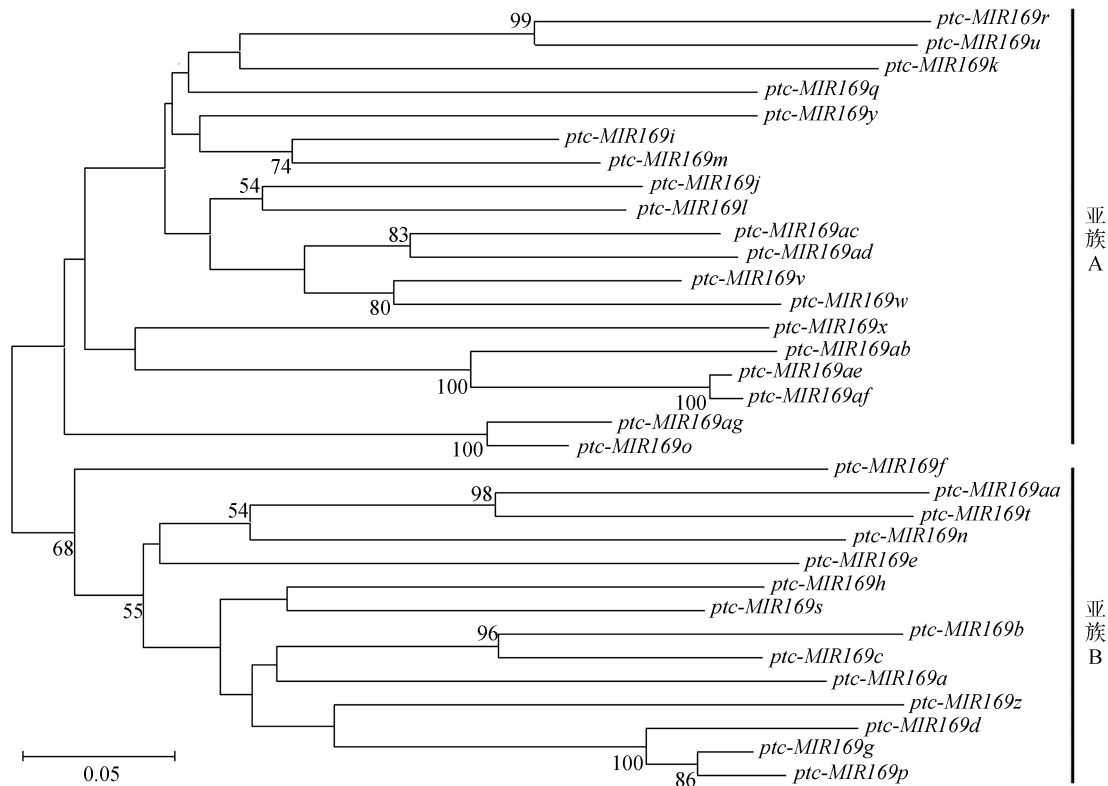


图 3 邻接法构建的毛果杨 *MIR169* 基因家族分子系统进化树

表 3 杨树 *MIR169* 基因家族成员在各组织中的表达序列标签数量

基因	各组织中的表达序列标签数量			
	形成层	吸水种子	休眠芽	花
<i>ptc-MIR169ae/af</i>	3	-	-	-
<i>ptc-MIR169b</i>	-	3	-	-
<i>ptc-MIR169c</i>	-	1	-	-
<i>ptc-MIR169k</i>	-	-	1	5

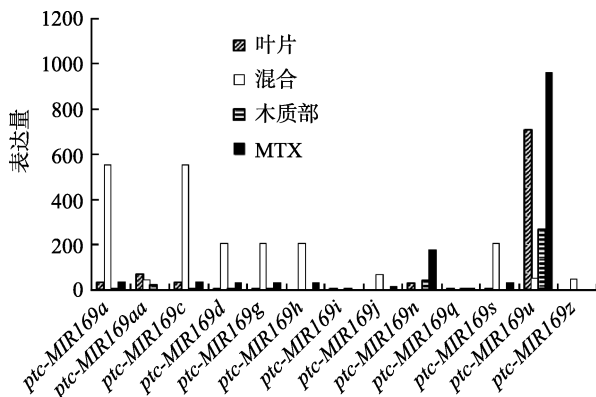


图 4 *MIR169* 基因家族成员在杨树不同组织的表达水平 miRNA 基因表达量数据计算方法为：原始读数/与基因组匹配的总读数 $\times 10^8$ ；其中“混合”样品为营养生长茎尖、雄花、雌花、雌花顶芽、雄花顶芽和侧芽组织的混合，MTX(Mechanically treated xylem)为机械胁迫处理(压力和拉力)的木质部。

因预测，得分 3 的靶基因如表 4 所示。在毛果杨中共有来自 31 个基因座的 74 个转录本是 *MIR169* 的靶基因，杨树 *MIR169* 基因家族的靶基因包括转录因子、蛋白激酶、抗病蛋白、泛素连接酶、载体蛋白等，这说明 *MIR169* 广泛参与了杨树多种生命活动的调控。其中转录因子 NF-YA 是杨树 *MIR169* 主要的靶基因，共有来自 7 个基因座的 33 个转录本为转录因子 NF-YA。

3 讨论

在小立碗藓 (*Physcomitrella patens*) 和莱茵衣藻 (*Chlamydomonas reinhardtii*) 中没有发现 *MIR169* 基因 (图 5)，而在裸子植物中仅在高山松 (*Pinus densata*) 中发现 1 个 *MIR169* 基因 (这可能与测序深度有关，因为目前高山松中的 *MIR169* 基因是通过转录组测序发现的，而没有采用专门针对小 RNA 的深度测序)^[22]。而蕨类植物 miRNA 的研究资料目前还很少。植物 *MIR169* 基因家族可能起源于维管植物出现之后裸子植物出现之前，即大约 320~600 MYA^[23]。经过漫长的进化过程，现存的单子叶和双子叶植物 *MIR169* 基因家族成员大多为 20 个以下。而唯有杨

表 4 杨树 *MIR169* 家族成员预测的靶基因

靶基因功能	靶基因序列号
氨基转移酶	Potri.001G099400.1/2
NF-YA (CBF-B/HAP2) 转录因子	Potri.001G257600.1/2; Potri.001G266000.1/2/3/4/5; Potri.006G053500.1/2/3/4; Potri.006G145100.1/2/3/4/5/6; Potri.009G052900.1/2/3/4/5/7/8; Potri.009G060600.1/2/3/4/5/6; Potri.018G064700.1/2/3
含 CCT 基序蛋白质	Potri.012G014000.1/2/5/6/7
含 tify 结构域和 Divergent CCT 基序蛋白	Potri.008G133400.10/11
亚铁离子转运蛋白; GTP 结合蛋白	Potri.002G134900.1
热激转录因子	Potri.014G141400.1/2
NF-YC(CBF-C/HAP5)转录因子	Potri.012G098500.1/2; Potri.015G097400.1
水解酶家族蛋白	Potri.010G216000.1
富含亮氨酸重复蛋白	Potri.006G061000.1; Potri.006G061200.1
MYB-like 蛋白	Potri.001G036000.1/2
Ndr 家族蛋白	Potri.018G054900.1/2
蛋白酶体亚基	Potri.002G148300.3
蛋白激酶	Potri.017G075400.1
PUF 家族蛋白	Potri.001G298000.1
核糖体蛋白	Potri.014G068600.1
蔗糖转运蛋白	Potri.019G061900.1
Tfb4 转录因子	Potri.005G176600.2/3
主要协助转运子超家族蛋白	Potri.002G016200.1/2; Potri.005G245900.1/2
泛素结合酶	Potri.013G130800.1/2/3
锌指蛋白	Potri.004G047000.1; Potri.014G134400.2/3; Potri.001G325400.1/2

树 *MIR169* 基因家族成员数扩大到 33 个。

杨树基因组经历了全基因组重复、染色体重排和串联重复等复杂的进化过程^[5]。本文所获得的结果表明大片段重复和串联重复在 *MIR169* 基因家族的扩张中都发挥了重要的作用。其中大片段重复发生的时间大多与杨柳科重复事件接近; 而且从基因组定位来看, 这些重复基因也多位于杨柳科重复所产生的染色体重复片段上。另外, 侧翼保守蛋白质基因的 K_s 值大多分布于 0.2~0.3, 这与利用 EST 分析杨树基因组重复事件所得 K_s 值一致^[24], 进一步证实推动杨树 *MIR169* 基因家族扩张的大片段重复大多就是杨柳科重复事件的一部分。

目前对 miRNA 基因产生与消亡的进化动力学机制尚不清楚^[12]。研究表明, 植物 miRNA 基因产生与消失的频率都很高^[25]。比较 11 种植物的 miRNA 基因表明, 许多古老 miRNA 基因家族在植物进化过程中消失, 说明保守的 miRNA 基因家族在进化过程中也是动态的^[26]。而保守 miRNA 基因往往能形成复杂的调控网络, 其功能也是多样的^[12]。在杨树中, 与

MIR169 基因家族起源时间相近, 甚至起源更早、更保守的 miRNA 基因家族, 例如 *MIR156*、*MIR159*、*MIR160*、*MIR166*、*MIR171*、*MIR172*、*MIR395*、*MIR398*、*MIR408* 等, 其基因家族成员数却少于 *MIR169* 基因家族(图 5)。

基因组进化过程中对重复基因的保留是具有偏好性的^[27]。杨树基因组中转录因子、激酶等相关的重复基因往往更倾向于保留下来, 这种选择性保留重复基因可能是杨树进化的机制之一^[5]。目前研究表明基因自身特性影响蛋白质重复基因的保留^[28], 但对 miRNA 重复基因的保留机制目前还缺乏研究。是什么机制导致杨树基因组中保留庞大的 *MIR169* 基因家族, 还有待进一步的研究。

MIR169 对靶基因 NF-YA 转录因子的作用及其功能在蒺藜苜蓿^[29]、拟南芥^[30]、水稻^[31]、胡杨^[32]中已得到验证, *MIR169* 通过对转录因子 NF-YA 的作用调控豆科植物根瘤的发育以及植物对干旱胁迫、盐胁迫的响应。根据生物信息学预测, 杨树 *MIR169* 基因家族的靶基因包括转录因子、蛋白激酶、抗病

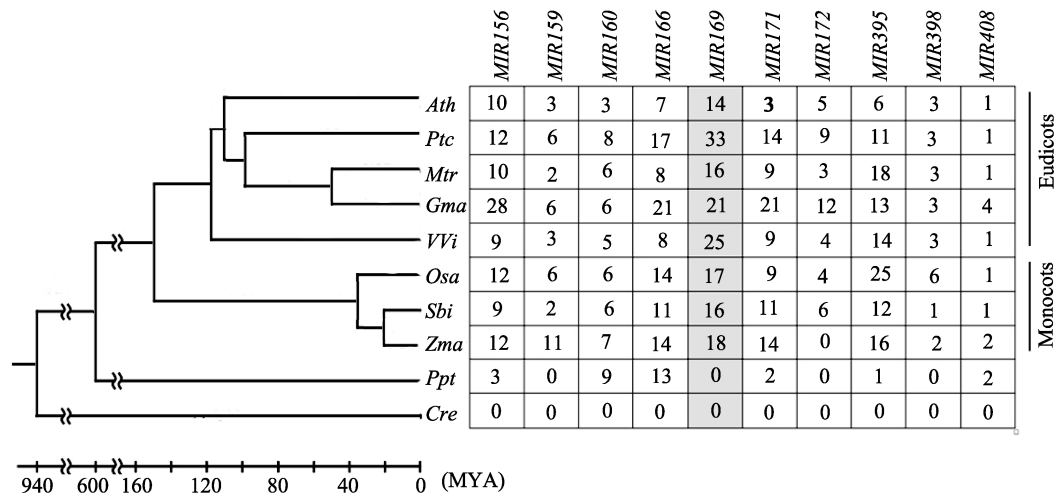


图 5 部分植物 10 个保守 miRNA 基因家族成员数

方框内的数字表示基因家族成员数量; 下方标尺为进化年代, 单位为百万年(MYA)。物种缩写如下: *Ath*: 拟南芥(*Arabidopsis thaliana*); *Ptc*: 毛果杨(*Populus trichocarpa*); *Mtr*: 蒺藜苜蓿(*Medicago truncatula*); *Gma*: 大豆(*Glycine max*); *Vvi*: 葡萄(*Vitis vinifera*); *Osa*: 水稻(*Oryza sativa*); *Sbi*: 高粱(*Sorghum bicolor*); *Zma*: 玉米(*Zea mays*); *Ppt*: 小立碗藓(*Physcomitrella patens*); *Cre*: 莱茵衣藻(*Chlamydomonas reinhardtii*)。

蛋白、泛素连接酶、载体蛋白等, 广泛参与多种生命活动的调节, 这也从另一方面说明了 *MIR169* 基因家族对杨树具有重要的作用。因此, 研究杨树中 *MIR169* 基因家族的功能, 尤其是与多年生木本植物相关的生命现象, 例如周期性生长、环境适应等方面的调节功能, 具有重要的理论与实际意义。

参考文献(References):

- [1] Jones-Rhoades MW, Bartel DP, Bartel B. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol*, 2006, 57(1): 19–53. [DOI](#)
- [2] Chen XM. Small RNAs and their roles in plant development. *Annu Rev Cell Dev Biol*, 2009, 25(1): 21–44. [DOI](#)
- [3] Shukla LI, Chinnusamy V, Sunkar R. The role of microRNAs and other endogenous small RNAs in plant stress responses. *Biochim Biophys Acta*, 2008, 1779(11): 743–748. [DOI](#)
- [4] Ruiz-Ferrer V, Voinnet O. Roles of plant small RNAs in biotic stress responses. *Annu Rev Plant Biol*, 2009, 60(1): 485–510. [DOI](#)
- [5] Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroev S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehrling J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, 2006, 313(5793): 1596–1604. [DOI](#)
- [6] Yang XH, Kalluri UC, DiFazio SP, Wullschlegler SD, Tschaplinski TJ, Cheng ZM, Tuskan GA. Poplar genomics: state of the science. *Crit Rev Plant Sci*, 2009, 28(5): 285–308. [DOI](#)
- [7] Taylor G. *Populus: Arabidopsis* for forestry. Do we need a model tree? *Ann Bot*, 2002, 90(6): 681–689. [DOI](#)
- [8] Jansson S, Douglas CJ. *Populus: A model system for plant biology*. *Annu Rev Plant Biol*, 2007, 58(1): 435–458. [DOI](#)
- [9] Lu SF, Sun YH, Chiang VL. Stress-responsive microRNAs

- in *Populus*. *Plant J*, 2008, 55(1): 131–151. [\[DOI\]](#)
- [10] Lu SF, Sun YH, Shi R, Clark C, Li L, Chiang VL. Novel and mechanical stress-responsive MicroRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *Plant Cell*, 2005, 17(8): 2186–2203. [\[DOI\]](#)
- [11] Puzeý JR, Karger A, Axtell M, Kramer EM. Deep annotation of *Populus trichocarpa* microRNAs from diverse tissue sets. *PLoS ONE*, 2012, 7(3): e33034. [\[DOI\]](#)
- [12] 魏强, 梁永宏, 李广林. 植物 miRNA 的进化. *遗传*, 2013, 25(3): 315–323. [\[DOI\]](#)
- [13] Li AL, Mao L. Evolution of plant microRNA gene families. *Cell Res*, 2007, 17(3): 212–218. [\[DOI\]](#)
- [14] Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*, 2011, 39(Suppl. 1): D152–D157. [\[DOI\]](#)
- [15] Maher C, Stein L, Ware D. Evolution of *Arabidopsis* microRNA families through duplication events. *Genome Res*, 2006, 16(4): 510–519. [\[DOI\]](#)
- [16] Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Geno Prot Bioinfo*, 2006, 4(4): 259–263. [\[DOI\]](#)
- [17] Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*, 2000, 17(1): 32–43. [\[DOI\]](#)
- [18] Blanc G, Wolfe KH. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*, 2004, 16(7): 1667–1678. [\[DOI\]](#)
- [19] Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science*, 2000, 290(5494): 1151–1155. [\[DOI\]](#)
- [20] Koch MA, Haubold B, Mitchell-Olds T. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol Biol Evol*, 2000, 17(10): 1483–1498. [\[DOI\]](#)
- [21] Li B, Qin Y, Duan H, Yin W, Xia X. Genome-wide characterization of new and drought stress responsive microRNAs in *Populus euphratica*. *J Exp Bot*, 2011, 62(11): 3765–3779. [\[DOI\]](#)
- [22] Wan LC, Zhang HY, Lu SF, Zhang L, Qiu ZB, Zhao YY, Zeng QY, Lin JX. Transcriptome-wide identification and characterization of miRNAs from *Pinus densata*. *BMC Genomics*, 2012, 13(1): 132. [\[DOI\]](#)
- [23] 方炎明. 陆地植物新系统树之诠释与简评. *南京林业大学学报 (自然科学版)*, 2009, 33(4): 1–7. [\[DOI\]](#)
- [24] Sterck L, Rombauts S, Jansson S, Sterky F, Rouze P, Van de Peer Y. EST data suggest that poplar is an ancient polyploid. *New Phytol*, 2005, 167(1): 165–170. [\[DOI\]](#)
- [25] Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS ONE*, 2007, 2(2): e219. [\[DOI\]](#)
- [26] Nozawa M, Miura S, Nei M. Origins and evolution of microRNA genes in plant species. *Genome Biol Evol*, 2012, 4(3): 230–239. [\[DOI\]](#)
- [27] De Bodt S, Maere S, van de Peer Y. Genome duplication and the origin of angiosperms. *Trends Ecol Evol*, 2005, 20(11): 591–597. [\[DOI\]](#)
- [28] Jiang WK, Liu YL, Xia EH, Gao LZ. Prevalent role of gene features in determining evolutionary fates of whole-genome duplication duplicated genes in flowering plants. *Plant Physiol*, 2013, 161(4): 1844–1861. [\[DOI\]](#)
- [29] Combier J P, Frugier F, de Billy F, Boualem A, El-Yahyaoui F, Moreau S, Vernié T, Ott T, Gamas P, Crespi M, Niebel A. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes Dev*, 2006, 20(22): 3084–3088. [\[DOI\]](#)
- [30] Li WX, Oono Y, Zhu JH, He XJ, Wu JM, Iida K, Lu XY, Cui XP, Jin HL, Zhu JK. The *Arabidopsis* NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. *Plant Cell*, 2008, 20(8): 2238–2251. [\[DOI\]](#)
- [31] Zhao BT, Ge LF, Liang RQ, Li W, Ruan KC, Lin HX, Jin YX. Members of miR-169 family are induced by high salinity and transiently inhibit the NF-YA transcription factor. *BMC Mol Biol*, 2009, 10(1): 29. [\[DOI\]](#)
- [32] Qin YR, Duan ZX, Xia XL, Yin WL. Expression profiles of precursor and mature microRNAs under dehydration and high salinity shock in *Populus euphratica*. *Plant Cell Rep*, 2011, 30(10): 1893–1907. [\[DOI\]](#)