

刘志祥,曾超珍,周永青,等. 杨树 *MIR156* 基因家族的进化与功能分化[J]. 江苏农业科学,2014,42(1):26-29.

杨树 *MIR156* 基因家族的进化与功能分化

刘志祥^{1,2}, 曾超珍¹, 周永青¹, 谭晓风²

(1. 中南林业科技大学生命科学与技术学院, 湖南长沙 410004;

2. 经济林育种与栽培国家林业局重点实验室/中南林业科技大学林学院, 湖南长沙 410004)

摘要:为阐明杨树 *MIR156* 基因家族的进化过程和功能分化,研究了杨树 *MIR156* 基因家族的扩张模式、倍增时间、系统发育、表达方式、启动子元件和靶基因。结果表明:杨树 *MIR156* 基因家族主要通过染色体大片段重复实现扩张,而串联重复对此没有贡献;不同家族成员表达方式已分化,其中 *ptc-MIR156g/h/i/j* 表达活性最高;家族成员启动子顺式作用元件存在差异;杨树 *MIR156* 的靶基因包括 29 个 SBP 结构域蛋白质,由于成熟序列存在差异,家族成员调控的靶基因也表现出差异。说明杨树 *MIR156* 基因家族成员的功能已经分化,在杨树中可能形成了复杂的调控网络。

关键词:杨树;microRNA;*MIR156* 基因家族;分子进化;功能分化

中图分类号: Q943 **文献标志码:** A **文章编号:** 1002-1302(2014)01-0026-04

杨树是全球范围内广泛栽培的速生丰产用材树种和生物质能源树种,具有重要经济价值,同时在生态环境保护中也具有重要价值^[1]。自毛果杨(*Populus trichocarpa*)基因组完成测序后^[2],杨树作为一个重要的模式系统广泛用于木材形成、周期性生长、适应性等方面的研究^[3]。

植物微 RNA(microRNA, miRNA)是一类具有调控作用的非编码小分子 RNA。miRNA 由基因组编码,通过 RNA 聚合酶 II 转录,通过 DCL 等蛋白的剪切加工形成长约 21 nt 的成熟序列,并在多种蛋白质的参与下识别与其序列互补的靶 mRNA 并介导其翻译抑制或剪切,从而实现对靶基因的转录后调控。由于 miRNA 的靶基因往往是具有调控功能的转录因子等,所以 miRNA 对植物生长发育和胁迫应答具有十分重要的调控功能,在植物分子育种中具有良好的应用前景^[4]。

MIR156 是一个古老的 miRNA 基因家族,对植物生长发育具有非常重要的调控作用,如赤霉素途径^[5]、育性^[6]、发育阶段转变^[7-8]、叶原基间隔长度与器官形态^[9]等。已有的研究主要以拟南芥(*Arabidopsis thaliana*)为材料,而 *MIR156* 在木本植物中的功能尚未见报道。本研究通过对杨树 *MIR156* 基因家族的扩张模式、倍增时间、系统发育、表达方式、启动子元件和靶基因的研究,初步阐明杨树 *MIR156* 基因家族的进化过程和功能分化,以期对杨树 *MIR156* 基因的功能研究以及在分子育种中的应用奠定基础。

1 材料与方法

1.1 基因组定位与基因倍增模式分析

从 miRBase 数据库(www.mirbase.org, Release 19)中获取

ptc-MIR156 基因家族序列和在杨树基因组 JGI_Poptr 2.0 中的坐标,以 FASTA 格式将杨树基因组数据提交给 Phytozome 数据库(www.phytozome.net),经 BLAST 比对获取其在杨树基因组 JGI_Poptr 3.0 上的坐标。采用 Maher 等的方法^[10]对杨树基因组数据(JGI_Poptr 3.0)进行基因家族扩张模式分析,包括串联重复和染色体大片段重复。

1.2 基因倍增时间估算

对倍增块中保守的蛋白质编码基因采用 Clustal X 进行氨基酸序列比对,然后以氨基酸序列比对结果为参照,进行编码序列比对,使用 KaKs_Calculator 计算序列间同义替换率 [K_s (次/个)]; K_s = 替换次数/同义替换位点个数,算法选择 YN 法。求得 K_s 平均值以用于计算倍增时间,计算公式为: $D = K_s / (2E)$ 。式中: D 表示时间(年),[次/(个·年)]; E [E = 替换次数/(同义替换位点个数 × 时间)]表示分子替换速率。杨树分子进化速率约为拟南芥的 1/6^[2],在拟南芥中 $E = 1.5 \times 10^{-8}$ 次/(个·年)^[11],所以杨树 E 取 2.5×10^{-9} 次/(个·年)。

1.3 分子系统发育树构建

将杨树 *MIR156* 基因家族序列进行 MUSCLE 比对,比对结果输入到 MEGA5 软件中,分别采用最大似然法和邻接法构建系统进化树,采用自展法检验进化树,重复值设置为 1 000。

1.4 启动子顺式作用元件分析

参考 Cui 等的方法^[12],从 Phytozome 数据库(http://www.phytozome.net, JGI_Poptr 3.0)中获取 *ptc-MIR156* 基因家族上游启动子序列。将所获取的序列提交给 PlantCARE(http://bioinformatics.psb.ugent.be/webtools/plantcare/html/),分析其中的顺式作用元件。

1.5 表达分析

从毛果杨小 RNA 高通量测序结果^[13]中提取 *MIR156* 基因家族成员原始测序读数,并进行均一化处理计算表达量:表达量 = 原始读数/与基因组匹配的总读数 $\times 10^6$ 。

1.6 靶基因分析

杨树 *MIR156* 基因家族靶基因预测采用 psRNATarget

收稿日期:2013-08-13

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

作者简介:刘志祥(1978—),男,湖南湘阴人,博士研究生,讲师,主要从事植物分子遗传学研究。E-mail:liuzx0927@foxmail.com。

通信作者:谭晓风,博士,教授,主要从事林业生物技术研究。Tel:(0731)85623406;E-mail:tanxiaofengcn@126.com。

(<http://plantgrn.noble.org/psRNATarget/#>), 将 miRNA 成熟序列提交给 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 基因组定位与基因倍增模式分析

表 1 杨树 *MIR156* 基因家族成员的基因组定位情况

<i>MIR156</i> 基因家庭成员	基因组定位 (JGI_Poptr2. 0)	基因组定位 (JGI_Poptr3. 0)	证据
<i>ptc - MIR156h</i>	scaffold_4:7938832 - 7938934[-]	—	克隆 ^[2,14-15]
<i>ptc - MIR156l</i>	scaffold_1:34445840 - 34445933[-]	Chr01:34292488 - 34292581[+]	小 RNA 深度测序 ^[13]
<i>ptc - MIR156a</i>	scaffold_6:19089414 - 19089514[-]	Chr06:20084161 - 20084261[-]	与 <i>ath - MIR156b</i> 序列相似 ^[2]
<i>ptc - MIR156b</i>	scaffold_6:22869899 - 22869999[-]	Chr06:23735954 - 23736054[-]	与 <i>ath - MIR156f</i> 序列相似 ^[2,15]
<i>ptc - MIR156c</i>	scaffold_6:25766668 - 25766767[+]	Chr06:26631525 - 26631624[+]	与 <i>ath - MIR156a</i> 序列相似 ^[2]
<i>ptc - MIR156k</i>	scaffold_11:11956115 - 11956215[-]	Chr11:11382228 - 11382328[-]	与 <i>ath - MIR156a</i> 序列相似 ^[2]
<i>ptc - MIR156i</i>	scaffold_12:4332062 - 4332160[-]	Chr12:4939241 - 4939339[+]	克隆 ^[2,14-15]
<i>ptc - MIR156j</i>	scaffold_15:4013699 - 4013798[+]	Chr15:4227002 - 4227101[-]	克隆 ^[2,14-15]
<i>ptc - MIR156g</i>	scaffold_5:11875533 - 11875635[-]	Chr17:13418464 - 13418566[+]	克隆 ^[2,14-15]
<i>ptc - MIR156f</i>	scaffold_18:12448484 - 12448584[-]	Chr18:13566875 - 13566975[-]	与 <i>ath - MIR156b</i> 序列相似 ^[2]
<i>ptc - MIR156e</i>	scaffold_18:2189691 - 2189790[+]	Chr18:1583358 - 1583457[-]	与 <i>ptc - MIR156j</i> 序列相似 ^[2,14]
<i>ptc - MIR156d</i>	scaffold_18:8350103 - 8350203[+]	scaffold_28:190974 - 191074[+]	与 <i>ptc - MIR156g</i> 序列相似 ^[2,14-15]

通过分析各成员的侧翼保守蛋白质基因, 共发现 5 对基因为染色体大片段重复的产物, 对其中侧翼保守蛋白质基因数量大于 4 的倍增基因对进行了倍增时间估算, 结果见表 2。表 2 显示, 产生倍增基因对 *ptc - MIR156a* 与 *ptc - MIR156f*、

通过查询 miRBase 和 Phytozome 数据库获取杨树 *MIR156* 基因家族在杨树基因组中的位置信息(表 1)。杨树 *MIR156* 基因家族共有 12 个成员, 除 *ptc - MIR156h* 在 JGI_Poptr 3. 0 中未找到定位, 其余 11 个成员均在 JGI_Poptr 3. 0 找到了位置。从定位情况可知, 杨树 *MIR156* 基因家族成员在基因组上均为分散分布, 未见成簇分布, 说明杨树 *MIR156* 基因家族成员间没有串联重复关系。

ptc - MIR156c 与 *ptc - MIR156e* 的染色体大片段重复发生的时间分别约在 4 598 万、5 605 万年前, 这一时间与杨树进化过程中最近所经历的一次全基因组重复——杨柳科重复在时间上较接近。

表 2 染色体大片段重复及其时间估算

倍增基因 1	倍增基因 2	保守侧翼蛋白质基因对数	平均 Ks	Ks 标准差	倍增时间(万年)
<i>ptc - MIR156a</i>	<i>ptc - MIR156f</i>	9	0. 280 256 444	0. 096 534 035	5 605
<i>ptc - MIR156c</i>	<i>ptc - MIR156e</i>	9	0. 229 8821 11	0. 034 782 328	4 598
<i>ptc - MIR156i</i>	<i>ptc - MIR156j</i>	2	—	—	—
<i>ptc - MIR156a</i>	<i>ptc - MIR156e</i>	1	—	—	—
<i>ptc - MIR156g</i>	<i>ptc - MIR156l</i>	1	—	—	—

由此可见, 杨树 *MIR156* 基因家族主要通过染色体大片段重复实现基因家族的扩张, 而串联重复对杨树 *MIR156* 基因家族的扩张并没有贡献。

2.2 分子系统发育树构建

分别采用邻接法和最大似然法构建杨树 *MIR156* 基因家族的分子系统发育树, 这 2 种方法所建进化树结构相似, 且大部分节点的 Bootstrap 值大于 70, 说明所建分子系统发育树是可靠的, 其中邻接树如图 1 所示。将分子系统发育树与基因倍增模式分析结果进行比较, 结果发现, 通过染色体大片段重复所产生的重复基因如 *ptc - MIR156a* 与 *ptc - MIR156f*、*ptc - MIR156c* 与 *ptc - MIR156e*、*ptc - MIR156i* 与 *ptc - MIR156j* 在进化树上位于分支末端, 说明这些染色体大片段重复导致的基因家族扩张是杨树 *MIR156* 基因家族进化过程中较晚的事件; *ptc - MIR156b* 与 *ptc - MIR156d*、*ptc - MIR156g* 与 *ptc - MIR156h* 在进化树中位于分支末端, 但它们不是串联重复和染色体大片段重复的产物, 说明它们可能是通过转座等其他方式进行扩张的产物^[16]。

2.3 表达分析

基于小 RNA 高通量测序的表达分析结果如图 2 所示。从图 2 可知, 在叶片、木质部、机械胁迫木质部和混合组织(包括营养生长茎尖、雄花、雌花、雌株顶芽、雄株顶芽和侧芽)中, *ptc - MIR156g/h/i/j* 表达活性最高, 其次是 *ptc - MIR156a/b/c/d/e/f*、*ptc - MIR156k* 与 *ptc - MIR156l* 表达活性均很低, 在混合组织中甚至没有检测到。

杨树 *MIR156* 基因家族在叶片和木质部中表达活性较高, 而在以花和芽组成的混合组织中表达活性较低。另外, 杨树 *MIR156* 基因家族各成员中的表达水平在不同组织中呈现出多样性, 说明不同成员的表达方式已经分化。

2.4 启动子顺式作用元件分析

采用 PlantCARE 分析杨树 *MIR156* 家族 11 个成员上游启动子所含的顺式作用元件, 部分元件如表 2 所示。结果(表 3)显示, 家族成员上游均具有 CAAT - box、TATA - box 等基本元件, 可能都具有转录活性。同时, 在不同成员上游发现了多个与激素调控、胁迫响应及光响应相关的元件, 说明 *MIR156* 可能参与这些过程的调控。表 3 还显示, 不同成员的上游元件存在差异, 而这将导致不同成员的表达存在时空差异。

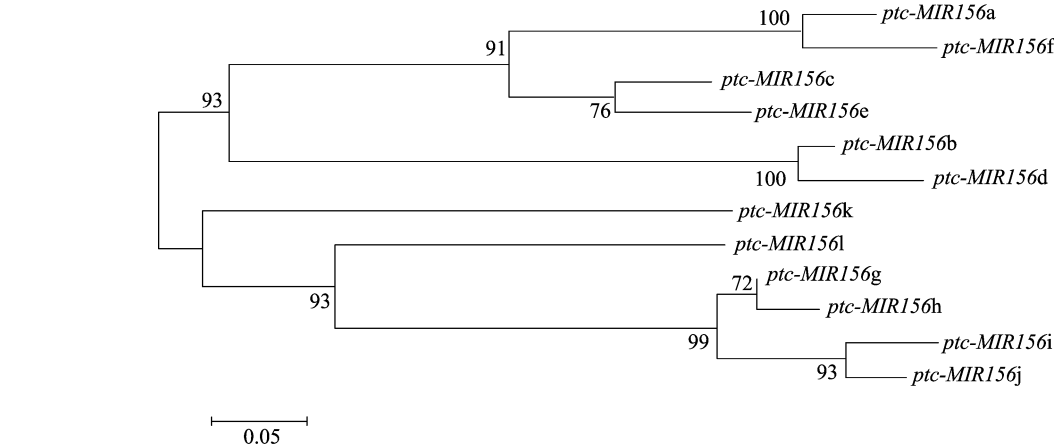


图1 杨树 *MIR156* 基因家族的邻接树

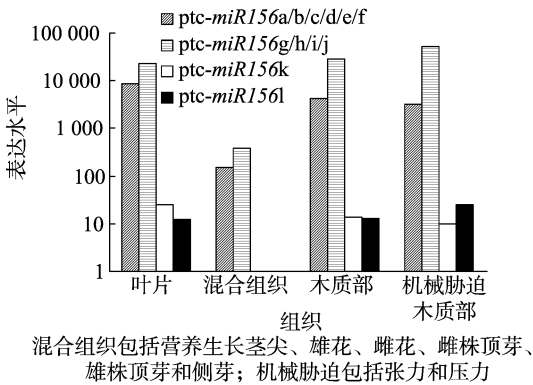


图2 杨树 *MIR156* 基因家族组织表达情况

图 3 所示,12 个家族成员产生的成熟序列只有 4 种,分别为 *ptc - MIR156a/b/c/d/e/f*、*ptc - MIR156g/h/i/j*、*ptc - MIR156k* 和 *ptc - MIR156l*。4 种成熟序列存在差异,可能导致其识别的靶基因有所差异。

通过 psRNATarget 预测分析共找到杨树 *MIR156* 的靶基因 29 个,均为 *SQUAMOSA* 启动子结合蛋白 (*SQUAMOSA* promoter binding protein, SBP)。SBP 是植物特有的一类转录因子,对植物的生长发育具有重要的调控功能,如花与果实发育、赤霉素途径、铜应答过程等^[17]。但由于成熟序列存在差异,导致不同成员的靶基因存在一定差异(图 4)。其中 20 个基因是全部成员的预测靶基因,而 *ptc - MIR156a/b/c/d/e/f* 和 *ptc - MIR156k* 分别独有 2 个靶基因。这说明由于杨树 *MIR156* 基因家族成员间成熟序列的差异,可能导致它们调控的靶基因也已经出现了差异,即不同成员间已出现了功能分化。

2.5 靶基因分析

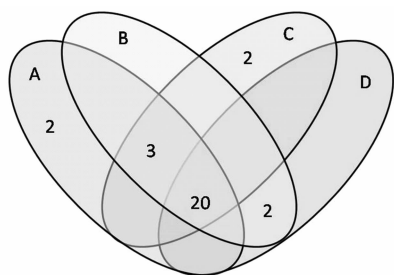
将杨树 *MIR156* 基因家族的成熟序列进行比对,结果如

表 3 杨树 *MIR156* 基因家族成员启动子中的顺式作用元件

功能	顺式作用元件	<i>MIR156a</i>	<i>MIR156b</i>	<i>MIR156c</i>	<i>MIR156d</i>	<i>MIR156e</i>	<i>MIR156f</i>	<i>MIR156g</i>	<i>MIR156i</i>	<i>MIR156j</i>	<i>MIR156k</i>	<i>MIR156l</i>
基本元件	CAAT - box	○	○	○	○	○	○	○	○	○	○	○
	TATA - box	○	○	○	○	○	○	○	○	○	○	○
激素调控	CGTCA - motif	-	○	○	-	○	○	○	-	○	-	○
	ERE	○	-	-	○	-	○	-	-	○	-	-
	GARE - motif	-	○	-	○	-	-	○	-	-	-	-
	P - box	-	-	-	-	○	-	-	-	-	-	○
	TATC - box	-	-	-	-	-	-	○	-	-	-	○
	TCA - element	○	○	○	○	○	○	○	-	○	-	-
	TGA - element	-	-	-	-	-	○	-	-	○	-	○
	TGACG - motif	-	○	○	-	○	○	○	-	○	-	○
胁迫响应	ABRE	○	○	-	-	○	-	○	-	○	-	-
	ARE	○	○	○	○	○	○	-	-	○	○	-
	Box - W1	-	-	-	○	○	-	-	○	-	○	○
	HSE	○	○	○	○	-	○	-	-	○	-	-
	MBS	○	○	-	○	-	○	○	○	○	-	○
	TC - rich repeats	○	○	○	○	○	-	○	-	-	-	-
光响应	Box 4	-	○	○	○	○	○	○	○	○	-	○
	G - Box	○	○	-	○	○	-	○	○	○	-	○
	Sp1	○	○	○	○	○	○	-	○	-	-	○

注:“○”表示有,“-”表示无。

<i>ptc-MIR156a</i>	.TGACAGAAGAGAGTCAGCAC
<i>ptc-MIR156b</i>	.TGACAGAAGAGAGTCAGCAC
<i>ptc-MIR156c</i>	.TGACAGAAGAGAGTCAGCAC
<i>ptc-MIR156d</i>	.TGACAGAAGAGAGTCAGCAC
<i>ptc-MIR156e</i>	.TGACAGAAGAGAGTCAGCAC
<i>ptc-MIR156f</i>	.TGACAGAAGAGAGTCAGCAC
<i>ptc-MIR156g</i>	TTGACAGAAGATAGACAGCAC
<i>ptc-MIR156h</i>	TTGACAGAAGATAGACAGCAC
<i>ptc-MIR156i</i>	TTGACAGAAGATAGACAGCAC
<i>ptc-MIR156j</i>	TTGACAGAAGATAGACAGCAC
<i>ptc-MIR156k</i>	.TGACAGAAGAGAGTCAGCAC
<i>ptc-MIR156l</i>	TTGACAGAAGATAGACAGCAC
一致序列	tgacagaaga g gagcac

图3 杨树 *MIR156* 基因成熟序列比对情况

A—*ptc-MIR156a/b/c/d/e/f*; B—*ptc-MIR156g/h/i/j*;
C—*ptc-MIR156k*; D—*ptc-MIR156l*

图4 杨树 *MIR156* 基因家族靶基因分布的预测情况

3 结论

miRNA 通过靶基因的反向重复或随机序列而起源,通过染色体片段重复或串联重复实现家族扩张^[16,18]。扩张模式分析结果表明,杨树 *MIR156* 基因家族主要通过染色体大片段重复实现扩增,而串联重复对此没有贡献。基于 miRNA 深度测序的表达分析结果表明,杨树 *MIR156* 基因家族成员的表达方式已经分化,而且不同成员的上游顺式作用元件及调控的靶基因也表现出差异,说明家族成员间功能已经分化,*MIR156* 基因家族在杨树中可能已经形成了复杂的调控网络,参与不同生命活动的调控。杨树为多年生木本植物,其生长发育与拟南芥等一年生或二年生植物差异显著,所以进一步研究 *MIR156* 及其靶基因 SBP 转录因子对杨树生长发育的调控具有重要意义。

参考文献:

- [1] Yang X H, Kalluri U C, DiFazio S P, et al. Poplar genomics: state of the science[J]. Crit Rev Plant Sci, 2009, 28(5): 285–308.
- [2] Tuskan G A, DiFazio S, Jansson S, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)[J]. Science, 2006, 313

(5793): 1596–1604.

- [3] Jansson S, Douglas C J. *Populus*: a model system for plant biology[J]. Annu Rev Plant Biol, 2007, 58: 435–458.
- [4] Chen X M. Small RNAs and their roles in plant development[J]. Annu Rev Cell Dev Biol, 2009, 25: 21–44.
- [5] Yu S, Galvão V C, Zhang Y C, et al. Gibberellin regulates the *Arabidopsis* floral transition through *MIR156* – targeted SQUAMOSA promoter binding – like transcription factors[J]. Plant Cell, 2012, 24(8): 3320–3332.
- [6] Xing S, Salinas M, Höhmann S, et al. *MIR156* – targeted and nontargeted SBP – box transcription factors act in concert to secure male fertility in *Arabidopsis*[J]. Plant Cell, 2010, 22(12): 3935–3950.
- [7] Wang J W, Czech B, Weigel D. *MIR156* – Regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*[J]. Cell, 2009, 138(4): 738–749.
- [8] Wu G, Park M Y, Conway S R, et al. The sequential action of *MIR156* and *MIR172* regulates developmental timing in *Arabidopsis*[J]. Cell, 2009, 138(4): 750–759.
- [9] Wang J W, Schwab R, Czech B, et al. Dual effects of *MIR156* – targeted SPL genes and *CYP78A5/KLUH* on plastochron length and organ size in *Arabidopsis thaliana*[J]. Plant Cell, 2008, 20(5): 1231–1243.
- [10] Maher C, Stein L, Ware D. Evolution of *Arabidopsis* microRNA families through duplication events[J]. Genome Res, 2006, 16(4): 510–519.
- [11] Koch M A, Haubold B, Mitchell – Olds T. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae)[J]. Mol Biol Evol, 2000, 17(10): 1483–1498.
- [12] Cui X, Xu S M, Mu D S, et al. Genomic analysis of rice microRNA promoters and clusters[J]. Gene, 2009, 431(1/2): 61–66.
- [13] Puzey J R, Karger A, Axtell M, et al. Deep annotation of *Populus trichocarpa* microRNAs from diverse tissue sets[J]. Mol Biol Evol, 2000, 17(10): 1483–1498.
- [14] Lu S, Sun Y H, Shi R, et al. Novel and mechanical stress – responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*[J]. Plant Cell, 2005, 17(8): 2186–2203.
- [15] Zhang B, Pan X, Cannon C H, et al. Conservation and divergence of plant microRNA genes[J]. Plant J, 2006, 46(2): 243–259.
- [16] Nozawa M, Miura S, Nei M. Origins and evolution of microRNA genes in plant species[J]. Genome Biol Evol, 2012, 4(3): 230–239.
- [17] 葛安静, 张春华, 董清华, 等. 草莓 *SBP* 基因家族生物信息学初步分析[J]. 中国农学通报, 2012, 28(13): 215–220.
- [18] 魏 强, 梁永宏, 李广林. 植物 miRNA 的进化[J]. 遗传, 2013, 35(3): 315–323.