

巨桉AGO基因家族的生物信息学分析

范春节, 闫慧芳, 裘珍飞, 曾炳山*, 刘英, 李湘阳, 王象军, 郭光生

(中国林业科学研究院热带林业研究所, 广州 510520)

摘要: 为了解巨桉(*Eucalyptus grandis*)中 Argonaute (AGO)的功能,经全基因组序列分析,巨桉中存在 14 个 AGO 基因家族成员,基因长度为 2676~3225 bp,编码的蛋白质具有保守的 Piwi、PAZ、DUF1785 结构域。巨桉 *EgrAGO*s 基因可分为 3 组,内含子和外显子结构具有明显的组织特异性。组内成员核苷酸序列和氨基酸序列保守性较高,同源性分别达到 88.14% 和 82.97%。*EgrAGO* 基因分布于第 2、4、7、8、10、11 条染色体上,在进化过程中存在着串联复制和大片段基因复制机制。预测巨桉的大部分 AGO 蛋白定位于细胞核和胞质中,表现出偏碱性和亲水性,具有较高的脂溶指数。基因表达分析表明,桉树 *EgrAGO* 成员在不同组织中的表达有明显差异,与木材形成相关的组织/器官中有较高的表达。这些为深入研究 *EgrAGO* 基因的功能奠定了基础。

关键词: 桉树; AGO 蛋白; 基因家族; 生物信息

doi: 10.11926/j.issn.1005-3395.2015.04.001

Bioinformatics Analysis of AGO Gene Family in *Eucalyptus grandis* Genome

FAN Chun-jie, YAN Hui-fang, QIU Zhen-fei, ZENG Bing-shan*, LIU Ying, LI Xiang-yang, WANG Xiang-jun, GUO Guang-sheng

(Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou 510520, China)

Abstract: In order to understand the function of Argonaute (AGO) in *Eucalyptus grandis*, the phylogeny, gene structure and expression pattern of *EgrAGO* were analyzed in genome-wide. The results showed that there were 14 *EgrAGOs* in *E. grandis*, which length ranged from 2676 to 3225 bp and *EgrAGOs* contain Piwi, PAZ and DUF1785 domains. *EgrAGO* genes could be divided into 3 clades, and intron and exon structures had obvious tissue specificity. The nucleotide sequence of *EgrAGO* and encoding amino acid sequence in each clade had homologies for 88.14% and 82.97%, respectively. Meanwhile, *EgrAGO* genes distribute in 2, 4, 7, 8, 10 and 11 chromosomes with tandem duplication and segmental duplication in the process of evolution. Most of *EgrAGO* proteins localize in nucleus and cytoplasm, showing basicity, hydrophilia and high aliphatic index. Furthermore, the expression of *EgrAGOs* in different tissues had significant differences, and which was high in tissues or organs related with wood forming. So, these would lay a foundation for studying *EgrAGO* genes function in future.

Key words: *Eucalyptus*; AGO protein; Gene family; Bioinformation

AGO 蛋白(Argonaute protein)是小 RNA 介导的 RNA 沉默通路中 RNA 诱导的沉默复合物(RNA-induced silencing complex, RISC)的核心成分。AGO

蛋白通过与 miRNAs (microRNAs)、siRNAs (small interfering RNAs)、piRNAs (Piwi-interacting RNAs) 等不同类型的小非编码 RNA (small non-coding

收稿日期: 2014-10-09

接受日期: 2015-01-09

基金项目: 国家“863” 高技术研究发展计划(2013AA102705); 国家自然科学基金项目(31400554); 中央级公益性科研院所基本科研业务费专项(RITFYWZX201304)资助

作者简介: 范春节,男,博士,研究方向为桉树木材形成分子调控机理。E-mail: fancy0417@163.com

* 通信作者 Corresponding author. E-mail: b.s.zeng@vip.tom.com

RNA)结合,在这些小RNA的引导下,AGO蛋白可以特异地停留在与小RNA互补的靶基因mRNA上,通过AGO蛋白自身的内切酶对目标靶基因进行切割以引起靶基因的沉默^[1-3]。根据sRNA的5'末端核苷酸,不同的AGO蛋白选择性地与sRNA结合。如5'-U的sRNA优先选择停靠在AGO1蛋白上,5'-A的sRNA则会选择停靠在AGO2、AGO4、AGO6、AGO9上,而5'-C的sRNA则选择与AGO5优先结合^[4]。在拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)等模式植物中,AGO蛋白与小RNA形成复合体,在调控植物生长发育、对逆境的胁迫响应以及防御病毒入侵中起到重要的作用^[5]。拟南芥的AGO基因家族共有10个成员,主要分为3类:第一类包括*AtAGO1*、*AtAGO5*、*AtAGO10*,第二类包括*AtAGO2*、*AtAGO3*、*AtAGO7*,第三类包括*AtAGO4*、*AtAGO6*、*AtAGO8*、*AtAGO9*^[6]。通过结合的小RNA推断出他们编码蛋白的主要功能和调控机制,*AtAGO1*参与miRNA通路调控的转录后基因沉默通路来调控植物的生长发育,以及转基因引起的siRNA、病毒siRNA、ta-siRNA调控途径^[7-8]。*AtAGO10*与*AtAGO1*存在部分相同或相似的功能,可互补甚至替代*AtAGO1*的功能^[9]。*AtAGO4*主要通过24nt-siRNA结合,调控重复序列、转座子和异染色质区域等表观遗传学沉默区域以调控植物的生长发育^[10],而*AtAGO6*和*AtAGO9*在DNA甲基化的缺失和sRNA类型的积累上存在差异,表明他们具有独特的功能机制,而*AtAGO4*和*AtAGO6*还存在着功能上的部分冗余^[11]。*AtAGO7*主要结合miR390,进一步引起起源于TAS3前体的tasiRNA,这些tasiRNA主要靶基因为编码激素响应相关蛋白的mRNA,也即为主要通过调控这些mRNA而参与到调控植物的生长发育中^[12-13]。另外不同类的AGOs蛋白也通过协同作用共同调控植物的生长发育^[14]。

桉树(*Eucalyptus sp.*)作为一种重要的造林树种,目前已经完成巨桉(*E. grandis*)的基因组测序^[15],有大量功能基因需要挖掘和应用,而目前尚未对巨桉中*EgrAGO*基因家族的结构及其功能进行相关研究。本研究以分析*EgrAGO*的表达模式和*EgrAGO*蛋白的性质为重点,借助已明晰的拟南芥相关基因的结构及功能,运用生物信息、比较基因组学、功能基因组学方法,在巨桉中找到相关基因,并分析其结构、功能及进化过程。通过对

*EgrAGO*基因在不同部位的表达进行分析,选择合适的基因进行分子克隆,为后期探索*EgrAGO*基因参与木材发育及其他生长发育的功能机制研究奠定基础。

1 材料和方法

1.1 材料

以巨桉(*Eucalyptus grandis*)为研究对象,AGO基因序列信息来源于桉树基因组数据库(版本V1.0),包括美国国立生物技术信息中心(NCBI)(www.ncbi.nlm.nih.gov)、Phytozome桉树基因组数据(<http://www.phytozome.net/eucalyptus.php>)、桉树基因组数据中心 EucGenIE (<http://www.eucgenie.org/>)。

1.2 AGOs基因的查询与确认

在拟南芥数据库TAIR (<http://www.arabidopsis.org>)搜索全部*AtAGO*基因的信息,在Phytozome V9.0 (<http://www.phytozome.net/search.php>)巨桉基因组中查找每一个*AtAGO*蛋白的同源序列,作为候选蛋白。得到这些蛋白序列相似的巨桉AGO基因后,将其作为二次查询序列,再次检索和反复查找补充,确定巨桉中所有候选的AGO家族基因。将所有的候选基因在PFAM (<http://pfam.sanger.ac.uk/>)和SMART (<http://smart.embl-heidelberg.de/>)蛋白结构域在线分析预测网站上分析候选蛋白结构域,若存在PF08699、PF02170、PF02171序列则认为该候选蛋白是AGO蛋白家族,最后得到的基因即为桉树所有的*EgrAGOs*基因家族成员。除去重复名称序列,将下载得到的所有核苷酸序列和氨基酸序列编辑为FASTA格式存储于txt文本中。

1.3 *EgrAGO*基因的染色体定位和复制分析

在Phytozome中对*EgrAGO*基因的Genomic view进行染色体定位。基因组的大片段复制(Segmental genome duplication)和串联重复基因(Tandemly duplicated genes)使用新的桉树基因组复制分析数据。重复基因之间的氨基酸和核苷酸相似度(Homology)采用DNAMAN软件进行计算。

1.4 *EgrAGO*蛋白的生物信息学分析

巨桉AGOs蛋白相关基因的DNA长度、转录长度、CDS序列、外显子、内含子结构以及翻译的

蛋白序列均从 Phytozome 的巨桉数据库中获得。巨桉 AGOs 蛋白的等电点、分子量、脂溶性、疏水性和稳定性均采用在线工具 ExPASy ProtParam (<http://www.expasy.org/tools/protparam.html>)进行分析。对 AGOs 蛋白的亚细胞定位使用 WoLF PSORT (<http://psort.nibb.ac.jp/>)的扩展程序 PSORTII (<http://www.psort.org>)进行预测。

1.5 序列比对和进化树分析

对巨桉和拟南芥 AGO 蛋白家族的氨基酸序列进行在线分析(<http://phylogeny.lirmm.fr/>)并构建进化树。其中多序列比对采用 Muscle 和 Gblock 程序,具体参数设置为:Muscle 采用 full processing mode 运行模块, Gblock 设置为 Min. seq. for flank pos.: 85%、Max. contig. nonconserved pos.: 8、Min. block length: 10、Gaps in final blocks: no。构建进化树采用基于最大似然法(Maximum likelihood)的 PhyML 程序和进化树分析动态作图标记软件 TreeDyn198.3,具体参数设置如下:其中 PhyML 参数设置为 Model: Default、Statistical test: aIrt、Number of categories: 4、Gamma: estimated、Invariable sites: estimated, TreeDyn 参数设置为 Conformation: rectangular、Branch annotation: bootstrap。

表 1 巨桉 AGO 基因家族的结构

Table 1 Structure of *EgrAGO* genes in *Eucalyptus grandis*

基因 Gene	基因座位 Locus	序列长度 Sequence length (bp)	转录长度 Transcription length (bp)	编码序列长度 CDS length (bp)	内含子数量 Number of introns
<i>EgrAGO1</i>	Eucgr.K02304.1	9143	3777	3225	21
<i>EgrAGO1a</i>	Eucgr.H00615.1	8251	3560	3156	21
<i>EgrAGO2</i>	Eucgr.D01097.1	3810	2907	2907	2
<i>EgrAGO3a</i>	Eucgr.D00105.1	3096	2676	2676	3
<i>EgrAGO3b</i>	Eucgr.D00106.1	3893	2868	2868	2
<i>EgrAGO3c</i>	Eucgr.D00109.1	3607	2877	2877	2
<i>EgrAGO4a</i>	Eucgr.B03780.1	6626	3157	2757	22
<i>EgrAGO4b</i>	Eucgr.J00634.1	8324	3905	2850	22
<i>EgrAGO5a</i>	Eucgr.B03907.1	7746	3522	3015	21
<i>EgrAGO5b</i>	Eucgr.B03909.1	7035	3205	3003	21
<i>EgrAGO6a</i>	Eucgr.K00735.1	7871	3250	2715	22
<i>EgrAGO6b</i>	Eucgr.K00736.1	6479	2980	2643	21
<i>EgrAGO7</i>	Eucgr.G02476.1	3896	3329	3051	2
<i>EgrAGO10</i>	Eucgr.H00532.1	9788	3491	2943	21

1.6 基因表达分析

将巨桉 AGO 蛋白的转录号输入到 EucGenIE 数据库中进行查询,查找其在不同组织中的表达量,构建热图,分析其在不同组织中的表达模式以及相对表达量。

2 结果和分析

2.1 巨桉中 *EgrAGO* 基因家族数量

利用拟南芥中的 10 个 AGO 蛋白进行同源比对分析,在 phytozome 的桉树数据库中共查到 14 个 *EgrAGO* 基因序列(表 1)。*EgrAGO* 基因的序列长度差异较大,其中最小的为 *EgrAGO3a*,只有 3096 bp,而 *EgrAGO10* 有 9788 bp。*EgrAGO* 转录的长度差异比基因组序列的小,所有 *EgrAGO* 转录长度为 2676~3905 bp,最短的是 *EgrAGO3a*,最长的为 *EgrAGO4b*。而编码区长度为 2645~3225 bp,最小的为 *EgrAGO6b*,最大的为 *EgrAGO1*。可以看出, *EgrAGO1* 和 *EgrAGO10* 基因有大量的非编码区,而 *EgrAGO2*、*EgrAGO3a*、*EgrAGO3b*、*EgrAGO3c* 和 *EgrAGO7* 的相对较少。*EgrAGO2*、*EgrAGO3a*、*EgrAGO3b*、*EgrAGO3c* 和 *EgrAGO7* 的内含子为 2~3 个,基因序列长度为 3000~4000 bp,

而其他家族成员大多在 6000 bp 以上,内含子为 21~22 个。这意味着这些基因通过结构上的差异行使着不同的功能或具有不同的作用模式。值得注意的是, *EgrAGO2*、*EgrAGO3a*、*EgrAGO3b* 和 *EgrAGO3c* 没有 UTR 序列,而其他家族成员大多有 3' UTR 或 5' UTR 序列,或者两者兼有之。

2.2 *EgrAGO* 基因在染色体中的分布

14 个 *EgrAGO* 基因选择性地分布在巨桉第 2、4、7、8、10 和 116 条染色体上(图 1),在 2、4、11 染色体上呈集中分布,其中第 4 条染色体上分布的最多,有 *EgrAGO3a*、*EgrAGO3b*、*EgrAGO3c*、*EgrAGO2* 等 4 个基因。而在第 2 和 11 条染色体上各有 3 个,第 8 条染色体上有 2 个,第 7 和 10 条染色体上各有 1 个基因。

可以明显看出, *EgrAGO* 基因家族中存在着重复基因,这可能是因为这些基因通过串联复制方式来增加数量,如第 2 条染色体上的 *EgrAGO5a* 和 *EgrAGO5b* 基因,第 4 条染色体上的 *EgrAGO3a*、*EgrAGO3b* 和 *EgrAGO3c* 基因,第 8 条染色体上的 *EgrAGO10* 和 *EgrAGO1a* 基因,第 11 条染色体上的 *EgrAGO6a* 和 *EgrAGO6b* 基因。另外约在 109.9 百

万年前桉树发生了一次全基因组的大规模加倍事件^[15],所以在不同染色体的大片段复制区 *EgrAGO*s 基因也存在重复基因,如第 2、10 染色体上的 *EgrAGO4a* 和 *EgrAGO4b* 基因在大片段复制区有 2 对重复基因,第 8、11 条染色体上也存在 *EgrAGO1a*、*EgrAGO1* 和 *EgrAGO10* 基因的重复基因。可见,在第 2、8 条染色体上既有大片段复制的基因也存在串联重复基因的基因簇。

2.3 *EgrAGO*s 基因家族的进化分析

对 14 个巨桉 *EgrAGO* 基因和 10 个拟南芥 *AtAGO* 基因编码的氨基酸序列构建进化树。从图 2 可见,这 24 个 *AGO* 基因家族成员可分为 3 类,每类都有非常高的支持率。A 类成员有 8 个,其中巨桉有 5 个,而拟南芥只有 3 个;B 类成员有 8 个,其中巨桉的 4 个成员聚在同一个分支上;C 类成员有 8 个,其中巨桉有 4 个,拟南芥有 4 个。由于每一类中均有来自拟南芥和巨桉的 *AGO* 基因,因此推测巨桉的祖先物种中至少含有 3 个 *AGO* 基因,并且在桃金娘科植物分化之后没有发生明显的基因丢失事件,这些也说明 *AGO* 基因作为一类在植物生长发育中具有重要功能的基因家族在进化上是保

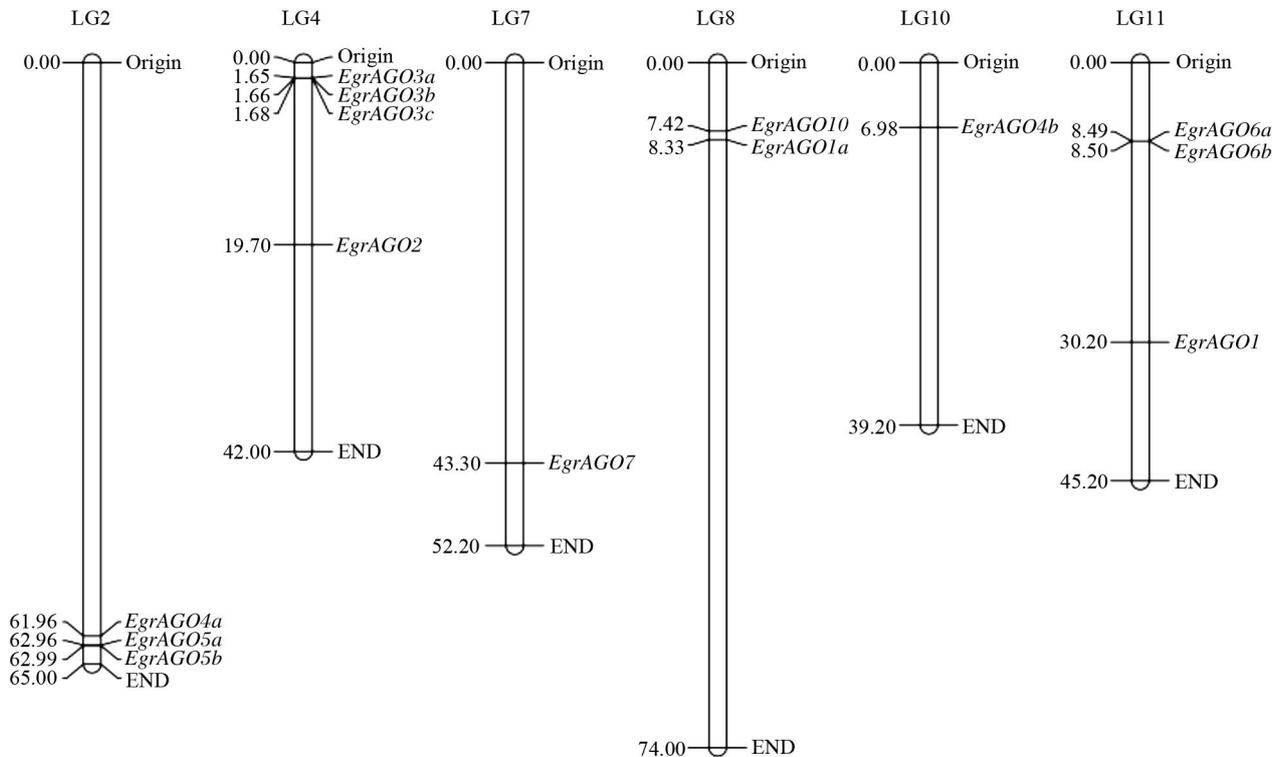


图 1 巨桉 *EgrAGO* 家族基因在染色体上的定位

Fig. 1 Distribution of *EgrAGO* genes on chromosomes of *Eucalyptus grandis*

守的。A类中的 Eucgr.B03907 与 Eucgr.B03909 基因座位，B类中的 Eucgr.D00105、Eucgr.D00106、Eucgr.D00109 和 Eucgr.D01097 基因座位，C类中的 Eucgr.K00735 和 Eucgr.K00736 基因座位分别具有较高的重复性，且分别聚类在小的分支上，这些都说明桉树从祖先物种分化之后发生了基因重复事件。同时所有的 AGO 蛋白都存在 3 个功能域：Piwi、PAZ、DUF1785。

应用 DNAMAN 对巨桉 AGO 蛋白家族的氨基酸序列进行多序列比对(表 2)，结果表明，A类和C类成员间核苷酸和氨基酸序列的一致性较高，均超过 50.0%，而B类成员间核苷酸序列的一致性最低，为 43.66%，氨基酸序列的一致性只有 30.75% (表 2)。同类成员间氨基酸的高度保守性也说明了他们可能存在基因功能的互补作用。同时由于 AGO

基因家族作为特异与小 RNA 结合进行目标靶基因切割，需要氨基酸序列和蛋白结构的高度精确，从而行使切割功能。A类和B类成员间的亲缘关系更近，氨基酸序列的一致性为 25.47%~36.56%；而B类和C类成员间的氨基酸序列一致性低于 28.0%。序列差异较大也意味着 AGO 基因家族成员中的 3 类基因可能具有各自独特的功能和作用方式。同时，基因核苷酸序列的一致性比对可以作为 AGO 蛋白家族成员分类的重要依据。从依据 AGO 蛋白的氨基酸序列构建的系统进化树看出，每一类成员往往聚集成群，同时同一类蛋白的基因内含子数量、相似程度以及序列长度、结构都保持着高度的一致性，这也说明同一类 AGO 蛋白的基因起源与其结构和功能有一定的关联。

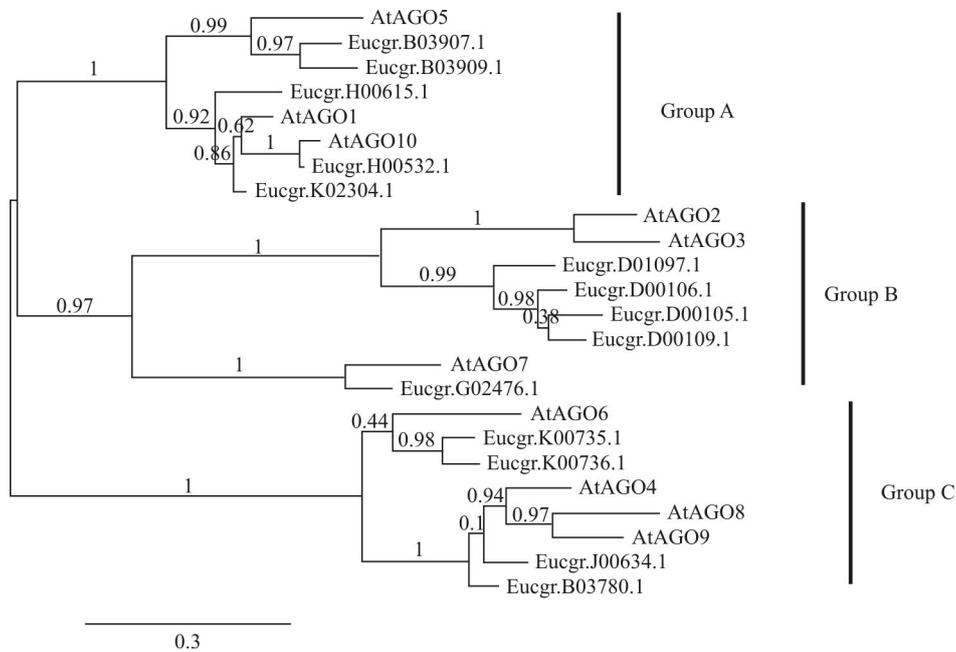


图 2 巨桉和拟南芥 AGO 蛋白家族的进化树。分支上的数字为支持率。

Fig. 2 Phylogenetic tree of AGO family in *Eucalyptus grandis* and *Arabidopsis thaliana*. Data in branches present bootstrap.

表 2 巨桉 AGO 家族氨基酸序列和核苷酸序列的一致性(%)

Table 2 Homology (%) of amino acid (nucleotide) sequences among three clades of AGO in *Eucalyptus grandis*

	A	B	C
A	50.30~78.73 (52.98~83.22)		
B	25.47~36.56 (39.06~46.91)	30.75~82.97 (43.66~88.14)	
C	25.92~31.80 (40.24~45.01)	23.04~27.46 (39.41~43.95)	55.67~82.43 (58.15~86.64)

2.4 EgrAGO蛋白的理化性质

对巨桉 AGO 蛋白的理化性质以及亚细胞定位进行了预测分析(表 3)。结果表明,所有 EgrAGO 家族成员的氨基酸序列均较长,含 880~1074 个氨基酸。EgrAGO 蛋白的等电点(PI)为 8.6~9.43,说明 EgrAGO 蛋白带正电荷。EgrAGO 蛋白的亲水性为 70.17~84.77,表现为亲水性。除了 EgrAGO3a 蛋白的稳定性指数小于 40,表现为稳定外,其他家族成员的稳定性指数均大于 40,表现为不稳定。EgrAGO 蛋白的脂溶指数(Aliphatic index)较高,为 -0.588~-0.260。预测大部分的 EgrAGO 蛋白定位于细胞核和细胞质中,这与其行使功能的位置保持一致;而 EgrAGO5a 蛋白定位于叶绿体-线粒体, EgrAGO5b 定位于叶绿体,意味着他们可能需在特定的位置和区域行使功能和作用。

2.5 EgrAGO 基因的表达

通过 EucGenIE 网站获得巨桉 AGO 家族基因

在未成熟木质部、木质部、韧皮部、茎尖、幼叶和成熟叶等 6 个组织中的转录表达数据,根据 *EgrAGOs* 基因在不同组织的表达数据做出热图(图 3)。结果表明, A 类和 B 类成员具有较高的转录丰度;而 C 类成员的转录丰度较低,其中 *EgrAGO2* 和 *EgrAGO3a* 在 6 个组织中均未有转录的现象。另外, *EgrAGO1*、*EgrAGO1a*、*EgrAGO4a.1*、*EgrAGO4b.1* 和 *EgrAGO4b.2* 在与木材形成相关的木质部和未成熟木质部中有较高的转录。几乎所有的家族成员都存在着差异表达,而且大多数基因在韧皮部和茎尖有着较高的表达量。值得关注的是 *EgrAGO4b.1* 在木质部中的表达较高,而在其他组织中表达量低,这意味着 *EgrAGO4b.1* 可能以某种方式参与巨桉木材形成的分子调控过程中。而 *EgrAGO10.1*、*EgrAGO10.2* 在茎尖和幼叶中的表达量远远高于其他组织的,说明其主要作用部位为茎尖和幼叶。而 *EgrAGO3c* 主要在韧皮部中表达,其可能参与韧皮部的发育过程。

表 3 EgrAGO 蛋白的理化性质和亚细胞定位

Table 3 Physical and chemical properties and subcellular localization of EgrAGO proteins

蛋白 Protein	氨基酸数量 Number of amino acids	分子量 Molecular weight (Da)	PI	亚细胞定位 Subcellular localization	稳定性 Stability	亲水性 Hydropathicity	脂溶指数 Aliphatic index
EgrAGO1	1074	119403.4	9.35	nucl: 9.0, cyto: 3.0, chlo: 1.0	不稳定 Unstable	70.17	-0.588
EgrAGO1a	1015	116567.1	9.35	nucl: 9.0, cyto: 2.0, chlo: 1.0, cysk: 1.0	不稳定 Unstable	73.00	-0.536
EgrAGO2	968	111836.0	9.21	nucl: 7.0, cyto: 4.0, mito: 2.0	不稳定 Unstable	79.21	-0.541
EgrAGO3a	891	101972.7	9.12	nucl: 5.0, cyto: 3.0, chlo: 2.0, vacu: 2.0, plas: 1.0	稳定 Stable	84.77	-0.260
EgrAGO3b	955	109764.5	9.06	nucl: 9.0, cyto: 2.0, chlo: 1.0, mito: 1.0	不稳定 Unstable	75.72	-0.532
EgrAGO3c	958	109030.5	8.60	nucl: 9.0, cyto: 2.0, chlo: 1.0, mito: 1.0	不稳定 Unstable	75.91	-0.473
EgrAGO4a	918	102315.2	9.09	nucl:12.0, cyto:2.0	不稳定 Unstable	83.40	-0.376
EgrAGO4b	949	105807.8	8.95	nucl: 8.0, vacu: 3.0, cyto: 2.0	不稳定 Unstable	81.83	-0.330
EgrAGO5a	1004	110689.5	9.50	chlo_mito: 6.3, mito: 6.0, chlo: 5.5, cyto_mito: 3.8	不稳定 Unstable	79.29	-0.382
EgrAGO5b	1000	110912.5	9.43	chlo: 6.0, mito: 5.0, nucl: 2.0	不稳定 Unstable	79.98	-0.382
EgrAGO6a	904	101338.3	9.21	nucl: 10.0, cyto: 2.0, chlo: 1.0	不稳定 Unstable	84.60	-0.360
EgrAGO6b	880	98870.6	9.22	nucl: 9.0, cyto: 3.0, chlo: 1.0	不稳定 Unstable	85.06	-0.378
EgrAGO7	1016	114973.0	9.34	nucl: 11.0, cyto: 1.0, vacu: 1.0	不稳定 Unstable	82.96	-0.450
EgrAGO10	980	110045.3	9.39	cyto: 9.0, nucl: 3.0, chlo: 1.0	不稳定 Unstable	79.79	-0.442

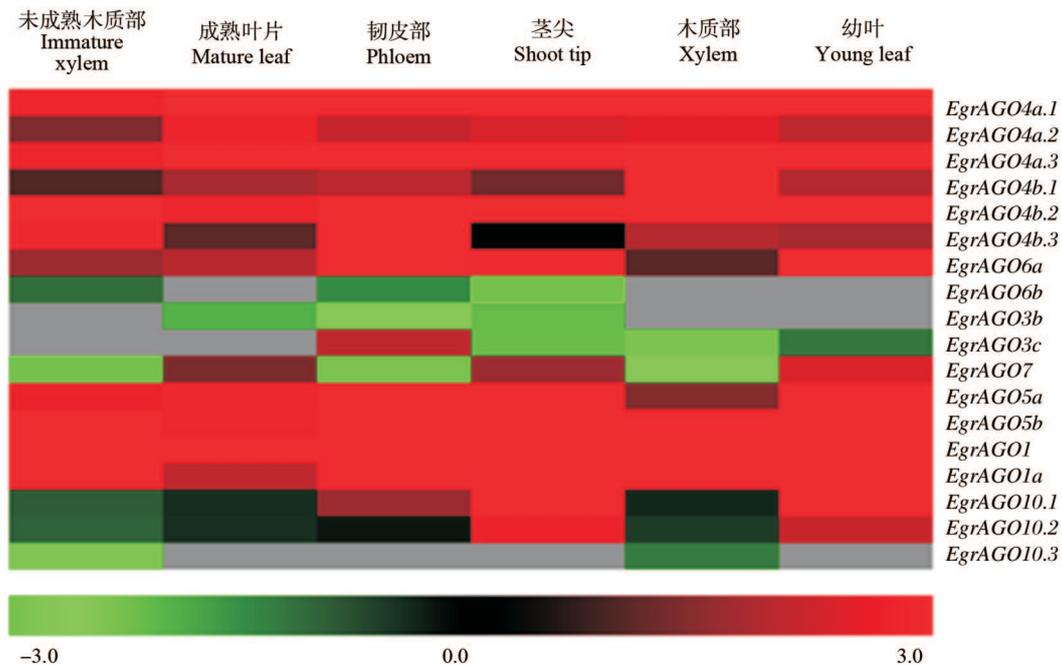


图 3 巨桉 *AGO* 家族基因在不同组织中的表达。不同颜色的小方块代表基因表达谱数据, 一个小块代表一个数据, 红色和绿色反映基因在同等条件下转录水平上调或下调, 颜色的深浅表示数据绝对值的大小。

Fig. 3 Expression of *EgrAGO* genes in different tissues or organs of *Eucalyptus grandis*. The difference in gene expression is listed in square brackets. A square bracket represents a value of expression of *EgrAGO* genes in a tissue or organ. Red represents up-regulation, and blue represents down-regulation. The depth of colour of square brackets indicate the value of gene expression.

3 讨论

本研究利用已知的拟南芥 AGOs 蛋白序列在桉树基因组数据库中进行搜索, 共找到 14 个巨桉 *EgrAGO* 基因, 通过对 *EgrAGOs* 基因序列和结构分析, 结果表明这些基因包含 3 个基因簇, 这些基因簇是在植物进化过程中通过染色体复制产生的。另外, 基因的外显子数量、大小在不同簇内保持较高的一致性, 这些与拟南芥和水稻中 *AGO* 基因家族的结构和特征一致^[6,16], 预示他们的基因起源甚至功能有一定的关联性。尽管如此, 与拟南芥相比, 巨桉 *EgrAGO* 基因的 A 类成员有 5 个, 比拟南芥多 2 个, 是否存在功能上的冗余, 或者这些新成员具有不同的功能或调控机制, 还需要进一步深入研究。

植物 *AGO* 蛋白功能的鉴定主要在模式植物如拟南芥中进行, 根据其结合的 sRNA 是否参与到子叶的分化、花器官形成、茎尖分生组织等的发育调控以及病毒侵染植物过程中的 RNAi 调控^[9,17-19]来鉴定。目前对 *AGO* 家族中亲缘关系最近的两个成员 *AGO1* 和 *AGO10* 的研究最多, 其功能主要是参与茎尖分生组织的维持, 而且通过

RNAi 机制调控 *LFY* (*LEAFY*)、*API* (*APETALA1*) 和 *AG* (*AGAMOUS*) 基因的表达, 最终影响植物的分生组织决定、开花转换和花器官的确定, 另外还通过 *KNOX*、*AG* 和 *AP3* 作用于叶片和花粉管的分化^[20-21]。*AGO1* 和 *AGO10* 还可以通过分别调控 miR172 和 miR165/166 以引起成花干细胞的终止^[17]。而 *AGO1* 作为 miR168 的靶基因, 近期的研究表明其与 miR168 通过协同调控在胁迫响应以及信号传导中起着重要的作用, 而且这种作用机制是保守的^[22]。表达分析表明, *AGO1* 广泛存在于各种组织中, 而 *AGO10* 则仅在分生组织、导管组织和侧生器官原基的近轴端^[23]。而在巨桉中, *AGO1* 广泛存在于不同组织中且表达丰度很高, 而 *AGO10* 主要在茎尖和幼嫩叶片中表达。这些说明 *AGO1* 和 *AGO10* 在进化上属于高度保守的基因, 也预示着他们在巨桉中的功能以及作用机制可能与拟南芥相似。

另外, 水稻中有 19 个 *AGO* 家族基因, 且主要在生殖生长阶段表达, 尤其在圆锥花序阶段的表达量升高, 其中 6 个基因能够响应激素处理, 且大多数基因与 *DCL* (*Dicer-like*)、*DRB* (*Double-stranded*

RNA binding)、RDR (RNA-dependent RNA polymerase)等参与 RNA 沉默相关的基因在生长发育以及非生物胁迫中存在着共表达现象^[16]。番茄 (*Lycopersicon esculentum*)中有 15 个 *SIAGO* 基因,大部分基因在番茄黄叶卷曲病毒侵染和非生物胁迫中表达量明显升高^[24]。*SIAGO1* 蛋白和 *miR168* 在决定植物生长速度、发育阶段改变、叶的向上性、果实的起始和膨大以及其他发育阶段起着重要的作用^[25]。

尽管在模式植物中对 AGO 蛋白家族成员的功能和作用机制的研究取得了很大进展^[22,25-28],但在木本植物中对 AGO 蛋白家族的基因研究较少。从杨树(*Populus sp.*)中克隆了 *PeAGO5a* 和 *PdAGO5b* 基因,其编码的氨基酸序列含有 3 个保守的结构域 DUF1785、PAZ 和 Piwi,差异表达分析表明其可能参与杨树硬枝插穗不定根发育的调控过程^[26]。次生长作为木本植物重要特性,AGO 是否通过 sRNA 调控途径参与这一过程值得关注。近期的研究表明,AGO1 和 AGO10 在不同部位通过 *miR165/166* 作用于茎尖分生组织的分化^[9,17],而木本植物形成层的分化已经证实受到 *miR165/166* 的调控^[27-28],而 AGO 蛋白家族成员是否作用于这个过程或者家族成员中的哪一个成员作用于这个过程,还需要进一步的研究。同时本研究结果表明, *EgrAGO4b.1* 在木质部中存在高表达,而在其他组织的表达量低甚至不表达。而 AGO4 可以通过 RNA-directed DNA methylation (RdDM)途径与 Heterochromatic siRNAs (hc-siRNAs)作用调控 DNA 的甲基化,在胞质中组装形成复合体,然后通过核孔运输到核内行使其功能^[29]。在冰晶花 (*Mesembryanthemum crystallinum*)中 *AGO4* 参与盐胁迫下 UBC8 的泛素化过程^[30]。而 *EgrAGO4* 是否通过调控 DNA 甲基化途径来调控巨桉的木材形成发育过程,还需深入研究。

参考文献

- [1] Tolia N H, Joshua-Tor L. Slicer and the argonautes [J]. *Nat Chem Biol*, 2007, 3(1): 36–43.
- [2] Ghildiyal M, Zamore P D. Small silencing RNAs: An expanding universe [J]. *Nat Rev Genet*, 2009, 10(2): 94–108.
- [3] Cenik E S, Zamore P D. Argonaute proteins [J]. *Curr Biol*, 2011, 21(12): R446–R449.
- [4] Mi S J, Cai T, Hu Y G, et al. Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5' terminal nucleotide [J]. *Cell*, 2008, 133(1): 116–127.
- [5] Mallory A, Vaucheret H. Form, function, and regulation of ARGONAUTE proteins [J]. *Plant Cell*, 2010, 22(12): 3879–3889.
- [6] Vaucheret H. Plant ARGONAUTES [J]. *Trends Plant Sci*, 2008, 13(7): 350–358.
- [7] Sha A H, Zhao J P, Yin K Q, et al. Virus-based MicroRNA silencing in plants [J]. *Plant Physiol*, 2014, 164(1): 36–47.
- [8] Fernández-Nohales P, Domenech M J, de Alba A E M, et al. AGO1 controls arabidopsis inflorescence architecture possibly by regulating *TFL1* expression [J]. *Ann Bot*, 2014, 114(7): 1471–1481.
- [9] Zhu H L, Hu F Q, Wang R H, et al. *Arabidopsis* argonaute10 specifically sequesters *miR166/165* to regulate shoot apical meristem development [J]. *Cell*, 2011, 145(2): 242–256.
- [10] Li C F, Pontes O, El-Shami M, et al. An ARGONAUTE4-containing nuclear processing center colocalized with cajal bodies in *Arabidopsis thaliana* [J]. *Cell*, 2006, 126(1): 93–106.
- [11] Zheng X W, Zhu J H, Kapoor A, et al. Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing [J]. *EMBO J*, 2007, 26(6): 1691–1701.
- [12] Cuperus J T, Fahlgren N, Carrington J C. Evolution and functional diversification of *MIRNA* genes [J]. *Plant Cell*, 2011, 23(2): 431–442.
- [13] Montgomery T A, Howell M D, Cuperus J T, et al. Specificity of ARGONAUTE7-miR390 interaction and dual functionality in *TAS3* trans-acting siRNA formation [J]. *Cell*, 2008, 133(1): 128–141.
- [14] Wang H, Zhang X R, Liu J, et al. Deep sequencing of small RNAs specifically associated with *Arabidopsis* AGO1 and AGO4 uncovers new AGO functions [J]. *Plant J*, 2011, 67(2): 292–304.
- [15] Myburg A A, Grattapaglia D, Tuskan G A, et al. The genome of *Eucalyptus grandis* [J]. *Nature*, 2014, 510(7505): 356–362.
- [16] Yang Y, Zhong J, Ouyang Y D, et al. The integrative expression and co-expression analysis of the *AGO* gene family in rice [J]. *Gene*, 2013, 528(2): 221–235.
- [17] Ji L J, Liu X G, Yan J, et al. *ARGONAUTE10* and *ARGONAUTE1* regulate the termination of floral stem cells through two microRNAs in *Arabidopsis* [J]. *PLoS Genet*, 2011, 7(3): e1001358.
- [18] Minoia S, Carbonell A, Di Serio F, et al. Specific ARGONAUTES selectively bind small RNAs derived from potato spindle tuber viroid and attenuate viroid accumulation *in vivo* [J]. *J Virol*, 2014, 88(20): 11933–11945.
- [19] Baumberger N, Baulcombe D C. *Arabidopsis* ARGONAUTE1 is an RNA slicer that selectively recruits microRNAs and short interfering RNAs [J]. *Proc Natl Acad Sci USA*, 2005, 102(33): 11928–11933.
- [20] Kidner C A, Martienssen R A. The role of ARGONAUTE1

- (AGO1) in meristem formation and identity [J]. *Dev Biol*, 2005, 280(2): 504–517.
- [21] Yang L, Huang W, Wang H, et al. Characterizations of a hypomorphic argonaute1 mutant reveal novel AGO1 functions in *Arabidopsis* lateral organ development [J]. *Plant Mol Biol*, 2006, 61(1/2): 63–78.
- [22] Li W, Cui X, Meng Z L, et al. Transcriptional regulation of *Arabidopsis* MIR168a and *ARGONAUTE1* homeostasis in abscisic acid and abiotic stress responses [J]. *Plant Physiol*, 2012, 158(3): 1279–1292.
- [23] Lynn K, Fernandez A, Aida M, et al. The *PINHEAD/ZWILLE* gene acts pleiotropically in *Arabidopsis* development and has overlapping functions with the *ARGONAUTE1* gene [J]. *Development*, 1999, 126(3): 469–481.
- [24] Bai M, Yang G S, Chen W T, et al. Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in *Solanum lycopersicum* [J]. *Gene*, 2012, 501(1): 52–62.
- [25] Xian Z Q, Huang W, Yang Y W, et al. miR168 influences phase transition, leaf epinasty, and fruit development via *SIAGO1s* in tomato [J]. *J Exp Bot*, 2014, 65(22): 6655–6666.
- [26] Xu M, Xie W F, Pan H X, et al. Cloning and characterization of *ARGONAUTE* genes in *Populus* [J]. *Sci Silv Sin*, 2011, 47(3): 46–51.
胥猛 谢雯凡, 潘惠新, 等. 杨树*ARGONAUTE*基因的克隆及序列分析 [J]. *林业科学*, 2011, 47(3): 46–51.
- [27] Robischon M, Du J, Miura E, et al. The *Populus* class III HD ZIP, *popREVOLUTA*, influences cambium initiation and patterning of woody stems [J]. *Plant Physiol*, 2011, 155(3): 1214–1225.
- [28] Ohashi-Ito K, Fukuda H. Transcriptional regulation of vascular cell fates [J]. *Curr Opin Plant Biol*, 2010, 13(6): 670–676.
- [29] Ye R Q, Wang W, Iki T, et al. Cytoplasmic assembly and selective nuclear import of *Arabidopsis* ARGONAUTE4/siRNA complexes [J]. *Mol Cell*, 2012, 46(6): 859–870.
- [30] Li C H, Chiang C P, Yang J Y, et al. RING-type ubiquitin ligase McCPN1 catalyzes UBC8-dependent protein ubiquitination and interacts with Argonaute 4 in halophyte ice plant [J]. *Plant Physiol Biochem*, 2014, 80: 211–219.