

## 棉花LIM蛋白基因家族的进化及表达特征分析

杨洋, 李波, 胡文冉, 张经博, 范玲\*

新疆农业科学院核技术生物技术研究所, 乌鲁木齐830091

**摘要:** 植物LIM蛋白参与细胞骨架的调节及调控次生细胞壁的发育。本研究在雷蒙德氏棉(D5)、亚洲棉(A2)基因组和陆地棉的A、D亚基因组中分别鉴定出14个LIM基因。陆地棉*GhLIM*基因在2个亚组间的染色体上的分布及基因的复制模式是一致的。推测其复制发生在其祖先同可(6个LIM)分开之后, A2、D5基因组分化之前; 陆地棉LIM除多倍体化使其数目加倍外, 没有基因复制与丢失事件发生。蛋白系统进化分析表明, 所有被子植物LIM蛋白可以划分为4个亚组(部分具有小亚组):  $\alpha$ LIM1 (FLIM1、WLIM1、PLIM1)、 $\beta$ LIM1、 $\gamma$ LIM2 (WLIM2)、 $\delta$ LIM2 (PLIM2-I、PLIM2-II)。陆地棉*GhLIM*的同源复制基因分别在WLIM2和PLIM2-I、PLIM2-II中聚为独立的分支。通过分析深度测序表达谱, 发现陆地棉*GhLIM*有3种表达类型: 组成型表达的*GhLIM1*、2 (FLIM), *GhLIM3* (WLIM1), *GhLIM13* ( $\beta$ LIM1)和*GhLIM4*、8、14 (WLIM2); 雄蕊中优势表达的*GhLIM7*、9 (PLIM2-I), *GhLIM5*、6、10、11 (PLIM2-II); 在雄蕊和纤维中优势表达的*GhLIM12* (PLIM2-I)。实时荧光定量PCR分析表明, *GhLIM*基因各成员在不同的纤维发育阶段具有不同的表达特征, *GhLIM1*、13、14在纤维的快速伸长期(5~15 DPA)上调表达, 在15 DPA表达量最高; *GhLIM2*则在纤维次生壁发育期(18~28 DPA)显著上调表达, 在28 DPA表达量最高。而*GhLIM12*在10~23 DPA持续高表达, 在28 DPA则没有表达。上述结果说明*GhLIM*家族基因在纤维发育的不同阶段发挥着不同的功能。

**关键词:** 棉花; LIM; 纤维发育; 基因复制; 基因表达模式

## Evolution and Expression Analysis of the LIM Protein Gene Family in Cotton

YANG Yang, LI Bo, HU Wen-Ran, ZHANG Jing-Bo, FAN Ling\*

Institute of Nuclear and Biological Technologies, Xinjiang Academy of Agricultural Sciences, Urumuqi 830091, China

**Abstract:** Plant LIM acts as an actin bundler to facilitate elongation of cells and also functions as a transcription factor involved in secondary cell wall development. In this study, 14 putative LIM genes were identified from the whole genomes of *Gossypium raimondii* (D5), *G. arboreum* (A2), and 2 subgenomes (A, D) of *G. hirsutum*, respectively. In *G. hirsutum*, the distribution and the duplication patterns of LIM family members on each couple homology chromosomes were consistent, suggesting that the LIM gene may had duplicated since the ancestors of cacao and diploid cottons had diverged, and before the divergence of the D5 and A2 diploid cottons, and *GhLIMs* may not experienced a single duplicate or gene loss in *G. hirsutum*, except the chromosome redoubling. The phylogenetic analysis suggested that the angiosperm LIM proteins could clustered into four groups (subgroups are gathered within groups):  $\alpha$ LIM1 (FLIM1, WLIM1, PLIM1),  $\beta$ LIM1,  $\gamma$ LIM2 (WLIM2),  $\delta$ LIM2 (PLIM2-I, PLIM2-II); the duplicated homologs of *GhLIM* were clusted in WLIM2, PLIM2-I, PLIM2-II subgroups respectively. The *GhLIM* genes expressing profiling analysis showed that, there were 3 types of expressing patterns: *GhLIM1/2* (FLIM), *GhLIM3* (WLIM1), *GhLIM13* ( $\beta$ LIM1) and *GhLIM4/8/14* (WLIM2) were widely expressed in different tissues; *GhLIM7/9* (PLIM2-I) and *GhLIM5/6/10/11* (PLIM2-II) were preferentially expressed in the stamen; *LIM12* (PLIM2-I) were strongly expressed in stamen and fiber. Quantitative PCR results of *GhLIM* exhibited different expression patterns during the process of fiber development, the expressing of *GhLIM1* was up-regulated in elongation stage, and peaked in 15 DPA; *GhLIM2* was redominant up-regulated at stage of secondary wall biosynthesis, and peaked in 28 DPA; and the expression level of *GhLIM12* was constantly high during 10~23 DPA, and disappeared in 28 DPA, these result indicating different subfunctionalisations of *GhLIM* genes in different fiber development phases.

**Key words:** cotton; LIM; fiber development; gene duplication; gene expression profiling

收稿 2015-09-18 修定 2015-11-17

资助 新疆维吾尔自治区高技术项目(201111116)、国家高技术研究发展计划项目(“863”项目)(2012AA101108-05-04)和新疆维吾尔自治区自然科学基金(2015211B013)。

\* 通讯作者(E-mail: fanling@xaas.ac.cn; Tel: 0991-4527003)。

LIM蛋白是一类半胱氨酸富集蛋白(Cys-rich proteins, CRPs), 其命名源于LIN11、ISL1和MEC3三个从动物中鉴定的含有LIM结构域蛋白的首字母缩写(Freyd等1990; Karlsson等1990; Way和Chalfie 1988)。LIM结构域在真核生物中广泛存在且高度保守, 其保守基序为[C-X<sub>2</sub>-C-X<sub>17</sub>-H-X<sub>2</sub>-C]-X<sub>2</sub>-[C-X<sub>2</sub>-C-X<sub>17</sub>-CX<sub>2</sub>-H] (Arnaud等2007)。植物中首先从向日葵(*Helianthus annuus* L.)花粉中分离出了HaP-LIM1 (Baltz等1992a, b), 随后又在烟草(*Nicotiana tabacum* L.) (Thomas等2006)、百合(*Lilium longiflorum* Thunb.) (Wang等2008)、拟南芥[*Arabidopsis thaliana* (L.) Heynh.] (Papuga等2010)及棉花(*Gossypium hirsutum* L.) (罗明等2003; Li等2013; Han等2013)等物种中陆续发现了LIM蛋白。植物LIM蛋白家族一般含有2个被40~50个氨基酸残基分隔的LIM结构域。Eliasson等(2000)将植物LIM分为在花粉中特异表达的 PLIM1、PLIM2和组成型表达的WLIM1, WLIM2。Arnaud等(2007)进一步将植物LIM蛋白分为四大亚组 $\alpha$ LIM1、 $\beta$ LIM1、 $\gamma$ LIM2和 $\delta$ LIM2; 其中 $\alpha$ LIM1具有PLIM1、WLIM1和FLIM三个小亚组,  $\beta$ LIM1为新发现的LIM蛋白类型, 具体功能不详,  $\gamma$ LIM2和 $\delta$ LIM2分别包括了WLIM2和PLIM2。植物LIM蛋白的亚细胞定位和动物一样, 具有细胞质、细胞核及核质同时定位3种情况。在动物中, 定位于细胞核的LIM蛋白主要参与发育基因的转录调控(Thaler等2002), 而定位于细胞质的LIM蛋白主要作为ABP (actin-binding protein)参与调节肌动蛋白细胞骨架形成(Tran等2005)。烟草NtWLIM1、2, 向日葵WLIM1和拟南芥LIM蛋白家族都表现出在细胞核质同时定位的特性(Brière等2003), 烟草NtWLIM1蛋白能够特异性的结合PAL-box激活苯丙烷代谢途径关键酶, 参与调控木质素生物合成(Kawaoka等2000; Kawaoka和Ebinuma 2001)。NtWLIM2通过激活组蛋白基因的表达参与调控细胞增殖与细胞周期进程(Moes等2013)。而向日葵HaPLIM1在花粉小孢子和成熟花粉的外皮层都具有定位, 主要同actin蛋白互作参与细胞骨架调控, 在多种类型细胞的核中没有发现HaPLIM1, 并且没有发现该蛋白具有特异的DNA或RNA结合活性(Baltz等1996)。

棉花是重要的经济作物, 而我国棉花纤维品

质仍需进一步的提高(张天真和郭旺珍2007)。为此, 研究者对棉纤维发育的分子机理进行了大量研究。Li等(2013)发现, 陆地棉*GhWLIM5*在纤维发育的快速伸长期优势表达, 主要功能是在细胞质中作为F-actin的结合蛋白, 使细胞骨架更加稳定。Han等(2013)发现, 陆地棉*GhWLIM1a*基因既作为actin蛋白的结合体能促进纤维的伸长生长, 也作为转录因子能激活PAL-box基因从而参与苯丙烷类化合物的合成影响次生细胞壁的发育。在基因组层面对棉花LIM蛋白基因家族的研究未见报道, 本研究利用棉花基因组测序数据, 在基因组层面分析陆地棉LIM蛋白基因家族, 从基因家族各成员的结构、系统发育关系、基因的染色体分布、基因复制及基因在棉花不同组织及棉纤维发育各阶段的表达谱来研究棉花LIM基因家族。为进一步鉴定该基因家族在棉纤维发育中的功能, 发掘新型的纤维品质改良基因奠定基础。

## 材料与方法

### 1 材料

陆地棉(*Gossypium hirsutum* L.)基因组及蛋白数据来自(<http://mascotton.njau.edu.cn>) (Zhang等2015)、亚洲棉(*Gossypium arboreum* L.)数据来自(<http://cgp.genomics.org.cn>) (Li等2014), 雷蒙德氏棉(*Gossypium raimondii* Ulbr.) (Paterson等2012)、可可(*Theobroma cacao* L.)、番木瓜(*Carica papaya* L.)、葡萄(*Vitis vinifera* L.)、大豆(*Glycine max* Merr.)、烟草(*Nicotiana tabacum* L.)、拟南芥[*Arabidopsis thaliana* (L.) Heynh.]、毛果杨(*Populus trichocarpa* Torr. & A.Gray)、水稻(*Oryza sativa* L.)、玉米(*Zea mays* L.)、耧斗菜(*Aquilegia coerulea*)、无油樟(*Amborella trichopoda* Baill.)、云杉(*Picea abies* H. Karst.)、卷柏(*Selaginella moellendorffii* Hieron.)、小立碗藓(*Physcomitrella patens* Bruch & Schimp.)数据均来自Phytozome数据库(<http://phytozome.jgi.doe.gov/pz/portal.html>)。棉花RNA-seq数据来自NCBI网站BioProject (PRJNA248163)中的SRA数据包, 包括陆地棉(TM-1): 根、茎、叶、花托、花瓣、雄蕊、雌蕊、副萼(SRX797899-SRX797906), 0 DPA (day post anthesis)胚珠(SRX797909), 5、10、20、25 DPA纤维(SRX797917-797920)。

陆地棉品种‘新陆早36号’于2014年种植于新疆玛纳斯试验田, 开花当天标记后, 采集5、10、15、18、21、23、28 DPA棉纤维样品, 每个样品均取3次重复, 用液氮速冻并保存, 用于RNA提取。

## 2 棉花及其他物种LIM蛋白家族的识别

从Pfam (pfam.xfam.org)下载LIM结构域种子序列, 使用隐马尔可夫模型(Hidden Markov Models)方法进行本地蛋白质数据库的检索, 候选LIM同源蛋白应符合两个条件: (1)具有2个完整的LIM保守结构域和(2) E值 $<1 \times e^{-5}$ 。从相应物种数据库中下载候选LIM蛋白的CDS序列及基因组序列作为候选。将候选基因的CDS序列在NCBI数据库中进行比对, 检索相应的基因及EST序列。蛋白保守基序(motif)的鉴定用MEME (<http://meme.sdsc.edu/meme/meme-intro.html>)软件进行(陈秀玲等2014)。

## 3 蛋白多序列比对与系统树构建

使用Clustal X软件在默认参数下进行LIM蛋白的多序列比对, 用BioEdit手动调整优化序列。通过ProtTest软件分析, 得出构建进化树的最适模型为JTT+G, 随后利用Phylips软件包(<http://evolution.genetics.washington.edu/phylip.html>), 选择JTT+G模型, bootstrap值设定为1 000, 采用邻接法(neighbor-join, NJ)构建植物LIM蛋白家族的系统发育树。

## 4 基因结构与基因重复分析

在GSDS网站(<http://gsds.cbi.pku.edu.cn/>), 将棉花LIM基因的编码区序列(CDS)与全基因组序列比对, 获得各基因内含子、外显子的个数和排布情况(易吉明等2015)。棉花LIM蛋白基因倍增模式分析参照Maher等(2006)的方法。使用MCScanX软件对棉花LIM蛋白基因家族成员所在基因组区段进行基因复制共线性分析, 采用Circos软件绘制LIM蛋白基因家族在陆地棉基因组中的分布与复制情况。

## 5 LIM基因在棉花不同组织及棉纤维不同发育时期的表达谱分析

采用Tophat和Cufflinks软件包对棉纤维转录组数据进行分析, 提取其中GhLIM基因的基因表达量(fragments per kilobase of exon per million fragments mapped, FPKM值) (Trapnell等2012), 进行对数转换和标准化, 并去除表达量较低的LIM基因

(FPKM $<1$ ), 利用层次聚类算法分析LIM基因在陆地棉和海岛棉棉纤维不同发育时期的表达数据, 通过软件HemI展现层次聚类结果(Deng等2014)。

采用热硼酸/蛋白酶K法(Wan和Wilkins 2014)提取各纤维样品RNA, 各样品RNA浓度均一化后, 利用M-MLV反转录酶(Promega, USA)合成cDNA。选取在纤维中优势表达的GhLIM基因进行实时荧光定量PCR分析(表1)。以棉花泛素基因Ubiquitin (EU604080.1)作为内参, 以 $2^{-\Delta\Delta C_t}$ 法计算基因的相对表达量(Livak和Schmittgen 2001)。

表1 LIM基因定量引物

Table 1 The primers for LIM gene quantitative PCR analysis

| 基因      | 引物序列(5'→3')                                       |
|---------|---|
| GhLIM1  | F: AAGCCTGTTTCAGATGCC<br>R: CACCGAATAAAATTCACCACT |
| GhLIM2  | F: TGGAGGGTGTACCATTAG<br>R: CGAGCTGGCTGTAGTTTC    |
| GhLIM12 | F: GAAGGGATGCTACAGTCA<br>R: CACCAGTCTTCTCAGATT    |
| GhLIM13 | F: GCTCCTTTGAGGGTGTCT<br>R: CCTGTGATATGAAGTGCC    |
| GhLIM14 | F: TCATTTGAGCAACTATTC<br>R: CTCCACCGTTACCTTCT     |

F表示正向引物, R表示反向引物。

## 实验结果

### 1 棉花LIM蛋白基因家族鉴定

通过对棉花基因组数据库的HMMsearch, 从雷蒙德氏棉、亚洲棉和陆地棉A、D两个亚基因组各鉴定了LIM蛋白14个。经过BLAST和结构域分析, 确定这些基因的编码蛋白均具有2个串联的LIM保守结构域(图1)。对棉花LIM基因进行系统进化分析(图1-A), 发现进化树由14个进化亚支组成, 每个亚支均由4个分别来自雷蒙德氏棉、亚洲棉和陆地棉A、D亚基因组的LIM直系同源体(orthologs)组成, 各直系同源蛋白间的相似度在95%以上, 表明棉花LIM家族有14个成员(图1-A)。陆地棉LIM基因命名为GhLIM1~14, 并在最后添加字母“a”或“d”来区分各基因的AD亚基因组归属, 对于亚洲棉和雷蒙德氏分别命名为GaLIM1~14和GrLIM1~14。

对棉花LIM蛋白家族的保守基序进行分析(图



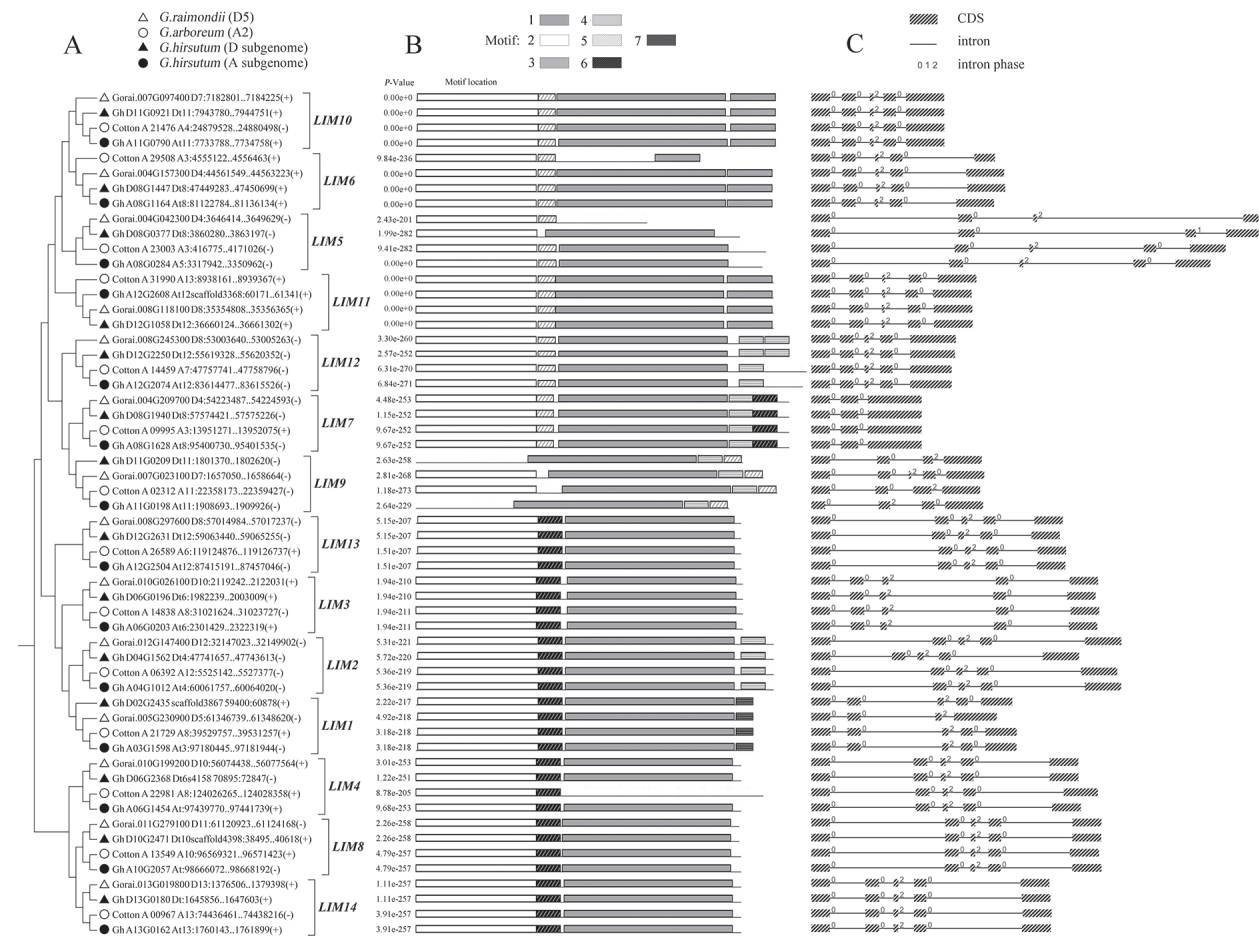


图1 棉花LIM蛋白保守系统进化(A)、保守结构域(B)及内含子外显子结构(C)

Fig.1 The phylogenetic (A), conserved domain (B) and gene intro-extro structure (C) of LIM gene family in cotton (*G. raimondii*, *G. arboreum*, and *G. hirsutum*)

1-B, 表2), 发现该蛋白家族除具有2个LIM保守结构域外, 大多数在C-端还有可变区。对棉花LIM基因的内含子和外显子的数量和分布情况进行分析(图1-C), 大多数棉花LIM基因含有4个内含子5个外显子。第1内含子和第1、2、4外显子长度高度保

守, 而第4内含子和第5外显子长度具有较高的变异。大多数直系同源基因间的内含子和外显子的数目和长度是保守的, 但个别基因的内含子和外显子数目和长度发生了变异, 如*GhLIM2d*、*GhLIM5a*。

表2 棉花LIM蛋白的保守基序

Table 2 The motifs of LIM proteins of cotton

| 基序编号 | 基序长度 | 一致序列   |
|------|------|--|
| 1    | 74   | Kx <sub>4</sub> FxGTx <sub>2</sub> KCx <sub>2</sub> CxKTxYPxKxVx <sub>2</sub> Gx <sub>3</sub> HKxCF(K/R)CxHGGCx <sub>4</sub> SxYx <sub>2</sub> Ax <sub>4</sub> LYCKHHx <sub>2</sub> QLFx <sub>2</sub> KG |
| 2    | 70   | (M/A)xFxGTx <sub>2</sub> KCxACxKTx <sub>2</sub> V(D/E)x <sub>4</sub> (D/E)(G/N)x <sub>3</sub> HKxCF(K/R)CxHCKGxLx <sub>4</sub> YxS(M/F)<br>(D/E)GVLYC(K/R)PHF(D/E)QLFKx <sub>2</sub> G                   |
| 3    | 26   | (S/A)Sx <sub>4</sub> Ex <sub>2</sub> Ax <sub>3</sub> Ax <sub>6</sub> Ex <sub>3</sub>   |
| 4    | 14   | (P/x)X <sub>3</sub> SX <sub>6</sub> E  |
| 5    | 10   | (F/x)(S/x)KNF(Q/x)(T/x)x(K/x)x   |
| 6    | 11   | (S/N)x <sub>2</sub> KxF(x)X <sub>3</sub> (P/A)K  |
| 7    | 10   | LTSMEIAAES   |

## 2 植物LIM蛋白的系统进化分析

将小立碗藓LIM基因作为外群体(out group)对LIM蛋白系统进化分析, 结果表明(图2): 所有陆生植物LIM蛋白家族起源于共同的祖先, 从小立碗藓之后, 植物LIM蛋白可划分为LIM1和LIM2两大亚组, 而从无油樟开始, 所有被子植物的LIM1和LIM2又各自分开, 形成4个亚组。根据文献(Ar-

naud等2007), 将4个亚组命名为 $\alpha$ LIM1、 $\beta$ LIM1和 $\gamma$ LIM2、 $\delta$ LIM2。

(1)  $\alpha$ LIM1亚组包括WLIM1、PLIM1及FLIM1三个小亚组(Arnaud等2007)。棉花在WLIM1分支中有1个成员GhLIM3, 即是已被证明在棉纤维发育中有重要功能的GhWLIM1a (Han等2013)。FLIM1基因家族具有在纤维组织中优势表达的特征

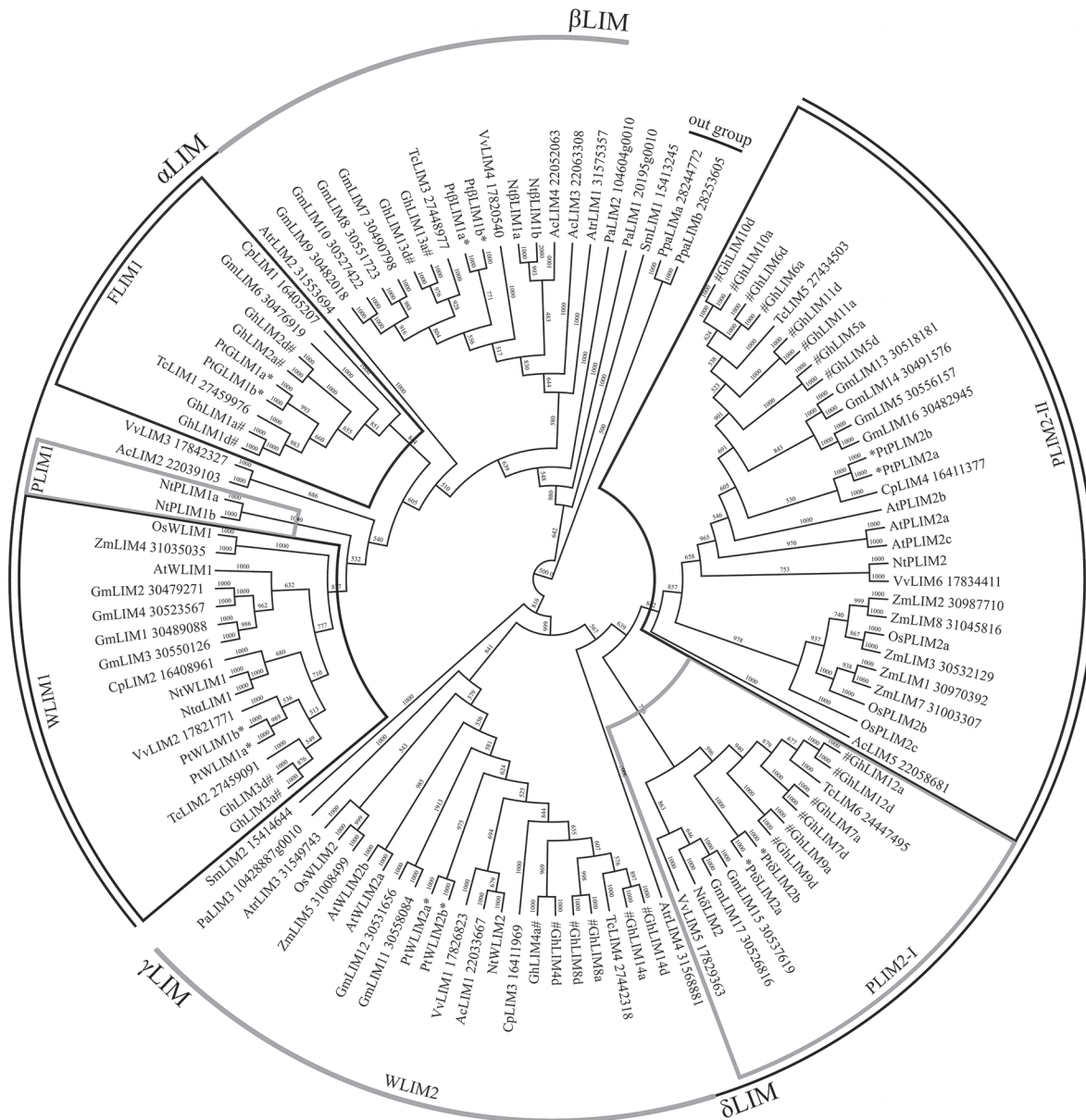


图2 陆地棉与其他物种LIM蛋白系统发育分析

Fig.2 The phylogenetic tree of LIM homologs from *G. hirsutum* and other plant

Ac: 矮斗菜, At: 拟南芥, Atr: 无油樟, Cp: 番木瓜, #Gh: 陆地棉, Gm: 大豆, Nt: 烟草, Os: 水稻, Pa: 云杉, Ppa: 小立碗藓, \*Pt: 毛果杨, Sm: 卷柏, Tc: 可可, Vv: 葡萄, Zm: 玉米; 拟南芥、毛果杨、水稻的LIM蛋白来自文献(Arnaud等2007); 其他物种LIM基因名称后的数字为基因组数据中的PACID。

(Arnaud等2007), 该分支中杨树的PtXLIM1a和PtXLIM1b在木质部发育过程中优势表达, 陆地棉在该分支中有2个成员GhLIM1和GhLIM2。(2)  $\beta$ LIM1是一个较小的分支, 单子叶植物及拟南芥没有这一类型的LIM蛋白。棉花具有1个成员GhLIM13。(3)  $\gamma$ LIM2亚组中主要为WLIM2。WLIM2亚组蛋白的保守性比较高。陆地棉在该亚组有3个成员: GhLIM4、8、14。(4)  $\delta$ LIM2: 该亚组形成两个大

分支, 分支I中有4个来自棉花: GhLIM5、6、10、11。分支II中没有单子叶植物, 棉花在该分支有3个成员: GhLIM7、9、12。

### 3 棉花LIM基因的共线性复制分析

陆地棉28个LIM基因在染色体上的分布情况如图3所示, LIM在A和D两个亚组的染色体分布具有对应关系(除GhLIM1a、d分别位于A03和D02染色体之外), A/D亚组的4、10、13号染色体上各具

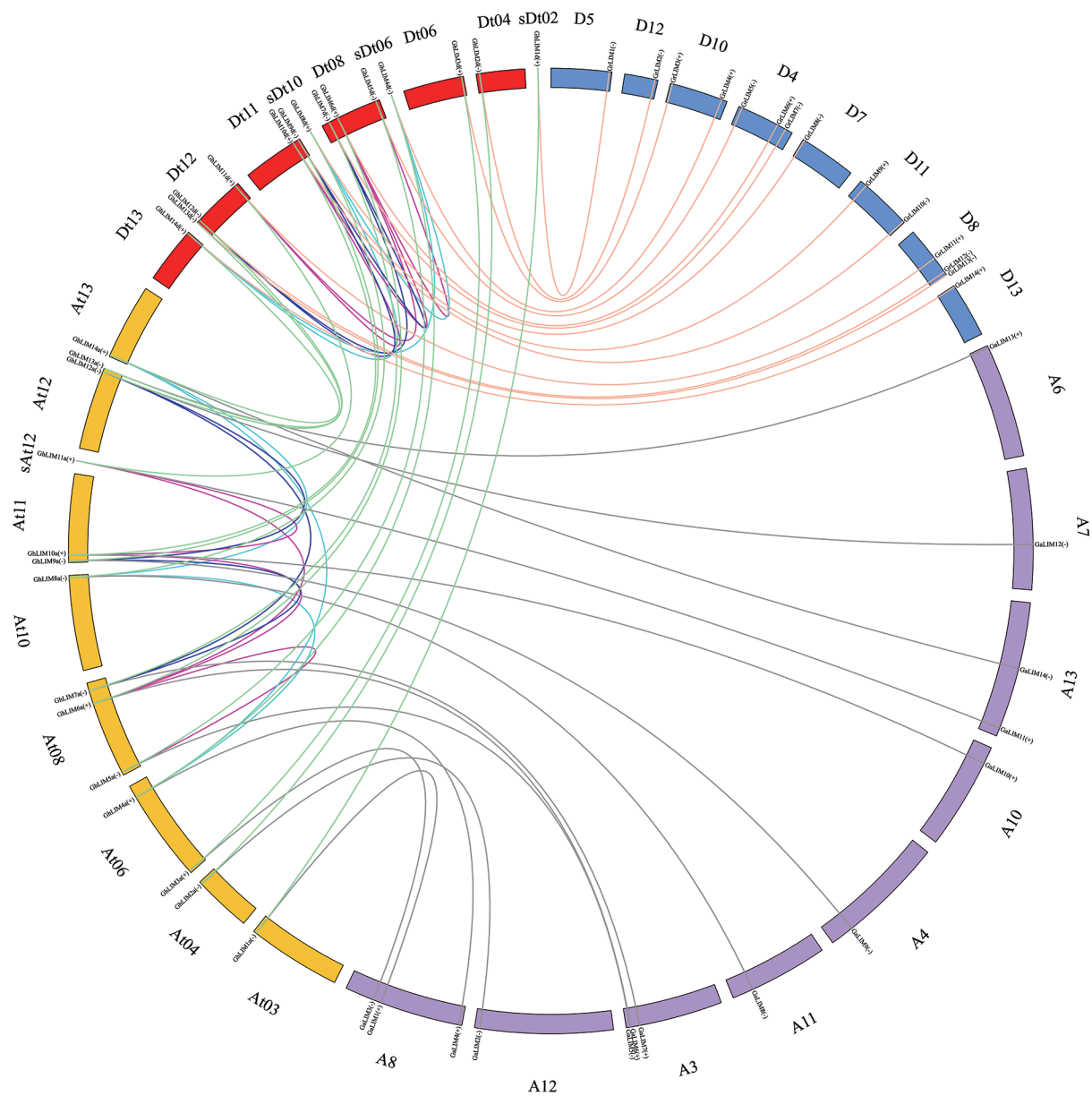


图3 陆地棉LIM基因家族的染色体分布与基因共线性分析

Fig.3 The chromosome location and gene duplication of LIM gene family in *G. hirsutum*

D: 雷蒙德氏棉染色体; A: 亚洲棉染色体; Dt: 陆地棉D亚组染色体; At: 陆地棉A亚组染色体; s: scaffold.

有1个对应的基因(*GhLIM2*、8、14), 6和11号染色体上各有2个基因(*GhLIM3*、4, *GhLIM9*、10), 8和12号染色体上各有3个基因(*GhLIM5*、6、7, *GhLIM11*、12、13)。在亚组内, *LIM*基因的分布是不均匀的, 具有成簇分布的现象。

通过MCScanX软件分析, 发现陆地棉*LIM*基因家族成员所在的基因组区域间存在显著的共线性关系(图3), 有10个棉花*LIM*家族成员位于3个染色体

倍增块上(表3), 并且不同亚组的*LIM*基因复制模式是一致的。结合进化分析(图2)可以看出,  $\gamma$ *LIM2*和 $\delta$ *LIM2*中的*GhLIM*基因发生了多次复制, 并且每一组复制基因都在进化树上和一个来自可可的*LIM*蛋白聚为一个分支。推测棉花*LIM*基因家族复制的时间发生在其祖先同可可分开之后, A2、D5基因组分化之前; 陆地棉中, 除多倍体化造成*LIM*基因的数目加倍外, 该基因没有其他的基因复制与丢失事件发生。

表3 *GhLIM*基因复制模式分析Table 3 Gene duplication of the LIM family in *G. hirsutum*

| 复制基因对 | 基因名称         | 染色体位置  | 复制模式    |
|-------|--------------|--|---------|
| 1     | <i>LIM5</i>  | A08:3317942-3321244; D08:3860280-3863197           | 染色体片段复制 |
|       | <i>LIM6</i>  | A08:81122784-81124119; D08:47449283-47450699       |         |
|       | <i>LIM10</i> | A11:7733788-7734758; D11:7943780-7944751           |         |
|       | <i>LIM11</i> | A12scaffold3368:60171-61341; D12:59063440-59065255 |         |
| 2     | <i>LIM7</i>  | A08:95400730-95401535; D08:57574421-57575526       | 染色体片段复制 |
|       | <i>LIM9</i>  | A11:1908693-1909926; D11:1801374-1802618           |         |
|       | <i>LIM12</i> | A12:8361447-8365526; D12:55619328-55620352         |         |
| 3     | <i>LIM4</i>  | A06:97439770-97441739; D06scaffold4158:70895-72847 | 染色体片段重复 |
|       | <i>LIM8</i>  | A10:98666072-98668192; D10scaffold4398:38499-40617 |         |
|       | <i>LIM14</i> | A13:1760147-1761896; D13:1645856-1647603           |         |

#### 4 陆地棉*LIM*基因家族的表达模式分析

对陆地棉(TM-1)不同发育阶段的纤维, 及其他不同组织的RNA-Seq数据进行分析, 获得了*GhLIM*基因表达量的FPKM值, 利用Hemi软件绘制出*GhLIM*基因表达模式图。结果如图4所示, 28个*GhLIM*基因均有表达。根据基因在不同组织中的表达模式, 可将其分为3种类型, (1)组成型表达: *GhLIM1*、2、3 ( $\alpha$ *LIM1*), *GhLIM13* ( $\beta$ *LIM1*)和*GhLIM4*、8、14 ( $\gamma$ *LIM2*), (2)在雄蕊中优势表达: *LIM5*、6、7、9、11 ( $\delta$ *LIM2*), (3)在雄蕊和纤维中优势表达: *LIM12* ( $\delta$ *LIM2*)。其中*GhLIM1*、2、12、13、14在纤维中优势表达。

棉纤维的发育主要可以分为5个不同的发育阶段: 起始(0~5 DPA), 快速伸长(5~18 DPA), 初次生壁发育转换(18~21 DPA), 次生壁发育(21~30 DPA)和脱水成熟。本研究进一步通过实时荧光定量PCR对*GhLIM1*、2、12、13、14在棉纤维不同发育阶段(5、10、15、18、21、23、28 DPA)的表达量进行检测, 结果如图5所示: *GhLIM1*、13、14均在纤维的快速伸长期(5~15 DPA)上调表达, 在

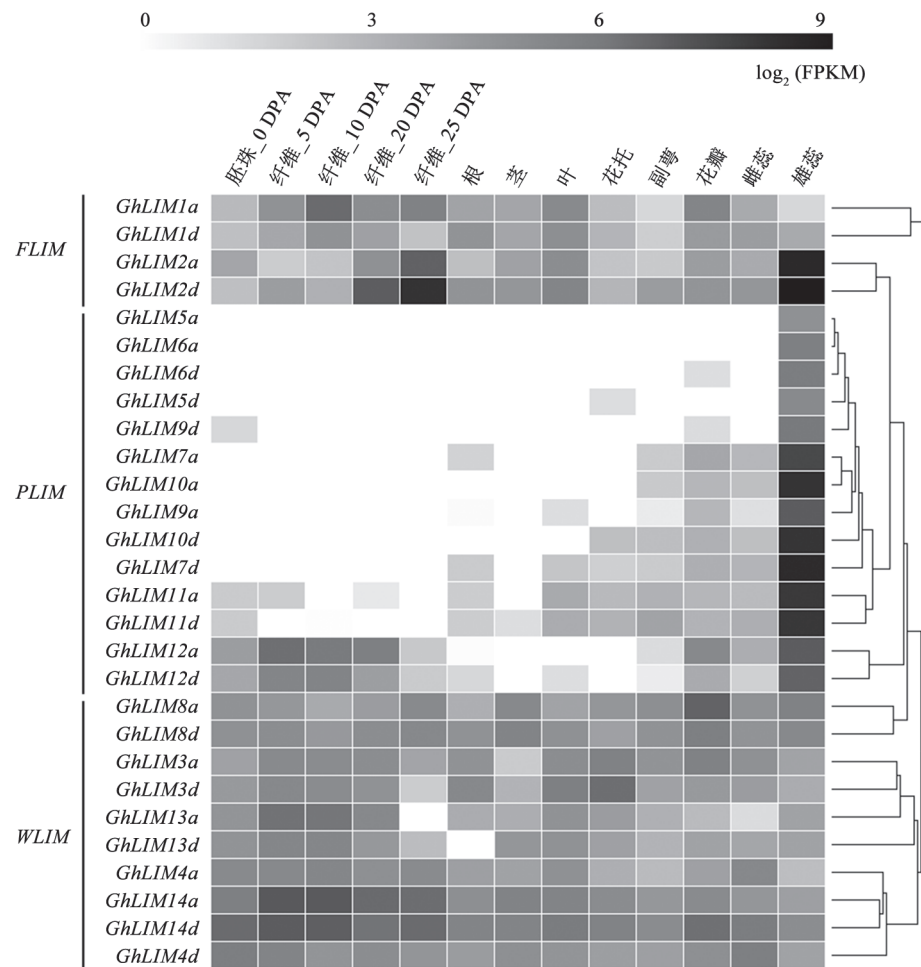
10~15 DPA表达量最高; 不同的是*GhLIM1*在之后的18~28 DPA持续较高表达, 而*GhLIM13*、14的表达水平则降低。*GhLIM2*从纤维18 DPA (转换期)开始显著上调表达, 至28 DPA表达量最高。而属于PLIM2-I亚组的*GhLIM12*则在10~23 DPA持续高表达, 在28 DPA (次生壁发育)表达水平很低。上述结果说明不同*GhLIM*家族基因在纤维发育的不同阶段可能发挥着不同的功能。

## 讨 论

### 1 棉花*LIM*基因家族的复制与进化

本研究的结果表明, 棉花*LIM*基因家族的大多数成员具有4内含子5外显子, 与大多数植物的*LIM*基因结构相似(Baltz等1992a, b; Thomas等2006; Arnaud等2007; Paguga等2010)。而Arnaud等(2007)发现杨树、拟南芥和水稻*LIM*基因家族的第1、2、4外显子和第1、4内含子高度保守, 本研究中, 棉花*LIM*的第1、2、4外显子和第1内含子同样高度保守, 但第5个外显子和第4个内含子具有较多的长度变异(200~2 000 bp), 并且对应其蛋白的C-端具



图4 *GhLIM* 基因家族在陆地棉中的表达模式Fig.4 The expression patterns of *GhLIM* genes in *G. hirsutum*

有可变的结构域, 推测棉花LIM基因家族在进化中功能产生了分化。

棉属的共同祖先在与可可分化之后(33.7 MYA), 经历了特有的基因组复制事件(Wang等2012; Paterson等2012; Li等2014, 2015; Zhang等2015), 本研究*GhLIM1a*、*d*分别位于A03和D02染色体上, 通过比较分析, 发现其恰好位于陆地棉A亚基因组上最大的染色体异位区上(A02–A03)(Zhang等2015), 可见该基因家族在陆地棉A、D两个亚基因组上的染色体分布是一致的。

## 2 棉花LIM基因的表达模式

植物LIM蛋白参与细胞骨架的调节及调控次生细胞壁的发育(Baltz等1992a, b; Papuga等2010), 两种WLIM蛋白在棉纤维发育中的功能已得到了证实(Han等2013; Li等2013)。本研究的结果表明,

棉花LIM基因家族共有14个成员, 其中在发育的棉纤维中优势表达的有五种(*GhLIM1*、2、12、13、14), 且它们在纤维发育过程中的表达模式有很大的不同, 表明它们在纤维发育中可能具有不同的功能。

Arnaud等(2012)对杨树PtLIM家族进行了系统进化和组织表达分析, 发现 $\alpha$ LIM1家族中的PtLIM基因可以划分为PtWLIM1和PtFLIM1两类, PtFLIM1主要在木质部分化和扩张时期表达, 推测其功能与次生壁中木质素的合成有关(Arnaud等2007, 2012), 并发现棉纤维EST文库中也有FLIM1类基因的大量表达。本研究发现, 棉花中的FLIM1有两个成员, *GhLIM1*在在纤维的快速伸长期(5~15 DPA)上调表达, 在10~15 DPA表达量最高; 并且在之后的18~28 DPA持续具有较高表达, *GhLIM2*从



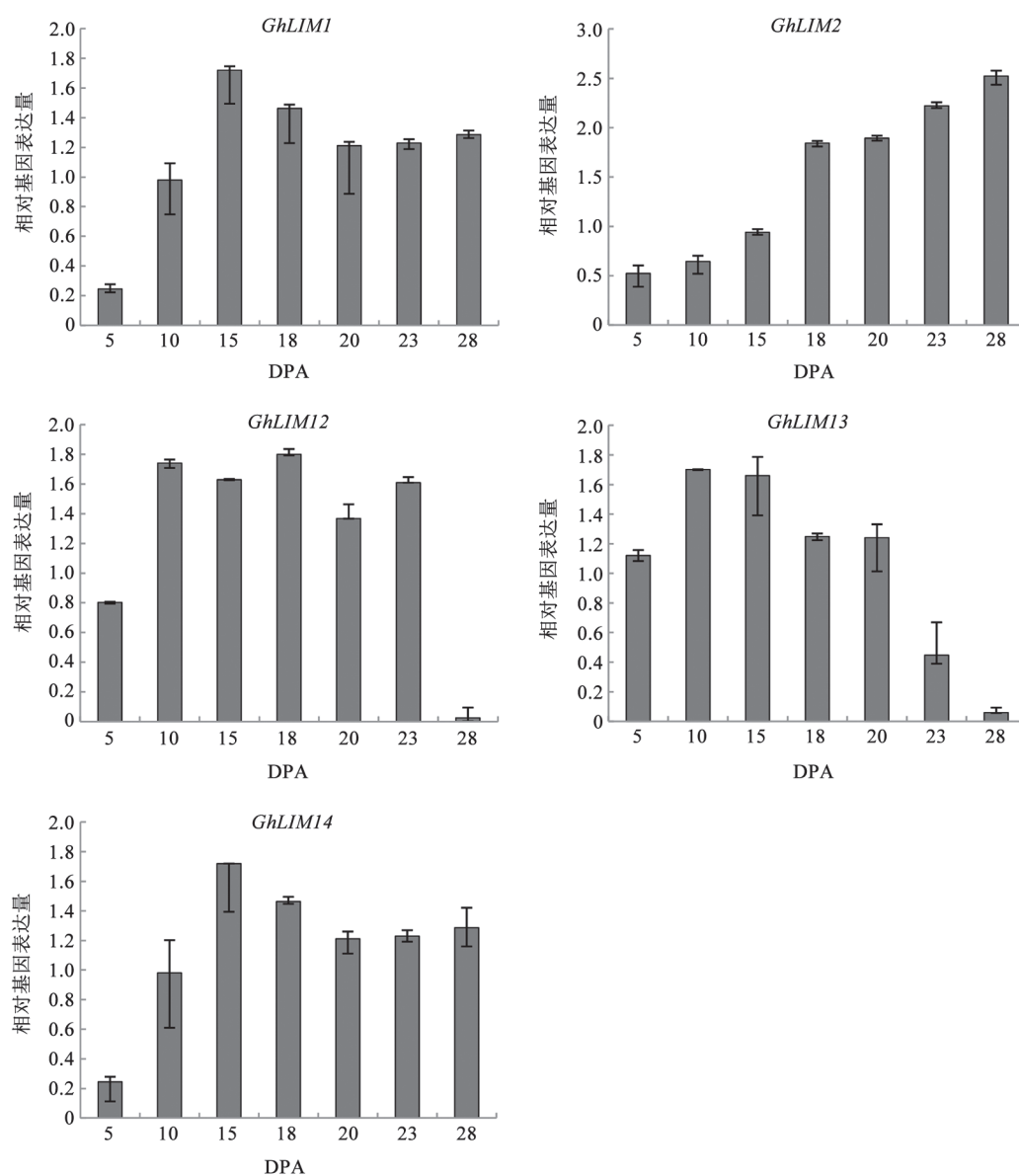


图5 GhLIM基因家族在棉纤维发育过程中的表达模式

Fig.5 The expression patterns of GhLIM genes during the process of fiber development

纤维18 DPA (转换期)开始显著上调表达,至28 DPA表达量最高。暗示这两个基因可能参与棉纤维次生壁的发育。

本研究中植物 $\delta$ LIM2亚组可以分为PLIM2-I和PLIM2-II,拟南芥*AtPLIM2*在细胞伸长过程中,具有对肌动蛋白的结合、稳定和成束调节能力(Paguga等2010)。杨树的纤维状花絮同棉花纤维具有相似的起源(Arnaud等2012)。在杨树中,*Pt-PLIM*也可划分为两种:属于PLIM2-II的*PtPLIM2a*、*b*主要在花粉中优势表达,而属于PLIM-I的

*PtPLIM2c*、*d*除在花粉中表达外,还在杨絮和荚膜中表达(Arnaud等2012)。本研究中棉花*GhLIM12*同杨树*PtPLIMc*、*d*同属PLIM2-I亚组,*GhLIM12*表达特征与其余棉花*GhLIM*不同,除在雄蕊中优势表达外,在纤维发育的5~23 DPA也有较高的表达,暗示该基因在棉纤维发育过程中有重要的作用。

#### 参考文献

- 陈秀玲,王傲雪,张珍珠,李景富(2014). 番茄*NAC*转录因子家族的鉴定及生物信息学分析. 植物生理学报, 50 (4): 461~470  
 罗明,肖月华,侯磊,罗小英,李德谋,裴炎(2003). 棉花LIM结构域

- 基因(*GhLIM1*)的克隆和表达分析. 遗传学报, 30 (2): 175~182
- 易吉明, 黄婷, 黄勇, 陈东红(2015). 小立碗蕨MADS-box基因家族的系统进化分析. 植物生理学报, 51 (2): 197~206
- 张天真, 郭旺珍(2007). 棉花分子育种的现状、问题与展望. 中国农业科技导报, 9 (2): 19~25
- Arnaud D, Déjardin A, Leplé JC, Lesage-Descauses MC, Boizot N, Villar M, Bénédicti H, Pilate G (2012). Expression analysis of *LIM* gene family in poplar, toward an updated phylogenetic classification. *BMC Res Notes*, 5: 102
- Arnaud D, Déjardin A, Leplé JC, Lesage-Descauses MC, Pilate G (2007). Genome-wide analysis of *LIM* gene family in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*. *DNA Res*, 14 (3): 103~116
- Baltz R, Domon C, Pillay DT, Steinmetz A (1992a). Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger protein. *Plant J*, 2 (5): 713~721
- Baltz R, Evrard JL, Bourdon V, Steinmetz A (1996). The pollen-specific LIM protein PLIM-1 from sunflower binds nucleic acids *in vitro*. *Sex Plant Rep*, 9 (5): 264~268
- Baltz R, Evrard JL, Domon C, Steinmetz A (1992b). A LIM motif is present in a pollen-specific protein. *Plant Cell*, 4 (12): 1465~1466
- Brière C, Bordel AC, Barthou H, Jauneau A, Steinmetz A, Alibert G, Petitprez M (2003). Is the LIM-domain protein HaWLIM1 associated with cortical microtubules in sunflower protoplasts? *Plant Cell Physiol*, 44 (10): 1055~1063
- Deng W, Wang Y, Liu Z, Cheng H, Xue Y (2014). HemI: a toolkit for illustrating heatmaps. *PLoS One*, 9 (11): e111988
- Eliasson A, Gass N, Mundel C, Baltz R, Kräuter R, Evrard JL, Steinmetz A (2000). Molecular and expression analysis of a LIM protein gene family from flowering plants. *Mol Gen Genet*, 264 (3): 257~267
- Freyd G, Kim SK, Horvitz HR (1990). Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*. *Nature*, 344 (6269): 876~879
- Han LB, Li YB, Wang HY, Wu XM, Li CL, Luo M, Wu SJ, Kong ZS, Pei Y, Jiao GL et al (2013). The dual functions of *WLIM1a* in cell elongation and secondary wall formation in developing cotton fibers. *Plant Cell*, 25 (11): 4421~4438
- Karlsson O, Thor S, Norberg T, Ohlsson H, Edlund T (1990). Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. *Nature*, 344 (6269): 879~882
- Kawaoka A, Ebinuma H (2001). Transcriptional control of lignin biosynthesis by tobacco LIM protein. *Phytochemistry*, 57 (7): 1149~1157
- Kawaoka A, Kaothien P, Yoshida K, Endo S, Yamada K, Ebinuma H (2000). Functional analysis of tobacco LIM protein Ntlim1 involved in lignin biosynthesis. *Plant J*, 22 (4): 289~301
- Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J et al (2015). Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotechnol*, 33 (5): 524~530
- Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C et al (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet*, 46 (6): 567~572
- Li Y, Jiang J, Li L, Wang XL, Wang NN, Li DD, Li XB (2013). A cotton LIM domain-containing protein (GhWLM5) is involved in bundling actin filaments. *Plant Physiol Biochem*, 66: 34~40
- 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 (4): 402~408
- Maher C, Stein L, Ware D (2006). Evolution of *Arabidopsis* microRNA families through duplication events. *Genome Res*, 16 (4): 510~519
- Moes D, Gatti S, Hoffmann C, Dieterle M, Moreau F, Neumann K, Schumacher M, Diederich M, Grill E, Shen WH et al (2013). A LIM domain protein from tobacco involved in actin-bundling and histone gene transcription. *Mol Plant*, 6 (2): 483~502
- Papuga J, Hoffmann C, Dieterle M, Moes D, Moreau F, Tholl S, Steinmetz A, Thomas C (2010). *Arabidopsis* LIM proteins: a family of actin bundlers with distinct expression patterns and modes of regulation. *Plant Cell*, 22 (9): 3034~3052
- Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J et al (2012). Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*, 492 (7429): 423~427
- Thaler JP, Lee SK, Jurata LW, Gill GN, Pfaff SL (2002). LIM factor *Lhx3* contributes to the specification of motor neuron and interneuron identity through cell-type-specific protein-protein interactions. *Cell*, 110 (2): 237~249
- Thomas C, Hoffmann C, Dieterle M, Van Troys M, Ampe C, Steinmetz A (2006). Tobacco WLM1 is a novel F-actin binding protein involved in actin cytoskeleton remodeling. *Plant Cell*, 18 (9): 2194~2206
- Tran TC, Singleton C, Fraley TS, Greenwood JA (2005). Cysteine-rich protein 1 (CRP1) regulates actin filament bundling. *BMC Cell Biol*, 6: 45
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 7 (3): 562~578
- Wan CY, Wilkins TA (1994). A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). *Anal Biochem*, 223 (1): 7~12
- Wang HJ, Wan AR, Jauh GY (2008). An actin-binding protein, LLIM1, mediates calcium and hydrogen regulation of actin dynamics in pollen tubes. *Plant Physiol*, 147 (4): 1619~1636
- Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S et al (2012). The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet*, 44 (10): 1098~1103
- Way JC, Chalfie M (1988). *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell*, 54 (1): 5~16
- Zhang TZ, Hu Y, Jiang WK, Fang L, Guan XY, Chen JD, Zhang JB, Sasaki C, Scheffler BE, Stelly DM et al (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fibre improvement. *Nat Biotechnol*, 33 (5): 531~537