

海岛棉 *GAUT* 基因家族的鉴定及其在棉纤维发育中的表达分析

张松雨, 刘正文, 张艳, 杨君, 马峙英, 王省芬

(教育部华北作物种质资源重点实验室/河北省作物种质资源重点实验室/河北农业大学, 保定 071000)

摘要:本研究鉴定了海岛棉全基因组中的 *GAUT* 基因家族, 对其进行了生物信息学及纤维发育过程中的表达分析。结果表明, 海岛棉 *GAUT* 家族包含 37 个成员, 分为 3 个亚组, 分布在海岛棉 18 条染色体上。*GAUT* 氨基酸序列保守基序有 4 个, 保守性较强, 所有 *GAUT* 蛋白均定位在高尔基体膜。根据 *GAUTs* 在海岛棉纤维发育不同时期的表达变化, 将其分为起始期高表达、纤维伸长期高表达、次生壁增厚期高表达、全时期低表达等 4 类模式。其中, *GAUT05*、*GAUT06*、*GAUT07*、*GAUT23*、*GAUT24*、*GAUT26* 等在纤维发育某个时期具有优势表达的基因可能在纤维发育过程中发挥着重要作用。上述结果为研究海岛棉 *GAUT* 基因家族在棉纤维发育中的功能提供了参考依据。

关键词: 海岛棉; 半乳糖醛酸转移酶; 生物信息学; 纤维发育; 基因表达

Genome-Wide Identification of *GAUT* Gene Family in *Gossypium barbadense* L. and Expression Analysis in Fiber Developmental Stages

ZHANG Song-yu, LIU Zheng-wen, ZHANG Yan, YANG Jun, MA Zhi-ying, WANG Xing-fen

(North China Key Laboratory for Crop Germplasm Resources of Education Ministry/Key Laboratory for Crop Germplasm Resources of Hebei/Hebei Agricultural University, Baoding 071000)

Abstract: In this study the alpha-galacturonosyltransferase (*GAUT*) family genes in the genome of *Gossypium barbadense* L. were identified and the expression analysis during the fiber development stages were carried out. The results showed that a total of 37 *GAUT* genes were identified from *G. barbadense* L. and classified into three subfamilies. They were distributed over 18 chromosomes of *G. barbadense* L. and had four conserved motifs. By predicting subcellular localization all *GAUT* proteins were located in the Golgi membrane. According to the transcriptional profiles of the *GAUTs* expression at the different stages of fiber development, they were divided into four expression patterns including high expression pattern of initial stage, high expression pattern of fiber elongation stage, high expression pattern of secondary wall thickening period and low expression pattern in the whole period. The differentially expressed genes *GAUT05*, *GAUT06*, *GAUT07*, *GAUT23*, *GAUT24*, *GAUT26* might play important roles in fiber development. These results provided a useful reference for understanding the function of the *GAUT* gene family in the development of fiber in *G. barbadense* L.

Key words: *Gossypium barbadense* L.; galacturonosyltransferase; bioinformatics; fiber development; gene expression

棉花是世界上第一大天然纤维作物, 具有重要的经济价值。中国主要种植陆地棉 (*Gossypium hirsutum* L.) 和海岛棉 (*G. barbadense* L.), 其中陆地棉栽培面积和产量在棉花总产中占主导地位, 但纤维品质不如海岛棉。海岛棉的纤维长度、比强度、细度

等表现优良, 是纺高支纱的必要原材料^[1-2]。因此, 海岛棉对于陆地棉纤维品质改良具有重要意义。

棉花纤维品质主要由纤维发育决定。棉纤维细胞作为一种特化的表皮细胞, 其发育模式与植物细胞壁有很多相似之处, 植物细胞壁主要包括纤维素、

收稿日期: 2017-12-19 修回日期: 2018-01-15 网络出版日期: 2018-05-16

URL: <http://kns.cnki.net/kcms/detail/11.4996.S.20180515.1543.001.html>

基金项目: 国家转基因生物新品种培育科技重大专项 (2016ZX08005-003); 河北农业大学作物学科梯队建设基金 (TD2016C201)

第一作者研究方向为作物遗传育种。E-mail: zhangsongyu407@163.com

通信作者: 王省芬, 研究方向为棉花遗传育种。E-mail: cotton@hebau.edu.cn

半纤维素、木质素、果胶和少量蛋白质等^[3]。棉纤维初生壁及次生壁成分均源于多糖^[4],初生壁主要由木葡聚糖、木聚糖和果胶等非纤维素物质组成,影响纤维长度;次生壁主要包括大量的纤维素多糖,直接决定纤维强度^[5]。果胶中有 60% 以上是多聚同型半乳糖醛酸,同聚半乳糖醛酸是半乳糖醛酸的同聚物。半乳糖醛酸参与果胶多聚物的构成,所以在植物中含有大量的半乳糖基转移酶(*GAUT* α -galacturonosyltransferase EC2. 4. 1. 43)^[4]。在拟南芥 *GAUT* 蛋白家族中,共鉴定出 15 名成员,分为 3 个分支:*GAUT*-A(*GAUT* 1-7)、*GAUT*-B(*GAUT* 8-11)和 *GAUT*-C(*GAUT* 12-15)^[6]。目前,*GAUT1* 是第一个在拟南芥中克隆并进行了生化和功能验证的同型半乳糖基转移酶基因,其编码产物为高尔基体 II 型跨膜蛋白,催化寡聚半乳糖醛酸的伸长^[7]。在拟南芥中,*GAUT* 6、*GAUT* 8、*GAUT* 9、*GAUT* 10、*GAUT* 11、*GAUT* 12、*GAUT* 13 和 *GAUT* 14 基因的突变影响果胶和/或木聚糖的生物合成^[8],Yin 等^[9]也证明该家族的多个半乳糖基转移酶成员都参与果胶和/或木聚糖的生物合成,*AtGAUT12* (*irx8*) 参与细胞壁多糖的合成,该基因的突变体导致多聚半乳糖醛酸和葡萄糖醛酸含量显著降低^[10]。Peña 等^[11]证明 *GTS*、*irx8* 和 *irx9* 与维管束的形态及细胞壁的厚度、细胞壁中的纤维素和木糖的含量相关。在番茄中,*gaut4* 基因在发育的器官中表达水平较高,沉默该基因后,植株的营养生长量与栅栏组织扩大量均明显增多^[12]。本课题组前期从海岛棉中克隆得到 *GbGAUT1* 和 *GbGAUT2* 基因,*GbGAUT1* 在海岛棉纤维发育过程中出现两个表达高峰,分别出现在开花后 10d 和 35d,也就是纤维细胞快速伸长期和次生细胞壁增厚期^[13]; *GbGAUT2* 在不同棉花品种中呈现不同的表达模式,在陆地棉中棉所 8 号和海岛棉 Pima90-53 两个品种中,*GbGAUT2* 均在 10DPA 出现表达高峰。随着棉纤维的发育,在中棉所 8 号中 *GbGAUT2* 的表达量呈下降趋势,但是在 Pima90-53 中表达量下降较少,并且在 25DPA 又出现一个表达高峰,将反义 *GbGAUT2* 转化拟南芥,导致叶片表皮毛变长^[4]。上述结果预示着 *GAUT* 基因家族可能在棉纤维发育中具有重要作用。

目前,已经公布的海岛棉基因组测序结果为挖掘纤维发育重要基因提供了重要基因组资源^[14]。本研究基于海岛棉基因组测序数据库,利用 HMMER^[15] 搜索 *GAUT* 保守序列,并进行系统分析,从全基因组水平鉴定 *GAUT* 基因家族成员,分析其基因结构、系统

进化、理化性质、蛋白序列保守性及保守位点等^[16]。结合课题组获得的海岛棉纤维发育转录组数据,分析 *GAUT* 基因家族成员在棉纤维发育中的表达规律,为今后研究 *GAUT* 基因家族成员的功能提供理论依据。

1 材料与方法

1.1 *GAUT* 基因家族序列搜索

从 TAIR (<http://www.arabidopsis.org/>) 网站搜索拟南芥 *GAUT* 基因,得到 15 个基因,再分别在 NCBI (<https://www.ncbi.nlm.nih.gov/>) 中下载其 DNA 序列、氨基酸序列。利用已知海岛棉 *GAUT* 基因的 DNA 序列,在 CottonFGD (<https://cottonfgd.org/>) 搜索并下载其 DNA、氨基酸序列^[14],在 Pfam (<http://pfam.xfam.org/search>) 网站分析其保守域,保守域 ID 为 PF01501 (*Glyco_transf_8*), *Glyco_transf_8* 家族的酶均具有糖基转移酶的活性^[9],借助 NCBI 网站鉴定是否包含完整的 *Glyco_transf_8* 保守域,再利用 Smart (<http://smart.embl-heidelberg.de/>) 和 NCBI 中 Conserved Domains 程序 (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) 对家族成员进行进一步确认。

1.2 海岛棉 *GAUT* 基因家族的染色体定位和基因结构分析

根据 *GAUT* 家族成员的染色体和位置信息,利用 Mapchart 软件绘制基因定位图。将鉴定出的海岛棉 *GAUT* 基因序列通过 MEGA 中的 Clustal W 进行比对,采用相邻连接法 (NJ, Neighbor-joining) 构建系统发育树^[17]。基因结构图的绘制借助 GSDS (<http://gsds.cbi.pku.edu.cn/>) 完成。

1.3 海岛棉 *GAUTs* 家族保守域分析

利用 Web Logo (<http://weblogo.berkeley.edu/logo.cgi>) 绘制海岛棉 *GAUT* 氨基酸序列 LOGO^[18] 图。利用 MEME (<http://meme-suite.org/tools/meme>) 分析氨基酸序列^[12]。

1.4 海岛棉与拟南芥 *GAUTs* 的系统进化分析

37 个海岛棉 *GAUTs* 氨基酸序列和 15 个拟南芥 *GAUTs* 氨基酸序列利用 MEGA 的 Clustal W 进行比对,采用邻接法构建系统发育树^[17]。

1.5 海岛棉 *GAUTs* 理化性质和亚细胞定位

GAUT 基因家族成员编码产物的氨基酸数量、分子量、等电点等理化性质分析由 ProtParam (<http://web.expasy.org/protparam/>) 进行预测^[19]。利用 ProtComp9.0 (<http://www.softberry.com/berry.phtml?topic=protcom&ppl&group=programs&subgroup=proloc>)、

CELLO v. 2.5 (<http://cello.life.nctu.edu.tw/>) 对蛋白质进行亚细胞定位预测^[20], 借助在线网站 TMHMM(<http://www.cbs.dtu.dk/services/TMHMM-2.0>) 蛋白跨膜分析软件检测肽段的跨膜区。

1.6 GAUT 基因家族在海岛棉纤维发育过程中的表达分析

本实验室前期完成了海岛棉品种 Pima 90-53 和海 7124 纤维发育的 7 个时间点 (DPA, 0、5、10、15、20、25、30 days post-anthesis) 的转录组测序, 得到 2 个品种在不同时间点的 RPKM (Reads Per Kilo bases per Million reads)。海岛棉 Pima90-53 和海 7124 两个品种的 RPKM 及表达规律一致, 因此本研究在进行基因表达分析时, 先计算海岛棉 Pima90-53 和海 7124 两个品种的 RPKM 平均值, 在平均值的基础上再进行 $\log_2(1 + \text{RPKM})$ 处理。表达量热图的绘制借助 GPS (<http://hemi.biocuckoo.org/>) 完成, 即利用不同的颜色展示 GAUT 基因家族成员不同的表达量。根据海岛棉 GAUTs 在纤维发育各时期表达量的不同, 将 37 个 GAUTs 进行分类, 并分别绘制表达趋势图, 探究 GAUTs 在海岛棉纤维发育不同时期的表达规律。

2 结果与分析

2.1 海岛棉 GAUT 家族全基因组鉴定

通过 BLAST 比对, 利用 Smart 在线软件剔除不含 Glyco_transf_8 保守结构域的序列, 最终在海岛棉基因组中鉴定出 37 个 GAUT 家族成员, 按照它们在染色体上的位置顺序, 命名为 GbGAUT01-37 (表 1), 其中 GAUT37 位于没有组装到染色体的大片段上。从基因在染色体上的分布来看, GAUT 基因家族成员分布在 18 条不同的染色体上。除 A 亚组的 7 号染色体和 D 亚组的 7 号染色体基因数量较多外, 其他染色体上只分布着 1~3 个 GAUT 基因 (图 1)。

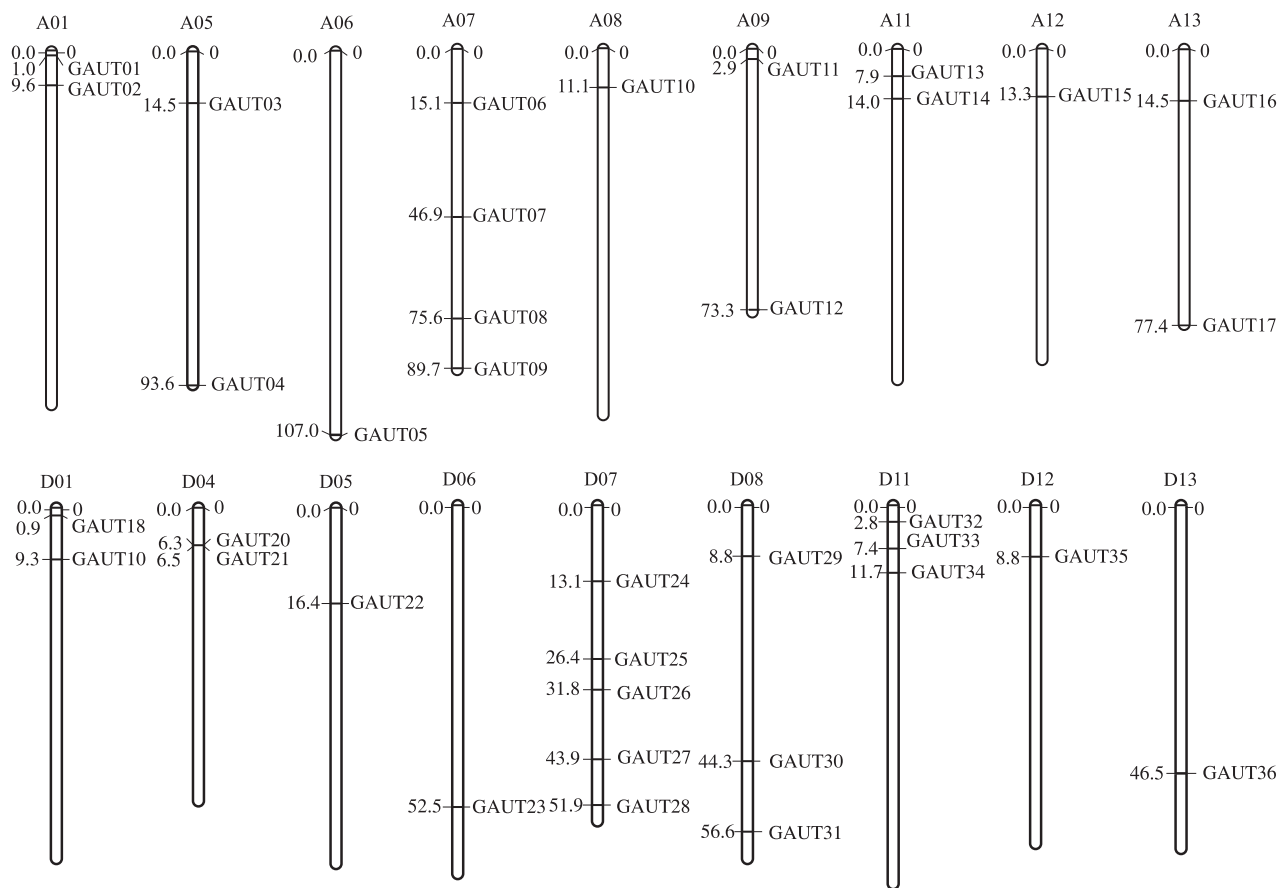
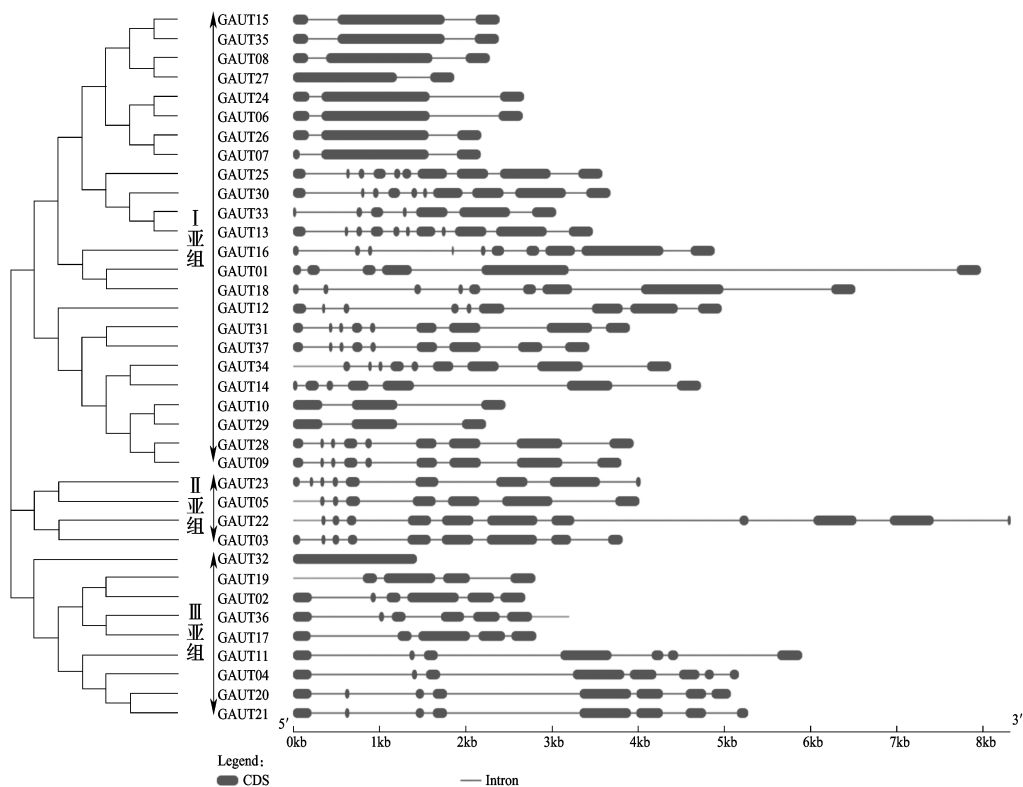
2.2 GAUT 家族基因结构分析

利用 37 条海岛棉 GAUT 基因的核苷酸序列构建系统进化树 (图 2), 可以将其分为 3 个亚组, I、II、III 亚组分别包含 24 个、4 个和 9 个基因。基因结构图 (图 2) 显示, 大多数 GAUT 基因的外显子个数在 3~11 之间, 以 3 个或 9 个外显子居多, 但 GAUT32 只有 1 个外显子。此外, 基因结构与亚组存在一定的联系, 如 I 亚组以含有 3 个外显子的基因居多, 而 II 亚组和 III 亚组中则以含有 9 个外显子和 6 个外显子的基因居多, 总的来看, 在进化树中处于相近分支的基因具有相似的基因结构。

表 1 海岛棉 GAUT 基因家族信息

Table1 Information of GAUT gene family in *G. barbadense* L.

基因名称 Gene	基因 ID Gene ID	染色体定位 Chromosome location
GAUT01	GOBAR_AA01456	A01:1016334-1024307
GAUT02	GOBAR_AA00301	A01:9613451-9616141
GAUT03	GOBAR_AA25346	A05:14497863-14501682
GAUT04	GOBAR_AA13922	A05:93596407-93601574
GAUT05	GOBAR_AA10943	A06:107027362-107031375
GAUT06	GOBAR_AA37671	A07:15099056-15101716
GAUT07	GOBAR_AA14268	A07:46927015-46929189
GAUT08	GOBAR_AA01514	A07:75555669-75557947
GAUT09	GOBAR_AA06647	A07:89670406-89674210
GAUT10	GOBAR_AA04077	A08:11065254-11067714
GAUT11	GOBAR_AA07870	A09:2874750-2880650
GAUT12	GOBAR_AA34869	A09:73319234-73324201
GAUT13	GOBAR_AA25184	A11:7886584-7890058
GAUT14	GOBAR_AA36945	A11:13994303-13999030
GAUT15	GOBAR_AA22791	A12:13348599-13350991
GAUT16	GOBAR_AA03897	A13:14455836-14460722
GAUT17	GOBAR_AA05453	A13:77398681-77401498
GAUT18	GOBAR_DD05031	D01:933831-940348
GAUT19	GOBAR_DD13138	D01:9304556-9307362
GAUT20	GOBAR_DD37379	D04:6278421-6283494
GAUT21	GOBAR_DD20347	D04:6480662-6485936
GAUT22	GOBAR_DD28302	D05:16373039-16381357
GAUT23	GOBAR_DD32865	D06:52531030-52535058
GAUT24	GOBAR_DD14647	D07:13131930-13134606
GAUT25	GOBAR_DD05509	D07:26427319-26430903
GAUT26	GOBAR_DD27569	D07:31847170-31849350
GAUT27	GOBAR_DD33929	D07:43934795-43936661
GAUT28	GOBAR_DD34963	D07:51871252-51875198
GAUT29	GOBAR_DD07609	D08:8776221-8778456
GAUT30	GOBAR_DD04306	D08:44255956-44259633
GAUT31	GOBAR_DD34943	D08:56586461-56590364
GAUT32	GOBAR_DD22761	D11:2807344-2808777
GAUT33	GOBAR_DD24858	D11:7360685-7363733
GAUT34	GOBAR_DD05442	D11:11746553-11750934
GAUT35	GOBAR_DD34814	D12:8771758-8774141
GAUT36	GOBAR_DD11145	D13:46541744-46544946
GAUT37	GOBAR_AA14759	scaffold_0069. UKA:114937-118369

图1 海岛棉 *GAUT* 基因家族的染色体定位Fig. 1 Chromosomal location of *GAUT* family members in *G. barbadense* L.图2 海岛棉 *GAUT* 基因家族进化树和基因结构Fig. 2 Phylogenetic tree and gene structure of the *GAUT* gene family in *G. barbadense* L.

2.3 海岛棉 *GAUT* 氨基酸序列保守域分析

保守结构域分析表明, *GAUT* 家族结构域有多个保守性的氨基酸残基, 如 D41、L49、P79、H83、C84 等, 部分保守性氨基酸残基组成 *GAUT* 家族特有的保守基序(图3)。*GAUT* 家族存在4个保守基序(表2), 第1个是 DPSLYHYALFSDNVLAASV-VVNSTIKNAKEPEKHVFHVTD, 整体表示为 [DN][PN][SD]LYH[YF]A[LIV]FSDN[VI]LAA[SA]VVVNS[TA][IV][KS][NSH]AKEP[ESW]K[HI]VFH[VI][VI]TD, 图3中用红线标出。第2个保守基序是 LNHLRFYLPEIFPKLNKILFLDDDIVVQK-DLTGLWSIDLKGVNGAVETC, 是 *GAUT* 家族的特征基序, 整体表示为 LNHLR[FI]YLP[ED][ILMV][FY]P[KA]L[NDH]K[IV][LV][FL]LD[DH]D[IV]V[VI]Q[KR]DL[TS][GP]LWS[IV]D[LM]

KGVNGAVETC, 图3中用绿线标出。第3个保守基序是 DPNACAWAYGMNIFDLDAWRRKNI-TETY-HYWQNLNLDRTLW, 整体表示为 [DN][PA][NK][AE]C[AGT]WA[YF]GMNIFDL[AE]WR[RK][KTE][NK][ICL]TE[TL]YHY[WY][QL][NQ]L[NG][LE][DK][RS]TLW, 在图3中用黄线标出。第4个保守基序是 IENAAVIHYNGNM-KPWLDIAINKYKPYWTKY, 整体表示为 IE[NR]AAVIH[YF][ND]GNMKPWL[DE][IL][AG][IM][NP][KQ][YL][KR]P[YL]W[TS]K[YH], 图3中用黑线标出。37个 *GbGAUT* 基因有4个相同的保守基序, 存在共同的氨基酸残基, 保守性较强; 但基因间又存在差异, 这些差异可能是导致基因功能发生分化的主要原因。

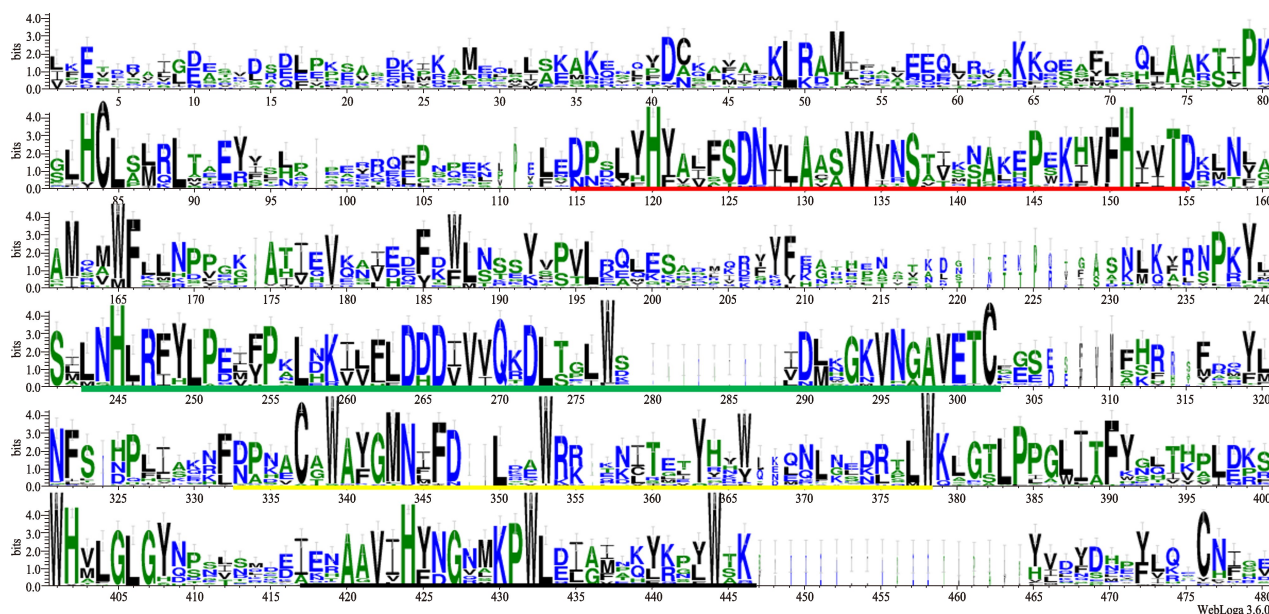


图3 海岛棉 37 个 *GAUT* 家族氨基酸序列保守性分析

Fig. 3 Conservation analysis of amino acid sequences of 37 *GAUT* genes in *G. barbadense* L.

表2 海岛棉 *GAUT* 基因家族保守基序

Table 2 Conservative motifs of *GAUT* gene family in *G. barbadense* L.

保守基序 Conservative motif	整体表示 Whole representation
DPSLYHYALFSDNVLAASV-VVNSTIKNAKEPEKHVFHVTD	[DN][PN][SD]LYH[YF]A[LIV]FSDN[VI]LAA[SA]VVVNS[TA][IV][KS][NSH]AKEP[ESW]K[HI]VFH[VI][VI]TD
LNHLRFYLPEIFPKLNKILFLDDDIVVQK-DLTGLWSIDLKGVNGAVETC	LNHLR[FI]YLP[ED][ILMV][FY]P[KA]L[NDH]K[IV][LV][FL]LD[DH]D[IV]V[VI]Q[KR]DL[TS][GP]LWS[IV]D[LM]KGVNGAVETC
DPNACAWAYGMNIFDLDAWRRKNI-TETY-HYWQNLNLDRTLW	[DN][PA][NK][AE]C[AGT]WA[YF]GMNIFDL[AE]WR[RK][KTE][NK][ICL]TE[TL]YHY[WY][QL][NQ]L[NG][LE][DK][RS]TLW
IENAAVIHYNGNMKPWLDAINKYKPYWTKY	IE[NR]AAVIH[YF][ND]GNMKPWL[DE][IL][AG][IM][NP][KQ][YL][KR]P[YL]W[TS]K[YH]

2.4 海岛棉与拟南芥 *GAUTs* 的系统进化分析

海岛棉与拟南芥的 *GAUTs* 进化树(图4)显示,海岛棉和拟南芥的 *GAUTs* 没有因为物种差异而单独分为两类。课题组前期从海岛棉中克隆了 *GbGAUT2*, *GbGAUT2* 在 10 DPA 出现表达高峰,而且转反义 *GbGAUT2* 拟南芥的叶片表皮毛变长^[4,21],本研究中 *GAUT04* 与其核苷酸序列相似度达到了

98.38%,推测 *GAUT04* 可能参与棉纤维伸长。在拟南芥中, *AtGAUT6*、*AtGAUT8*、*AtGAUT9*、*AtGAUT10*、*AtGAUT11*、*AtGAUT12*、*AtGAUT13* 和 *AtGAUT14* 基因的突变影响果胶和/或木聚糖的生物合成^[8],而果胶、木聚糖又是组成棉纤维初生壁的重要成分,参与棉纤维的伸长,因此,推测与这些基因位于同一分支的海岛棉 *GAUTs* 可能也参与棉纤维发育。



图4 拟南芥和海岛棉 *GAUTs* 基因家族进化树分析

Fig. 4 Phylogenetic tree of *GAUTs* family members in *Arabidopsis* and *G. barbadense* L.

2.5 海岛棉 *GAUT* 基因家族的理化性质和亚细胞定位

对 *GAUT* 蛋白理化性质分析表明,海岛棉 *GAUT* 家族成员编码的肽链平均由 569.9 个氨基酸组成,分子量介于 43288.14 ~ 107765.85 D, 平均 65481.82 D,理论等电点介于 5.46 ~ 9.51 之间,平均值为 8.34,多数偏碱性(表3)。蛋白不稳定指数平均值为 42.17,有 15 个 *GAUT* 成员为稳定蛋白(蛋白不稳定指数 < 40);其余的 22 个基因为不稳定蛋白(蛋白不稳定指数 > 40)。除 *GAUT04*、*GAUT26* 外,其余 *GAUT* 基因家族成员的脂溶指数 AI 都小于 100,平均值是 88.36,具有较高的脂溶性;所有的 *GAUTs* 蛋白序列的亲疏水性 GRAVY 都是负数,均表现为亲水性。

对 37 个 *GAUTs* 蛋白进行亚细胞定位预测(表3),显示所有家族成员均定位到高尔基体膜(Golgi membrane)。蛋白跨膜区分析结果表明,11

个 *GAUTs* 不含跨膜区,其余 26 个有跨膜区, *GAUT27* 有 2 个跨膜结构域。

2.6 *GAUT* 基因家族在海岛棉纤维发育过程中的表达分析

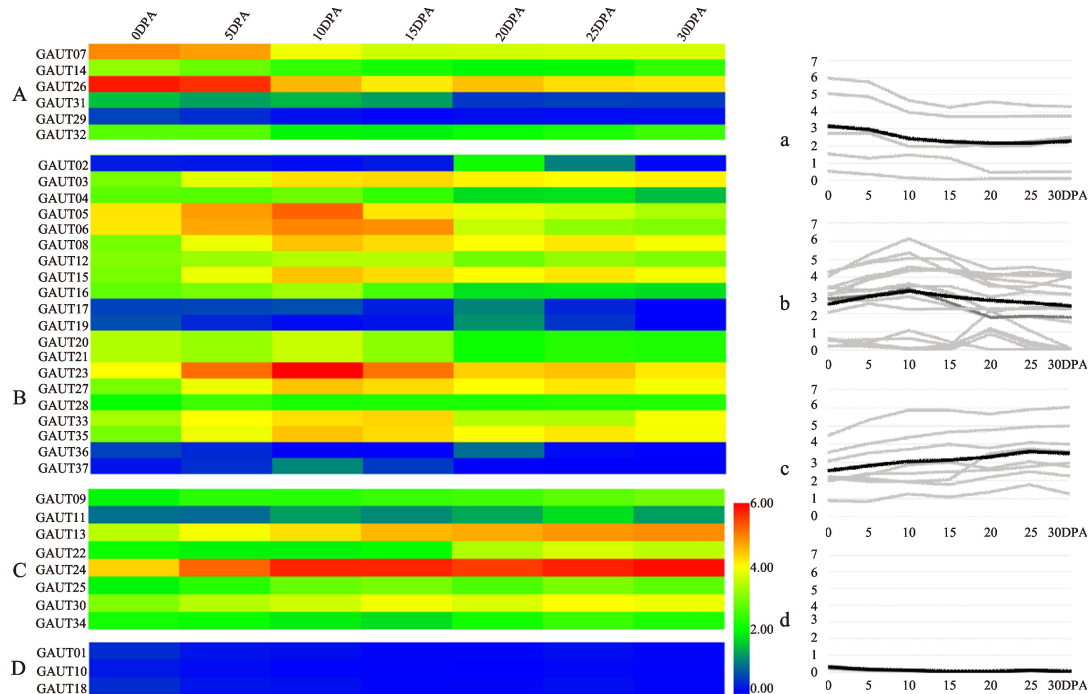
根据本实验室提供的棉纤维发育转录组数据,对两个海岛棉品种的 RPKM 平均值进行 $\log_2(1 + \text{RPKM})$ 处理,可以将 *GAUT* 家族不同成员的表达分为:起始期高表达、纤维伸长期高表达、次生壁增厚期高表达、全时期低表达等 4 类表达模式。*GAUT07*、*GAUT14*、*GAUT26*、*GAUT29*、*GAUT31*、*GAUT32* 等 6 个基因的表达峰值均出现在 0DPA,随着纤维发育天数的增加其表达量下降,为起始期高表达模式基因,推测其主要在纤维发育起始期发挥作用(图 5A, 5a)。*GAUT02*、*GAUT03*、*GAUT04*、*GAUT05*、*GAUT06* 等 20 个基因的表达峰值出现在 5DPA、10DPA、15DPA、20DPA 的某一时间点,但表达趋势图在 10DPA 达到表达高峰,为伸长期高表达基因,推测

表 3 海岛棉 *GAUT* 基因家族理化性质及亚细胞定位Table 3 Physicochemical properties and subcellular location of *GAUT* proteins in *G. barbadense* L.

基因名称 Gene name	氨基酸数量 Number of amino acids	分子量(D) Molecular weight	理论等电点 pI	蛋白不稳 定指数 II	脂溶指数 AI	总平均疏水性 GRAVY	亚细胞定位 Subcellular location	跨膜区数量 Numbers of transmembrane domain
<i>GAUT01</i>	599	69171.51	9.33	40.22	85.64	-0.447	Golgi membrane	1
<i>GAUT02</i>	616	70407.79	9.48	49.62	97.18	-0.154	Golgi membrane	0
<i>GAUT03</i>	377	43288.14	5.53	49.77	89.79	-0.219	Golgi membrane	1
<i>GAUT04</i>	477	54653.77	6.63	43.92	101.01	-0.119	Golgi membrane	1
<i>GAUT05</i>	494	56815.41	9.40	36.55	85.43	-0.397	Golgi membrane	0
<i>GAUT06</i>	688	79515.46	6.05	41.19	89.13	-0.389	Golgi membrane	1
<i>GAUT07</i>	585	66938.66	9.09	47.73	99.15	-0.086	Golgi membrane	1
<i>GAUT08</i>	580	66934.38	7.25	49.89	85.4	-0.465	Golgi membrane	1
<i>GAUT09</i>	936	107765.85	9.51	44.30	87.32	-0.443	Golgi membrane	0
<i>GAUT10</i>	491	56853.22	8.04	36.97	82.40	-0.438	Golgi membrane	1
<i>GAUT11</i>	564	64941.74	9.19	31.70	80.25	-0.337	Golgi membrane	1
<i>GAUT12</i>	695	79494.3	9.06	36.86	86.83	-0.359	Golgi membrane	1
<i>GAUT13</i>	423	48303.02	8.63	37.97	94.75	-0.073	Golgi membrane	1
<i>GAUT14</i>	586	67510.39	6.34	47.22	79.52	-0.541	Golgi membrane	1
<i>GAUT15</i>	448	50942.82	7.28	44.04	97.12	-0.121	Golgi membrane	1
<i>GAUT16</i>	560	64603.46	9.41	34.52	83.07	-0.322	Golgi membrane	1
<i>GAUT17</i>	535	61279.09	9.06	45.52	89.50	-0.436	Golgi membrane	0
<i>GAUT18</i>	569	65214.91	9.03	32.66	82.13	-0.294	Golgi membrane	1
<i>GAUT19</i>	560	64559.3	9.34	33.47	83.07	-0.333	Golgi membrane	1
<i>GAUT20</i>	534	60759.56	9.04	45.02	99.7	-0.055	Golgi membrane	1
<i>GAUT21</i>	529	60994.11	9.22	32.39	79.28	-0.423	Golgi membrane	0
<i>GAUT22</i>	599	69215.37	9.22	45.25	83.54	-0.472	Golgi membrane	1
<i>GAUT23</i>	512	58594.97	8.86	36.85	99.80	-0.036	Golgi membrane	0
<i>GAUT24</i>	498	56887.67	7.66	50.05	84.20	-0.53	Golgi membrane	1
<i>GAUT25</i>	609	68864.9	8.98	43.16	85.91	-0.347	Golgi membrane	0
<i>GAUT26</i>	512	58605.62	6.49	44.43	101.09	-0.065	Golgi membrane	1
<i>GAUT27</i>	377	43347.21	5.46	51.90	90.56	-0.218	Golgi membrane	2
<i>GAUT28</i>	662	76329.03	6.54	39.84	87.78	-0.485	Golgi membrane	0
<i>GAUT29</i>	628	72452.24	9.15	47.84	92.68	-0.377	Golgi membrane	0
<i>GAUT30</i>	672	76998.33	9.15	39.74	88.94	-0.425	Golgi membrane	0
<i>GAUT31</i>	557	63753.21	8.89	39.27	83.32	-0.339	Golgi membrane	1
<i>GAUT32</i>	583	67181.86	9.37	49.34	96.67	-0.213	Golgi membrane	0
<i>GAUT33</i>	576	66497.69	9.16	43.32	86.84	-0.563	Golgi membrane	1
<i>GAUT34</i>	551	63682.37	6.45	49.97	80.34	-0.508	Golgi membrane	1
<i>GAUT35</i>	687	79528.6	8.74	45.91	85.20	-0.486	Golgi membrane	1
<i>GAUT36</i>	649	74580.34	9.34	38.97	84.25	-0.471	Golgi membrane	1
<i>GAUT37</i>	569	65361.02	9.09	33.08	80.42	-0.316	Golgi membrane	1

这类基因可能影响着纤维长度(图 5B, 5b)。*GAUT09*、*GAUT11*、*GAUT13*、*GAUT22*、*GAUT24*、*GAUT25*、*GAUT30*、*GAUT34* 等 8 个基因的表达峰值出现在 25DPA 或 30DPA, 表达趋势图总体呈上升趋势(图 5C, 5c), 为次生壁增厚期高表达基因, 这类基

因可能影响着纤维强度。*GAUT01*、*GAUT10*、*GAUT18* 这 3 个基因在各个时期的表达量都很低, 为全时期低表达模式基因, 推测这些基因在棉纤维发育过程中可能作用较小(图 5D, 5d)。



A、B、C、D 分别代表一类 *GAUTs* 基因的表达量热图; a、b、c、d 为相应于 A、B、C、D 中基因的表达量趋势图

a、b、c、d 横坐标为棉纤维发育天数 (DPA), 纵坐标为棉纤维表达量 ($\log_2(1 + \text{RPKM})$)

A、B、C、D represent the expression heatmap of a kind of *GAUTs*, respectively. a、b、c、d represent the expression trend of a kind of *GAUTs* corresponding to A、B、C、D, respectively

The horizontal coordinates of a、b、c、d are the cotton fiber development days (DPA), and the vertical coordinates are the cotton fiber expression ($\log_2(1 + \text{RPKM})$)

图 5 海岛棉 *GAUT* 基因家族表达热图和趋势图

Fig. 5 The heatmap and trend plots of *GAUT* gene family expression in *G. barbadense* L.

3 讨论

本研究通过对海岛棉进行全基因组分析,共鉴定出 37 个海岛棉 *GAUT* 家族成员,均含有 Glyco_transf_8 保守结构域,所有家族成员均具有糖基转移酶的活性。全部 *GAUT* 定位于高尔基体膜。*GAUT* 基因家族的外显子个数都在 3 ~ 11 之间,以 3 个或 9 个外显子居多,但 *GAUT32* 只有 1 个外显子, *GAUT27* 含有 2 个外显子,这可能是由基因的进化和变异引起的。

在植物中含有大量的半乳糖基转移酶,其功能具有多样性。在番茄中 *GAUT* 影响植株的营养生长以及果胶的生物合成^[13]。拟南芥中 *GAUT6*、*GAUT8*、*GAUT9*、*GAUT10*、*GAUT11*、*GAUT12*、*GAUT13* 和 *GAUT14* 基因的突变影响果胶和/或木聚糖的生物合成^[8], *AtGAUT12* (*irx8*) 参与细胞壁多糖的合成,该基因的突变体导致多聚半乳糖醛酸和葡萄糖醛酸的含量显著降低^[10]。推测与这些基因位于同一分支的 *GbGAUTs* 可能也影响细胞壁的组成。本研究

中 *GAUT* 家族成员位于进化树的不同分支,也暗示了其功能的多样性。

本研究利用海岛棉转录组数据,将 *GAUT* 基因家族的表达模式分为 4 类,其中,在起始期高表达、纤维伸长期高表达、次生壁增厚期高表达的 3 类基因可能影响着纤维长度和强度。本研究中 *GAUT05*、*GAUT06*、*GAUT07*、*GAUT23*、*GAUT24*、*GAUT26* 等在纤维发育某个时期优势表达,系统进化树结果显示这些基因与 *AtGAUT8*、*GAUT9*、*GAUT11* 位于同一亚组,推测可能调节棉纤维中果胶/木聚糖的生物合成,进而影响棉纤维的长度。课题组前期从海岛棉中克隆了 *GbGAUT1* 和 *GbGAUT2*,本研究中 *GAUT32* 和 *GAUT04* 与其核苷酸序列相似度分别达到了 99.92% 和 98.38%, *GbGAUT1* 和 *GbGAUT2* 均在 10DPA 出现表达高峰,而且转反义 *GbGAUT2* 拟南芥的叶片表皮毛变长^[4,13,21],推测 *GAUT32* 和 *GAUT04* 可能参与棉纤维伸长。目前对 *GAUT* 基因家族在棉纤维发育过程中功能的了解十分有限。因此,系统研究 *GAUT* 基因家族可为棉纤

维品质分子改良提供重要的候选基因。

参考文献

- [1] 姚贺盛,张亚黎,易小平,薛军,罗毅,罗宏海,张旺锋. 海岛棉和陆地棉叶片光合特性、冠层结构及物质生产的差异. 中国农业科学,2015,48(2):251-261
- [2] 刘艳. 海岛棉枯萎病抗性相关基因的克隆及功能验证. 乌鲁木齐:新疆农业大学,2013
- [3] Cosgrove D J. Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology*,2005,6(11):850-861
- [4] 王敬敬. 棉纤维发育果胶多糖及半纤维素合成相关基因的克隆及表达分析. 保定:河北农业大学,2012
- [5] 陈良兵,李永起. 棉花纤维发育的分子研究进展. 分子植物育种,2004,2(1):105-111
- [6] Sterling J D, Atmodjo M A, Inwood S E, Kolli V S K, Quigley H F, Hahn M G, Mohnen D. Functional identification of an Arabidopsis pectin biosynthetic homogalacturonan galacturonosyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*,2006,103(13):5236-5241
- [7] Wang L, Wang W, Wang Y Q, Liu Y Y, Wang J X, Zhang X Q, Ye D, Chen L Q. Arabidopsis galacturonosyltransferase (GAUT) 13 and GAUT14 have redundant functions in pollen tube growth. *Molecular Plant*,2013,6(4):1131-1148
- [8] Caffall K H, Mohnen D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate Research*,2009,344:1879-1900
- [9] Yin Y, Chen H, Hahn M G, Mohnen D, Xu Y. Evolution and function of the plant cell wall synthesis-related glycosyltransferase family 8. *Plant Physiology*,2010,153(4):1729-1746
- [10] Persson S, Caffall K H, Freshour G, Hilley M T, Bauer S, Poindexter P, Hahn M G, Mohnen D, Somerville C. The arabidopsis irregular xylem8 mutant is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity. *The Plant Cell*,2007,19(1):237-255
- [11] Peña M J, Zhong R, Zhou G K, Richardson E A, O'Neill M A, Darvill A G, York W S, Ye Z H. Arabidopsis irregular xylem8 and irregular xylem9: implications for the complexity of glucuronoxylan biosynthesis. *The Plant Cell*,2007,19(2):549-563
- [12] Godoy F D, Bermúdez L, Lira B S, Souza A P D, Elbl P, Demarco D, Alseekh S, Insani M, Buckeridge M, Almeida J, Grigioni G, Fernie A R, Carrari F, Rossi M. Galacturonosyltransferase 4 silencing alters pectin composition and carbon partitioning in tomato. *Journal of Experimental Botany*,2013,64(8):2449-2466
- [13] Chi J N, Han G Y, Wang X F, Zhang G Y, Sun Y X, Ma Z Y. Isolation and molecular characterization of a novel homogalacturonan galacturonosyl-transferase gene (*GbGAUT1*) from *Gossypium barbadense*. *African Journal of Biotechnology*,2009,8(19):4755-4764
- [14] Liu X, Zhao B, Zheng H J, Hu Y, Lu G, Yang C Q, Chen J D, Chen J J, Chen D Y, Zhang L, Zhou Y, Wang L J, Guo W Z, Bai Y L, Ruan J X, Shangguang X X, Mao Y B, Shan C M, Jiang J P, Zhu Y Q, Jin L, Kang H, Chen S T, He X L, Wang R, Wang Y Z, Chen J, Wang L J, Yu S T, Wang B Y, Wei J, Song S C, Lu X Y, Gao Z C, Gu W Y, Deng X, Ma D, Wang S, Liang W H, Fang L, Cai C P, Zhu X F, Zhou B L, Chen Z J, Xu S H, Zhang Y G, Wang S Y, Zhang T Z, Zhao G P, Chen X Y. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Scientific Reports*,2015,30(5):14139
- [15] Finn R D, Clements J, Eddy S R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*,2011,39:29-37
- [16] 郭庆勋,张春雨,王晶莹,周连霞,王彦涛. 九台晚李 *PGIP* 基因的克隆及生物信息学分析. 植物遗传资源学报,2010,11(5):650-653
- [17] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*,2013,30(12):2725-2729
- [18] Crooks G E, Hon G, Chandonia J M, Brenner S E. WebLogo: a sequence logo generator. *Genome Research*,2004,14(6):1188-1190
- [19] 樊连梅,刘更森,刘成连,原永兵. 苹果金属硫蛋白基因 *Md-FjMT2* 克隆及生物信息学分析. 植物遗传资源学报,2011,12(5):782-789
- [20] 沈知临,许磊,陈文,吴楠,蔡永萍,林毅,高俊山. 亚洲棉和雷蒙德氏棉 *MATE* 基因家族生物信息学及其同源基因在陆地棉中的表达分析. 棉花学报,2016,28(3):215-226
- [21] 张华. 棉纤维发育相关基因 *GbGAUT2* 和 *GbUGP1* 的功能研究. 保定:河北农业大学,2013