

赵 亮, 狄佳春, 陈旭升. 棉花基因组数据库中 *CPS&KS* 基因的查找与分析[J]. 江苏农业学报, 2016, 32(1): 27-33.
doi: 10.3969/j.issn.1000-4440.2016.01.004

棉花基因组数据库中 *CPS&KS* 基因的查找与分析

赵 亮, 狄佳春, 陈旭升

(江苏省农业科学院经济作物研究所/农业部长江下游棉花与油菜重点实验室, 江苏 南京 210014)

摘要: 为了探究赤霉素合成关键酶 *CPS&KS* 类基因在棉花中的分布情况, 分析了已发表的拟南芥 *CPS&KS* 基因编码的蛋白质氨基酸序列, 发现该类基因存在 2 个保守结构域, *Terpene_synth* 和 *Terpene_synth_C*, 其在 Pfam 数据库中的种子文件分别为 PF01397 和 PF03936。利用软件 HMMER 中的 Hmmssearch 程序在雷蒙德氏棉蛋白质数据库中调取 72 条序列, 通过与拟南芥中的 *CPS&KS* 基因序列进行比对, 最终在雷蒙德氏棉基因组中筛选到 5 个 *CPS&KS* 候选基因, 这 5 个基因与四倍体棉花 TM-1 基因组 At 和 Dt 亚组的 10 个基因具有同源关系。通过对赤霉素敏感的棉花超矮秆突变体赤霉素处理前后转录组数据库分析比对, 最终确定 2 个 *KS* 候选基因在转录组中表现为上调作用。

关键词: 棉花; 赤霉素合成酶; 柯巴基焦磷酸合酶(*CPS*); 内根-贝壳杉烯合酶(*KS*)

中图分类号: S562.035.3 **文献标识码:** A **文章编号:** 1000-4440(2016)01-0027-07

Analysis of *ent*-copalyl diphosphate synthase and *ent*-kaurene synthase (*CPS&KS*) gene family in cotton genome databases

ZHAO Liang, DI Jia-chun, CEHN Xu-sheng

(Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Key Laboratory of Cotton and Rape in the Lower Reaches of the Yangtze River, Ministry of Agriculture, Nanjing 210014, China)

Abstract: The study focused on the distribution of *ent*-copalyl diphosphate synthase and *ent*-kaurene synthase (*CPS&KS*), the key enzymes for biosynthesis of GA, in cotton genome. The *CPS&KS* gene family has two conservative structure domains, *Terpene_synth* and *Terpene_synth_C*, based on the analysis of their amino acid sequences from *Arabidopsis thaliana*, and the seed files in Pfam database are PF01397 and PF03936, respectively. By using Hmmssearch of HMMER, 72 sequences were harvested in the genome database from *Gossypium raimondii*. Five genes exhibiting homologies with 10 genes in subgenomes At and Dt of genome TM-1 of tetraploid cotton were identified. Based on the alignment of transcriptome database of an ultra-dwarf mutant with GA sensitivity before and after GA treatment, two *KS* genes were found up-regulated.

Key words: cotton; GA biosynthetase; *ent*-copalyl diphosphate synthase (*CPS*); *ent*-kaurene synthase (*KS*)

收稿日期: 2015-06-09

基金项目: 国家青年科学基金项目(31401429); 江苏省青年基金项目(BK20140747)

作者简介: 赵 亮(1984-), 男, 吉林农安人, 博士, 助理研究员, 从事棉花遗传育种研究工作。(Tel) 025-84390617; (E-mail) jaascotton@jaas.ac.cn

通讯作者: 陈旭升, (Tel) 025-84390371; (E-mail) njcxs@126.com

赤霉素(GA)是一种二萜类的植物激素, 在植物整个生命周期中调控植物的生长与发育。二十世纪 30 年代, 日本科学家从真菌中提取出了赤霉素混合物晶体, 并命名为赤霉素 A; 到了二十世纪 50 年代, 在美国、英国和日本科学家联合努力下赤霉素 A 的化学成分被确定为 GA₁、GA₂ 和 GA₃ 的混合物, 其

中 GA_3 是主要成分^[1], 也被称为赤霉素。赤霉素是一种调控多种发育进程的植物内源激素^[2]。尽管在高等植物和真菌中已经鉴定出 136 种赤霉素^[3], 但是只有少数具有生物学活性, 例如, GA_1 、 GA_3 、 GA_4 和 GA_7 ^[4]。在拟南芥中, 赤霉素对于种子的萌发, 叶片和根的生长, 成花诱导(短日照条件下), 抽薹, 花药和花瓣的发育及种子的发育等起着关键的作用^[5]。

通过气相色谱-质谱分析法对赤霉素成分的分析, 赤霉素代谢酶的纯化, 赤霉素缺陷突变体的分离及相应基因的克隆使赤霉素的生物合成和分解代谢途径得到深入研究^[4,6]。高等植物体内赤霉素生物合成包括 3 个阶段: 第 1 阶段, 在原质体中由牻牛儿苗牻牛儿焦磷酸(GGDP)合成内根-贝壳杉烯(*ent-kaurene*); 第 2 阶段, 通过细胞色素 P450 单加氧酶将内根-贝壳杉烯转变为 GA_{12} ; 第 3 阶段为 C_{20} -GAs 和 C_{19} -GAs 的合成。其中第 2 和第 3 阶段是在细胞质中完成^[5,7-8]。赤霉素代谢途径中, GGDP 被转变为内根-贝壳杉烯需要 2 步环化反应, 首先在柯巴基焦磷酸合酶(*ent-copalyl diphosphate synthase*)的催化下, GGDP 环化形成内根柯巴基焦磷酸(*ent-copalyl diphosphate*, *ent-CDP*), 后者在内根-贝壳杉烯合酶(*ent-kaurene synthase*)催化下环化成为赤霉素的前身——内根-贝壳杉烯^[5]。

参与赤霉素第 1 阶段合成的柯巴基焦磷酸合酶和内根-贝壳杉烯合酶同属于萜类合酶。而植物的萜类合酶主要分为 2 类^[9]: 类型 I 包含有一个天冬氨酸富集基元(DDXXD), 这个基元具有结合金属离子的作用; 而类型 II 拥有一个 DXDD 基元, 该基元在序列中的位置有别于第一类的 DDXXD 基元^[10]。第 1 个类型 II 萜类合酶被克隆后^[11], 该酶相应的环化机制和基元被认为与鲨烯-藿烯环化酶十分相似^[12]。因此, 类型 II 与二萜(萜类合酶家族)和三萜(羊毛甾醇合酶家族)环化酶不相关^[13]。

本研究利用生物信息学方法从二倍体棉花雷蒙德氏棉及四倍体棉花 TM-1 基因组数据库中查找可能编码参与赤霉素合成第一阶段中的柯巴基焦磷酸合酶和内根-贝壳杉烯合酶基因, 并根据与所发表的拟南芥相关基因的聚类结果, 筛选候选基因, 为研究棉花赤霉素代谢途径奠定基础。

1 材料与方法

1.1 数据库来源

棉花二倍体 D 基因组雷蒙德氏棉(*G. Raimondii*)的基因组数据库来源于 <http://www.phytozome.net>, 棉花四倍体 AADD 基因组 TM-1 的基因组数据库来源于 <http://mascotton.njau.edu.cn>, 拟南芥相关基因来自于 TAIR 数据库(<http://www.arabidopsis.org/>)。

1.2 基因注释

CPS&KS 基因注释根据 Sonnhammer 等^[14]的方法, HMMER 版本为 3.0; CPS&KS 种子文件来自于 Pfam 数据库(<http://pfam.sanger.ac.uk/>), 种子编号为: PF01397 和 PF03936。蛋白质氨基酸序列保守结构域的分析采用在线的 SMART(<http://smart.embl-heidelberg.de/>)。

1.3 序列比对及进化树构建

用 MEGA 5.05^[15] 构建进化树, 进化树构建方法利用最大似然函数, Bootstrap 实行 1 000 次重复。序列比对用 Clustal1.83 版本, 比对结果及处理用 GENDOC 软件完成。

1.4 棉花转录组数据库构建及测序

超矮秆突变体材料来自陆地棉 C119-2 的分离群体, 是一种自然突变体^[15]。收集超矮秆突变体及经赤霉素处理后正常植株相同部位的叶片, 在液氮中迅速冷冻。液氮处理后的材料交由北京诺禾致源生物信息科技有限公司进行 RNA 的提取、测序及数据分析。构建数据库所用材料都具有 3 次生物学重复。

2 结果与分析

2.1 雷蒙德氏棉全基因组 CPS&KS 基因的注释

分析已经公布的拟南芥 1 个 CPS 基因(*At4g02780*)和 1 个 KS 基因(*At1g79460*)编码的蛋白质氨基酸序列分析, 发现这 2 个基因含有 2 个共同的结构域, 即 *Terpene_synth* (PF01397) 和 *Terpene_synth_C* (PF03936)。通过 HMMER3.0 版本中的 Hmsearch 程序及 2 个种子文件, 最终在雷蒙德氏棉蛋白质数据库中得到 72 条有注释结果的蛋白质氨基酸序列。其中最小的蛋白质含有 66 个氨基酸, 最大的含有 847 个氨基酸; 这些基因不均匀地分布在雷蒙德氏棉的 13 条染色体上, 其中 D5 染色体上分布最多, 含有 25 个基因; 有 5 条染色体没有注释结果, 分别为 D2、D3、D6、D7 和 D12 染色体(图 1)。

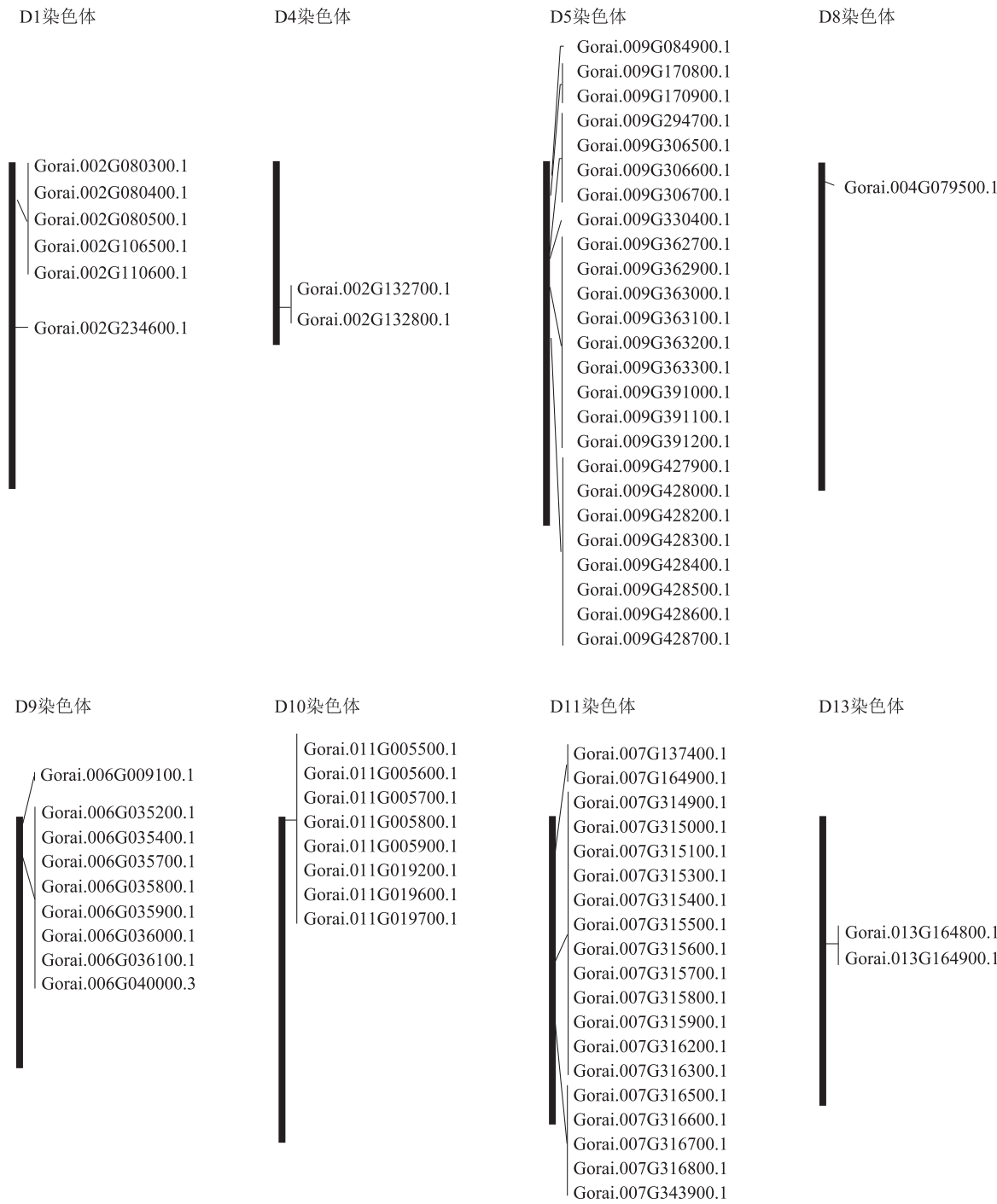


图 1 *CPS&KS* 基因在 D 组染色体上的位置
Fig.1 The loci of *CPS&KS* genes in D genome chromosomes

2.2 *CPS&KS* 基因系统进化分析

为进一步减少候选基因的数量,将注释得到的 72 条蛋白质氨基酸序列与拟南芥 *CPS* 基因 (*AtCPS*, *At4g02780*) 和 *KS* 基因 (*AtKS*, *At1g79460*) 编码的蛋白质氨基酸序列进行聚类分析,最终确定 5 个基因为最

有可能的候选基因。在这 5 个基因中,2 个属于 *CPS* 基因的候选基因 (Gorai. 004G079500. 1、Gorai. 007G137400.1),3 个属于 *KS* 基因的候选基因 (Gorai. 002G110600. 1、Gorai. 009G170800. 1、Gorai. 009G170900. 1), 分别分布在 D5 (Gorai. 009G170800. 1、Gorai.

009G170900.1), D1 (Gorai.002G110600.1)、D8 (Gorai.004G079500.1) 和 D11 (Gorai.007G137400.1) 染色体上 (图 2)。

2.3 TM-1 基因组中与雷蒙德氏棉 *CPS&KS* 基因同源的基因

利用以上得到的 5 个候选基因的核苷酸序列, 通过 cross_match 程序, 最终在 TM-1 基因组数据库中找到 10 个同源基因。这 10 个同源基因分别来源于 TM-1 的 At 和 Dt 亚组 (其中 Gh_Sca006732G01

不清楚其来源) (表 1)。为了验证这些基因的同源性关系, 我们利用 ClustX1.83 版本对序列进行了可视化的比较分析。分析结果表明: 这些同源基因之间存在着较高的同源性, 而通过构建进化树发现, 同源基因能够保证在同一分支上。相比于 TM-1 At 亚组核苷酸序列, 来自于雷蒙德氏棉中的核苷酸序列与 TM-1 Dt 亚组的核苷酸序列具有更高的同源性 (图 3)。

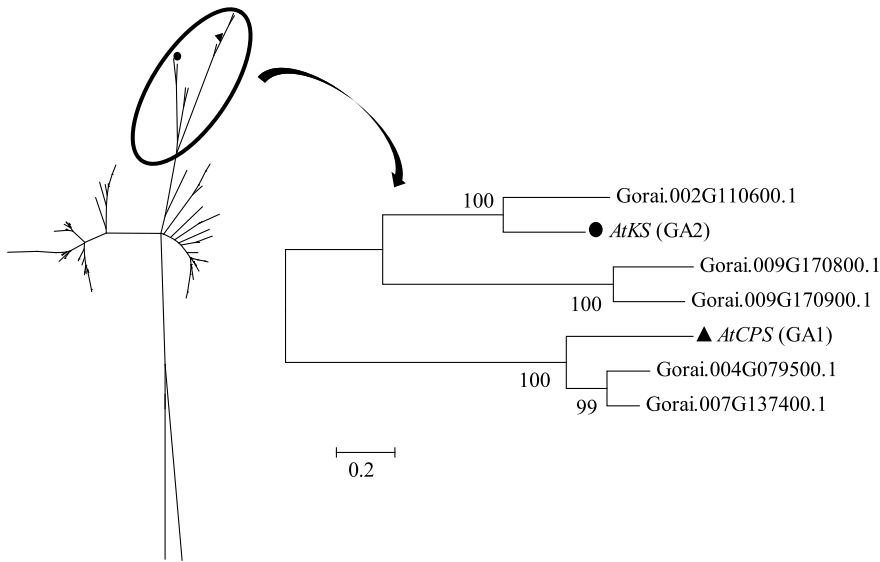
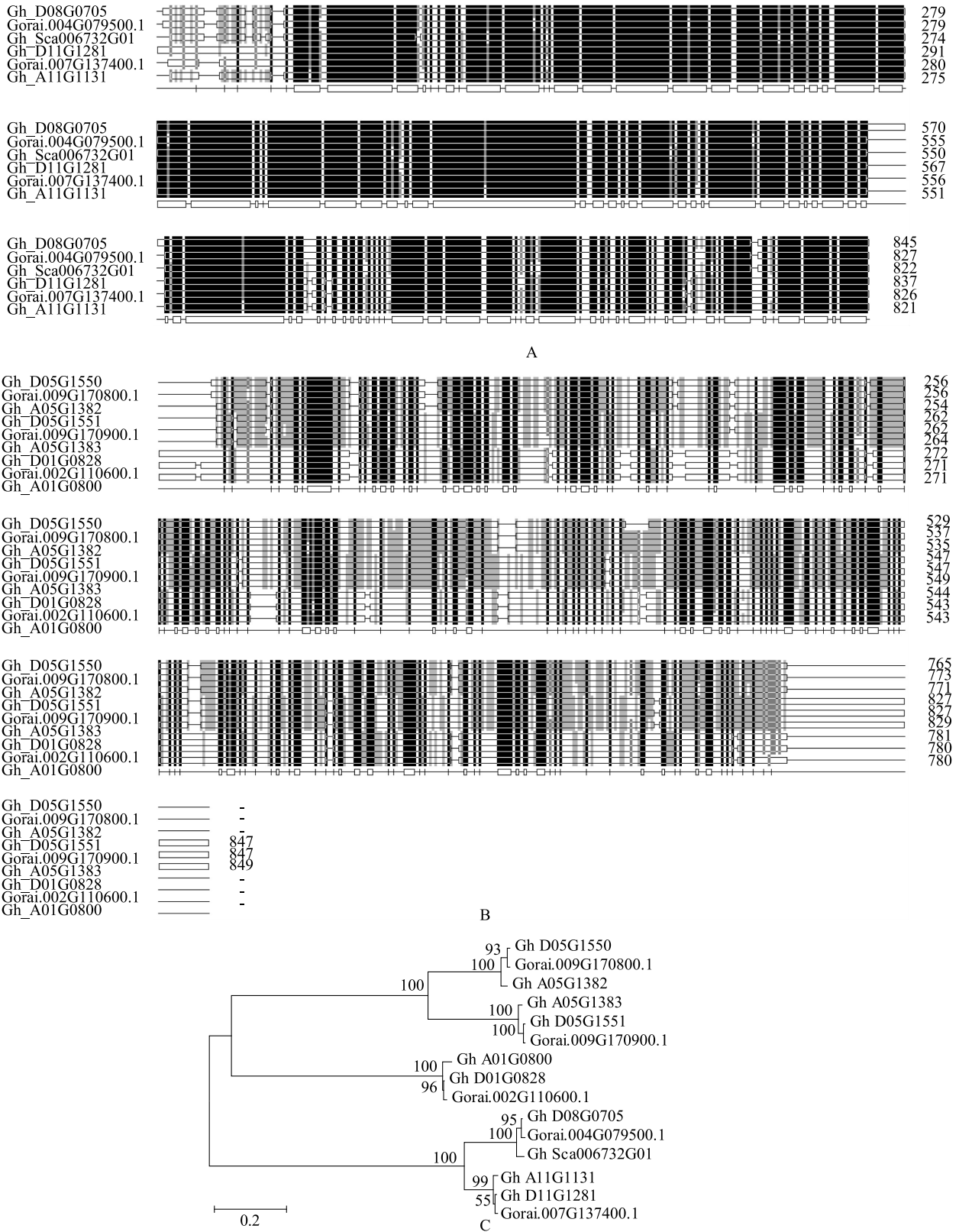


图 2 *CPS&KS* 基因的系统进化树
Fig.2 The phylogenetic tree of *CPS&KS* genes

表 1 来自于 TM-1 和雷蒙德氏棉基因组数据库中的同源基因

Table 1 The homologous genes from TM-1 and *Gossypium Raimondii* genome databases

TM-1 数据库中的基因				雷蒙德氏棉数据库中的基因		
基因序列号	亚组	染色体	氨基酸序列长度 (aa)	基因序列号	染色体	氨基酸序列长度 (aa)
Gh_D08G0705	Dt	D08 (Chr.24)	845	Gorai.004G079500.1	D08	827
Gh_Sca006732G01	未知	未知	822			
Gh_D11G1281	Dt	D11 (Chr.21)	837	Gorai.007G137400.1	D11	826
Gh_A11G1131	At	A11 (Chr.11)	821			
Gh_D01G0828	Dt	D01 (Chr.15)	781	Gorai.002G110600.1	D01	780
Gh_A01G0800	At	A01 (Chr.1)	780			
Gh_D05G1550	Dt	D05 (Chr.19)	765	Gorai.009G170800.1	D05	773
Gh_A05G1382	At	A05 (Chr.5)	771			
Gh_D05G1551	Dt	D05 (Chr.19)	847	Gorai.009G170900.1	D05	847
Gh_A05G1383	At	A05 (Chr.5)	849			



A: CPS 基因同源比对; B: KS 基因同源比对; C: CPS&KS 基因系统进化树。

图 3 来自于 2 个基因组数据库的同源基因之间的关系

Fig.3 The relationship between homologous genes from two genome databases

2.4 候选 *CPS&KS* 基因在转录组中的表达分析

陈旭升等^[15]曾报道一个对赤霉素敏感的陆地棉突变体材料,该材料在不喷施赤霉素的情况下表现为极端矮化,当喷施赤霉素后能够成长为正常植株。提取该突变体和喷施赤霉素后的正常植株的 RNA,构建转录组文库。利用得到的候选基因在转录组文库中查找同源基因,发现雷蒙德氏棉的 5 个 *CPS&KS* 基因中,只有 2 个 *KS* 基因能够

在转录组差异基因中比对到同源基因,这 2 个基因都位于雷蒙德氏棉的 D5 染色体上,分别为 Gorai.009G170800.1 和 Gorai.009G170900.1,并且这 2 个基因在转录组中表现为上调作用(表 2)。同时,我们也检测了所有在雷蒙德氏棉中有 *CPS&KS* 注释结果的基因,发现也仅仅是 Gorai.009G170800.1 和 Gorai.009G170900.1 这 2 个 *KS* 基因在转录组差异基因中有比对结果。

表 2 候选 *CPS&KS* 基因在转录组中的表达

Table 2 The expression of candidate *CPS&KS* genes in transcriptome database

基因号	GA+读数 ¹⁾	突变体读数	表达倍数	<i>p</i> 值	<i>p</i> 校正值
Gorai.009G170800.1	789.24	123.98	2.67	2.77×10^{-13}	2.38×10^{-10}
Gorai.009G170900.1	253.67	24.9	3.35	3.03×10^{-26}	2.89×10^{-22}

¹⁾ 突变体喷施赤霉素后的转录组读数。

3 讨论

CPS&KS 是参与赤霉素第一阶段合成的关键酶,在原质体中参与由牻牛儿苗牻牛儿苗焦磷酸(GGDP)合成内根-贝壳杉烯(Ent-kaurene)这一过程^[4-5]。通过对已经公布的拟南芥 *CPS* 基因^[11](At4g02780)和 *KS* 基因^[17](At1g79460)编码的蛋白质氨基酸序列分析,发现这 2 个基因在 Pfam 数据库中含有 2 个共同的结构域,即 Terpene_synth(PF01397)和 Terpene_synth_C(PF03936)。利用生物信息学方法在雷蒙德氏棉蛋白质数据库中得到 72 条有注释结果的蛋白质氨基酸序列。这些有注释结果的基因不均匀地分布在雷蒙德氏棉的 13 条染色体上,其中 D05 染色体上分布最多,含有 25 个基因,有 5 条染色体没有注释结果,分别为 D2、D3、D6、D7 和 D12 染色体。为了减少后期对该类基因研究的工作量,我们通过与拟南芥中该类基因所构建的系统进化树进行聚类分析以减少候选基因的数量,最终在雷蒙德氏棉中得到 5 个可信度较高的基因。棉花四倍体基因组中 Dt 亚组被认为是来自于二倍体 D 组的雷蒙德氏棉,同时四倍体中 At 和 Dt 亚组具有较高的同源关系^[18-19],因此,来自于雷蒙德氏棉的每个 *CPS&KS* 候选基因能够在 TM-1 数据库中找到 2 个同源的基因,并且这些基因在碱基序列上具有较高的同源性。

将得到的候选基因在棉花超矮秆突变体及赤霉素处理后正常植株的转录组数据库中进行查找分

析,发现只有位于 D5 染色体上的 2 个 *KS* 基因能够在转录组差异基因中找到同源基因,并且这 2 个基因在赤霉素处理后表现为上调作用。这可能是由于其余基因并不参与调控该突变体性状的形成,因此在转录数据库中不存在差异表达。景超等^[20]通过 SSR 分子标记将超矮秆突变体基因 *du* 定位在棉花 A6 染色体上。但是本研究预测的所有 *CPS&KS* 基因并未在 A6 或 D6 染色体上出现,因此可以推断 *CPS&KS* 基因并不是控制该性状的主要基因,但是控制该性状的基因却可以影响 *KS* 基因的表达。而吴巧娟等^[21]曾报道过一个半显性矮秆突变体 *Sdu*,利用 SSR 分子标记初步将该基因定位在陆地棉第 19 染色体(D5 染色体)上。而本研究中位于 D5 染色体上的 2 个 *KS* 基因是否与半显性矮秆突变体相关联,需要进一步研究。

参考文献:

- [1] TAKAHASHI N, PHINNEY B O, MACMILLAN J. Gibberellins [M]. New York: Springer, 1991.
- [2] MACMILLAN J. Occurrence of Gibberellins in Vascular Plants, Fungi, and Bacteria [J]. Journal of Plant Growth Regulation, 2001, 20(4): 387-442.
- [3] HEDDEN P, PHILLIPS A L. Gibberellin metabolism: new insights revealed by the genes [J]. Trends Plant Sci, 2000, 5(12): 523-530.
- [4] SUN T P. Gibberellin metabolism, perception and signaling pathways in Arabidopsis [J]. Arabidopsis Book, 2008, 6: 103.
- [5] YAMAGUCHI S. Gibberellin metabolism and its regulation [J]. Annu Rev Plant Biol, 2008, 59: 225-251.

- [6] HEDDEN P, PROEBSTING W M. Genetic analysis of gibberellin biosynthesis[J]. *Plant Physiol*, 1999, 119(2): 365-370.
- [7] PHILLIPS A L. Gibberellins in Arabidopsis[J]. *Plant Physiology and Biochemistry*, 1998, 36(1): 115-124.
- [8] BOHLMANN J, MEYER-GAUEN G, CROTEAU R. Plant terpenoid synthases: molecular biology and phylogenetic analysis[J]. *Proc Natl Acad Sci USA*, 1998, 95(8): 4126-4133.
- [9] PETERS R J, CROTEAU R B. Abietadiene synthase catalysis: conserved residues involved in protonation-initiated cyclization of geranylgeranyl diphosphate to (+)-copalyl diphosphate[J]. *Biochemistry*, 2002, 41(6): 1836-1842.
- [10] SUN T P, KAMIYA Y. The Arabidopsis GA1 locus encodes the cyclase ent-kaurene synthetase A of gibberellin biosynthesis[J]. *Plant Cell*, 1994, 6(10): 1509-1518.
- [11] OCHS D, KALETTA, ENTIAN K D, et al. Cloning, expression, and sequencing of squalene-hopene cyclase, a key enzyme in triterpenoid metabolism[J]. *Journal of bacteriology*, 1992, 174(1): 298-302.
- [12] WENDT K U, SCHULZ G E. Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes[J]. *Structure*, 1998, 6(2): 127-133.
- [13] SONNHAMMER E L, EDDY S R, BIRNEY E, et al. Pfam: multiple sequence alignments and HMM-profiles of protein domains[J]. *Nucleic Acids Res*, 1998, 26(1): 320-322.
- [14] TAMURA K, PETERSON D, PETERSON N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. *Mol Biol Evol*, 2011, 28(10): 2731-2739.
- [15] 陈旭升,狄佳春,许乃银,等. 陆地棉超矮秆突变性状质量遗传规律分析[J]. *遗传*, 2007, 29(4): 471-474.
- [16] YAMAGUCHI S, SUN T, KAWAIDE H, et al. The GA2 locus of *Arabidopsis thaliana* encodes ent-kaurene synthase of gibberellin biosynthesis[J]. *Plant Physiol*, 1998, 116(4): 1271-1278.
- [17] WENDEL J F, CRONN R C. Polyploidy and the evolutionary history of cotton [J]. *Advances in Agronomy*, 2003, 78(2): 139-186.
- [18] Paterson A H. *Cotton Genomics*[M]. Berlin Heidelberg: Springer, 2010:45-63.
- [19] 景 超,马晓杰,狄佳春,等. 陆地棉超矮秆突变体基因的初步定位[J]. *遗传*, 2011, 33(12): 1393-1397.
- [20] 吴巧娟,肖松华,刘剑光,等. 棉花半矮秆基因的定位[J]. *江苏农业学报*, 2012, 28(1): 214-215.

(责任编辑:张震林)