

海岛棉 WRKY 转录因子的全基因组鉴定及 响应黄萎病菌侵染的表达分析

傅明川, 李 浩, 陈义珍, 王立国, 柳展基, 刘任重

(农业部黄淮海棉花遗传改良与栽培生理重点实验室 / 山东棉花研究中心, 济南 250100)

摘要: WRKY 蛋白是植物中一类重要的转录因子, 其在很多生物过程中都发挥有重要作用。目前, 关于海岛棉 (*Gossypium barbadense* L.) WRKY 转录因子的研究相对较少。本研究利用生物信息学方法在海岛棉基因组中鉴定出 180 个 WRKY 转录因子 (GbWRKY)。根据基因结构及系统进化关系, 可将 GbWRKY 基因分为 3 组 (I~III), II 组又可细分为 5 个亚组 (IIa~IIe)。GbWRKY 基因在染色体上分布不均匀。基因重复事件分析表明, 全基因组复制及区段重复事件是 GbWRKY 基因数量增多的主要原因。表达分析发现, 有 61 个 GbWRKY 基因在黄萎病菌 (*Verticillium dahliae*) 侵染条件下呈现差异表达, 表明它们在黄萎病菌胁迫应答过程中发挥作用。本研究结果揭示了海岛棉 WRKY 转录因子的结构、进化及表达特征, 为进一步研究棉花 WRKY 基因的功能提供了理论指导。

关键词: WRKY 转录因子; 棉花; 黄萎病; 生物信息学

Genome-wide Investigation of WRKY Transcription Factors in *Gossypium barbadense* and Their Expression Patterns in Response to *Verticillium dahliae* Infection

FU Ming-chuan, LI Hao, CHEN Yi-zhen, WANG Li-guo, LIU Zhan-ji, LIU Ren-zhong

(Key Laboratory of Cotton Breeding and Cultivation in Huang-Huai-Hai Plain, Ministry of Agriculture/

Cotton Research Center of Shandong Academy of Agricultural Sciences, Jinan 250100)

Abstract: WRKY proteins, as one of the most important transcription factor families in plant, play important roles involved in many biological processes. However, studies of WRKY proteins in *Gossypium barbadense* L. remain scarce. By deploying bioinformatic methods to perform a genome-wide characterization in *G. barbadense* genome, 180 WRKY transcription factors were identified. According to their structural features and topology of the phylogenetic tree, the GbWRKY genes were classified into three major groups (I-III). In addition, group II is further classified into five subgroups (IIa-IIe). The GbWRKY genes distributed unevenly on the 26 chromosomes. Analysis of gene duplication events demonstrated that the expansion of GbWRKY genes is mainly due to whole genome duplication and segmental duplication events. 61 GbWRKY genes showed significantly differential expression under *Verticillium dahliae* Kleb. infection, implying that they might play important roles in interaction

收稿日期: 2018-12-10 修回日期: 2019-01-14 网络出版日期: 2019-02-14

URL: <http://doi.org/10.13430/j.cnki.jpgr.20181210004>

第一作者研究方向为生物信息学、作物遗传育种, E-mail: fumingchuan@shandong.cn

通信作者: 柳展基, 研究方向为作物遗传育种, E-mail: liuzhanji@shandong.cn

刘任重, 研究方向为作物遗传育种, E-mail: liurenzh@163.com

基金项目: 农业部黄淮海棉花遗传改良与栽培生理重点实验室开放课题 (2016KL04); 山东省农业科学院农业科技创新工程 (CXGC2018B01, CXGC2018E06)

Foundation project: The Open Project of Key Laboratory of Cotton Breeding and Cultivation in Huang-Huai-Hai Plain, Ministry of Agriculture (2016KL04), The Agricultural Science and Technology Innovation Project of Shandong Academy of Agricultural Sciences (CXGC2018B01, CXGC2018E06)

to *V.dahlia* Kleb. infection. Therefore, this study provided the structure, evolution and expression pattern of *GbWRKY* genes, which will help further characterizing the functions of *GbWRKY* genes in *Gossypium barbadense* L..

Key words: WRKY transcription factor; cotton; Verticillium wilt; bioinformatics

WRKY 转录因子是植物中最重要的蛋白家族之一。WRKY 蛋白家族含有 1 个或 2 个保守的、由 60 个氨基酸组成的结构域,因其 N 端存在一段高度保守的 WRKYGQK 氨基酸序列,因此将其命名为 WRKY 结构域。此外,WRKY 蛋白还包含一锌指结构域。根据 WRKY 结构域数量及锌指结构域类型的不同,通常可将 WRKY 转录因子分为 3 组(I、II、III)。其中,I 组 WRKY 蛋白包含 2 个 WRKY 结构域,II、III 组包含 1 个 WRKY 结构域;I、II 组的锌指结构域为 C₂H₂ 类型(C-X₄₋₅-C-X₂₂₋₂₃-H-X₁-H),III 组为 C₂HC 类型(C-X₇-C-X₂₃-H-X₁-C)。II 组又可进一步分为 IIa~IIe 5 个亚组。WRKY 转录因子可与顺式作用元件 W box (TTGACC/T) 特异性结合,从而发挥转录调控作用。

WRKY 蛋白最早发现于红薯 (*Ipomoea batatas* (L.) Lam.) 中^[1],其后在野燕麦 (*Avena fatua* L.)^[2]和欧芹 (*Petroselinum crispum* (Mill.) Fuss)^[3]中也有相关报道。目前研究表明,WRKY 转录因子参与调控植物中众多基因的表达,其在植物生长发育、新陈代谢、种子萌发/发育/休眠、叶片衰老、抵御外界生物/非生物胁迫等过程中都发挥有重要作用^[4-8]。例如,在拟南芥 (*Arabidopsis thaliana* (L.) Heynh.) 中,AtWRKY38 和 AtWRKY62 在植株抵御病原菌侵染胁迫过程中起调控作用^[9];大豆 (*Glycine max* (L.) Merr.) GmWRKY13、GmWRKY21 和 GmWRKY54 转录因子在植株响应低温、干旱、盐等非生物胁迫过程中发挥有重要作用^[10];OsWRKY42 转录因子可抑制水稻 (*Oryza sativa* L.) 中 *OsMT1d* 基因介导的 ROS 代谢过程,从而促进植株叶片的衰老^[11];近日研究发现,小麦 (*Triticum aestivum* L.) 中 TaWRKY51 转录因子通过抑制乙烯合成途径中 *TaACS2*、*TaACS7* 及 *TaACS8* 基因的表达促进侧根形成,从而影响植株对水分和养分的吸收^[12]。

棉花 (*Gossypium* spp.) 是世界范围内重要的纤维及油料作物之一,其在我国经济发展中占有十分重要的地位。研究表明,WRKY 转录因子在调控棉花纤维发育、响应逆境胁迫等过程中也发挥有重要作用^[13-16]。亚洲棉 (*Gossypium arboreum* L.) GaWRKY1 转录因子参与调控棉酚合成途径,其可激活 *CAD1-A* 基因的表达,从而促进棉花中倍半萜

烯的生物合成^[13];陆地棉 (*Gossypium hirsutum* L.) GhWRKY3 转录因子在植物激素诱导、受伤胁迫、病原菌侵染等处理条件下表达水平发生显著变化,表明其在生物/非生物胁迫应答过程中发挥重要作用^[14];Zhang 等^[17]分析了陆地棉在黄萎病菌侵染 24 h 后的转录组数据,结果共发现 4794 个差异表达基因,其中 GhWRKY2 及 GhWRKY29 为棉花抵御黄萎病菌侵染的正调控因子,而 GhWRKY13 为负调控因子;陆地棉 *GhWRKY42* 基因受茉莉酸甲酯 (MeJA, methyl jasmonate)、脱落酸 (ABA, abscisic acid)、干旱、盐等胁迫诱导表达,且其表达水平与植株高度和叶片衰老程度相关,表明 GhWRKY42 蛋白参与调控棉花非生物胁迫应答及植株衰老、茎部生长等过程^[18]。

近年来,随着棉花基因组测序工作地陆续完成,为我们从全基因组水平研究棉花 WRKY 转录因子提供了基础。Cai 等^[19]在雷蒙德氏棉 (*Gossypium raimondii* Ulbr.) 基因组中鉴定出 120 个 WRKY 转录因子,其中分别有 12 个、14 个、16 个 WRKY 基因在盐、干旱及黄萎病菌胁迫条件下出现显著表达变化,表明其在棉花生长发育和抗逆过程中发挥调控作用。Ding 等^[20]分别在雷蒙德氏棉和亚洲棉中鉴定出 112 个和 109 个 WRKY 基因,转录组分析表明,WRKY 转录因子在棉纤维发育起始、伸长、成熟等各个生长阶段都发挥重要调控作用;可变剪切分析发现,进化树中 I 组 WRKY 基因存在最多可变剪切事件。

黄萎病是一种影响棉花高产优质的重要限制因素^[21-22]。与广泛种植的陆地棉相比,海岛棉 (*Gossypium barbadense* L.) 在抵御黄萎病菌 (*Verticillium dahliae* Kleb.) 侵染方面具有明显优势,因此从分子水平研究海岛棉的抗病机制、挖掘其抗病基因,对于棉花育种工作具有重要意义。本研究对海岛棉基因组中的 WRKY 转录因子进行了系统鉴定,并对其物理化学特性、系统进化、基因结构、染色体定位、基因重复事件等进行了详细分析。通过分析高抗海岛棉品种海 7124 在黄萎病菌侵染不同时间点下的 RNA-Seq 数据,研究了 WRKY 转录因子在棉花响应黄萎病菌侵染过程中的表达模式。该研究可帮助我们进一步了解海岛棉 WRKY 蛋白家

族的特征及功能,从而为将来选育高抗优质棉花新品种提供理论基础。

1 材料与方法

1.1 海岛棉 WRKY 转录因子 (GbWRKY) 的鉴定

使用 HMM 方法查找海岛棉基因组 (<http://database.chgc.sh.cn/cotton/index.htm>) 中的 WRKY 转录因子。首先,在 Pfam 数据库 (<https://pfam.xfam.org/>) 中下载 WRKY 结构域种子文件 (PF03106),然后利用 HMMER3 (<http://www.hmmerr.org/>) 中的 hmmsearch 程序查找海岛棉蛋白质组中的 WRKY 转录因子 (evalue=1e-10)。结果共找到 203 个备选 WRKY 蛋白。之后利用 SMART (<http://smart.embl-heidelberg.de/>) 和 InterProScan (<http://www.ebi.ac.uk/interpro/>) 对上述结果进行进一步验证,最终确定了 180 个 WRKY 蛋白用于之后的分析。

利用在线工具 ExPASy (http://web.expasy.org/compute_pi/) 对 WRKY 转录因子的分子量、等电点等物理化学特性进行分析,并使用 WoLF PSORT (<http://www.genscript.com/wolf-psort.html>) 进行亚细胞定位预测。此外,使用 WebLogo 在线网站 (<http://weblogo.berkeley.edu/logo.cgi>)^[23] 对 WRKY 结构域的氨基酸组成偏好进行分析。

1.2 GbWRKY 转录因子的系统进化分析

使用 ClustalX (默认参数) 对 GbWRKY 转录因子进行多序列比对,并使用 MUSCLE 软件验证比对结果。利用 MEGA 7 构建邻接树 (neighbor-joining tree), 参数设置如下: Possion model; pairwise deletion; 1000 bootstraps。

1.3 GbWRKY 转录因子的基因结构分析

利用在线工具 Gene Structure Display Server (<http://gsds.cbi.pku.edu.cn/>)^[24] 对 GbWRKY 基因的外显子/内含子构成进行分析。使用 MEME 软件包 (<http://meme-suite.org/>) 进行 motif 分析, 参数设置如下: maximum numbers of different motifs to find=10; minimum motif width=6; maximum motif width=300; evalue=1e-10。

1.4 GbWRKY 转录因子的染色体定位及基因重复事件

利用海岛棉 GFF3 基因组注释文件对 GbWRKY 基因进行染色体定位分析,并通过 MapInspect 软件绘制染色体定位图。根据染色体定位,将 GbWRKY 转录因子命名为 GbWRKY001~GbWRKY180。

利用本地 BLASTP (evalue=1e-10) 对 GbWRKY

转录因子进行基因重复事件分析,将序列一致性大于 60%,且比对到的序列长度占较长序列长度 80% 以上的比对结果选为同源基因对。结合染色体定位,将同源基因对分为串联重复 (tandem duplication) 和区段重复 (segmental duplication) 两类,并利用 Circos 软件^[25] 绘图。此外,利用 PAML (<http://abacus.gene.ucl.ac.uk/software/paml.html>) 中的 codeml 程序计算同源基因对的 Ka/Ks 值,其中 Ka/Ks<1 代表负向选择, Ka/Ks>1 代表正向选择。

1.5 GbWRKY 转录因子在黄萎病菌侵染条件下的表达模式分析

以 Chen 等^[26] 的 RNA-Seq 数据 (PRJNA234454) 为基础,分析了黄萎病菌侵染条件下 GbWRKY 转录因子的表达模式。本试验以高抗黄萎病品种海 7124 为材料,使用浸根法侵染生长 2 周的幼苗根部 10 min,然后分别选取侵染后 2 h、6 h、12 h、24 h、48 h、72 h 后的幼苗根部为材料,提取总 RNA 进行 RNA-Seq 分析 (另使用未被侵染的材料为对照)。使用 HISAT2^[27] 和 featureCounts^[28] 程序对 RNA-Seq 原始数据进行分析处理,然后利用 R 中 DESeq 软件包分析差异表达基因 ($P<0.001$)。通过 R 及 ClustVis (<http://biit.cs.ut.ee/clustvis/>) 绘制热图。

1.6 RNA 提取及 qRT-PCR 分析

以高抗黄萎病品种海 7124 为材料,选取长势一致的 2 周幼苗,将其从花盆中小心取出,用无菌蒸馏水将根部冲洗干净,从根尖剪去 1 cm,然后用浓度为 1×10^7 个/mL 的落叶型大丽轮枝菌 Vd8 孢子悬液侵染 10 min。分别选取 2 h、6 h、12 h、24 h、48 h、72 h 时间点的侵染和未被侵染 (CK) 幼苗根系为材料,利用 Trizol 法提取总 RNA (Aidlab 公司 EASYspin Plus Complex Plant RNA Kit 试剂盒),反转录为 cDNA (Aidlab 公司 TRUEscript RT Kit 试剂盒)。

选取部分基因,利用 qPCR 分析它们在黄萎病菌侵染下的表达特征。qPCR 使用 Thermo Fisher Scientific 的 QuantStudio 5 实时荧光定量 PCR 系统,试剂盒选用 Aidlab 公司的 2× Sybr Green qPCR Mix, 荧光染料为 SYBR Green,以棉花 β -actin 为内参基因。反应程序: 95 °C 预变性 3 min; 95 °C 15 s, 60 °C 15 s, 40 个循环; 熔解曲线程序: 95 °C 15 s, 60 °C 1 min, 95 °C 15 s。反应体系为: 2× SYBR qPCR Mix 10 μ L; DNA Template 0.8 μ L; Forward Primer 0.4 μ L; Reverse Primer 0.4 μ L; ddH₂O 8.4 μ L。设置 3 次生物学重复,基因相对表达量使用 $2^{-\Delta\Delta Ct}$ 法进行分析^[29]。qPCR 所用引物见表 1。

表 1 qRT-PCR 所用引物
Table 1 Sequences of the primers used in qRT-PCR

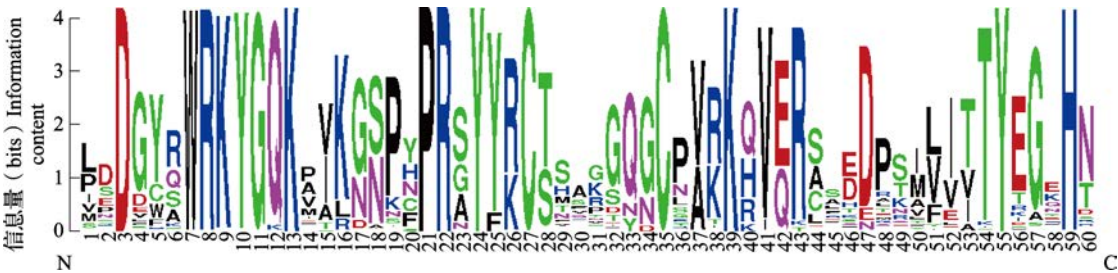
基因 Gene	正向引物 Forward primer	反向引物 Reverse primer
<i>GbWRKY002</i>	GGTGGGGCTGAGAACAAAACCT	ACATTGCATCCTTGATGCGTGC
<i>GbWRKY003</i>	TACTACTACTGCCGTCGCCGAT	TCCGACCATCGGAGGTGATGTT
<i>GbWRKY020</i>	TGGGCGCTTTGTTGGGTCTAAA	TTGCCCATATTTCTCCAGCGA
<i>GbWRKY038</i>	GGGTCACTGCTACAACGATTTGC	ACAGCTGGTCTCTGTGTTTCCA
<i>GbWRKY075</i>	TCGTCGGGTTCCAATGTTGCAT	GGTGTGATAAGGCAGCCGAGAG
<i>GbWRKY090</i>	TCTGCTTCACCCAACCCTAGAAGA	CTCCAAGCCCATACATCCGACG
<i>GbWRKY096</i>	GCAGCTTATGATCCCTTGACAACA	TTCAGACCCAACAAAGCTCCCA
<i>GbWRKY121</i>	CTCTTTTGCCCCAAGCTGTCCA	ACTTCACTTGGAGAAGGGCGTG
<i>GbWRKY146</i>	TCTTTACCATCGAACATGGCGCT	TCCGACGACAACATCCCATTGTA
<i>GbWRKY179</i>	CTTCACTCTCGGCTGCCTCATC	GCCCCAAGACACAGTGTGGAT
<i>β-actin</i>	GATTCCGTTGTCCAGAAGTCCT	TACGGTCTGCAATACCAGGGA

2 结果与分析

2.1 海岛棉 WRKY 转录因子 (GbWRKY) 的鉴定

首先使用 HMM 方法在海岛棉基因组中找到 203 个备选 WRKY 蛋白, 然后通过 SMART 和 InterProScan 进行进一步验证, 最终共确定了 180

个 WRKY 转录因子 (GbWRKY001~GbWRKY180) 用于之后的分析。位点组成偏好性分析表明, 海岛棉 WRKY 结构域的长度为 60 aa, 其中 7~13 位点的 W、R、K、Y、G、Q、K 高度保守; 此外, 还有其他一些位点的氨基酸组成也存在明显偏好 (图 1)。



每个位点的得分代表了该位点的信息量, 氨基酸字母的高度代表其在相应位点中的相对频率
The bit score indicates the information content for each position in the sequence. The height of symbols indicates the relative frequency of each amino acid at the corresponding position

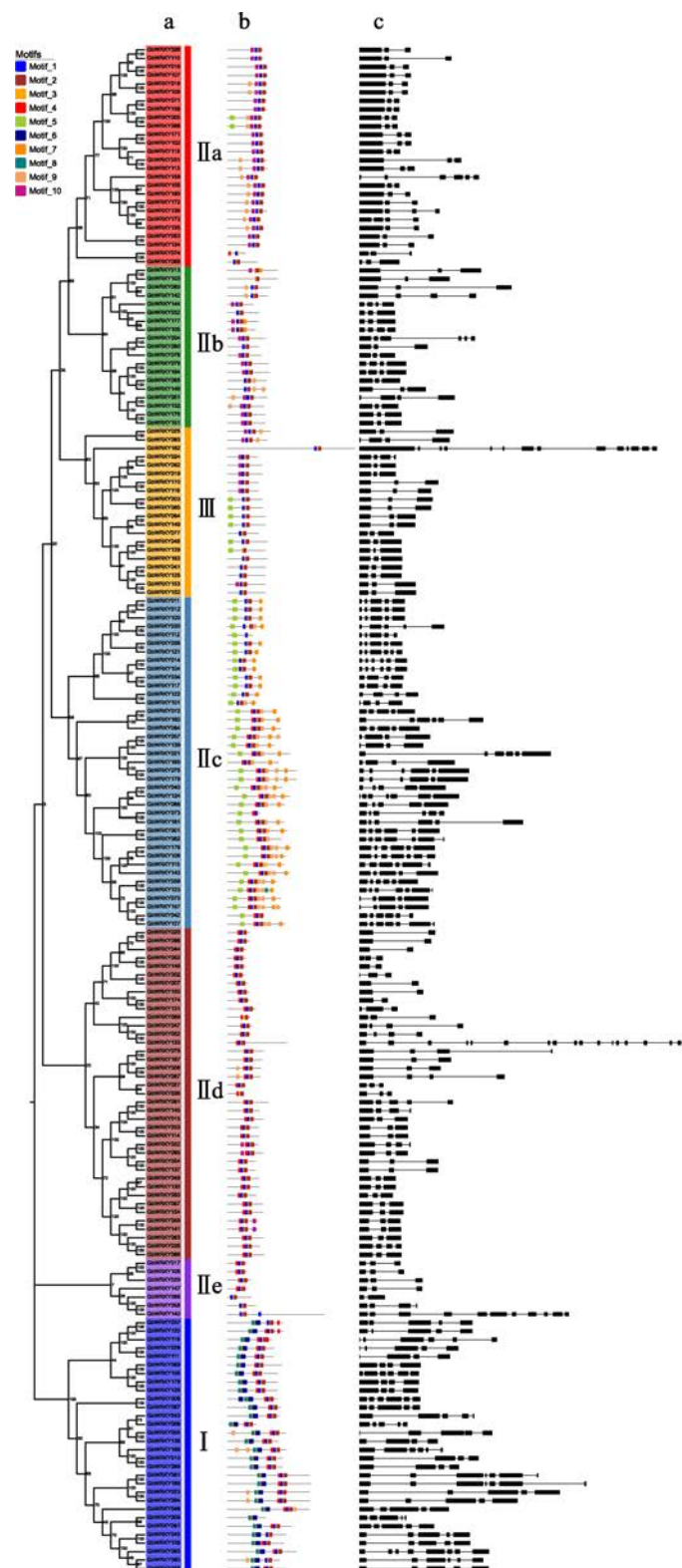
图 1 WRKY 结构域的位点保守性分析
Fig.1 Sequence conservation of the WRKY domain

物理化学特性分析表明, GbWRKY 转录因子的序列长度为 146~1131 aa; 理论 pI 为 4.59~10.06; 分子量为 16.86~128.62 kD。亚细胞定位分析发现, 94% 的 GbWRKY 蛋白定位于细胞核内, 符合转录因子通常为核蛋白的基本特征。

2.2 GbWRKY 转录因子的系统进化分析

使用邻接法 (NJ, neighbor-joining) 构建系统发育树来研究 GbWRKY 之间的进化关系。根据 WRKY 结构域数量及锌指结构域的不同, 可将

GbWRKY 蛋白分为 I、II、III 3 组。其中, I 组包含 2 个 WRKY 结构域, II、III 组包含 1 个 WRKY 结构域; I、II 组的锌指结构域为 C₂H₂ 类型, III 组为 C₂HC 类型。需要注意的是 GbWRKY009 蛋白, 虽然其只包含 1 个 WRKY 结构域, 但在进化树中与 I 组成员关系更近。在 3 个大组中, II 组包含的成员数量最多 (130 个), 根据基因结构及进化关系, 还可将其进一步分为 IIa~IIe 5 个亚组 (图 2a)。



a: GbWRKY 转录因子的系统进化树; 进化树由 MEGA 7 利用邻接法构建, bootstrap 值为 1000; 3 个大组 (I-III) 及 5 个亚组 (IIa-IIe) 由不同背景的颜色标示出来; b: *GbWRKY* 基因的 motif 构成; 不同 motif 由不同颜色方块表示; c: *GbWRKY* 基因的外显子 / 内含子构成; 黑色方块和实线分别代表外显子和内含子区域

a: The phylogenetic tree was constructed with MEGA 7 using the neighbor-joining (NJ) method with 1000 bootstrap replicates based on a multiple alignment of 180 amino acid sequences of *GbWRKY* genes from *G.barbadense*. The three major groups (I-III) and five subgroups (IIa-IIe) are marked with colorful backgrounds, b: Schematic diagram of the conserved motifs in the *GbWRKY* proteins. Each motif is represented by a number in the colored box, c: Exon/intron structures of the *GbWRKY* genes. Exons and introns are represented by boxes and black lines, respectively

图 2 GbWRKY 转录因子的系统进化关系及基因结构特征

Fig.2 Phylogenetic relationship and gene structures of the GbWRKY transcription factors

2.3 GbWRKY 转录因子的基因结构

通过使用 GSDS 在线软件分析 *GbWRKY* 基因的外显子/内含子构成发现, *GbWRKY* 基因的外显子数量为 3~18 个, 其中 3-2 (3 个外显子、2 个内含子) 的结构类型数量最多 (79 个)。结合系统发育树分析发现, 同一组内的基因通常具有相似的基因结构特征, 例如: I 组中有一半的基因为 5-4 类型, IIc 亚组主要由 5-4、6-5 类型组成, 而 III 组基本均为 3-2 类型 (图 2c)。

使用 MEME 软件分析了 *GbWRKY* 转录因子的 motif 构成特征。与外显子/内含子构成相类似, 系统发育树中同一组内的基因通常具有相似的 motif 构成特征。例如, I 组基因通常为 8-3-6-4-1-3-2 的构成模式 (数字代表相应的 motif), IIa 亚组中 (9)-10-1-3-2 为主要形式, 而 III 组基因的 motif 构成主要为 (5/10)-1-3-2 (图 2b)。

2.4 GbWRKY 基因的染色体定位及基因重复事件

根据海岛棉基因组注释文件, 分析了 *GbWRKY* 基因的染色体分布特征。其中, 一共有 169 个 *GbWRKY* 基因 (*GbWRKY001*~*GbWRKY169*) 可定位于 26 条染色体上, 另有 11 个 *GbWRKY* (*GbWRKY170*~*GbWRKY180*) 定位于 scaffolds 上。结果表明, *GbWRKY* 基因在染色体上分布不均匀, 其倾向定位于染色体的两端区域, 而中间区域的基因密度相对较低。在 26 条染色体中, 定位于 A05 上的基因最多 (15 个), 其次为 D11 (12 个), 而位于 A01、A03 上的 *GbWRKY* 基因最少 (2 个) (图 3)。

基因重复事件是新基因产生和基因功能进化的重要来源。基因重复事件可分为两类: 重复基因位于同一染色体上且它们之间没有其他基因的重复事件称为串联重复 (tandem duplication); 重复基因位于不同染色体、或同一染色体上相距较远位置的重复事件称为区段重复 (segmental duplication)。本研究分析了海岛棉 *GbWRKY* 基因的重复事件, 结果共发现 69 对重复基因, 其中有 57 对可定位到染色体上 (图 4)。在 57 对重复基因中, 有 5 对为串联重复 (*GbWRKY011*/*GbWRKY012*、*GbWRKY049*/*GbWRKY050*、*GbWRKY119*/*GbWRKY120*、*GbWRKY152*/*GbWRKY153*、*GbWRKY172*/*GbWRKY173*), 而其余 52 对均为区段重复, 表明区段重复事件是海岛棉 *GbWRKY* 基因数量增多的一个主要途径。

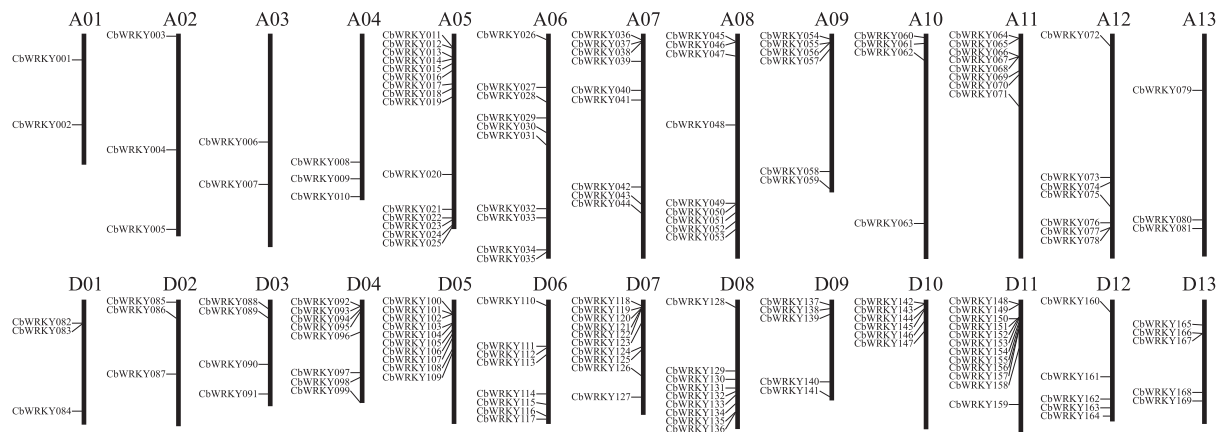
为进一步分析 *GbWRKY* 基因在进化过程中承受的选择压力, 计算了重复基因对的 Ka/Ks 值。结果表明, 绝大多数基因对的 Ka/Ks<1, 表明它们受负选择压力影响。此外, 有 11 对基因的 Ka/Ks>1, 说明它们可能经历了正向选择过程。

2.5 GbWRKY 基因响应黄萎病菌侵染的表达模式

通过分析高抗黄萎病海岛棉品种海 7124 在黄萎病菌侵染不同时间点下 (0 h、2 h、6 h、12 h、24 h、48 h、72 h) 的 RNA-Seq 数据, 研究了 *GbWRKY* 基因响应黄萎病菌侵染的表达特征。结果发现, 共有 61 个 *GbWRKY* 基因在黄萎病菌侵染不同时间点下出现显著表达变化 ($P<0.001$), 其中显著上调和下调的基因分别为 30 个和 31 个。差异表达基因最多的时间点为侵染后 24 h (46 个), 其次为侵染后 48 h (43 个), 表明 *GbWRKY* 转录因子主要在黄萎病菌侵染后期 (24~72 h) 发挥作用。有 13 个基因 (7 个上调表达、6 个下调表达) 只在侵染后某一个时间点出现显著表达变化, 表明它们可能在相应时间点发挥有特殊作用; 相反, 有 12 个基因 (8 个上调表达、4 个下调表达) 的表达水平在所有时间点都表现出显著变化, 暗示它们可能在整个侵染过程中都发挥有调控作用。在所有差异表达基因中, 24 h 下的 *GbWRKY020* 基因表现出最显著的上调表达 ($\log_2\text{Ratio}=8.86$, $P<0.001$), 而 48 h 下的 *GbWRKY038* 表现出最显著下调表达 ($\log_2\text{Ratio}=-9.42$, $P<0.001$)。

结合系统发育树分析发现, 差异表达基因不均匀分布在 3 个组中, 同一组内的差异基因通常表现出相似的变化趋势, 如: III 组中除 *GbWRKY019* 外其他所有基因均表现为显著下调; 在 IIb 亚组中, 几乎所有差异表达基因 (除 *GbWRKY051*) 均为显著上调。相反, 有些组中的差异表达基因表现出不同的表达模式, 如: 在 IIc 亚组 18 个差异表达基因中, 有 12 个表现为显著上调, 而另外 6 个为显著下调。另外, 通过比较重复基因对之间的表达模式发现, 同源基因之间通常具有相似的表达特征。

通过绘制热图, 进一步分析了 *GbWRKY* 基因在不同侵染时间点的表达特征 (图 5)。如图 5 所示, 根据表达模式不同, 处理后的 6 个时间点聚集在一起, 未处理组单独为一个分支。此外, 6 个时间点又可进一步划分为 2 组, 分别代表了黄萎病菌侵染前期 (2~12 h) 和侵染后期 (24~72 h)。系统树中关系较近的基因通常具有相似的表达模

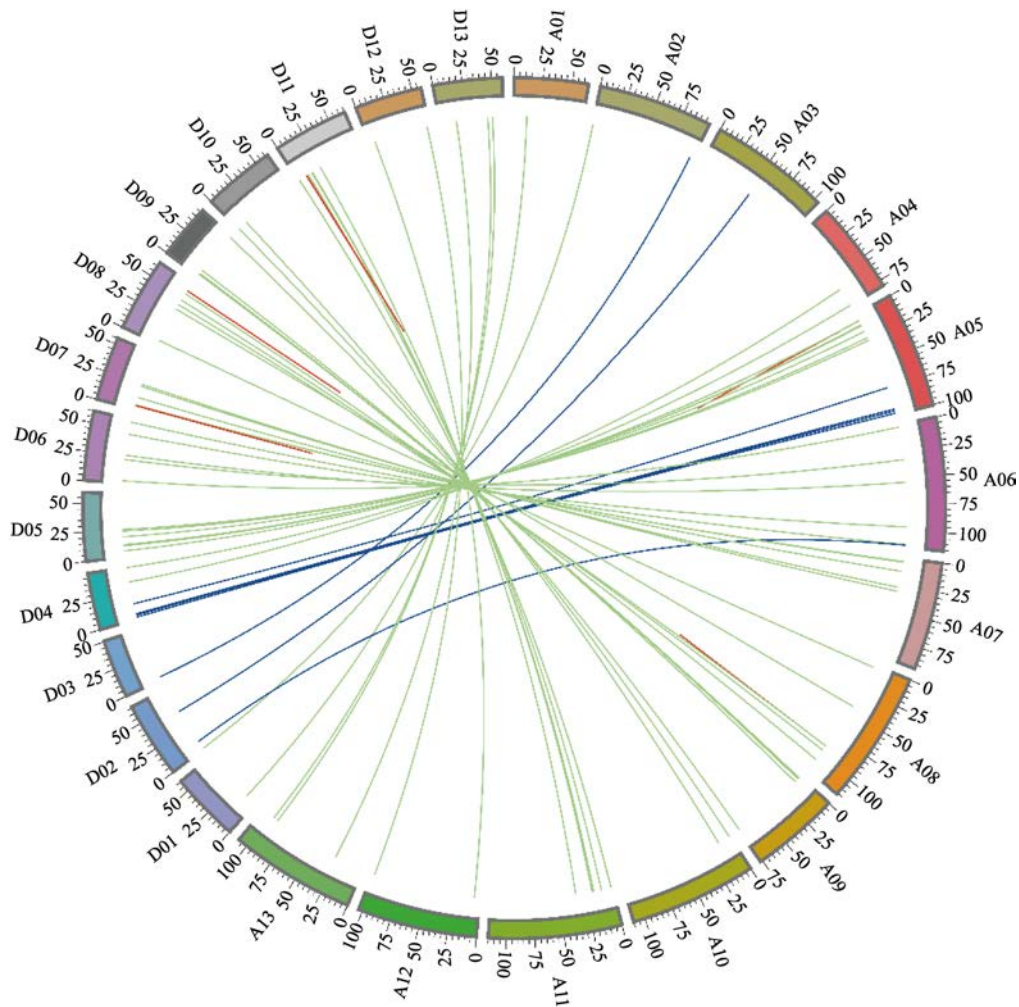


GbWRKY 基因的染色体位置根据海岛棉基因组注释文件得到。染色体序号标注于每条染色体上方

The chromosomal position of each *GbWRKY* gene was mapped according to the *G.barbadense* L. genome annotation file. The chromosome number is indicated at the top of each chromosome

图 3 *GbWRKY* 基因的染色体定位

Fig.3 Chromosomal distribution of the *GbWRKY* genes

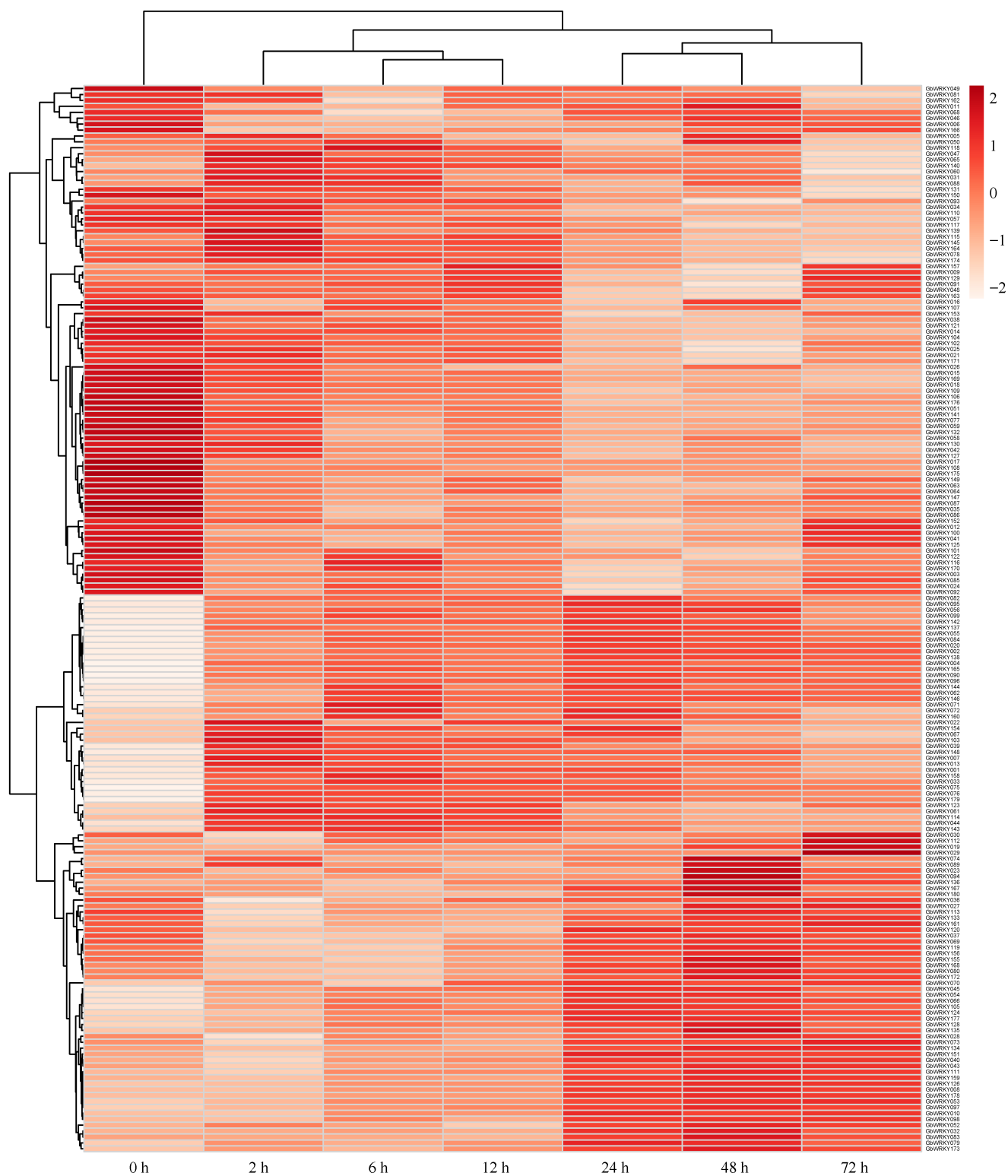


曲线代表海岛棉基因组中的 *GbWRKY* 同源基因对; 红线代表串联重复基因; 绿线代表 A、D 基因组相应染色体之间的区段重复基因 (如 A02/D02 之间的重复基因); 蓝线代表其他区段重复基因

The lines represent the *GbWRKY* homologous gene pairs of *G.barbadense*. The red lines indicate tandem duplication, the green lines indicate segmental duplication between A genome and the corresponding D genome (e.g. the gene pairs between A02 and D02), and the blue lines indicate other gene pairs

图 4 *GbWRKY* 的基因重复事件

Fig.4 The duplication events of the *GbWRKY* genes



GbWRKY 基因在黄萎病菌感染不同时间点下 (0 h、2 h、6 h、12 h、24 h、48 h、72 h) 的表达热图。表达水平使用不同深浅的红色表示。RNA-Seq 数据来源于 NCBI 数据库 (PRJNA234454), 原始数据经标准化后用于绘制热图

The heat map shows the expression patterns of *GbWRKY* genes under different time points (0 h, 2 h, 6 h, 12 h, 24 h, 48 h, 72 h) after *V. dahliae* infection. Expression levels are indicated by sequential red color. RNA-Seq data was obtained from the NCBI database (PRJNA234454). The

normalized RNA-Seq data was used to generate this heat map

图 5 *GbWRKY* 基因的表达热图

Fig.5 Expression profiles of the *GbWRKY* genes

式, 如: *GbWRKY037/GbWRKY120*、*GbWRKY020/GbWRKY096*; 但有时也会表现出差异较大的表达特征, 如: *GbWRKY029/GbWRKY058*、*GbWRKY038/GbWRKY112*。

为进一步验证 *GbWRKY* 基因响应黄萎病菌侵染的表达模式, 选取了 10 个基因进行 qRT-PCR 分析 (图 6)。与转录组数据相类似, *GbWRKY002*、*GbWRKY020*、*GbWRKY075* 等 7 个基因在黄萎病菌侵染后均发生上调表达; 而 *GbWRKY003*、*GbWRKY038*、*GbWRKY121* 出现下调表达。其中, *GbWRKY020* 上调最为明显, 其在侵染后 6 h 和 24 h 的表达量分别为对照的 7.20 倍和 7.15 倍; 而 *GbWRKY003* 下调最为明显, 其在 72 h 的表达量为对照的 0.11 倍。

3 讨论

近年来, 随着植物基因组测序工作的陆续完成, 为我们从全基因组水平研究 WRKY 转录因子提供了基础^[6, 30-34]。棉花作为世界范围内重要的纤维及油料作物, 对其 WRKY 转录因子的研究也有相关报道^[19-20, 35]。目前, 关于海岛棉全基因组 WRKY 转录因子的研究还相对较少。本研究以海岛棉为材料, 利用生物信息学方法在其基因组中共鉴定出 180 个 *GbWRKY* 转录因子, 数量多于雷蒙德氏棉 (112 个) 和亚洲棉 (109 个)^[20]。分析前人报道发现, WRKY 转录因子在高等植物基因组中所占比例大致相同, 其数量随基因组增大而增多。雷蒙德氏棉/亚洲棉与海岛棉之间发生了二倍体到四倍体的染色体加倍事件, 结合基因重复事件研究结果, 我们认为海岛棉中 WRKY 转录因子数量的增多主要源于全基因组复制及区段重复事件。

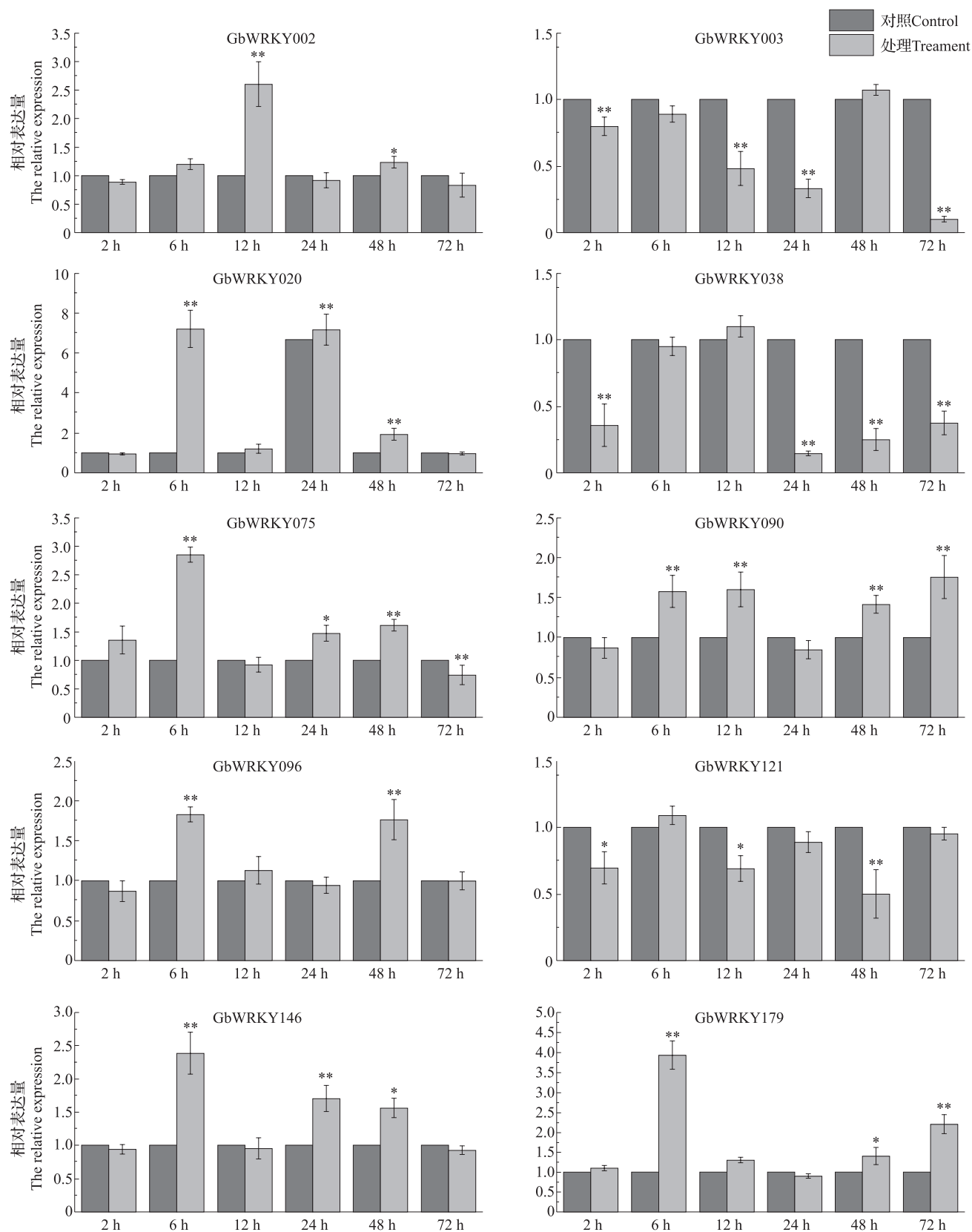
根据系统进化关系, 可将 *GbWRKY* 蛋白分为 I、II、III 3 组, II 组又可进一步分为 IIa~IIe 5 个亚组。需要注意的是 *GbWRKY009* 转录因子, 其只包含 1 个 WRKY 结构域, 但根据系统进化关系被分到 I 组内。通过分析 *GbWRKY009* 在拟南芥 (*AT2G38470.1*)、亚洲棉 (*Ga04G0565.1*)、雷蒙德氏棉 (*Gorai.012G119600.1*) 和陆地棉 (*Gh_D04G1318.1*) 中的直系同源蛋白发现, 它们均含有 2 个 WRKY 结构域。因此, 推测该基因可能在海岛棉进化过程中发生了片段缺失, 也可能是由于测序

和拼接错误导致的序列不完整。具体原因以及序列缺失是否会影响该基因的功能, 还需后续开展进一步的研究。

WRKY 蛋白作为植物中最大的转录因子家族之一, 参与调控众多生物过程, 特别在植物抵御外界生物/非生物胁迫过程中发挥重要作用^[36-37]。本研究分析了高抗海岛棉品种海 7124 在黄萎病菌侵染不同时间点下的 RNA-Seq 数据, 结果共找到 61 个差异表达 *GbWRKY* 基因, 表明 *GbWRKY* 转录因子参与调控黄萎病菌胁迫应答过程。前人研究表明, I 组 WRKY 基因起源最早, II、III 组 WRKY 基因是随着物种进化及基因复制过程扩大而来, 从而帮助高等植物增强对环境的适应性^[38]。本研究中 61 个差异表达基因主要分布在 II (75%)、III (20%) 组, I 组内最少 (5%), 表明 II、III 组 WRKY 基因更多参与植株对外界生物胁迫应答过程。该结果与 Dou 等^[35]对陆地棉 *GhWRKY* 转录因子响应非生物胁迫下的表达特征相类似, 进一步表明 II、III 组 WRKY 转录因子在植株抵御生物/非生物胁迫过程中发挥有更为重要的作用。

WRKY 转录因子可通过相互作用来调控其行使功能^[5]。通过分析海岛棉 *GbWRKY* 基因上游 1.5 kb 启动子区发现, 超过一半基因 (56%) 至少含有一个 W box 顺式作用元件, 表明 *GbWRKY* 基因家族内存在着自我调控机制。之前研究报道, 拟南芥中防御相关 *AtWRKY* 基因启动子区相比其他 *AtWRKY* 基因含有更多 W box 顺式作用元件, 从而通过家族内基因相互作用来激活/抑制其功能^[36]。本研究中的 61 个差异表达 *GbWRKY* 基因 (平均 0.6 个) 相比其他基因 (平均 1.0 个) 的 W box 数量没有明显差异, 表明海岛棉 *GbWRKY* 转录因子抗逆相关调控网络和调控机制与拟南芥之间可能存在着区别。

综上, WRKY 转录因子在进化过程中发展出不同的结构特征和生化特性, 从而使其具有多种不同的生物学功能。表达分析表明, 海岛棉 *GbWRKY* 基因在植株抵御黄萎病菌侵染过程中发挥作用, 下一步通过深入研究这些基因调控的下游基因, 构建其调控网络, 可帮助我们进一步了解 *GbWRKY* 基因的功能和抗病机制, 从而为将来棉花抗病育种工作提供有用的基因资源和理论指导。

图 6 10 个 *GbWRKY* 基因的 qRT-PCR 分析Fig.6 Expression analysis of 10 selected *GbWRKY* genes using qRT-PCR

参考文献

- [1] Ishiguro S, Nakamura K. Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and β -amylase from sweet potato. *Molecular and General Genetics*, 1994, 244(6): 563-571
- [2] Rushton P J, Macdonald H, Huttly A K, Lazarus C M, Hooley R. Members of a new family of DNA-binding proteins bind to a conserved *cis*-element in the promoters of α -Amy2 genes. *Plant Molecular Biology*, 1995, 29(4): 691-702
- [3] Rushton P J, Torres J T, Parniske M, Wernert P, Hahlbrock K, Somssich I E. Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley *PR1* genes. *The EMBO Journal*, 1996, 15(20): 5690-5700
- [4] Eulgem T, Rushton P J, Robatzek S, Somssich I E. The WRKY superfamily of plant transcription factors. *Trends in Plant Science*, 2000, 5(5): 199-206
- [5] Rushton P J, Somssich I E, Ringler P, Shen Q J. WRKY transcription factors. *Trends in Plant Science*, 2010, 15(5): 247-258
- [6] Ülker B, Somssich I E. WRKY transcription factors: from DNA binding towards biological function. *Current Opinion in Plant Biology*, 2004, 7(5): 491-498
- [7] Chen L, Song Y, Li S, Zhang L, Zou C, Yu D. The role of WRKY transcription factors in plant abiotic stresses. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 2012, 1819(2): 120-128
- [8] 王磊, 高晓清, 朱冬华, 周永力, 黎志康. 植物 WRKY 转录因子家族基因抗病相关功能的研究进展. *植物遗传资源学报*, 2011, 12(1): 80-85
Wang L, Gao X Q, Zhu L H, Zhou Y L, Li Z K. Advances in research on function of WRKY transcription factor genes in plant resistance. *Journal of Plant Genetic Resources*, 2011, 12(1): 80-85
- [9] Kim K C, Lai Z, Fan B, Chen Z. *Arabidopsis* WRKY38 and WRKY62 transcription factors interact with histone deacetylase 19 in basal defense. *The Plant Cell*, 2008, 20(9): 2357-2371
- [10] Zhou Q Y, Tian A G, Zou H F, Xie Z M, Lei G, Huang J, Wang C M, Wang H W, Zhang J S, Chen S Y. Soybean WRKY-type transcription factor genes, *GmWRKY13*, *GmWRKY21*, and *GmWRKY54*, confer differential tolerance to abiotic stresses in transgenic *Arabidopsis* plants. *Plant Biotechnology Journal*, 2008, 6(5): 486-503
- [11] Han M, Kim C Y, Lee J, Lee S K, Jeon J S. *OsWRKY42* represses *OsMT1d* and induces reactive oxygen species and leaf senescence in rice. *Molecules and Cells*, 2014, 37(7): 532-539
- [12] Hu Z, Wang R, Zheng M, Liu X, Meng F, Wu H, Yao Y, Xin M, Peng H, Ni Z, Sun Q. TaWRKY51 promotes lateral root formation through negative regulation of ethylene biosynthesis in wheat (*Triticum aestivum* L.). *The Plant Journal*, 2018, 96(2): 372-388
- [13] Xu Y H, Wang J W, Wang S, Wang J Y, Chen X Y. Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)- δ -cadinene synthase-A. *Plant Physiology*, 2004, 135(1): 507-515
- [14] Guo R, Yu F, Gao Z, An H, Cao X, Guo X. *GhWRKY3*, a novel cotton (*Gossypium hirsutum* L.) WRKY gene, is involved in diverse stress responses. *Molecular Biology Reports*, 2011, 38(1): 49-58
- [15] Yan H, Jia H, Chen X, Hao L, An H, Guo X. The cotton WRKY transcription factor GhWRKY17 functions in drought and salt stress in transgenic *Nicotiana benthamiana* through ABA signaling and the modulation of reactive oxygen species production. *Plant and Cell Physiology*, 2014, 55(12): 2060-2076
- [16] Shi W, Liu D, Hao L, Wu C, Guo X, Li H. GhWRKY39, a member of the WRKY transcription factor family in cotton, has a positive role in disease resistance and salt stress tolerance. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 2014, 118(1): 17-32
- [17] Zhang W, Zhang H, Liu K, Jian G, Qi F, Si N. Large-scale identification of *Gossypium hirsutum* genes associated with *Verticillium dahliae* by comparative transcriptomic and reverse genetics analysis. *PLoS One*, 2017, 12(8): e0181609
- [18] Gu L, Wei H, Wang H, Su J, Yu S. Characterization and functional analysis of *GhWRKY42*, a group IId WRKY gene, in upland cotton (*Gossypium hirsutum* L.). *BMC Genetics*, 2018, 19(1): 48
- [19] Cai C, Niu E, Du H, Zhao L, Feng Y, Guo W. Genome-wide analysis of the WRKY transcription factor gene family in *Gossypium raimondii* and the expression of orthologs in cultivated tetraploid cotton. *The Crop Journal*, 2014, 2(2): 87-101
- [20] Ding M, Chen J, Jiang Y, Lin L, Cao Y, Wang M, Zhang Y, Rong J, Ye W. Genome-wide investigation and transcriptome analysis of the WRKY gene family in *Gossypium*. *Molecular Genetics and Genomics*, 2015, 290(1): 151-171
- [21] 林玲, 张昕, 邓晟. 棉花黄萎病研究进展. *棉花学报*, 2014, 26(3): 260-267
Lin L, Zhang X, Deng S. Research advances in cotton *Verticillium* wilt. *Cotton Science*, 2014, 26(3): 260-267
- [22] 朱荷琴, 李志芳, 冯自力, 冯鸿杰, 魏锋, 赵丽红, 师勇强, 刘世超, 周京龙. 我国棉花黄萎病研究十年回顾及展望. *棉花学报*, 2017, 29(S): 37-50
Zhu H Q, Li Z F, Feng Z L, Feng H J, Wei F, Zhao L H, Shi Y Q, Liu S C, Zhou J L. Overview of cotton *Verticillium* wilt research over the past decade in China and its prospect in future. *Cotton Science*, 2017, 29(supplement): 37-50
- [23] Crooks G E, Hon G, Chandonia J M, Brenner S E. WebLogo: A sequence logo generator. *Genome Research*, 2004, 14(6): 1188-1190
- [24] Hu B, Jin J, Guo A Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, 2015, 31(8): 1296-1297
- [25] Krzywinski M I, Schein J E, Birol I, Connors J, Gascoyne R, Horsman D, Jones S J, Marra M A. Circos: An information aesthetic for comparative genomics. *Genome Research*, 2009, 19(9): 1639-1645
- [26] Chen J, Huang J, Li N, Ma X, Wang J, Liu C, Liu Y, Liang Y, Bao Y, Dai X. Genome-wide analysis of the gene families of resistance gene analogues in cotton and their response to *Verticillium* wilt. *BMC Plant Biology*, 2015, 15(1): 148
- [27] Kim D, Langmead B, Salzberg S L. HISAT: a fast spliced

- aligner with low memory requirements. *Nature Methods*, 2015, 12: 357
- [28] Liao Y, Smyth G K, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 2014, 30 (7): 923-930
- [29] Livak K J, Schmittgen T D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, 2001, 25 (4): 402-408
- [30] Ross C A, Liu Y, Shen Q J. The WRKY gene family in rice (*Oryza sativa*). *Journal of Integrative Plant Biology*, 2007, 49 (6): 827-842
- [31] Huang S, Gao Y, Liu J, Peng X, Niu X, Fei Z, Cao S, Liu Y. Genome-wide analysis of WRKY transcription factors in *Solanum lycopersicum*. *Molecular Genetics and Genomics*, 2012, 287 (6): 495-513
- [32] Wei K F, Chen J, Chen Y F, Wu L J, Xie D X. Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. *DNA Research*, 2012, 19 (2): 153-164
- [33] Xie T, Chen C, Li C, Liu J, Liu C, He Y. Genome-wide investigation of WRKY gene family in pineapple: evolution and expression profiles during development and stress. *BMC Genomics*, 2018, 19 (1): 490
- [34] Jue D, Sang X, Liu L, Shu B, Wang Y, Liu C, Xie J, Shi S. Identification of WRKY gene family from *Dimocarpus longan* and its expression analysis during flower induction and abiotic stress responses. *International Journal of Molecular Sciences*, 2018, 19 (8): 2169
- [35] Dou L, Zhang X, Pang C, Song M, Wei H, Fan S, Yu S. Genome-wide analysis of the WRKY gene family in cotton. *Molecular Genetics and Genomics*, 2014, 289 (6): 1103-1121
- [36] Dong J, Chen C, Chen Z. Expression profiles of the *Arabidopsis* WRKY gene superfamily during plant defense response. *Plant Molecular Biology*, 2003, 51 (1): 21-37
- [37] Ramamoorthy R, Jiang S Y, Kumar N, Venkatesh P N, Ramachandran S. A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormone treatments. *Plant and Cell Physiology*, 2008, 49 (6): 865-879
- [38] Zhang Y, Wang L. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evolutionary Biology*, 2005, 5 (1): 1

欢迎订阅 2020 年《植物科学学报》

《植物科学学报》是中国科学院主管、中科院武汉植物园主办、科学出版社出版、国内外公开发行的植物学综合性学术期刊,主要刊载植物学及各分支学科的原始研究论文。栏目设置:系统与进化、生态与生物地理、遗传与育种、生理与发育、资源与植物化学、技术与方法、专题综述、研究快报、学术讨论、重要书刊评介和学术动态等。读者对象:科研院所和高等院校从事植物科学研究的科研人员、教师和研究生,以及相关学科、交叉学科的科技工作者。

本刊为中国自然科学核心期刊,已被中国科学引文数据库核心库(CSCD)、《中文核心期刊要目总览》(北大)、中国科技论文与引文数据库(中信所)、中国生物学文献数据库、中国核心期刊(遴选)数据库、中国知识资源总库《中国科技期刊精品数据库》、中国期刊全文数据库、《中国药学文摘》、美国《化学文摘》、美国《生物学文摘》、美国《剑桥科学文摘:自然科学》、俄罗斯《文摘杂志》、日本《科学技术文献速报》、英国《国际农业与生物科学研究中心》(文摘)、波兰《哥白尼索引》、万方数据——数字化期刊群、中国学术期刊(光盘版)、超星期刊域出版平台和中文科技期刊数据库等国内外检索期刊、数据库作为核心期刊或统计源期刊收录。本刊相继获全国优秀科技期刊奖、中国科学院优秀期刊奖、湖北省优秀期刊奖。

双月刊,大 16 开,国内定价 50 元,全年 300 元。订阅方式:(1)全国各地邮局均可订阅(邮发代号:38-103);(2)通过科学出版社期刊发行部购买(电话 010-64017032, 64017539);(3)网上购买(搜淘宝店、微店店铺名称:中科期刊);(4)直接与本刊编辑部联系订阅(免收邮挂费)。

地址:武汉市东湖高新区九峰一路 201 号(九峰一路与光谷七路路口以北)中科院武汉植物园内《植物科学学报》编辑部

邮编:430074

电话:027-87700820

E-mail: editor@wbpcas.cn, zwkxbjb@wbpcas.cn

网址: <http://www.plantscience.cn>