

普通烟草 WRKY 基因家族的鉴定及表达分析

向小华^{1,2}, 吴新儒¹, 晁江涛¹, 杨明磊^{1,2}, 杨帆³, 陈果^{1,2},
刘贯山¹, 王元英¹

1. 中国农业科学院烟草研究所, 烟草行业烟草基因资源利用重点实验室, 青岛 266100;
2. 中国农业科学院研究生院, 北京 100081;
3. 云南省烟草公司大理州公司, 大理 671000

摘要: WRKY 基因家族编码产物是一类含有 WRKY 保守结构域的转录因子(Transcription factors, TFs), 在植物的生长发育、胁迫应答等过程中起着重要的作用。目前, 已在多种植物中鉴定了 WRKY 基因家族, 但在普通烟草(*Nicotiana tabacum* L.)中 WRKY 家族的系统鉴定与分析还未见报道。本研究利用 WRKY 保守域全蛋白序列 PF03106 检索普通烟草蛋白序列, 获得 WRKY 家族候选序列共 164 个。利用多种生物信息学软件对该家族成员进行了系统进化树、保守结构域、亚细胞定位预测、染色体定位和组织表达等分析。分组鉴定和进化树分析结果显示, 可将 164 个候选 NtWRKY 蛋白分为 I、II 和 III 3 个亚家族。其中, II 亚家族又可细分为 II-a、II-b、II-c、II-d 和 II-e 共 5 个亚组。WRKY 结构域分析结果显示, 多数基因由 2~5 个外显子组成, 各成员间的核心结构域高度保守, 含有 WRKY 盒和锌指结构。亚细胞定位预测结果显示, 大部分 NtWRKY 成员都定位在细胞核中, 参与核基因的转录调控; 而第 III 亚家族的大部分成员(74%)定位于细胞质基质中, 可能参与细胞质基因的转录调控。染色体定位结果显示, 共有 154 个 NtWRKY 基因能够定位于 24 条染色体中, 且呈不均匀分布, 6 号染色体上分布最多, 含 16 个候选基因, 10 号染色体上分布最少, 只有 1 个候选基因。组织表达分析结果显示, 大部分 NtWRKY 基因在烟草根、茎、叶中都有表达, 但不同基因间的表达模式存在差异, 这暗示了烟草 NtWRKY 家族的不同成员在功能上可能具有多样性。研究还发现基因 NtWRKY26、NtWRKY30、NtWRKY32 的表达模式受黑胫病原菌(*Phytophthora parasitica* var. *nicotianae*)诱导, 可能与烟草-黑胫病菌互作机制相关。本研究不仅为研究普通烟草 WRKY 转录因子在调控生长发育过程中的作用奠定了相关理论基础, 也为进一步分析普通烟草 WRKY 基因提供了有价值的信息。

关键词: 普通烟草; WRKY 基因家族; 转录因子; 生物信息学; 基因表达

收稿日期: 2016-01-12; 修回日期: 2013-05-16

基金项目: 公益性行业专项(编号: 201203091)和国家烟草专卖局重点项目(编号: 110201002002)资助[Supported by the Special Fund of Public Welfare (No. 201203091) and Key Project of the State Tobacco Monopoly Administration (No. 110201002002)]

作者简介: 向小华, 在读博士研究生, 研究方向: 烟草抗病育种。E-mail: xiangxiaohuacaas@163.com

通讯作者: 王元英, 博士, 研究员, 博士生导师, 研究方向: 烟草遗传育种。E-mail: wangyuanying@caas.cn

DOI: 10.16288/j.yczs.16-016

网络出版时间: 2016/7/14 15:50:59

URI: <http://www.cnki.net/kcms/detail/11.1913.R.20160714.1550.002.html>

Genome-wide identification and expression analysis of the WRKY gene family in common tobacco (*Nicotiana tabacum* L.)

Xiaohua Xiang^{1,2}, Xinru Wu¹, Jiangtao Chao¹, Minglei Yang^{1,2}, Fan Yang³, Guo Chen^{1,2}, Guanshan Liu¹, Yuanying Wang¹

1. Tobacco Research Institute of CAAS, Key Laboratory of Tobacco Genetic Improvement and Biotechnology, CAAS, Qingdao 266100, China;

2. Graduate School of Chinese Academy of Agricultural Sciences, Beijing 100081, China;

3. Yunnan Tobacco Company Dalizhou Branch, Dali 671000, China

Abstract: The coding products of WRKY gene family plays important roles in plant growth and development as well as in various stress responses. They have been identified in various plants, but only few in common tobacco (*Nicotiana tabacum* L.). In this study, 164 putative WRKY proteins in the common tobacco genome were identified by using the conserved WRKY sequence (PF03106) from the Pfam database. Phylogenetic trees, functional domain analysis, chromosomal localization, subcellular localization and tissue expression patterns were analyzed with the bioinformatics softwares, including DNAMAN 5.0, Weblogo 3, MEGA 5.1, MG2C and MEME. First of all, phylogenetic trees divided all the candidate genes into three subfamilies: I, II and III, respectively, and subfamily II could be further divided into five subgroups: group II-a, -b, -c, -d and -e. Secondly, the WRKY regions contained a highly conserved heptapeptide stretch WRKYGQK followed by a zinc-finger motif. Most of the *NtWRKY* genes contained 2–5 exons and a highly conserved gene structure. Thirdly, 154 out of 164 *NtWRKY* genes were distributed with different densities on 24 chromosomes, and each subfamily with different patterns and frequency. The largest number of *NtWRKY* genes was found on chromosome VI, and only one on chromosome X. Fourthly, the majority of *NtWRKY* members located in the nucleus, with 74 percent of subfamily III in the extracellular matrix. Lastly, the members in the same subfamily had different spatial and temporal expression profiles, with 11 *NtWRKY* genes in roots, stems and leaves expressed at various levels. The expression of genes *NtWRKY26*, *NtWRKY30* and *NtWRKY32* can be induced by *Phytophthora nicotianae*. Our research thus provides valuable information for *NtWRKY* gene cloning and functional characterization in common tobacco.

Keywords: *Nicotiana tabacum* L.; WRKY gene family; transcription factor; bioinformatics; gene expression

转录调控是生物在生长发育及代谢过程中调控细胞或组织中基因表达的重要机制^[1]。近年来,比较基因组研究结果表明动物和植物中广泛存在着与转录调控相关的基因,且这些转录基因家族可能与动植物的遗传进化和生物多样性密切相关^[2]。转录因子(Transcription factors, TFs)是一类锚定在目标基因启动子区域特异序列上的蛋白质,能对下游目标基因的转录进行促进或抑制作用,对生物体的正常生长发育和应对外界胁迫过程都具有重要的作用^[1,3]。WRKY 转录因子是一类植物特有的调控因子,其得名源于该类蛋白家族成员中含有一段约 60 个氨基酸残基组成的高度保守结构域,通常包括 1 个或 2 个 WRKYGQK 七肽序列和锌指结构(Zinc-finger motif),能够与靶基因上游启动子区域的顺式作用元件

如 W-box(T)(T)TGAC(C/T)^[4]或 SURE(糖响应顺式元件)^[5]区域特异性结合。近年来研究发现亮氨酸拉链结构(Leucine zippers)能够增加 WRKY 转录因子与 W-box 的结合特异性,从而发挥转录调控作用^[6]。Rushton 等^[7]根据所含 WRKYGQK 七肽序列的数目,首次将 WRKY 基因家族分为 Group I (含有 2 个 WRKY 结构域)和 Group II (含有 1 个 WRKY 结构域)两类。Eulgem 等^[4]根据 WRKY 保守结构域的数量和锌指结构的类型不同,进一步将 WRKY 基因家族分为 I、II 和 III 3 个亚家族共 5 组即 I、II-a+II-b、II-c、II-d+II-e 和 III,并提出高等植物 WRKY 基因家族 II-a+II-b、II-c 和 II-d+II-e 都是由 I 组的 C 端 WRKY 结构域进化而来的观点。Cormack 等^[8]研究还发现,WRKY 转录因子家族中除含有 WRKY

结构域和锌指结构外,还含有核定位信号(Nuclear localization signal, NLS)、TIR-NBS-LRR 和丝氨酸/苏氨酸域等结构。功能研究发现,WRKY 基因家族作为一类广谱调控因子,在植物形态发生^[9,10]、转录激活^[8,11]和逆境胁迫^[12-14]等多种生理生化过程发挥重要作用。

自首个编码 WRKY 转录因子 SPF1 从甘薯(*Ipomoea batatas*)^[6]中被发现以来,已陆续从拟南芥(*Arabidopsis thaliana*)^[15]、烟草(*Nicotiana tabacum*)^[16]和苹果(*Malus domestica*)^[17]等多种植物中鉴定出相应的 WRKY 基因。普通烟草(*Nicotiana tabacum* L.)作为重要的特种经济作物及重要的模式植物,对 WRKY 基因家族进行分析不仅可以系统地研究该家族成员间的相关性,也能为 WRKY 基因家族在不同物种中的克隆和进一步的功能分析奠定基础。目前,国内外在普通烟草 WRKY 转录因子研究方面已有报道,但从全基因组水平上的系统研究还未见报道。本研究以普通烟草全基因组中的 WRKY 转录因子家族为研究对象,利用多种生物信息学方法对普通烟草 NtWRKY 转录因子基因家族的进化关系、保守结构域、基因结构、染色体定位和组织表达等方面进行了分析,并分析了部分 NtWRKY 基因在不同抗性烟草材料接种黑胫病菌后的表达模式。本研究不仅为克隆普通烟草 NtWRKY 基因提供了重要信息,也为后续研究 NtWRKY 转录因子在调控植物应对外界胁迫提供了重要的理论基础。

1 材料和方法

1.1 材料与人工接种

普通烟草品种:红花大金元(记为 H)、小黄金 1025(记为 X)、革新三号(记为 G)、M691(由红花大金元经 EMS 诱变获得的黑胫病中抗病突变体,记为 M)。其中,红花大金元和小黄金 1025 为黑胫病感病材料,革新三号和 M691 为黑胫病抗病材料,以上供试材料均由中国农业科学院烟草研究所生物技术科提供。本研究中所使用的病原真菌为黑胫病 0 号生理小种强毒株由中国农业科学院烟草研究所植物保护科冯超老师馈赠,菌丝在谷子培养基上 28℃ 培养 7 d 后备用。种子经 15% 双氧水消毒后用

无菌水清洗 3 遍进行播种,待长出 6 片真叶时,参照苏振刚等^[18]建立的方法进行接种,接种后保持高温高湿环境(温度 30℃,湿度维持在 80% 以上)利于黑胫病菌的生长和繁殖。接种 7 d 后,分别取根、茎、叶组织经处理后于液氮中速冻,-80℃ 保存以备提取 RNA。

1.2 NtWRKY 序列鉴定与理化性质分析

在 Pfam 数据库中(<http://pfam.Sanger.ac.uk/>)下载 WRKY 基因家族保守序列 PF03106,并以此作为参考序列,利用 HMMER 软件(<http://hmmer.janelia.org/>) 在烟草基因组数据库(https://solgenomics.net/organism/Nicotiana_tabacum/genome)中进行相似序列检索。将下载的普通烟草基因组序列、蛋白序列及基因注释信息构建本地数据库,所得结果在 Pfam 数据库中进行结构域预测并去除假阳性序列,根据 Eulgem 等^[4]对 WRKY 基因家族分类方法,将候选序列归类到 I、II、III 3 个亚家族中。利用在线软件 ExPASy Proteomics(<http://expasy.org/>)对候选基因编码的氨基酸进行理化性质分析^[19]。

1.3 NtWRKY 家族保守域和进化分析

利用序列分析软件 DNAMAN5.0 对普通烟草基因组中预测得到的候选序列进行多序列比对,参数设置为默认值^[17]。选取保守结构域序列,利用在线软件 WebLogo 3(<http://weblogo.threeplusone.com/>)和 MEME(<http://meme.sdsc.edu/meme/meme-intro.html>)对候选基因编码的氨基酸保守结构域和锌指结构进行预测和展示分析。在使用 MEME 软件时,需要对参数基序重复的数量、基序的长度和预测基序的数量进行设置^[20,21]。

从 TAIR 数据库中(<http://www.arabidopsis.org/>)随机下载拟南芥 WRKY 转录因子家族中各亚家族所对应的成员 1~3 个作为参考序列。本研究共下载了 14 个 AtWRKY 家族转录因子,包括 I 亚家族(AtWRKY1、AtWRKY2 和 AtWRKY26)、II-a 亚组(AtWRKY18)、II-b 亚组(AtWRKY6 和 AtWRKY31)、II-c 亚组(AtWRKY13 和 AtWRKY23)、II-d 亚组(AtWRKY15 和 AtWRKY17)、II-e 亚组(AtWRKY35 和 AtWRKY69)和 III 亚家族(AtWRKY46 和 AtWRKY67)。利用软件 MEGA5.0 对候选 NtWRKY 蛋白序列与 14 个 AtWRKY 参考序列进行多序列比

对, 将比对结果按照 Neighbor-Joining 方法构建邻接树, Bootstrap 重复 1000 次^[22], 对普通烟草 NtWRKY 的保守域蛋白序列进行进化树分析。将下载的 NtWRKY 候选基因的核酸序列和 CDS 序列, 利用 GSDS2.0 软件(<http://gsds.cbi.pku.edu.cn/>)对候选基因进行结构分析^[23]。

1.4 NtWRKY 基因家族定位分析

将下载的候选 NtWRKY 基因家族各成员的氨基酸序列利用在线软件 ProtComp 9.0 (<http://linux1.softberry.com>)对 NtWRKY 基因家族成员进行亚细胞定位预测分析。利用 Perl 程序解析烟草基因组信息文件, 确定 NtWRKY 基因家族成员在普通烟草染色体上位置分布, 利用 MG2C 在线软件(http://mg2c.iask.in/mg2c_v2.0/)将 NtWRKY 成员定位到普通烟草 24 条染色体上并绘制图谱。

1.5 NtWRKY 基因家族表达模式分析与验证

为分析所预测的候选 NtWRKY 基因的表达模式, 本研究利用已有的烟草 EST 和基因芯片数据(<http://solgenomics.net/>)平台搜索与 NtWRKY 基因家族匹配的芯片数据, 对数据进行整理后利用 Cluster 3.0 软件进行聚类分析, 并利用 Treeview 软件对结果进行可视化分析^[24], 并对部分候选 NtWRKY 基因进行了组织特异性表达验证(引物序列见表 1)。使用 TRIZOL Kit 试剂盒提取总 RNA, 具体操作参照说

明书进行。RNA 经 DNase I 消化去除基因组 DNA 污染, 并进行定性定量检测确保其质量满足后续实验。取 1 μg RNA, 利用反转录试剂盒 PrimeScriptTM 1ST Strand cDNA Synthesis Kit 合成第一链 cDNA, 具体操作参照说明书进行, 将 cDNA 的浓度稀释至 25 ng/ μL 后备用。以 *Ntubc2* (Ubiquitin-conjugating enzyme E2, GenBank 登录号: AB026056)作为内参基因^[25], 反应体系组成为: 2.5 μL 10 \times Taq Reaction Buffer(含 Mg^{2+})、2 μL dNTPs(2.5 mmol/L)、1 μL 正反向引物(10 $\mu\text{mol/L}$)、0.5 μL Taq DNA 聚合酶 模板 cDNA 50 ng, 用 ddH₂O 补齐至 25 μL 。PCR 扩增程序为: 94 $^{\circ}\text{C}$ 预变性 5 min; 94 $^{\circ}\text{C}$ 变性 30 s, 59 $^{\circ}\text{C}$ 复性 30 s, 72 $^{\circ}\text{C}$ 延伸 30 s, 35 个循环; 扩增结束后再 72 $^{\circ}\text{C}$ 延伸 7 min。取 5 μL PCR 扩增产物进行 1% EB/琼脂糖凝胶电泳检测。

2 结果与分析

2.1 NtWRKY 序列鉴定及保守结构域分析

利用 WRKY 结构域序列 PF03106 作为种子序列在普通烟草基因组中进行 Blastp 比对分析, 并对候选序列进行筛选, 手动去除冗余及不完整编码序列, 从普通烟草基因组中共鉴定出 164 条普通烟草 WRKY 转录因子蛋白序列。候选 NtWRKY 成员保守结构域分析结果显示, 各成员间的 WRKY 结构域是高度保守的(图 1A)。利用 SWISS-MODEL 软件构

表 1 组织特异表达检测引物序列

Table 1 Primer sequences of the tissue-specific expression analysis of the NtWRKY genes

基因名称	正向引物 F(5' 3')	反向引物 R(5' 3')	<i>T_m</i> 值($^{\circ}\text{C}$)	产物大小(bp)
<i>NtWRKY07</i>	CAATGCCCTCGGAGCTACTA	TCATGCATGCGAATGTCAGG	59	209
<i>NtWRKY19</i>	TGGAGAGGCTGTTCACAAGT	TAGGCAGCAACTTAAGGGCT	59	173
<i>NtWRKY26</i>	CAGTCTCTCCGTCCCAAACT	CCCTTCACTTGTTTCTGCCC	59	250
<i>NtWRKY28</i>	GCGGATCTGAAAAGGGCAAT	AACTTCGGCGGTATCCTCAT	59	189
<i>NtWRKY30</i>	CACATTCCAATTCCGTGCCA	TGAACCTGCTTTGTCCACCG	59	166
<i>NtWRKY32</i>	GGGTCTCCATTTCAGGAA	TTTGGAATTGGTGCTGTCGG	59	177
<i>NtWRKY81</i>	ATCAGCCCTTGGATCCACAA	AGTATGCAGAGCCCAAGTGT	59	185
<i>NtWRKY114</i>	CGAGGACGCTACAAAAGAAGG	TGACTCGAAATAGTGCGGA	59	220
<i>NtWRKY141</i>	ACGTCTACCTCCTTTACCGC	GAAGAAGTAGAAAGCGGCGG	59	231
<i>NtWRKY147</i>	CTCAGTCTCAAGCAACAGCC	TTGTTGTGGAAGTGTG GCC	59	155
<i>NtWRKY163</i>	TTGCCGGAAAAAGTCATCAGC	ATCCTCATTAGACTCGCCGG	59	195
<i>Ntubc2</i>	CTGACATCTCCCGCACTCTTA	CATAGTCCATTCTAGTTGAGCA	59	179

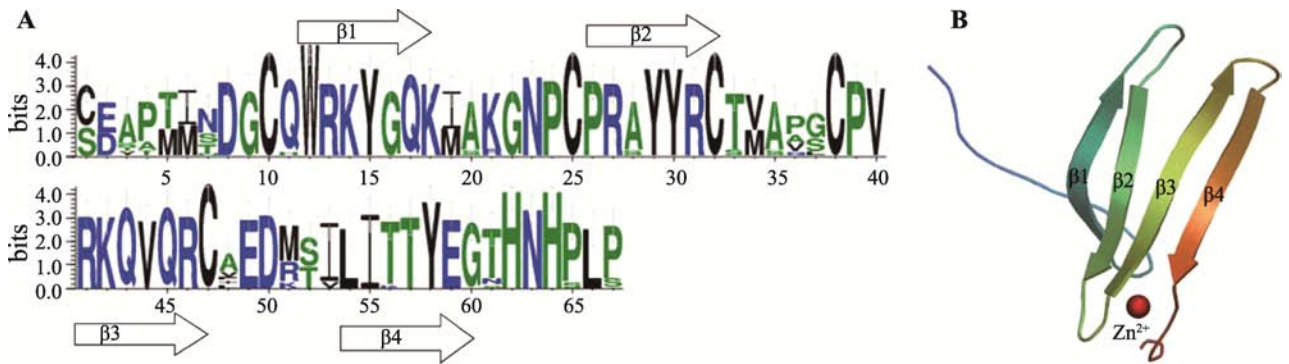


图 1 谱通烟草 WRKY 蛋白保守结构域序列(A)及 3D 结构模型(B)

Fig. 1 The conserved domain (A) and 3D structural model (B) of the WRKY conserved domains in common tobacco WRKY proteins

建了 NtWRKY 蛋白保守域的 3D 模型, 分析发现目标序列与 AtWRKY1 蛋白的 WRKY 结构域 3D 结构相似度较高, 都包括 4 个 β 折叠结构域和一个 Zn^{2+} 结合位点(图 1B)。

利用 DNAMAN5.0 软件对候选序列进行多序列比对, 并参考 Eulgem 等^[4]对 WRKY 保守域的描述及分类方法, 可将 164 个候选普通烟草 NtWRKY 蛋白序列分为 3 个亚家族(I、II 和 III), 其中 II 亚家族成员最多, 共有 102 个, I 和 III 亚家族各含有 31 个候选基因。保守结构域分析结果显示(图 2), I 亚家族成员中除 NtWRKY18 外, 其余成员都含有 2 个标准的 WRKY 保守结构域, 根据 WRKY 保守结构域在蛋白序列中的位置又可进一步分为 I-N 亚组和 I-C 亚组。进一步分析发现, I-N 亚组和 I-C 亚组中的氨基酸序列所含有 WRKYGQK 七肽序列的两侧基序的类型也存在较大的差异。I-N 亚组中除 NtWRKY7N 为 WRKYGEK 外, 其余成员都为 WRKYQK, 且锌指结构高度保守, 都为 $CX_4C_{22}HXXH$ 形式。I-C 亚组中除 NtWRKY18C 含有不完整的七肽域外, 其余成员都为 WRKYGQK; 锌指结构保守性较 I-N 亚组低, NtWRKY8C 和 NtWRKY26C 无锌指结构, 其余成员都为 $CX_4C_{23}HXXH$ 形式。研究还发现 NtWRKY26 含有 3 个 WRKYGQK 七肽结构域。以上结果初步表明, I 亚家族中的 I-N 和 I-C 亚组的 WRKY 结构域在起源或功能分化上可能不完全相同。

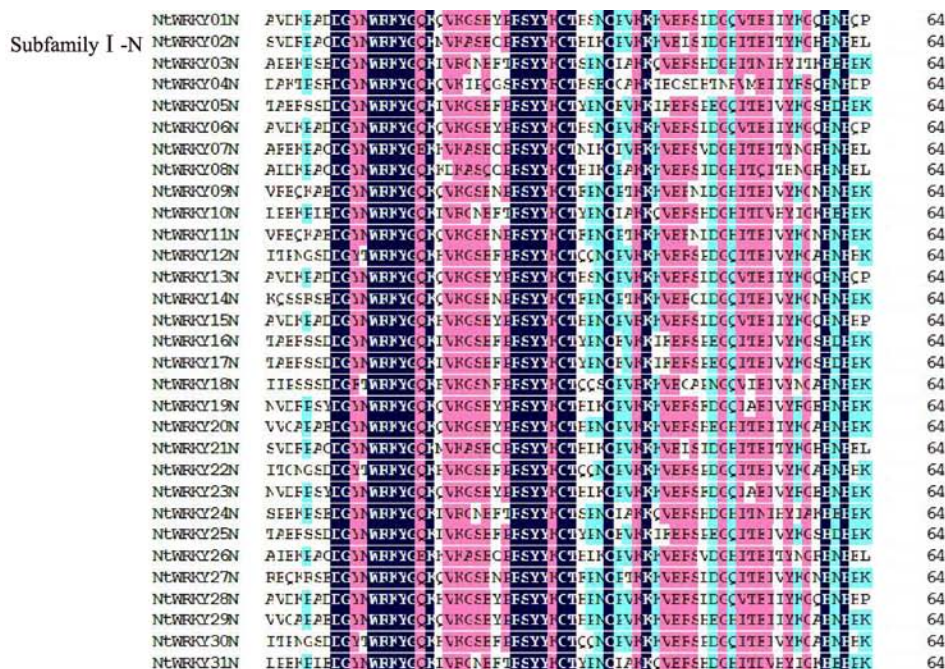
根据 II 亚家族成员各自的特点与差异, 又可细分为 II-a、II-b、II-c、II-d 和 II-e 共 5 个亚组, 分

别含有 4、19、45、15 和 19 个成员。其中, II-a、II-b、II-d 和 II-e 亚组 C 端保守七肽域和锌指结构相对保守, 分别为 WRKYGQK 和 $CX_5C_{23}HXXH$ 形式。II-c 亚组中各成员间锌指结构高度保守为 $CX_4C_{23}HXXH$, 而 C 端 WRKY 七肽域变异较大, 其中 45 个成员中有 32 个成员为原始 WRKYGQK 七肽结构域, 11 个成员为 WRKYGKK 类型变异, 2 个成员为 WRKYGMK 变异类型。III 组共有 31 个成员, 各成员间的 C 端七肽域和锌指结构高度保守, 分别为 WRKYGQK 和 $CX_7C_{23}HXXC$ 形式(表 2)。

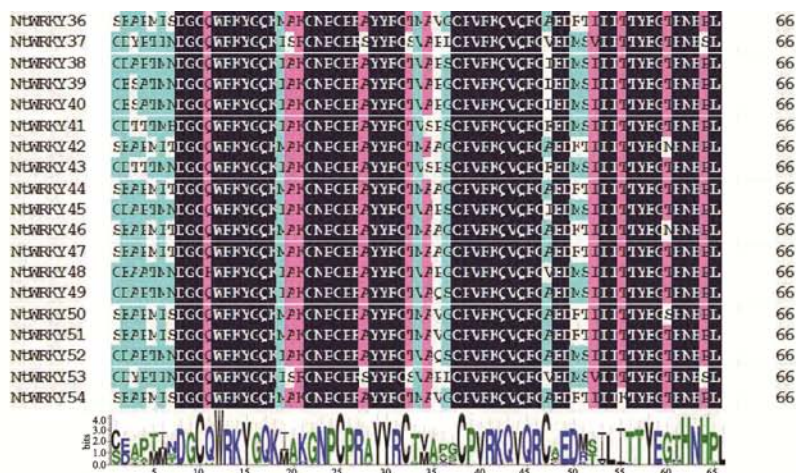
对候选 NtWRKY 基因家族各成员的理化特征进行分析发现, 不同亚家族之间的氨基酸序列长度分子量大小和等电点等变化呈现不同规律(附表 1)。WRKY 家族中各成员的平均长度为 375 个氨基酸, 但各成员间差异较大, 其中 NtWRKY35(II-a 亚组)候选序列为最长序列, 包含 1234 个氨基酸残基, NtWRKY110(II-d 亚组)为最短序列, 只有 159 个氨基酸残基。所有候选 WRKY 蛋白序列的等电点范围为 4.85(NtWRKY93, II-c 亚组)~10.32(NtWRKY102, II-c 亚组)不等。

2.2 NtWRKY 基因家族系统进化和结构分析

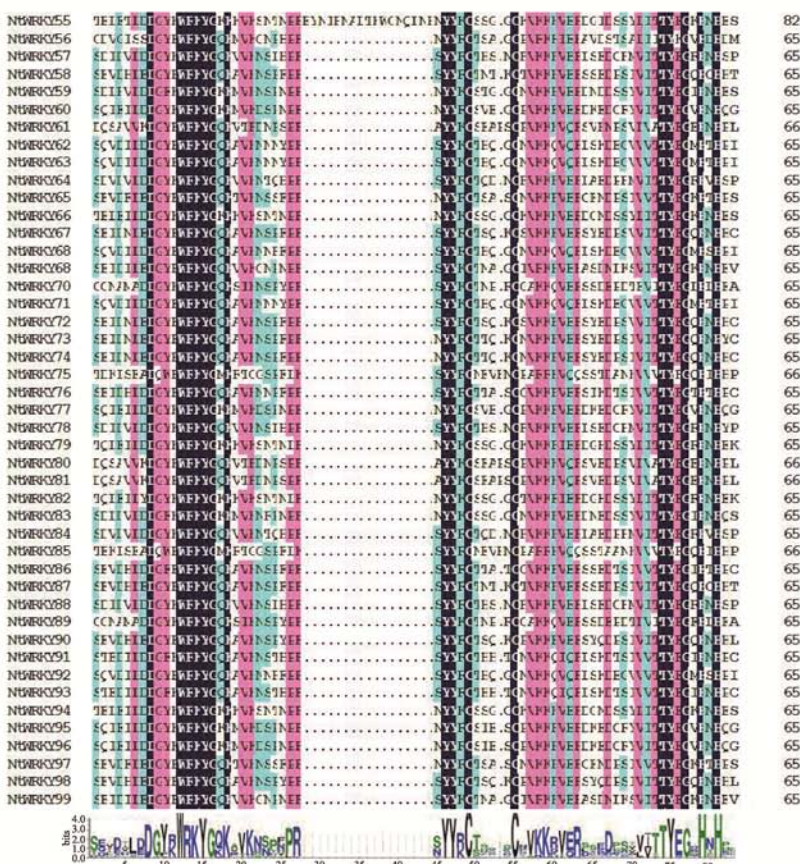
利用 MEGA5.0 软件对 NtWRKY 蛋白的保守结构域序列进行多序列比对并构建了系统进化树(图 3)。分析发现, 普通烟草 WRKY 基因家族各个亚组成员间相对保守, 且将所选定的 14 个 AtWRKY 参考序列与本研究预测的候选基因家族中各自对应的亚家族具有相同的进化关系。以上结果表明, 进化树分析结果与初步鉴定结果相吻合, 本研究检索到的 WRKY



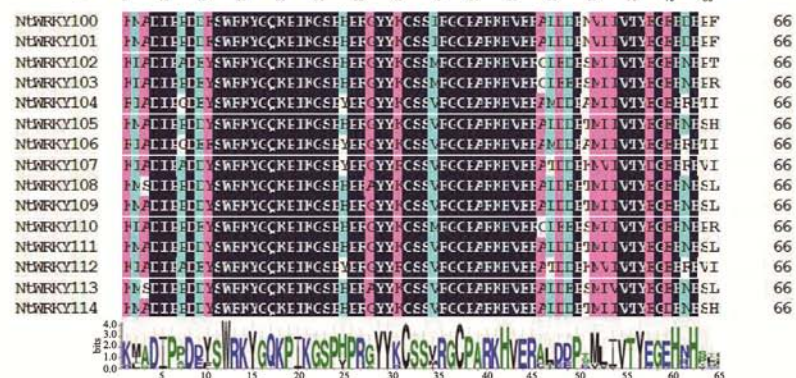
Subfamily II -b



Subfamily II -c



Subfamily II -d



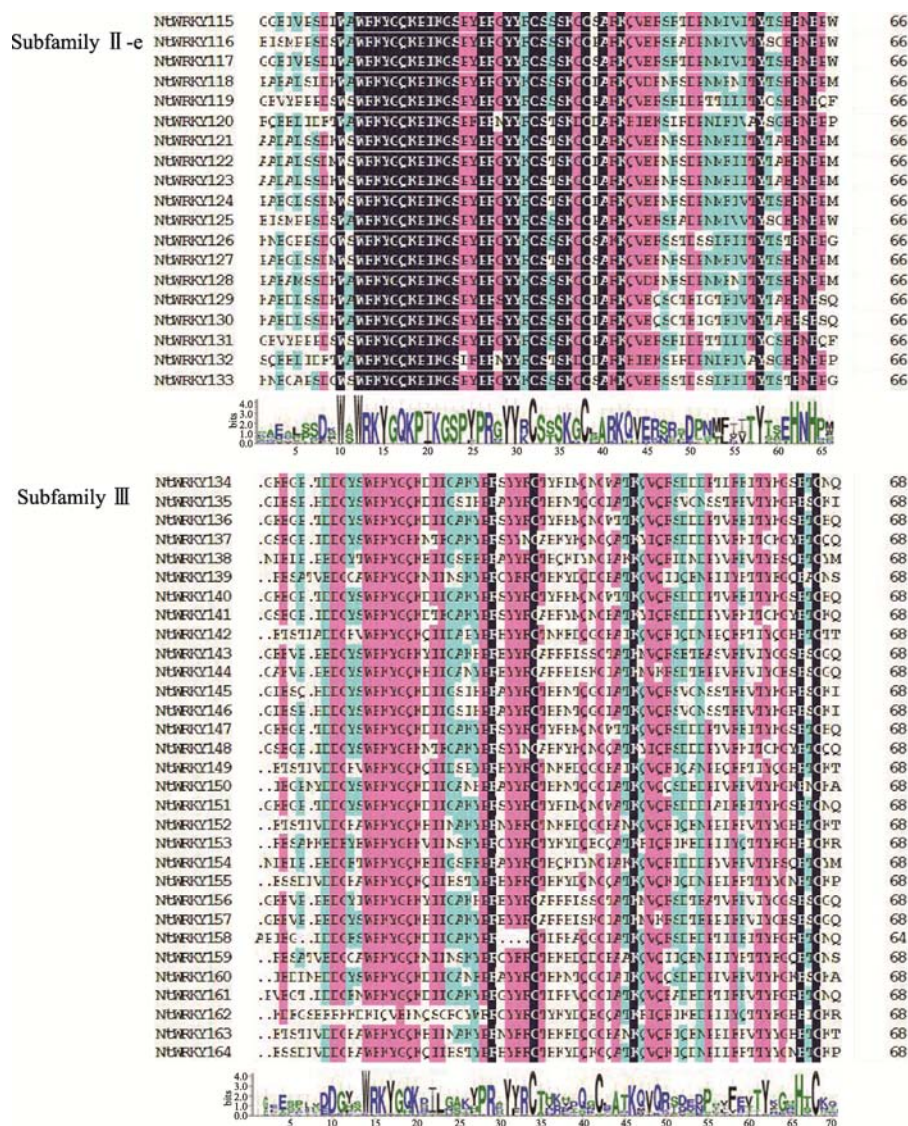


图2 普通烟草 NtWRKY 转录因子家族各亚家族保守结构域分析

Fig. 2 Sequence analysis of the WRKY conserved domains in the sub-family of NtWRKY proteins

基因家族成员可信度较高，可以进行后续的生物信息学分析。通过下载候选基因的核酸序列、CDS 序列及蛋白序列，利用 MEGA5.0 软件和 GSDS2.0 软件进行基因结构分析。数据分析结果显示，除 *NtWRKY17*、*NtWRKY151* 只含有 1 个外显子外，大多数普通烟草 WRKY 基因家族成员的基因结构中外显子的个数主要集中在 2~5 个之间(图 4A)。这也从侧面反映出同一亚家族蛋白成员及其编码基因在结构和组成上高度相似，同时也验证了系统发育分析的可靠性。

对编码 WRKY 结构域分析发现, 绝大多数 WRKY

保守结构域都是由基因的最后一个外显子翻译编码, 且在不同亚家族中的负责编码的外显子长度不同。保守结构域预测结果显示, 所有 NtWRKY 基因家族成员都包含有模体 1(WRKY 盒)和模体 2(锌指结构), 但进一步分析发现各亚家族中的模体组成也有各自的特点(图 4B)。I 亚家族中含有 2 个 WRKY 盒(模体 1 和模体 3)分别靠近 N 端和 C 端, 其编码产物主要由模体 1、2、3、5 和 6 组成。II 亚家族中各成员的模体组成差异较大, 但各亚组间具有一定的保守性。II-a 亚组主要包含模体 1、2、4 和 7; II-b 亚组主要包含模体 1、2、4、5、7 和 8, 但 NtWKRY43、NtWKRY53 和

表 2 NtWRKY 候选序列中七肽域和锌指结构的变异
Table 2 Variation of the heptapeptide and zinc-finger structures in the NtWRKY candidate proteins

变异类型	基因名称		分组
七肽变异	WRKYGHK	<i>NtWRKY137</i> 、 <i>NtWRKY143</i> 、 <i>NtWRKY148</i> 、 <i>NtWRKY153</i> 、 <i>NtWRKY156</i>	Ⅲ
	WRKYGKK	<i>NtWRKY55</i> 、 <i>NtWRKY59</i> 、 <i>NtWRKY60</i> 、 <i>NtWRKY66</i> 、 <i>NtWRKY77</i> 、 <i>NtWRKY79</i> 、 <i>NtWRKY82</i> 、 <i>NtWRKY83</i> 、 <i>NtWRKY94</i> 、 <i>NtWRKY95</i> 、 <i>NtWRKY96</i>	Ⅱ -c
	WRKYGМК	<i>NtWRKY75</i> 、 <i>NtWRKY85</i>	Ⅱ -c
	WRKYGЕK	<i>NtWRKY7N</i>	I
	None	<i>NtWRKY162</i>	Ⅲ
	None	<i>NtWRKY18C</i>	I
锌指结构变异	CX ₄ C ₂₃ HXH	<i>NtWRKY1C</i> - <i>NtWRKY7C</i> 、 <i>NtWRKY9C</i> - <i>NtWRKY25C</i> 、 <i>NtWRKY27C</i> - <i>NtWRKY31C</i>	I
	CX ₄ C ₂₃ HXH	<i>NtWRKY55</i> - <i>NtWRKY60</i> 、 <i>NtWRKY62</i> - <i>NtWRKY74</i> 、 <i>NtWRKY76</i> - <i>NtWRKY79</i> 、 <i>NtWRKY82</i> - <i>NtWRKY84</i> 、 <i>NtWRKY86</i> - <i>NtWRKY99</i>	Ⅱ -c
	CX ₄ C ₂₂ HXH	<i>NtWRKY1N</i> - <i>NtWRKY31N</i>	I
	CX ₇ X ₂₃ HXC	<i>NtWRKY153</i>	Ⅲ
	None	<i>NtWRKY8C</i>	I
	None	<i>NtWRKY26C</i>	I

2.3 NtWRKY 基因家族定位分析

亚细胞定位预测结果显示(图 5), 本研究所鉴定的 164 个 NtWRKY 转录因子中共有 118(71.95%) 个定位在细胞核中, 其余 46 个 NtWRKY 成员定位在细胞质中, 可能参与细胞质基因的转录调控。结果分析还发现, 编码亚细胞定位相关的结构域可能与进化关系相关, 如 I、Ⅱ亚家族成员主要定位在细胞核, Ⅲ亚家族共有 24 个成员都定位于细胞质中, 占该家族总数的 77.4%。

染色体定位结果显示, 164 个 NtWRKY 基因中有 154 个 NtWRKY 基因分布于 24 条染色体中, 且呈不均匀分布(图 6)。其中, 6 号染色体上分布最多, 共有 14 个家族成员; 16 号和 1 号染色体分布次之, 分别含有 11 个和 10 个候选基因; 10 号染色体上分布最少, 仅含有 1 个候选基因。另外, 共有 10 个基因(*NtWRKY17*、*NtWRKY29*、*NtWRKY40*、*NtWRKY57*、*NtWRKY76*、*NtWRKY110*、*NtWRKY111*、*NtWRKY120*、*NtWRKY153*、*NtWRKY164*)定位于 scaffold 上, 目前无法确定在染色体上的位置。通过各个亚家族在染色体上的分布定位图还发现, 同一家族各成员在染色体上的分布也存在差异。I 亚家族共有 29 个基因定位在染色体上, 除在 5、8、14、15、18、20 和 21 号染色体无分布, 其余染色体均有分布; Ⅱ亚家

族共有 96 个候选基因定位在染色体上, 且在 24 条染色体上均有分布; Ⅲ亚家族也同样有 29 个基因定位在除 4、5、8、10、12、13、14、17 和 24 以外的其余染色体上。

2.4 NtWRKY 基因表达模式分析

从普通烟草高通量转录组测序数据库中提取 WRKY 基因家族数据, 将数据进行聚类分析和可视化处理后发现, 同一家族的不同基因在组织特异性表达模式上存在差异, 大多数成员在根、茎和叶组织中均有表达, 且表达模式多样化(图 7)。其中, 有 156 个 NtWRKY 家族成员在根组织中有表达, 有 142 个 NtWRKY 家族成员茎组织中有表达, 有 133 个在叶片组织中有表达。分析还发现, 各个成员在不同组织中的表达情况也存在明显的差异, 这表明各基因的表达具有一定的组织特异性。I 亚家族中, *NtWRKY26* 在根组织中的表达量最高; *NtWRKY21* 在茎组织中的表达量最高; *NtWRKY28* 在叶片组织中的表达量最高; *NtWRKY8*、*NtWRKY30* 和 *NtWRKY31* 只在根中表达; *NtWRKY7* 和 *NtWRKY11* 在根、茎、叶中都不表达。Ⅱ亚家族中, *NtWRKY78* 在根组织中的表达量最高, *NtWRKY114* 在茎组织中的表达量最高; *NtWRKY81* 在叶片组织中的表达量最高; *NtWRKY51*、*NtWRKY68*、*NtWRKY73*、*NtWRKY103* 及 *NtWRKY111*

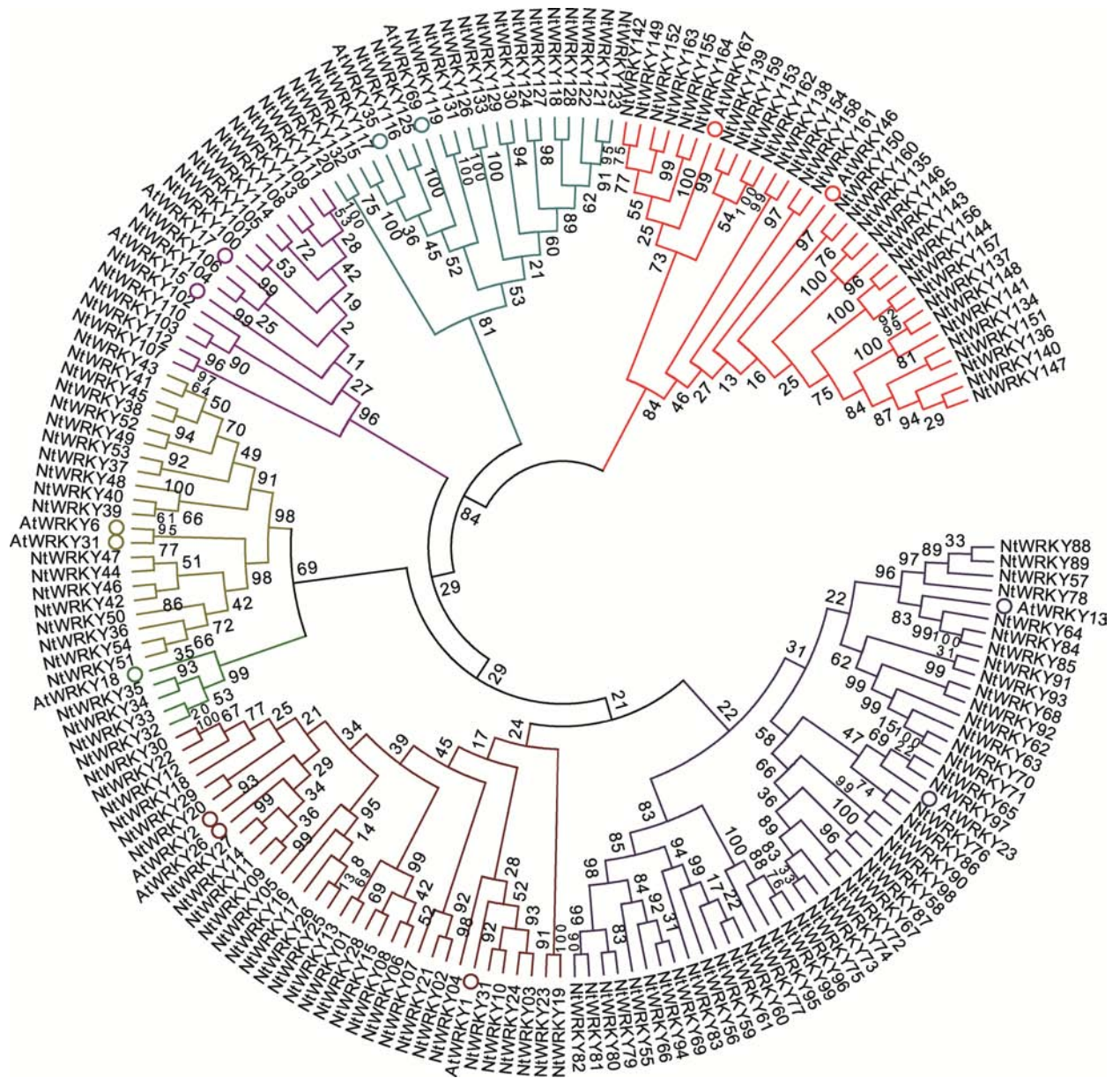


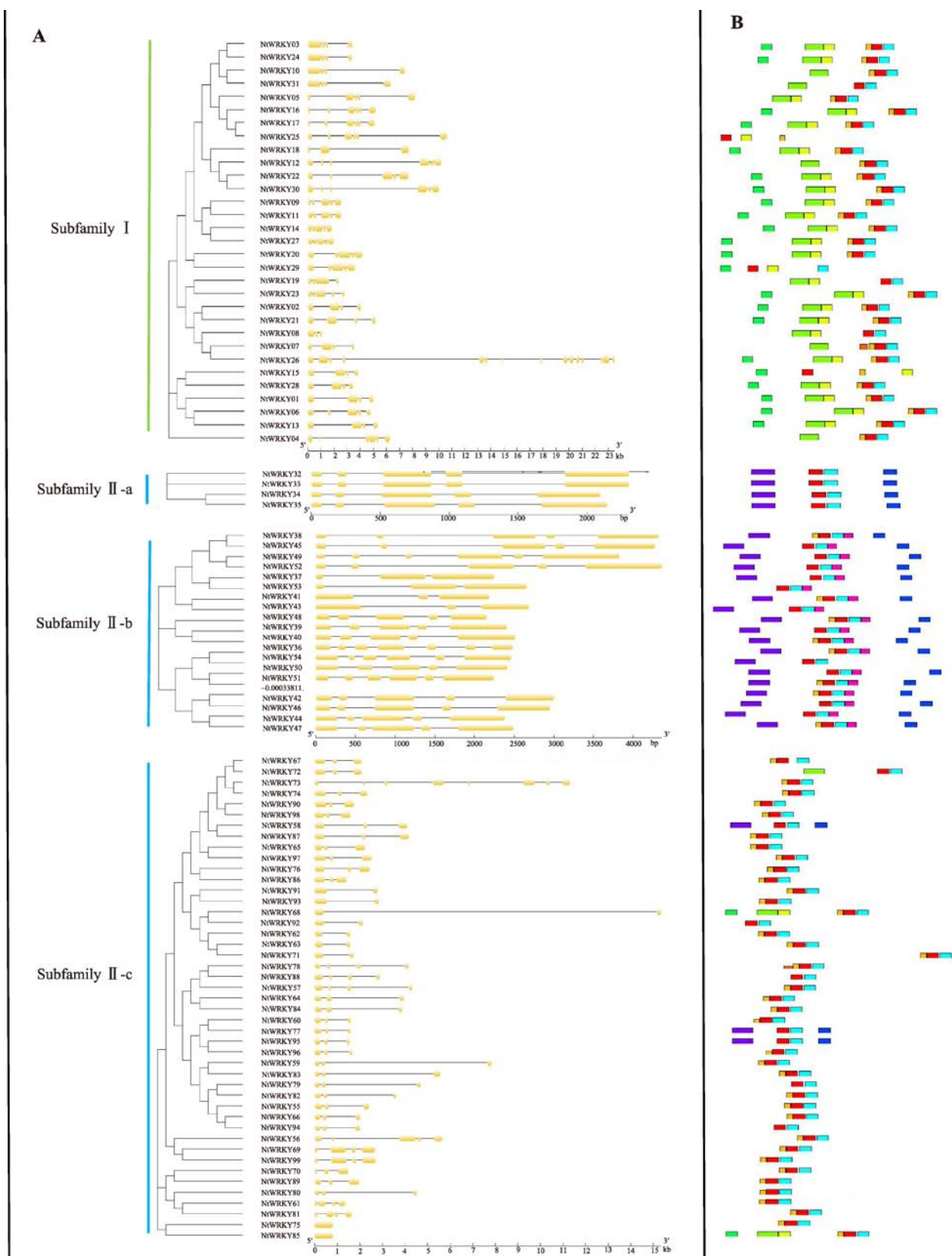
图 3 普通烟草 NtWRKY 蛋白的保守域进化关系

Fig. 3 Phylogenetic analysis of conserved domains in NtWRKY proteins

只在根中表达; *NtWRKY32*、*NtWRKY82*、*NtWRKY92* 和 *NtWRKY105* 在根、茎、叶中不表达或表达量较低。III 亚家族中, *NtWRKY141* 在根和茎组织中的表达量较高; *NtWRKY163* 在叶片中的表达量最高; *NtWRKY161* 只在叶片中检测到表达; *NtWRKY153*、*NtWRKY159* 只在根组织中表达; *NtWRKY147* 在根、茎、叶组织中都不表达。综上所述表明, NtWRKY 基因家族基因在不同材料或者同一材料不同组织的表达存在差异, 这可能是由于基因的表达模式受材

料基因型和环境因素的共同调控的结果。

为验证基因芯片数据分析的可靠性, 本研究选取 11 个 NtWRKY 基因在黑胄病不同抗、感材料的根、茎、叶组织进行了表达量验证(图 8)。验证结果表明, 所选取的除 *NtWRKY7* 和 *NtWRKY147* 这 2 个基因的表达不受外源病原菌的诱导外, 其余 NtWRKY 基因的表达量受外源病原菌的诱导调节, 且都表现为表达量上调。选候选基因在根、茎、叶组织中的总体表达量上调顺序为: 根>茎>叶。但也有



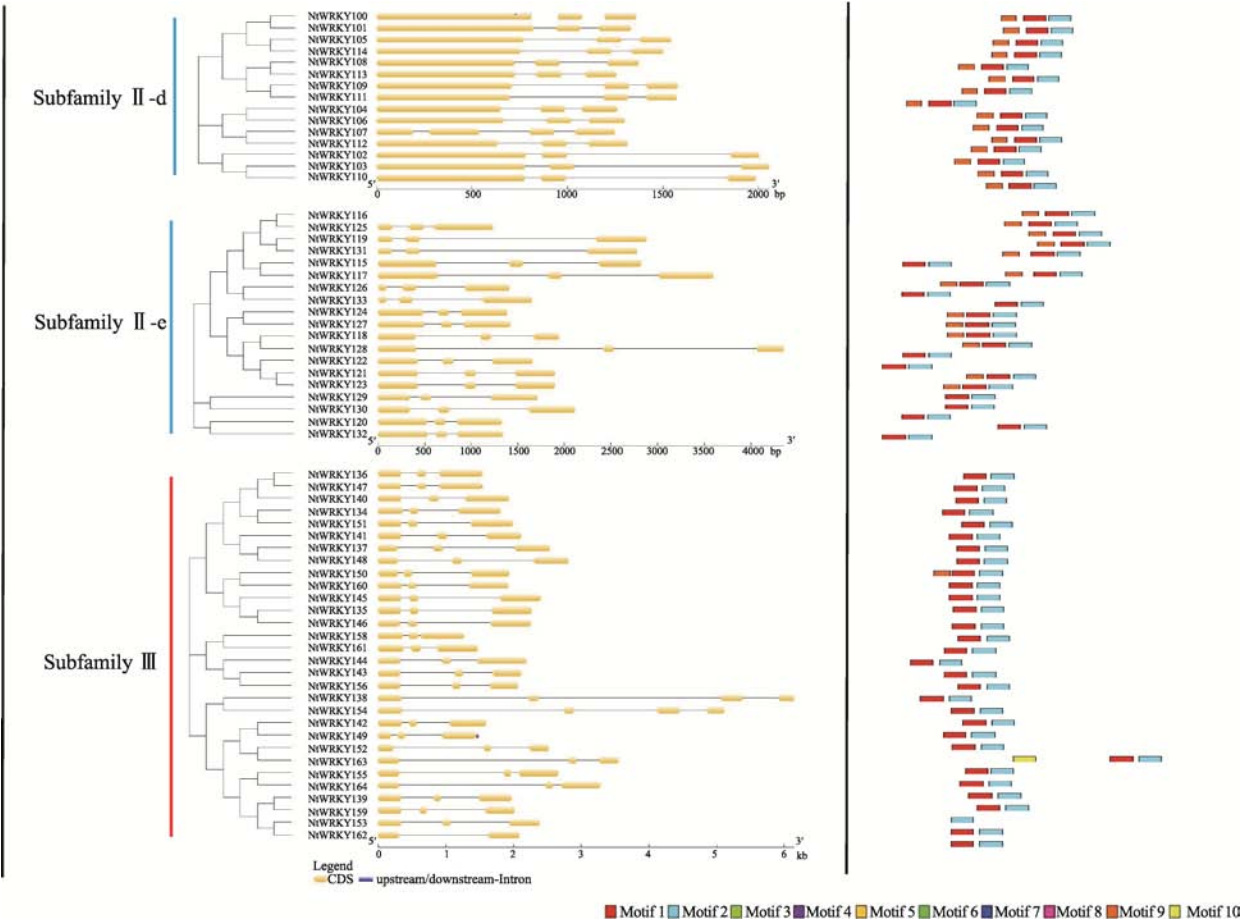


图 4 NtWRKY 基因家族基因结构(A)和模体组成(B)分析
Fig. 4 Gene structure (A) and motif component (B) analysis of the candidate genes of the NtWRKY gene family

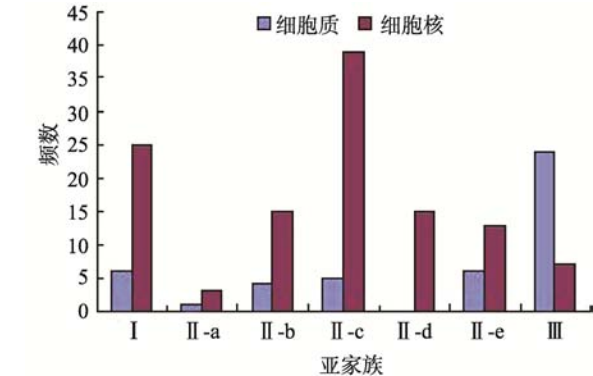


图 5 普通烟草 NtWRKY 基因家族的亚细胞定位预测
Fig. 5 Sub-cellular localization of NtWRKY gene family

个别基因的表达模式不同于其他基因，如 *NtWRKY32* 在根中表达模式受病原菌的诱导，在茎和叶片中没有检测到其表达；*NtWRKY81* 在叶片中的表达受外源病原菌的诱导，但未在根和茎中检测到其表达。

3 讨论

WRKY 基因家族是一类早已被证明能够调节植物多重生理过程的转录因子。自 Ishiguro 等^[6]从甘薯中克隆了第一个 WRKY 基因 *SPF1* 以来，已从水稻、玉米、蒺藜苜蓿等多种植物的根、叶片、花序、种子、微观组织以及多种胁迫中已有克隆了超过 500 个 WRKY 的 ESTs 序列，其生物学功能也得到广泛的挖掘和鉴定。本研究从普通烟草全基因组中共鉴定出 164 个 WRKY 基因家族成员，明显高于拟南芥基因组中的数目^[26]。普通烟草是异源四倍体，其基因组主要通过基因富集区测序获得，但因基因组组装的序列过短，准确性还有待于验证，因此可能会忽略了很多由于片段过短的 WRKY 候选基因。

根据其保守结构域和锌指结构的差异特征又可

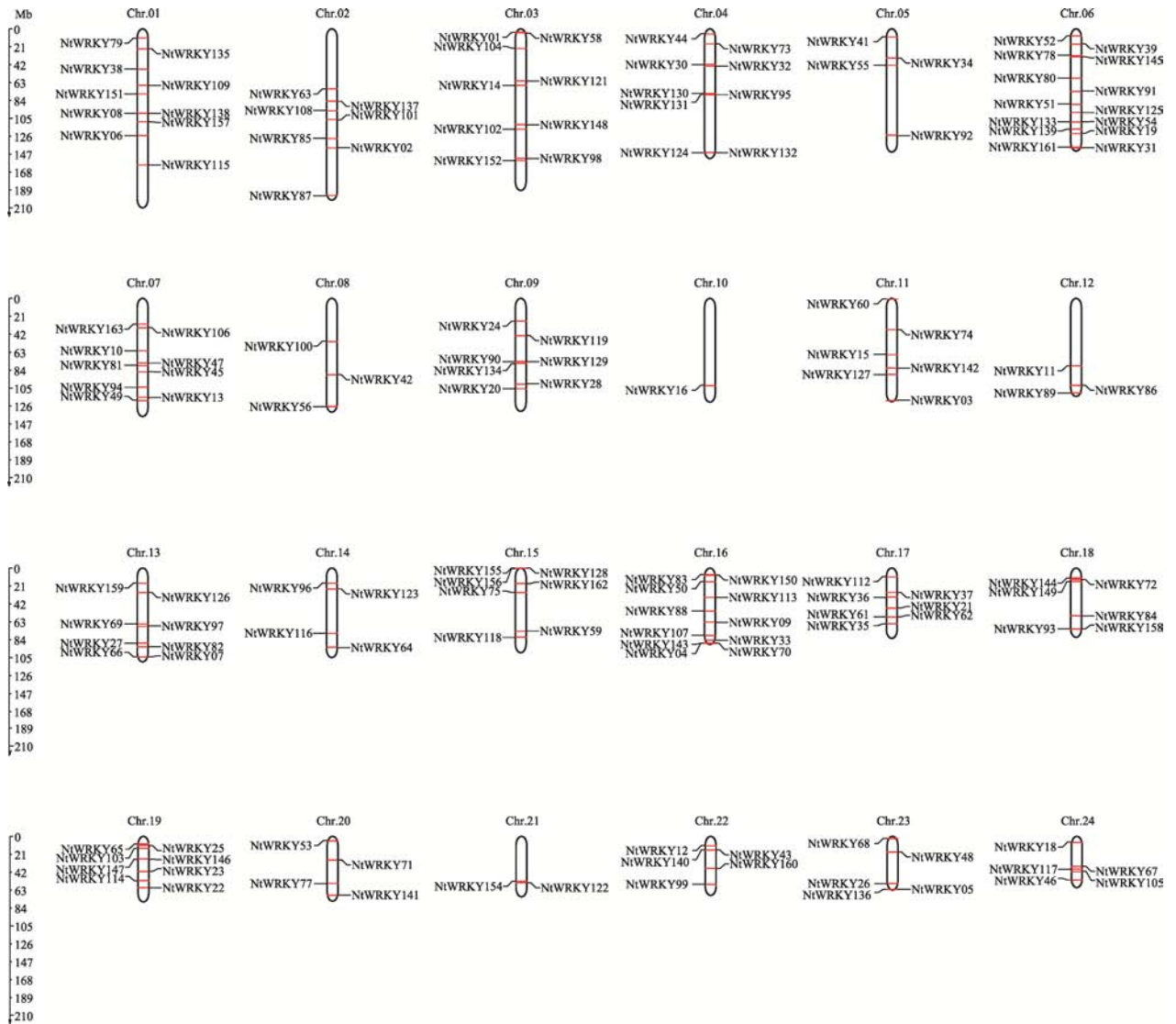


图 6 普通烟草 *NtWRKY* 基因家族的染色体定位结果
Fig. 6 Chromosomal locations of the *NtWRKY* gene family

将其细分为 I、II、III 共 3 个亚家族, 其组成分别为 I 亚家族为 31 个成员; II 亚家族为 102 个成员; III 亚家族为 31 个成员。其中, II 亚家族依据其结构的进化分析可进一步分为 5 个亚组: II-a、II-b、II-c、II-d 和 II-e。进化分析研究表明, 普通烟草 *NtWRKY* 基因家族进化过程中的多样性主要受各家族成员的保守 WRKYGQK 七肽结构域和 C2H2 锌指结构的变异等方面的影响。典型的 I 亚家族成员都含有 2 个 WRKYGQK 七肽序列和锌指结构域, 但是锌指结构略有差异, N 端为 $CX_4C_{23}HXH$ 结构, C 端为 $CX_4C_{22}HX$ 结构; II 亚家族成员包含 1 个 WRKYGQK

七肽序列和 C 端 $CX_5C_{23}HXH/CX_4C_{23}HXH$ 锌指结构; III 亚家族蛋白含有一个 WRKY 保守域, 但其锌指结构较为独特为 $CX_7C_{23}HXC$ 类型。以上结果表明, *NtWRKY* 基因家族各成员在结构上保守程度较高的, 但保守域的变异也使得 *NtWRKY* 基因家族在进化过程出现了多样性。分析结果支持了 Eisen 等^[27]的观点, 即 II 亚家族和 III 亚家族的由来是由 I 亚家族通过 N 端和 C 端 WRKY 保守结构域的变异于缺失演化而来的。

基因中内含子和外显子的组成可为基因在进化分析研究中提供重要证据^[28]。本研究对候选 *NtWRKY*

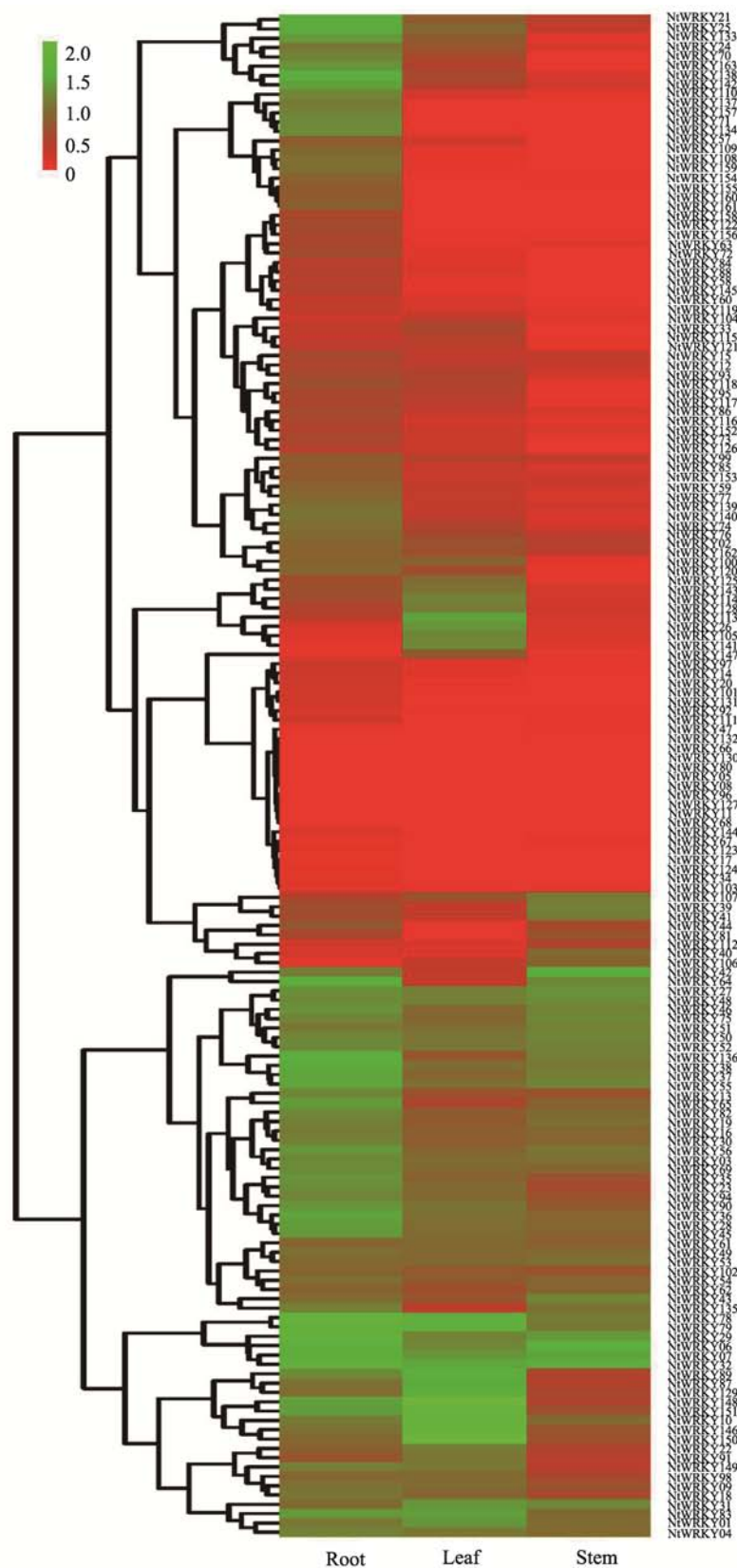


图 7 NtWRKY 家族基因在组织中的表达模式

Fig. 7 The pattern of tissue expression of NtWRKY gene family

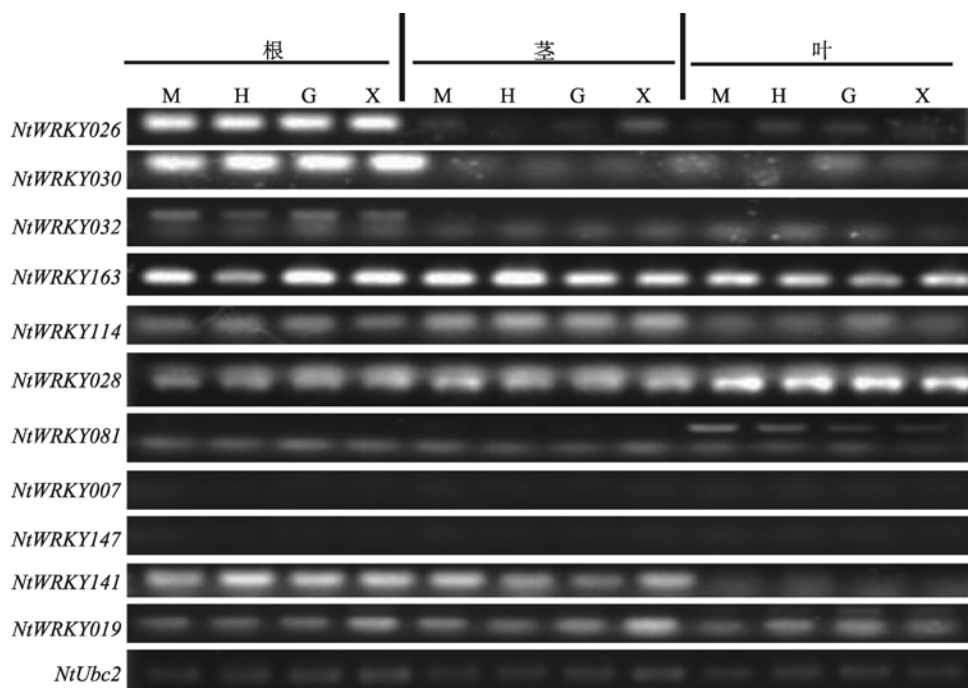


图 8 分 NtWRKY 基因在不同烟草组织和品种中的表达分析

Fig. 8 Expression analysis of several *NtWRKY* genes in different tobacco tissues and cultures

M: 抗黑胫病突变体; H: 红花大金元; G: 革新三号; X: 小黄金 1025。

基因家族各成员的基因结构分布和遗传进化进行了分析发现, *NtWRKY* 基因家族各成员的基因结构中大都包含 2~5 个外显子, 且负责编码 WRKY 保守结构域的最后一个外显子序列长度在各亚家族成员间存在差异。候选基因亚细胞定位发现, 在 III 组中有 24 个成员亚细胞定位结果位于细胞质中, 这似乎与转录因子在核基因发挥作用的常理相悖论。但笔者认为, 这些定位于细胞质中转录因子很可能参与到细胞质中的基因表达调控, 如线粒体、叶绿体中的相关基因表达。另外本研究结果指出 III 亚家族各成员间的锌指结构区域保守性较低, 我们推测这可能是由于核基因较细胞质基因更加稳定, 且核基因发生变异的几率也远小于质基因, 这也可能是对 WRKY 家族 III 亚家族成员编码蛋白保守程度低的一个推测, 但具体定位情况还需要通过后续更深入的研究来验证。

基因的表达模式是研究基因功能的重要方面, 前人研究结果表明同一个亚家族的基因具有相似的表达模式、基因结构、功能、进化起源等^[29,30]。WRKY 转录因子作为植物抗病基因转录调控的重要顺式作

用元件早被人所熟知, 其表达受生物胁迫和非生物胁迫等因素诱导^[4]。高通量基因芯片技术为全基因组范围内研究不同基因的表达模式提供了重要方法。本研究基于 EST 数据和高通量基因芯片数据, 通过生物信息学工具分析了普通烟草 *NtWRKY* 基因的表达模式并进行了聚类分析, 结果表明同一家族的大部分成员都表现出相似的表达模式, 但组织特异性差异较大。部分 *NtWRKY* 基因的验证结果表明, 所选取的 *NtWRKY* 基因在普通烟草的根、茎、叶组织中具有广泛的表达, 但在根和茎中的表达量高于叶片, 并且 *NtWRKY26*、*NtWRKY30* 和 *NtWRKY32* 明显受外源黑胫病菌侵染的诱导。因此, 我们推测这几个 *NtWRKY* 基因可能在黑胫病-烟草互作过程中发挥作用, 但具体生物学功能还需要进一步的验证。

参考文献(References):

- [1] Zhang YJ, Wang LJ. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol Biol*, 2005, 5(1): 1. [DOI]
- [2] Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR,

- Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu GL. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, 2000, 290(5499): 2105–2110. [DOI]
- [3] Martinez E. Multi-protein complexes in eukaryotic gene transcription. *Plant Mol Biol*, 2002, 50(6): 925–947. [DOI]
- [4] Eulgem T, Rushton PJ, Robatzek S, Somssich IE. The WRKY superfamily of plant transcription factors. *Trends Plant Sci*, 2000, 5(5): 199–206. [DOI]
- [5] Sun CX, Palmqvist S, Olsson H, Borén M, Ahlandsberg S, Jansson C. A novel WRKY transcription factor, *SUSIBA2*, participates in sugar signaling in barley by binding to the sugar-responsive elements of the *iso1* promoter. *Plant Cell*, 2003, 15(9): 2076–2092. [DOI]
- [6] 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. *Mol Gen Genet*, 1994, 244(6): 563–571. [DOI]
- [7] Rushton PJ, Somssich IE, Ringler P, Shen QJ. WRKY transcription factors. *Trends Plant Sci*, 2010, 15(5): 247–258. [DOI]
- [8] Cormack RS, Eulgem T, Rushton PJ, Köchner P, Hahlbrock K, Somssich IE. Leucine zipper-containing WRKY proteins widen the spectrum of immediate early elicitor-induced WRKY transcription factors in parsley. *Biochim Biophys Acta*, 2002, 1576(1–2): 92–100. [DOI]
- [9] Ishida T, Hattori S, Sano R, Inoue K, Shirano Y, Hayashi H, Shibata D, Sato S, Kato T, Tabata S, Okada K, Wada T. *Arabidopsis* *TRANSPARENT TESTA GLABRA2* is directly regulated by R2R3 MYB transcription factors and is involved in regulation of *GLABRA2* transcription in epidermal differentiation. *Plant Cell*, 2007, 19(8): 2531–2543. [DOI]
- [10] Lagacé M, Matton DP. Characterization of a WRKY transcription factor expressed in late torpedo-stage embryos of *Solanum chacoense*. *Planta*, 2004, 219(1): 185–189. [DOI]
- [11] Rushton PJ, Torres JT, Parniske M, Wernert P, Hahlbrock K, Somssich IE. Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley *PR1* genes. *EMBO J*, 1996, 15(20): 5690–5700. [DOI]
- [12] Cheong YH, Chang HS, Gupta R, Wang X, Zhu T, Luan S. Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*. *Plant Physiol*, 2002, 129(2): 661–677. [DOI]
- [13] Hara K, Yagi M, Kusano T, Sano H. Rapid systemic accumulation of transcripts encoding a tobacco WRKY transcription factor upon wounding. *Mol Gen Genet*, 2000, 263(1): 30–37. [DOI]
- [14] Yoda H, Ogawa M, Yamaguchi Y, Koizumi N, Kusano T, Sano H. Identification of early-responsive genes associated with the hypersensitive response to tobacco mosaic virus and characterization of a WRKY-type transcription factor in tobacco plants. *Mol Gen Genet*, 2002, 267(2): 154–161. [DOI]
- [15] de Pater S, Greco V, Pham K, Memelink J, Kijne J. Characterization of a zinc-dependent transcriptional activator from *Arabidopsis*. *Nucl Acids Res*, 1996, 24(23): 4624–4631. [DOI]
- [16] Chen CH, Chen ZX. Isolation and characterization of two pathogen-and salicylic acid-induced genes encoding WRKY DNA-binding proteins from tobacco. *Plant Mol Biol*, 2000, 42(2): 387–396. [DOI]
- [17] Gu YB, Ji ZR, Chi FM, Qiao Z, Xu CN, Zhang JX, Dong QL, Zhou ZS. Bioinformatics and expression analysis of the WRKY gene family in apple. *Sci Agric Sin*, 2015, 48(16): 3221–3238.
谷彦冰, 冀志蕊, 迟福梅, 乔壮, 徐成楠, 张俊祥, 董庆龙, 周宗山. 苹果 WRKY 基因家族生物信息学及表达分析. 中国农业科学, 2015, 48(16): 3221–3238. [DOI]
- [18] Su ZG, Yang AG, Sun YH, Luo CG, Liu GS, Zhou J, Li YY, Yang F, Zhao BY, Wang YY. Construction and analysis of tobacco SSH library induced by *Phytophthora parasitica* var. *nicotianae*. *Acta Agronom Sin*, 2011, 37(10): 1763–1770.
苏振刚, 杨爱国, 孙玉合, 罗成刚, 刘贯山, 周佳, 李元元, 杨帆, 赵百英, 王元英. 黑胙病菌诱导的烟草 SSH 文库构建及其分析. 作物学报, 2011, 37(10): 1763–1770. [DOI]
- [19] Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. ExPASy: SIB bioinformatics resource portal. *Nucl Acids Res*, 2012, 40(W1): W597–W603. [DOI]
- [20] Song H, Nan ZB. Genome-wide identification and analysis of WRKY transcription factors in *Medicago truncatula*. *Hereditas (Beijing)*, 2014, 36(2): 152–168.
宋辉, 南志标. 蒺藜苜蓿全基因组中 WRKY 转录因子的鉴定与分析. 遗传, 2014, 36(2): 152–168. [DOI]
- [21] Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. Menlo Park, California: AAAI Press, 1994: 28–36. [DOI]
- [22] Crooks GE, Hon G, Chandonia JM, Brenner SE. Web Logo: A sequence logo generator. *Genome Res*, 2004, 14(6): 1188–1190. [DOI]
- [23] Hu B, Jin JP, Guo AY, Zhang H, Luo JC, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinform-*

- tics, 2015, 31(8): 1296–1297. [DOI]
- [24] Sun TT, Wang DW, Gong DP, Chen L, Chen YQ, Sun YH. Genome-wide identification and bioinformatic analysis of B3 superfamily in tomato. *J Plant Genet Res*, 2015, 16(4): 806–814.
孙亭亭, 王大伟, 龚达平, 陈乐, 陈雅琼, 孙玉合. 番茄 B3 超家族成员鉴定及生物信息学分析. *植物遗传资源学报*, 2015, 16(4): 806–814. [DOI]
- [25] Schmidt GW, Delaney SK. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Mol Genet Genomics*, 2010, 283(3): 233–241. [DOI]
- [26] Eulgem T, Somssich IE. Networks of WRKY transcription factors in defense signaling. *Curr Opin Plant Biol*, 2007, 10(4): 366–371. [DOI]
- [27] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA*, 1998, 95(25): 14863–14868. [DOI]
- [28] Ramamoorthy R, Jiang SY, Kumar N, Venkatesh PN, Ramachandran S. A Comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and Phytohormone treatments. *Plant Cell Physiol*, 2008, 49(6): 865–879. [DOI]
- [29] Duan MR, Nan J, Liang YH, Mao P, Lu L, Li LF, Wei CH, Lai LH, Li Y, Su XD. DNA binding mechanism revealed by high resolution crystal structure of *Arabidopsis thaliana* WRKY1 protein. *Nucl Acids Res*, 2007, 35(4): 1145–1154. [DOI]
- [30] Ross CA, Liu Y, Shen QJ. The WRKY gene family in rice (*Oryza sativa*). *J Intergr Plant Biol*, 2007, 49(6): 827–842. [DOI]

(责任编辑: 夏先春)

• 综合信息 •

中国科技核心期刊、中国农业核心期刊 全国中文核心期刊、全国优秀农业期刊

《植物遗传资源学报》征订启事

《植物遗传资源学报》是中国农业科学院作物科学研究所和中国农学会主办的学术期刊, 为中国科技论文统计源期刊、中国科学引文数据库来源期刊(核心期刊)、中国核心期刊(遴选)数据库收录期刊、中国学术期刊综合评价数据库统计源期刊, 又被《中国生物学文摘》和中国生物学文献数据库、中文科技期刊数据库收录。2015 年本刊影响力大幅提升, 在农艺学类期刊排名中均提前。据《中国科技期刊引证报告》(核心版)统计: 《植物遗传资源学报》影响因子为 1.149, 比 2014 年提高了近 10%, 在农艺学类期刊中排名第 3。据 CNKI《中国学术期刊影响因子年报》统计: 《植物遗传资源学报》复合影响因子为 1.695, 在 48 种农艺学类期刊排名第 4, 期刊综合影响因子 1.146。

报道内容为大田、园艺作物, 观赏、药用植物, 林用植物、草类植物及其经济植物的有关植物遗传资源基础理论研究、应用研究方面的研究成果、创新性学术论文和高水平综述或评论。诸如, 种质资源的考察、收集、保存、评价、利用、创新, 信息学、管理学等; 起源、演化、分类等系统学; 基因发掘、鉴定、克隆、基因文库建立、遗传多样性研究。

双月刊, 大 16 开本, 196 页。定价 20 元, 全年 120 元。各地邮局发行。

邮发代号: 82-643。国内刊号 CN11-4996/S, 国际统一刊号 ISSN1672-1810。

本刊编辑部常年办理订阅手续, 如需邮挂每期另加 3 元。

地址: 北京市中关村南大街 12 号 《植物遗传资源学报》编辑部

邮编: 100081 电话: 010-82105794 010-82105796 (兼传真)

网址: www.zwyczy.cn

E-mail: zwyczyxb2003@163.com

zwyczyxb2003@caas.cn

附表 1 NtWRKY 候选基因信息汇总
Supplemental Table 1 Summary of NtWRKY candidate genes

基因名称	定位名称	染色体	定位位置	类型	WRKY 结构域	锌指结构	氨基酸数目	分子量(Da)	等电点(pI)
NtWRKY01	Ntab0088260	Chr.3	4011042..4016009	I	WRKYGQK	C2H2	525	57446.23	8.3361
NtWRKY02	Ntab0148910	Chr.2	147394494..147398529	I	WRKYGQK	C2H2	513	56386.11	6.9975
NtWRKY03	Ntab0157910	Chr.11	126156660..126160026	I	WRKYGQK	C2H2	535	59662.45	7.2584
NtWRKY04	Ntab0187610	Chr.16	93387437..93389373	I	WRKYGQK	C2H2	256	28645.27	10.1988
NtWRKY05	Ntab0199830	Chr.23	65636065..65637598	I	WRKYGQK	C2H2	361	40847.1	5.4077
NtWRKY06	Ntab0231610	Chr.01	132028031..132031028	I	WRKYGQK	C2H2	507	55522.73	6.2537
NtWRKY07	Ntab0238890	Chr.13	109872916..109873935	I	WRKYGQK	C2H2	243	27333.89	10.175
NtWRKY08	Ntab0244960	Chr.01	104283199..104289341	I	WRKYGQK	None	341	37939.3	6.9906
NtWRKY09	Ntab0340060	Chr.16	66899356..66902007	I	WRKYGQK	C2H2	498	55688.19	7.7084
NtWRKY10	Ntab0350920	Chr.07	64716391..64718897	I	WRKYGQK	C2H2	558	62225.08	8.7214
NtWRKY11	Ntab0353580	Chr.12	82995715..83005889	I	WRKYGQK	C2H2	624	69638.36	8.327
NtWRKY12	Ntab0354530	Chr.22	111699959..111705241	I	WRKYGQK	C2H2	525	57446.23	8.3361
NtWRKY13	Ntab0358650	Chr.07	122020056..122022433	I	WRKYGQK	C2H2	563	61008.24	5.8474
NtWRKY14	Ntab0413910	Chr.03	69888104..69889809	I	WRKYGQK	C2H2	182	20729.45	10.0737
NtWRKY15	Ntab0437710	Chr.11	69404524..69406587	I	WRKYGQK	C2H2	308	34480.27	6.62
NtWRKY16	Ntab0481340	Chr.10	107637318..107639633	I	WRKYGQK	C2H2	313	35165.98	7.0387
NtWRKY17	Ntab0506620	0	Chr.0:200549..201331	I	WRKYGQK	C2H2	260	29525.52	6.5752
NtWRKY18	Ntab0546740	Chr.24	7519056..7520300	I	NONE	C2H2	257	28310.94	9.7754
NtWRKY19	Ntab0660340	Chr.06	129598200..129600182	I	WRKYGQK	C2H2	347	38962.36	10.2618
NtWRKY20	Ntab0680470	Chr.09	111448234..111453385	I	WRKYGQK	C2H2	513	56428.13	7.604
NtWRKY21	Ntab0681110	Chr.17	49744894..49746463	I	WRKYGQK	C2H2	325	36192.02	10.0763
NtWRKY22	Ntab0708310	Chr.19	63564716..63567488	I	WRKYGQK	C2H2	474	52720.94	9.3565
NtWRKY23	Ntab0709710	Chr.19	43481228..43482709	I	WRKYGQK	C2H2	273	30771.12	6.9093
NtWRKY24	Ntab0782720	Chr.09	28125726..28128134	I	WRKYGQK	C2H2	518	56555.41	8.3681
NtWRKY25	Ntab0788190	Chr.19	11334708..11338860	I	WRKYGQK	C2H2	309	33385.67	5.4509
NtWRKY26	Ntab0797690	Chr.23	98343525..98345487	I	WRKYGQK	ALQL	545	60855.55	8.398
NtWRKY27	Ntab0800290	Chr.13	92802185..92804427	I	WRKYGQK	C2H2	508	55341.81	6.7518
NtWRKY28	Ntab0854710	Chr.09	105753332..105755982	I	WRKYGQK	C2H2	323	36761.49	5.2509
NtWRKY29	Ntab0894150	0	Chr.0:408040..410146	I	WRKYGQK	C2H2	185	21310.88	9.8288

续表

基因名称	定位名称	染色体	定位位置	类型	WRKY 结构域	锌指结构	氨基酸数目	分子量(Da)	等电点(pI)
NtWRKY30	NtAb0906780	Chr.04	44094887..44096143	I	WRKYGQK	C2H2	368	40591.1	5.1625
NtWRKY31	NtAb0925670	Chr.06	147400589..147402054	I	WRKYGQK	C2H2	361	40520.11	5.0495
NtWRKY32	NtAb0375720	Chr.04	46164689..46166005	II-a	WRKYGQK	C2H2	366	41089.25	6.2758
NtWRKY33	NtAb0573910	Chr.16	89235894..89237123	II-a	WRKYGQK	C2H2	305	34294.99	5.7602
NtWRKY34	NtAb0579930	Chr.05	36093830..36096779	II-a	WRKYGQK	C2H2	514	55960.2	6.7791
NtWRKY35	NtAb0791820	Chr.17	68792271..68815722	II-a	WRKYGQK	C2H2	1234	137362.01	9.2001
NtWRKY36	NtAb0041780	Chr.17	35949372..35951845	II-b	WRKYGQK	C2H2	422	46778.43	6.3278
NtWRKY37	NtAb0059360	Chr.17	30210728..30212972	II-b	WRKYGQK	C2H2	477	52133.8	6.6722
NtWRKY38	NtAb0075270	Chr.01	49941614..49945933	II-b	WRKYGQK	C2H2	538	59428.25	7.2242
NtWRKY39	NtAb0113090	Chr.06	18966402..18968804	II-b	WRKYGQK	C2H2	503	55391.09	6.4203
NtWRKY40	NtAb0159290	0	Chr.0:793814..795142	II-b	WRKYGQK	C2H2	255	29020.57	6.995
NtWRKY41	NtAb0177310	Chr.05	10163558..10165820	II-b	WRKYGQK	C2H2	344	38875.62	6.4077
NtWRKY42	NtAb0233930	Chr.08	194312067..194314059	II-b	WRKYGQK	C2H2	191	22077.23	6.8606
NtWRKY43	NtAb0291910	Chr.22	16922987..16924905	II-b	WRKYGQK	C2H2	361	40769.98	5.7992
NtWRKY44	NtAb0360900	Chr.04	6515530..6517831	II-b	WRKYGQK	C2H2	353	38785.12	8.141
NtWRKY45	NtAb0459790	Chr.07	90747148..90752291	II-b	WRKYGQK	C2H2	541	59073.99	7.4335
NtWRKY46	NtAb0587420	Chr.24	74091920..74093288	II-b	WRKYGQK	C2H2	334	37384.58	10.0366
NtWRKY47	NtAb0762650	Chr.07	79684375..79686522	II-b	WRKYGQK	C2H2	405	45447.43	6.4285
NtWRKY48	NtAb0764420	Chr.23	19488294..19490806	II-b	WRKYGQK	C2H2	197	22512.26	7.5326
NtWRKY49	NtAb0773110	Chr.07	126173124..126176436	II-b	WRKYGQK	C2H2	525	57856.23	7.2269
NtWRKY50	NtAb0785700	Chr.16	116902310..116912920	II-b	WRKYGQK	C2H2	609	65881.74	6.9754
NtWRKY51	NtAb0817240	Chr.06	93432461..93436821	II-b	WRKYGQK	C2H2	615	67170.13	7.0123
NtWRKY52	NtAb0825280	Chr.06	9023045..9025916	II-b	WRKYGQK	C2H2	242	27404.14	7.9001
NtWRKY53	NtAb0912770	Chr.20	5755563..5757472	II-b	WRKYGQK	C2H2	342	38662.79	5.1705
NtWRKY54	NtAb0936480	Chr.06	115510678..115512326	II-b	WRKYGQK	C2H2	245	27776.88	5.3939
NtWRKY55	NtAb0006340	Chr.05	45038988..45041366	II-c	WRKYGKK	C2H2	237	27222.8	6.9532
NtWRKY56	NtAb0023160	Chr.08	133452851..133458495	II-c	WRKYGQK	C2H2	511	56052.64	7.2352
NtWRKY57	NtAb0054340	0	Chr.0:1683522..1687823	II-c	WRKYGQK	C2H2	235	26555.23	7.8892
NtWRKY58	NtAb0088000	Chr.03	5129933..5134019	II-c	WRKYGQK	C2H2	309	33376.57	5.4473

续表

基因名称	定位名称	染色体	定位位置	类型	WRKY 结构域	锌指结构	氨基酸数目	分子量(Da)	等电点(pI)
NtWRKY59	NtAb0100000	Chr.15	78118737..78126560	II -c	WRKYGKK	C2H2	178	20502.95	7.9544
NtWRKY60	NtAb0131500	Chr.11	1131944..1133517	II -c	WRKYGKK	C2H2	178	20389.51	5.7473
NtWRKY61	NtAb0164470	Chr.17	60306374..60307935	II -c	WRKYGQK	C2H2	166	19136.69	9.9681
NtWRKY62	NtAb0164480	Chr.17	60458890..60460451	II -c	WRKYGQK	C2H2	166	19136.69	9.9681
NtWRKY63	NtAb0172660	Chr.02	74141812..74143865	II -c	WRKYGQK	C2H2	347	38939.31	10.2165
NtWRKY64	NtAb0196910	Chr.14	97959428..97967583	II -c	WRKYGQK	C2H2	500	54760.11	6.3306
NtWRKY65	NtAb0227710	Chr.19	9332393..9337160	II -c	WRKYGQK	C2H2	585	63653.26	8.3411
NtWRKY66	NtAb0238850	Chr.13	109798288..109801810	II -c	WRKYGKK	C2H2	427	47553.86	7.5571
NtWRKY67	NtAb0269220	Chr.24	37353489..37354747	II -c	WRKYGQK	C2H2	318	34843.66	10.1255
NtWRKY68	NtAb0317230	Chr.23	2975855..2977446	II -c	WRKYGQK	C2H2	327	37179.41	4.9732
NtWRKY69	NtAb0349760	Chr.13	69041727..69049122	II -c	WRKYGQK	C2H2	549	60407.91	6.7411
NtWRKY70	NtAb0386150	Chr.16	92757905..92760006	II -c	WRKYGQK	C2H2	280	31199.93	7.7477
NtWRKY71	NtAb0416170	Chr.20	29304558..29306209	II -c	WRKYGQK	C2H2	315	35149.87	6.6244
NtWRKY72	NtAb0438350	Chr.18	13982274..13984167	II -c	WRKYGQK	C2H2	316	35223.97	7.7983
NtWRKY73	NtAb0441810	Chr.04	18463846..18468118	II -c	WRKYGQK	C2H2	537	59483.29	6.3652
NtWRKY74	NtAb0499400	Chr.11	38927688..38932759	II -c	WRKYGQK	C2H2	540	58875.81	7.4557
NtWRKY75	NtAb0520030	Chr.15	30342253..30349941	II -c	WRKYGKM	C2H2	447	49683.61	8.2562
NtWRKY76	NtAb0562190	0	Chr.0:4244..5817	II -c	WRKYGQK	C2H2	178	20389.51	5.7473
NtWRKY77	NtAb0565330	Chr.20	58263673..58267817	II -c	WRKYGKK	C2H2	207	23459.83	8.2237
NtWRKY78	NtAb0568510	Chr.06	33143426..33145816	II -c	WRKYGQK	C2H2	346	38910.71	6.2687
NtWRKY79	NtAb0642070	Chr.01	11239767..11242090	II -c	WRKYGKK	C2H2	517	57445.39	9.5929
NtWRKY80	NtAb0668120	Chr.06	61161098..61162722	II -c	WRKYGQK	C2H2	258	29102.64	6.5108
NtWRKY81	NtAb0674890	Chr.07	82774162..82778298	II -c	WRKYGQK	C2H2	727	78975.8	6.5496
NtWRKY82	NtAb0695160	Chr.13	97681004..97688668	II -c	WRKYGKK	C2H2	683	75672.13	7.1103
NtWRKY83	NtAb0721170	Chr.16	8275813..8279671	II -c	WRKYGKK	C2H2	228	26116.39	9.1507
NtWRKY84	NtAb0726020	Chr.18	59079044..59081024	II -c	WRKYGQK	C2H2	355	40375.81	5.5842
NtWRKY85	NtAb0736630	Chr.02	135747837..135749237	II -c	WRKYGKM	C2H2	333	36812.29	8.5607
NtWRKY86	NtAb0748910	Chr.12	107346847..107348260	II -c	WRKYGQK	C2H2	363	39881.71	6.3089
NtWRKY87	NtAb0790970	Chr.02	206255083..206257224	II -c	WRKYGQK	C2H2	361	39762.38	8.5632

续表									
基因名称	定位名称	染色体	定位位置	类型	WRKY 结构域	锌指结构	氨基酸数目	分子量(Da)	等电点(pI)
NtWRKY88	Ntab0838650	Chr.16	53281562..53283939	II -c	WRKYGQK	C2H2	291	33480.68	7.4968
NtWRKY89	Ntab0879650	Chr.12	116838404..116840123	II -c	WRKYGQK	C2H2	352	39648.93	6.8497
NtWRKY90	Ntab0883680	Chr.09	77977024..77978731	II -c	WRKYGQK	C2H2	313	35582.64	4.8545
NtWRKY91	Ntab0889060	Chr.06	77508386..77511963	II -c	WRKYGQK	C2H2	727	79052.84	6.4395
NtWRKY92	Ntab0901290	Chr.05	131637539..131647541	II -c	WRKYGQK	C2H2	624	69638.36	8.327
NtWRKY93	Ntab0907520	Chr.18	75453604..75455714	II -c	WRKYGQK	C2H2	312	35229.04	4.8488
NtWRKY94	Ntab0918770	Chr.07	109448216..109450989	II -c	WRKYGKK	C2H2	270	29518.55	5.858
NtWRKY95	Ntab0923050	Chr.04	281437855..281439501	II -c	WRKYGKK	C2H2	168	19257.2	6.1228
NtWRKY96	Ntab0923310	Chr.14	18935195..18941494	II -c	WRKYGKK	C2H2	584	64438.21	7.0535
NtWRKY97	Ntab0937880	Chr.13	71983730..71985291	II -c	WRKYGQK	C2H2	308	34443.33	8.2221
NtWRKY98	Ntab0940660	Chr.03	260343557..260344867	II -c	WRKYGQK	C2H2	318	34698.5	10.1111
NtWRKY99	Ntab0975840	Chr.22	59613037..59614534	II -c	WRKYGQK	C2H2	345	38167.23	10.2616
NtWRKY100	Ntab0011560	Chr.08	53540628..53541984	II -d	WRKYGQK	C2H2	366	40215.75	10.1334
NtWRKY101	Ntab0117230	Chr.02	112156905..112158233	II -d	WRKYGQK	C2H2	369	40567.91	10.0232
NtWRKY102	Ntab0146100	Chr.03	224140245..224142244	II -d	WRKYGQK	C2H2	349	39365.74	10.3154
NtWRKY103	Ntab0174290	Chr.19	14766114..14768298	II -d	WRKYGQK	C2H2	403	44940.85	7.4181
NtWRKY104	Ntab0290230	Chr.03	24355415..24358098	II -d	WRKYGQK	C2H2	422	47258.29	6.5883
NtWRKY105	Ntab0300500	Chr.24	43744734..43760065	II -d	WRKYGQK	C2H2	185	21322.78	9.8183
NtWRKY106	Ntab0411170	Chr.07	36438649..36440437	II -d	WRKYGQK	C2H2	489	54810.29	7.7602
NtWRKY107	Ntab0556670	Chr.16	83376446..83378864	II -d	WRKYGQK	C2H2	339	38123.31	7.1777
NtWRKY108	Ntab0597310	Chr.02	101350419..101351821	II -d	WRKYGQK	C2H2	228	25824.97	5.395
NtWRKY109	Ntab0616620	Chr.01	69844857..69847103	II -d	WRKYGQK	C2H2	339	38228.82	5.692
NtWRKY110	Ntab0666420	0	Chr.0:37665..42158	II -d	WRKYGQK	C2H2	159	18649.78	8.6824
NtWRKY111	Ntab0690030	0	Chr.0:25739..29347	II -d	WRKYGQK	C2H2	185	21498.82	8.4477
NtWRKY112	Ntab0952100	Chr.17	10944410..10945663	II -d	WRKYGQK	C2H2	334	37246.33	9.9177
NtWRKY113	Ntab0958540	Chr.16	36399348..36402030	II -d	WRKYGQK	C2H2	498	55688.19	7.7084
NtWRKY114	Ntab0984540	Chr.19	54687269..54690810	II -d	WRKYGQK	C2H2	224	25918.17	8.0675
NtWRKY115	Ntab0051300	Chr.01	268460290..268463107	II -e	WRKYGQK	C2H2	403	44187.33	8.0161
NtWRKY116	Ntab0078310	Chr.14	80705578..80706831	II -e	WRKYGQK	C2H2	304	34305.04	6.0683

续表

基因名称	定位名称	染色体	定位位置	类型	WRKY 结构域	锌指结构	氨基酸数目	分子量(Da)	等电点(pI)
NtWRKY117	NtAb0102080	Chr.24	40746621..40750213	II -e	WRKYGQK	C2H2	450	48991.02	6.3522
NtWRKY118	NtAb0189140	Chr.15	85527443..85531363	II -e	WRKYGQK	C2H2	225	25914.17	9.3926
NtWRKY119	NtAb0318540	Chr.09	46315510..46318021	II -e	WRKYGQK	C2H2	500	55621.08	8.6155
NtWRKY120	NtAb0375810	0	Chr.0:330193..330193	II -e	WRKYGQK	C2H2	223	25154.15	6.6706
NtWRKY121	NtAb0400140	Chr.03	64386913..64388208	II -e	WRKYGQK	C2H2	322	35193.09	10.1551
NtWRKY122	NtAb0437000	Chr.21	57392952..57396764	II -e	WRKYGQK	C2H2	532	57475.34	7.4727
NtWRKY123	NtAb0438760	Chr.14	26122019..26133329	II -e	WRKYGQK	C2H2	615	68525.34	8.1436
NtWRKY124	NtAb0571100	Chr.04	153127933..153130234	II -e	WRKYGQK	C2H2	353	38785.12	8.141
NtWRKY125	NtAb0577580	Chr.06	203500650..203502743	II -e	WRKYGQK	C2H2	356	39125.67	7.7633
NtWRKY126	NtAb0605350	Chr.13	30418622..30420195	II -e	WRKYGQK	C2H2	329	36676.71	9.9926
NtWRKY127	NtAb0755110	Chr.11	93775274..93777758	II -e	WRKYGQK	C2H2	562	60990.17	5.9511
NtWRKY128	NtAb0870630	Chr.15	439887..441945	II -e	WRKYGQK	C2H2	280	31246.98	7.7309
NtWRKY129	NtAb0883710	Chr.09	78068775..78071543	II -e	WRKYGQK	C2H2	231	26240.76	9.2003
NtWRKY130	NtAb0909610	Chr.04	80064398..80066406	II -e	WRKYGQK	C2H2	285	33131.9	6.4389
NtWRKY131	NtAb0922970	Chr.04	281744098..281745623	II -e	WRKYGQK	C2H2	168	19257.2	6.1228
NtWRKY132	NtAb0927560	Chr.04	253018566..253020642	II -e	WRKYGQK	C2H2	246	28421.99	6.668
NtWRKY133	NtAb0936620	Chr.06	115205884..115208394	II -e	WRKYGQK	C2H2	337	37766.54	6.7901
NtWRKY134	NtAb0010790	Chr.09	80215216..80217014	III	WRKYGQK	C2HC	364	41189.62	5.797
NtWRKY135	NtAb0187140	Chr.01	24961019..24967260	III	WRKYGQK	C2HC	503	55187.42	6.6533
NtWRKY136	NtAb0199840	Chr.23	65690556..65693084	III	WRKYGQK	C2HC	304	34396.88	8.3066
NtWRKY137	NtAb0223500	Chr.02	89517041..89519247	III	WRKYGQK	C2HC	282	31756.7	7.9999
NtWRKY138	NtAb0244970	Chr.01	104152645..104154606	III	WRKYGQK	C2HC	300	34841.01	6.7375
NtWRKY139	NtAb0256880	Chr.06	123881798..123883861	III	WRKYGQK	C2HC	308	34480.27	6.62
NtWRKY140	NtAb0291920	Chr.22	16970338..16972442	III	WRKYGQK	C2HC	323	36505.33	6.794
NtWRKY141	NtAb0294710	Chr.20	72865516..72867056	III	WRKYGQK	C2HC	350	38733.84	10.2756
NtWRKY142	NtAb0317380	Chr.11	85868660..85871534	III	WRKYGQK	C2HC	273	29794.84	5.6302
NtWRKY143	NtAb0386190	Chr.16	92707822..92710006	III	WRKYGQK	C2HC	385	43287.21	7.8093
NtWRKY144	NtAb0394410	Chr.18	12132636..12134529	III	WRKYGQK	C2HC	316	35223.97	7.7983
NtWRKY145	NtAb0569060	Chr.06	34785316..34786698	III	WRKYGQK	C2HC	359	39559.47	6.4186

续表

基因名称	定位名称	染色体	定位位置	类型	WRKY 结构域	锌指结构	氨基酸数目	分子量(Da)	等电点(pI)
NtWRKY146	Ntab0627810	Chr.19	27951150..27952683	III	WRKYGQK	C2HC	361	40847.1	5.4077
NtWRKY147	Ntab0627820	Chr.19	28007778..28010587	III	WRKYGQK	C2HC	304	34396.88	8.3066
NtWRKY148	Ntab0634070	Chr.03	118548508..118553173	III	WRKYGQK	C2HC	160	18577.76	8.7721
NtWRKY149	Ntab0718970	Chr.18	16666627..16668553	III	WRKYGQK	C2HC	319	36819.53	5.3088
NtWRKY150	Ntab0721040	Chr.16	8991757..8997313	III	WRKYGQK	C2HC	227	26352.04	9.3864
NtWRKY151	Ntab0728350	Chr.01	80551800..80552585	III	WRKYGQK	C2HC	261	29730.77	6.4299
NtWRKY152	Ntab0771130	Chr.03	162957337..162961161	III	WRKYGQK	C2HC	633	69136.41	7.1404
NtWRKY153	Ntab0840710	0	Chr.0:237..5345	III	WRKYGQK	CGHC	341	37805.08	6.3135
NtWRKY154	Ntab0847390	Chr.21	55658980..55662400	III	WRKYGQK	C2HC	525	57172.28	8.0436
NtWRKY155	Ntab0870450	Chr.15	151251..155599	III	WRKYGQK	C2HC	266	29657.15	9.72
NtWRKY156	Ntab0870650	Chr.15	454270..462558	III	WRKYGQK	C2HC	591	66398.51	6.987
NtWRKY157	Ntab0876820	Chr.01	115042109..115044047	III	WRKYGQK	C2HC	287	32157.73	5.06
NtWRKY158	Ntab0907490	Chr.18	75479504..75482338	III	WRKYGQK	C2HC	230	26447.82	8.9211
NtWRKY159	Ntab0910320	Chr.13	18994640..18997293	III	WRKYGQK	C2HC	482	53182.02	8.2402
NtWRKY160	Ntab0915540	Chr.22	39625938..39627930	III	WRKYGQK	C2HC	191	22077.23	6.8606
NtWRKY161	Ntab0926010	Chr.06	146258160..146259494	III	WRKYGQK	C2HC	372	41725.13	6.5316
NtWRKY162	Ntab0931830	Chr.15	19071600..19074060	III	WRKYGQK	C2HC	498	55198.12	7.0315
NtWRKY163	Ntab0986740	Chr.07	31989332..31992607	III	WRKYGQK	C2HC	321	36330.1	5.1161
NtWRKY164	Ntab0159100	0	Chr.0:745141..747650	III	WRKYGQK	C2HC	511	56293.13	6.4277

注：将暂无定位信息的 NtWRKY 基因暂时定位在假设的 0 号染色体上。