



张元野,张利国,张明,等.工业大麻 WRKY 家族全基因组鉴定及生物信息学分析[J].黑龙江农业科学,2024(10):17-26.

工业大麻 WRKY 家族全基因组鉴定 及生物信息学分析

张元野,张利国,张明,房郁研,郑楠,闫博巍,隋月

(黑龙江省农业科学院 经济作物研究所,黑龙江 哈尔滨 150086)

摘要:WRKY 转录因子是一类具有多种生物学功能的基因,工业大麻中 WRKY 家族研究较少。本研究使用拟南芥 WRKY 家族蛋白序列与工业大麻全基因组进行比对,鉴定到 34 个 WRKY 家族成员,不均等分布于 9 条染色体上,同时可分为 Group I、Group II a~Group II e、Group III 共 7 个亚族,其中 Group I 亚族成员数量最多,数量为 11 个;CsWRKY15 包含的结构域最多,CsWRKY34 启动子区域的顺式作用元件最多,二者可能行使着较多生物学功能;工业大麻 WRKY 家族与拟南芥、水稻、玉米、大豆及大麻自身间共线性分析显示,CsWRKYs 存在丰富的直系同源基因进化关系。

关键词:工业大麻;WRKY;蛋白理化性质;同源性分析;生物信息学

工业大麻(*Cannabis sativa* L.),又名汉麻、火麻、线麻、寒麻等,为桑科(Moraceae)大麻亚科(Subfam Cannabioideae)一年生植物,在我国具有悠久的栽培历史^[1-2]。工业大麻的麻茎可沤麻提取纤维^[3-4],麻籽既可加工食品也能用于榨油^[5-6],花叶中蕴含的萜烯酚类化合物集医疗、美容、保健功能^[7-10]于一体,多样化的功能使得工业大麻成为经济效益优良的作物。

WRKY 转录因子是植物中最庞大的转录因子家族之一,普遍在氨基酸 N-端包含有 WRKYGQK 保守结构域,C-端存在 C₂H₂或 C₂HC 锌指结构^[11]。根据结构域和锌指结构的差异,WRKY 转录因子可分为:Group I、Group II、Group III 三类,Group II 进一步分为 a~e 5 个亚类^[12-13],此外,因 Group II 亚族部分差异较小,Rushton 等^[14]将其合并为 II a+II b、II c、II d+II e 等 3 个亚类。众多研究表明 WRKY 转录因子参与多种生理进程,包括:生物和非生物胁迫^[15],种子的形成、发育、休眠和萌发^[16-17],叶片发育和衰老^[18],生物碱类、萜类、黄酮类等次生代谢物合成^[19-20]等,目前在拟南芥^[15]、水稻^[16]、大豆^[17]、油菜^[18]、番茄^[19]、小麦^[20]、黄连^[21]等作物中均成功鉴定到不同数量的 WRKY 转录因子,生物学功能研究也较为丰富^[22-25]。大豆中过表达 *GmWRKY12* 提高了

抗旱性和耐盐性^[26],陆地棉中过表达 *GhWRKY33* 同样提升了抗旱能力^[27],玉米 *ZmWRKY101* 过表达植株的耐盐能力得到了显著增强^[28],拟南芥过表达 *AtWRKY25* 和 *AtWRKY26* 植株的耐热性增强^[29],*AtWRKY41* 功能缺失突变体的花青素含量显著提高^[30],三七中过表达 *PnWRKY1* 显著提高了三七皂苷的合成量^[31],WRKY 转录因子的深入研究,有助于提升植物抗逆性、次生代谢合成等能力。

工业大麻中 WRKY 转录因子的相关研究较少,Xin 等^[32]对响应非生物胁迫的 WRKY 家族成员进行分析,其中有 25 个 CsWRKYs 与非生物胁迫显著相关,在胁迫下呈现上调或下调。Wei 等^[33]选用籽用麻作为研究材料,鉴定到 39 个 CasWRKYs,发现 13 个 CasWRKYs 响应 GA₃。Iranbakhsh 等^[34]使用冷等离子体对大麻种子进行处理,30 d 后对幼苗进行处理和分析,与对照相比,转录因子 WRKY1 上调了 11.55 倍,CBDAS 转录水平提高了 12.4 倍,THCAS 转录水平提高了 25.6 倍,WRKY1 可能与大麻素合成呈现正相关。

顽拗型再生性使得工业大麻组培进展和分子育种技术的相关研究远不如其他作物丰富,许多其他作物中运用的分子技术很难在大麻中实

收稿日期:2024-04-10

基金项目:黑龙江省农业科技创新跨越工程——经济作物突破性新品种选育及产业化应用(GX23GG04)。

第一作者:张元野(1998—),男,硕士,研究实习员,从事工业大麻遗传育种研究。E-mail:1965263812@qq.com。

通信作者:张利国(1978—),男,博士,研究员,从事麻类遗传基础研究。E-mail:zlg86@aliyun.com。

现^[35],针对 WRKY 家族生物学功能的研究也少之又少。工业大麻 WRKY 转录因子家族的全基因鉴定,有助于在不依靠其他分子生物学手段的情况下提前了解 WRKY 家族的基因结构,并进行功能预测,对 WRKY 家族的生物学研究具有一定指导意义,进一步对大麻基因组深度剖析能够为育种提供新方向。

1 材料与方法

1.1 工业大麻 WRKY 转录因子的全基因组鉴定

拟南芥和工业大麻的全基因组序列(fasta 文件)及基因组注释文件(GFF3 格式)从 Ensembl Plants(<http://plants.ensembl.org/index.html>)数据库网站下载获得,拟南芥 WRKY 转录因子下载于植物转录因子数据库 PlantTFDB(<http://planttfdb.gao-lab.org/index.php>)。利用 TBtools 软件^[36]的 BLAST 功能将拟南芥 WRKY 家族氨基酸序列与工业大麻全基因组蛋白序列进行比对,设定为 E 值 $<1E-5$,获得比对文件 Blast.tab1;对 Blast.tab1 中的数据进行初步处理,排序并去除重复项和无义项。在 Pfam 数据(<http://pfam-legacy.xfam.org/>)库中,获得 WRKY 转录因子保守结构域序列比对文件,利用 HMMER 软件构建隐马尔可夫模型,与大麻全基因组蛋白序列进行比对,获得比对文件 Blast.tab2;对 Blast.tab2 中的数据进行初步处理,排序并去除重复项和无义项。取 Blast.tab1 和 Blast.tab2 的交集,并利用 NCBI 中的 Batch CDD search (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>)对结构域进行分析,剔除掉不包含 WRKY 家族结构域的蛋白序列,最终得到大麻 WRKY 转录因子家族的氨基酸序列文件。

1.2 WRKY 转录因子的蛋白理化性质预测、染色体定位及进化分析

通过 ExPASy 在线网站(<https://web.expasy.org/protparam/>)预测大麻 WRKY 转录因子的蛋白长度、分子质量、等电点、不稳定系数、亲水性等理化性质,通过在线网站 WOLF PSORT (<https://wolfsort.hgc.jp/>)预测亚细胞定位。根据鉴定出的工业大麻 WRKY 转录因子家族信息和基因组注释文件,利用 TBtools 的 Gene location visualize from GTF/GFF 功能进行染色体定位。使用 MAGE11 对大麻 WRKY 转录因

子家族的氨基酸序列进行比对,并构建系统进化树,使用在线网站 Evolview(<https://www.evolgenius.info/evolview/>)美化进化树。

1.3 大麻 WRKY 转录因子家族基因结构分析

利用 MEME (<https://meme-suite.org/meme/tools/meme>)对大麻 WRKY 家族成员蛋白序列进行保守 Motif 预测,设定 Motif 个数为 10,获得预测图。利用 NCBI 中的 Batch CDD Search 对大麻 WRKY 家族成员蛋白序列进行保守结构域分析。使用 TBtools 软件的 Gene Structure View 结合系统进化树、保守 Motif 预测、保守结构域和内含子-外显子结构作图,对大麻 WRKY 转录因子家族基因结构进行可视化。

1.4 大麻 WRKY 转录因子家族启动子顺式作用元件预测

通过 Plant CARE 网站(<https://bioinformatics.psb.ugent.be/>)的 search for care 工具对大麻 WRKY 转录因子启动子上游 2 000 bp 的基因序列进行启动子顺式作用元件预测,剔除重复元件、无名元件及 TATA/CAAT-box 等,并对筛选后的顺式作用元件进行注释,最后利用 TBtools 软件的 Simple BioSequence Viewer 功能进行可视化,辅助 Graphpad prism 8.0.2 完成顺式作用元件堆叠图,并对二者进行结合和美化。

1.5 大麻 WRKY 基因家族同源性分析

结合大麻、拟南芥、水稻、玉米和大豆的基因组文件及注释文件,使用 TBtools 软件的 One step MCScanX 功能得到与 4 种植物的直系同源基因信息,分别以“*At-N*”“*Os-N*”“*Zm-N*”“*Gm-N*”“*Cs-N*”表示拟南芥、水稻、玉米、大豆和大麻的染色体,N 为染色体编号,使用 TBtools 软件的 Dual systemy plot for MCScanX 功能绘制大麻与 4 种植物的直系同源基因图。同时,使用 TBtools 软件的 One step MCScanX 功能对大麻物种内直系同源基因信息进行分析,使用 Advanced Circos 功能对获得的数据进行可视化。

2 结果与分析

2.1 工业大麻 WRKY 转录因子的鉴定和蛋白理化性质分析

经过 BLAST 比对、隐马尔可夫模型比对和 WRKY 结构域鉴定,并剔除假基因后,最终在工业大麻基因组中鉴定到 34 个 WRKY 基因,为了

便于后续的分析,按照基因号顺序命名为 CsWRKY1~CsWRKY34,根据拟南芥 WRKY 分类标准,将 CsWRKYs 分为 5 个亚族。工业大麻 WRKY 转录因子家族理化性质详情如表 1 所示,其中 CsWRKY24 编码的序列最短,为 88 个氨基酸,分子量为 10 572.98 Da,CsWRKY13 编码的序列最长,达 745 个氨基酸,分子量为 83 259.77 Da;共有 15 个 CsWRKYs 为酸性蛋白,其余 19 个

CsWRKYs 为碱性蛋白;所有 CsWRKYs 蛋白的平均亲水系数均小于 0,全部是亲水性蛋白;所有工业大麻 WRKY 转录因子蛋白的不稳定系数均大于 30,证明其在植物体内稳定存在的时间相对较短;除 CsWRKY14 和 CsWRKY24 的亚细胞定位显示在细胞质,其他 CsWRKYs 成员亚细胞定位均预测在细胞核内。

表 1 大麻 WRKY 家族成员理化性质及基本信息

基因名称	氨基酸数目	分子量/Da	等电点 PI	不稳定系数	脂肪系数	平均亲水系数	亚细胞定位
CsWRKY1	345	38888.52	5.92	54.37	62.14	-0.607	细胞核
CsWRKY2	345	38845.40	5.70	53.68	63.28	-0.611	细胞核
CsWRKY3	344	38689.22	5.61	53.31	63.46	-0.600	细胞核
CsWRKY4	285	31392.60	8.61	41.43	47.58	-0.909	细胞核
CsWRKY5	639	69732.13	6.97	48.60	61.74	-0.748	细胞核
CsWRKY6	478	52140.70	8.13	48.53	61.95	-0.746	细胞核
CsWRKY7	279	30891.57	9.16	46.01	55.95	-0.910	细胞核
CsWRKY8	187	22302.42	5.41	53.60	58.29	-1.341	细胞核
CsWRKY9	208	23711.22	9.24	47.11	57.64	-0.961	细胞核
CsWRKY10	204	23141.60	9.15	45.91	58.77	-0.941	细胞核
CsWRKY11	207	23555.03	9.11	46.46	57.92	-0.944	细胞核
CsWRKY12	175	19669.51	9.72	59.30	58.46	-0.699	细胞核
CsWRKY13	745	83259.77	8.62	47.56	61.38	-0.767	细胞核
CsWRKY14	246	28988.19	8.56	36.86	85.61	-0.564	细胞质
CsWRKY15	299	33021.46	9.96	50.52	64.92	-0.625	细胞核
CsWRKY16	382	42566.28	9.29	46.93	59.50	-0.873	细胞核
CsWRKY17	562	61087.35	6.36	49.97	62.10	-0.729	细胞核
CsWRKY18	662	70701.58	6.05	57.79	51.47	-0.787	细胞核
CsWRKY19	473	53305.58	6.63	58.22	47.59	-1.039	细胞核
CsWRKY20	509	55499.94	7.72	61.93	49.63	-0.911	细胞核
CsWRKY21	508	55343.75	7.26	61.69	49.72	-0.904	细胞核
CsWRKY22	380	42655.03	5.39	44.58	72.34	-0.718	细胞核
CsWRKY23	259	28785.42	7.29	43.45	49.34	-0.854	细胞核
CsWRKY24	88	10572.98	9.17	51.90	59.66	-0.851	细胞质
CsWRKY25	224	25321.54	6.22	44.29	39.96	-1.054	细胞核
CsWRKY26	437	48152.51	8.90	48.53	59.34	-0.839	细胞核
CsWRKY27	279	31406.68	8.07	53.72	52.72	-1.137	细胞核
CsWRKY28	371	42688.71	4.97	32.41	62.24	-0.996	细胞核
CsWRKY29	370	42510.08	5.49	66.05	57.16	-0.906	细胞核
CsWRKY30	590	64376.87	6.72	59.86	48.31	-0.959	细胞核
CsWRKY31	375	40715.81	5.40	66.62	41.95	-1.004	细胞核
CsWRKY32	410	45662.16	6.04	55.81	48.59	-0.860	细胞核
CsWRKY33	192	21622.89	9.47	54.89	53.85	-0.817	细胞核
CsWRKY34	111	13326.10	9.70	42.27	47.30	-1.318	细胞核

2.2 工业大麻 WRKY 转录因子系统进化分析

本研究将工业大麻和拟南芥 WRKY 转录因子进行系统进化树分析,并按照拟南芥 WRKY 转录因子家族的分类方式进行分类,根据分类标准,将 CsWRKYs 分为 5 个亚族,如图 1 所示,Group I 亚族 11 个成员,包括 CsWRKY13/17/

18/19/20/21/25/26/27/30/31 等;Group II a+II b 亚族 6 个成员,包括 CsWRKY5/6/7/8/14/16; Group II c 亚族 8 个成员,包括 CsWRKY9/10/11/23/24/32/33/34 等;Group II d+II e 亚族 4 个成员,包括 CsWRKY4/12/15/28 等;Group III 亚族 5 个成员,包括 CsWRKY1/2/3/22/29 等。

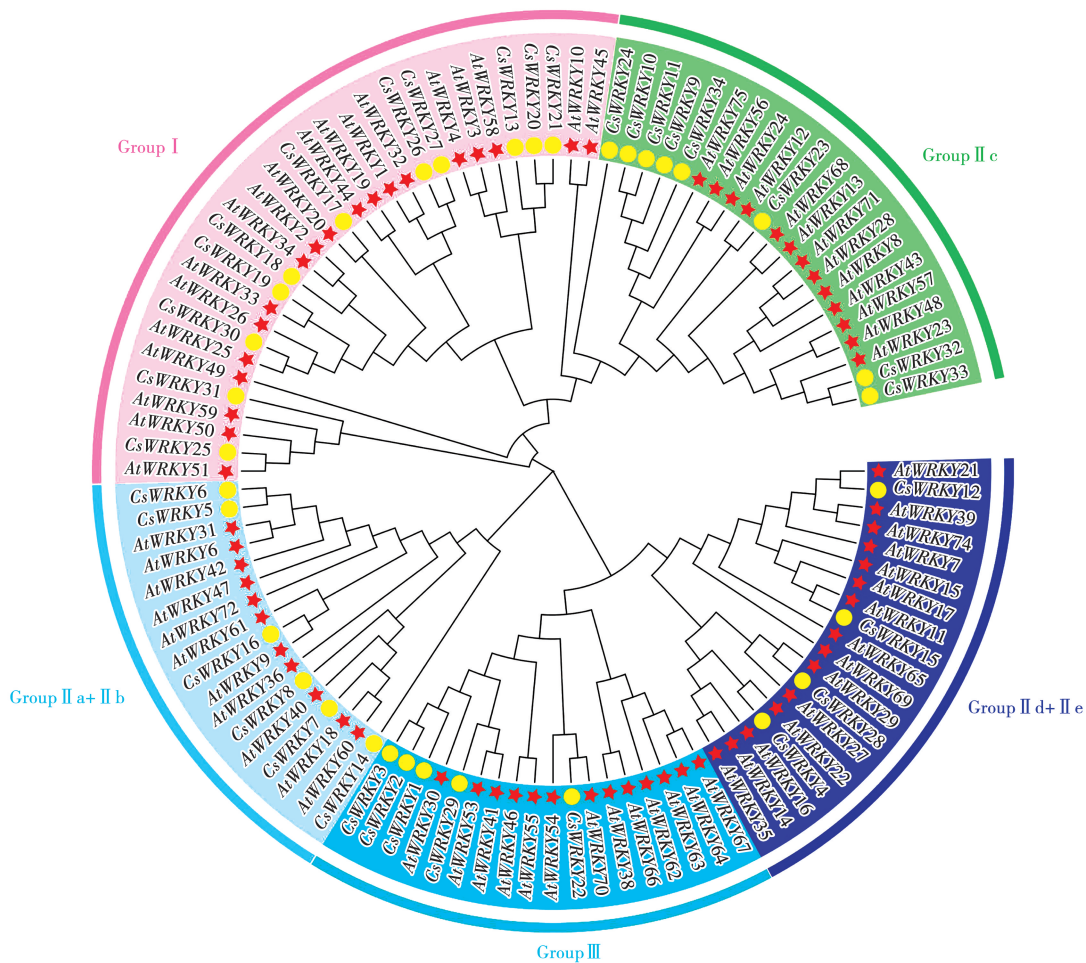


图 1 大麻和拟南芥 WRKYs 基因的系统进化分析

2.3 工业大麻 WRKY 家族的染色体定位

工业大麻共有 10 对染色体,WRKY 转录因子家族成员除 Chr7 染色体上无分布外,在其他染色体上均有分布,如图 2 所示,将大麻染色体依次命名为 Chr1~Chr10,Chr1 上分布有 3 个 CsWRKYs (CsWRKY23/32/33),Chr2 上分布有 8 个 CsWRKYs (CsWRKY1/2/3/4/5/6/7/22),Chr3 上分布有 3 个 CsWRKYs (CsWRKY24/26/27),Chr4 上分布有 5 个 CsWRKYs (CsWRKY8/9/10/11/25),Chr5

上分布有 3 个 CsWRKYs (CsWRKY12/28/29), Chr6 上分布有 4 个 CsWRKYs (CsWRKY13/14/30/31),Chr8 上分布有 4 个 CsWRKYs (CsWRKY15/16/17/18),Chr9 上分布有 1 个 CsWRKYs (CsWRKY19), Chr10 上分布有 3 个 CsWRKYs (CsWRKY20/21/34)。其中 CsWRKY8 较为靠近 Chr4 的端粒区,鉴于精度的问题,其他类似的定位于染色体末端的基因实际上距离端粒区有很长的一段序列。

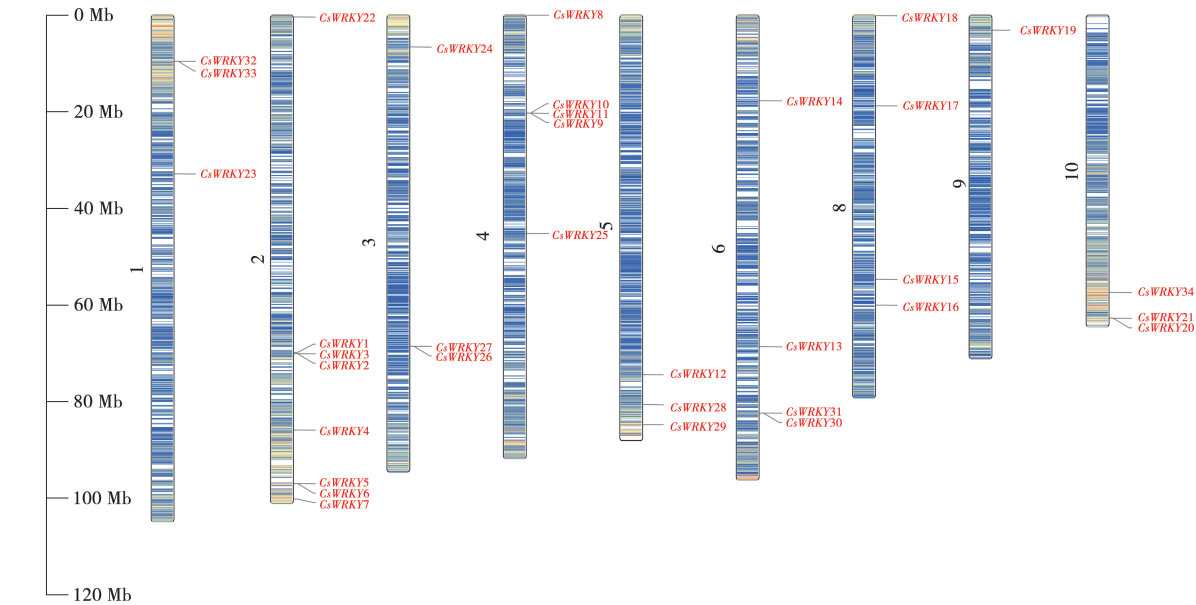
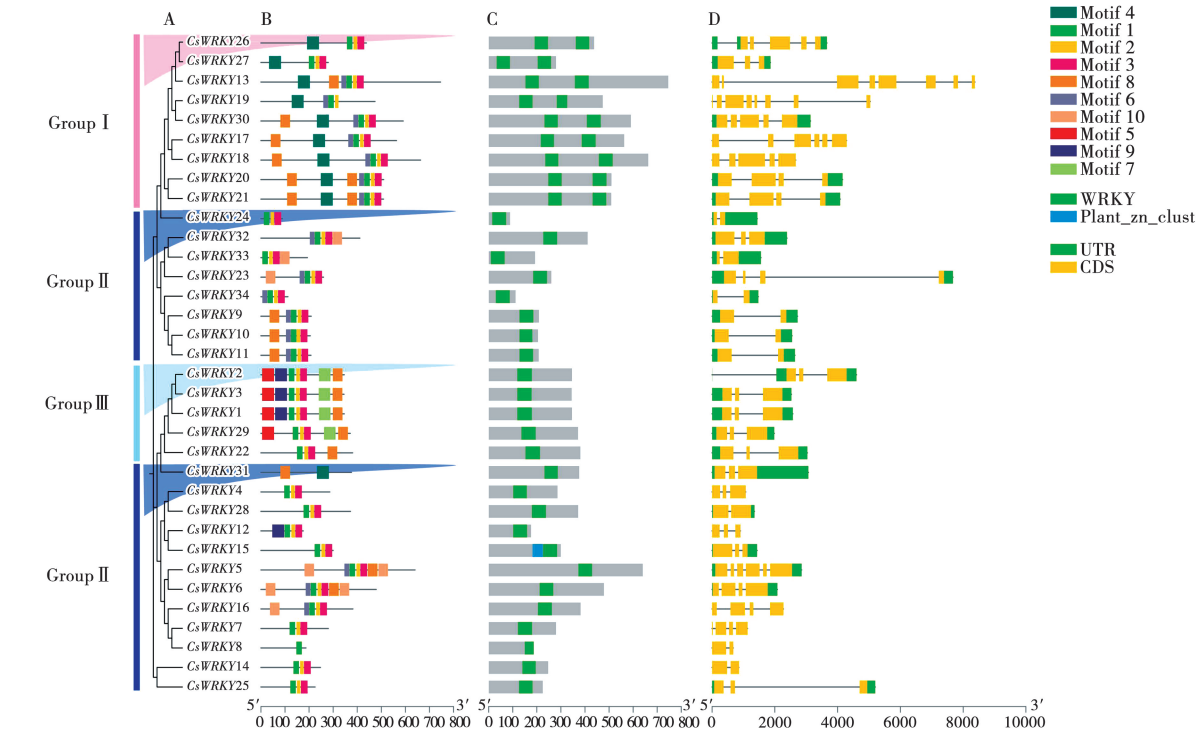


图 2 大麻 WRKY 家族的染色体定位

2.4 大麻 WRKY 家族基因结构分析

分别获得系统进化树、保守 Motif 预测、保守结构域和内含子-外显子的数据或图片后,使用 TBtools 进行整合,最终结果如图 3 所示。通过进化树对大麻 WRKY 转录因子家族进行了分类,共分为 3 个亚族;从 Motif 预测结果可以看出,不同的亚族之间的 Motif 结构,存在明显的差

异,Group I 表现出 Motif4 在前、其他 Motif 聚团在后的趋势,类似地,Group III 亚族表现为 Motif5 在前、其他 Motif 聚团在后的趋势,而 Group II 亚族基本表现为所有 Motif 聚团的状态。*CsWRKY31* 是 Group II 亚族特殊的一员,该成员的 Motif 排列与 Group I 家族前半部分相似。



A. 进化树;B. Motif 预测;C. 蛋白结构域;D. 基因结构。

图 3 大麻 WRKY 家族成员的蛋白质结构域和基因结构

Group I 的 WRKY 结构域与其他亚族差异明显,具有 2 个 WRKY 结构域,而 *CsWRKY31* 仅包含有 1 个 WRKY 结构域,结合图 B 和图 C 发现 *CsWRKY30*、*CsWRKY31* 的基因结构十分相似,二者在染色体上处于临近位置,前半部分氨基酸序列和内含子排列高度相似,且 *CsWRKY31* 的 CDS+UTR 长度与 *CsWRKY30* 差异较小,两者在进化过程中可能来自同一祖先,*CsWRKY31* 中间位置的碱基产生了突变使得后续的部分序列由 CDS 转变为 UTR,导致将其归类为 Group II 亚族;从 Motif 结构上看,*CsWRKY31* 具有 Group I 亚族特有的 Motif4 和 Motif8,同样证实其有 Group I 亚族的特征,*CsWRKY30*、*CsWRKY31* 可能是旁系同源基因。类似地,*CsWRKY20/21* 和 *CsWRKY9/10/11* 分别为旁系同源基因,2 个基因簇各自的 Motif 预测、WRKY 结构域及内含子-外显子结构相似度均较高,仅序列长短存在差异,可能在进化过程中缺失或增加了部分碱基;Group III 亚族的 5 个成员 *CsWRKY1/2/3/23/30* 的 Motif 预测、WRKY 结构域相似度较高,基因长度差异显著,但内含子-外显子结构差异较小,表现为 UTR+3 个内含子的结构。

Group II 5 个家族的基因长度、内含子-外显子结构差异丰富,Motif 预测各有特点,不同家族的 Motif 排列顺序存在差异。Group II c 亚族中存在一个特殊的成员 *CsWRKY15*,除 WRKY 蛋白结构域外,还包含 Plant_zn_clust 结构域,该结构域为植物锌簇域,与 WRKY 结构域相关性很高。此外,在早期的筛选过程中,删除了 10 个鉴

定到的包含 WRKY Superfamily 结构域但无 WRKY 结构域的成员,WRKY Superfamily 与 WRKY 结构域相似度很高,可能具有类似的功能,有待进一步分析鉴定。

2.5 大麻 WRKY 家族启动子区域顺势作用元件分析

通过 Plant CARE 网站对大麻 WRKY 转录因子家族启动子区域 2 000 bp 内的顺势作用元件进行了预测,其结果如图 4 所示,各 *CsWRKYs* 的顺势作用元件数量和类别差异显著。*CsWRKY20/21* 和 *CsWRKY9/10/11* 这两类核苷酸序列和蛋白结构域相似性较高的基因簇,在顺势作用元件的预测中相似性较低;如 *CsWRKY9/10/11* 在染色体上定位于相邻位置,甚至部分序列存在重叠,预测到的顺势作用元件同样存在差异,*CsWRKY9* 相较于 *CsWRKY10/11* 包含有 2 种特异性的元件:ATBP-1 结合位点(1 个)、胚乳调控(2 个),类似地,*CsWRKY10* 包含有 1 个玉米蛋白代谢元件,*CsWRKY11* 包含有 1 个昼夜节律调控元件;*CsWRKY20/21* 同样在染色体上定位于相邻位置,部分序列存在重叠,启动子区域顺势作用元件的差异较为显著,*CsWRKY20* 相较于 *CsWRKY21* 多了 5 种特异性元件:生长素响应(1 个)、水杨酸响应(3 个)、赤霉素响应(1 个)、玉米蛋白代谢(1 个)和栅栏细胞调节(1 个),而 *CsWRKY21* 包含茉莉酸甲酯响应(2 个)、昼夜节律(1 个)等两种特异性元件。此类旁系同源基因簇中的基因,可能在多种响应调节中扮演不同的角色。

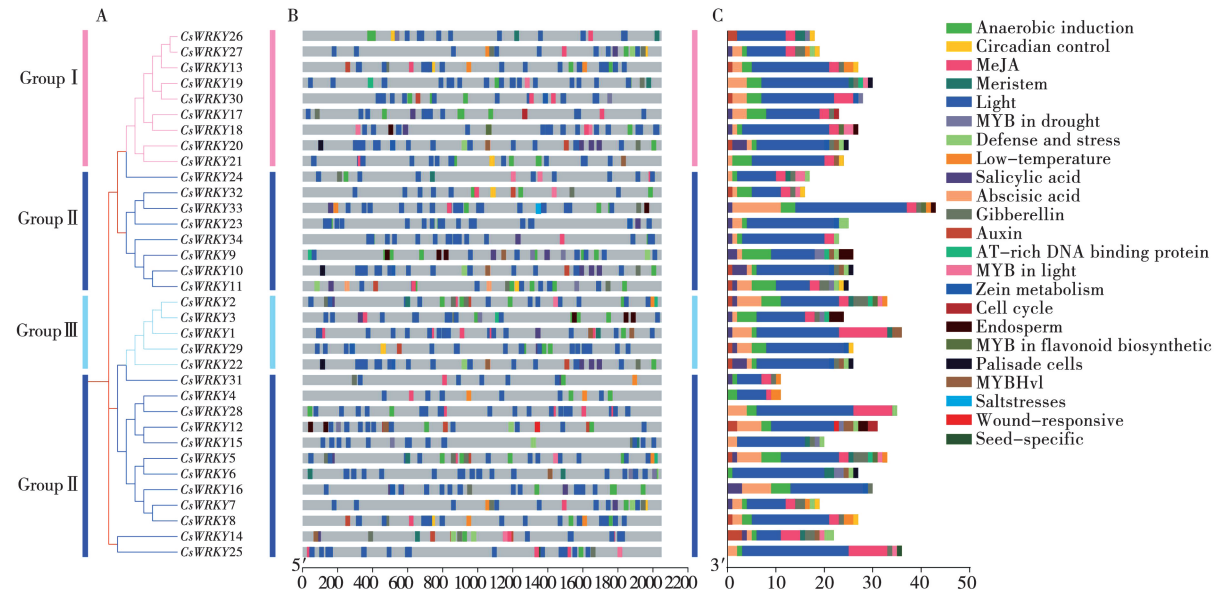


图 4 大麻 WRKY 家族成员启动子区顺势作用元件分析

大麻 WRKY 转录因子家族共预测到 869 个顺势作用元件,其中光响应元件为 452 个,占比为 52.01%,可见光照确实为影响植物生长生存的重要环境因素;生长素、水杨酸、脱落酸、茉莉酸甲酯及赤霉素等激素响应元件为 223 个,占比为 25.66%;厌氧反应、损伤反应、干旱、低温、昼夜节律及防御与胁迫等环境响应元件为 126 个,占比为 14.49%;玉米代谢、分生组织、类黄酮类合成、细胞周期及栅栏细胞等体内调节元件数量较少,共 32 个,仅占整体的 3.56%,此外,CsWRKY33 包含 1 个盐胁迫响应元件,其他 CsWRKYs 中并未出现,CsWRKY12 包含 1 个损伤反应响应元件,CsWRKY25 包含 1 个种子特异性调控元件。从响应元件的数量上预测,工业大麻 WRKY 转录因子家族主要参与对外界环境的响应工作。CsWRKY33 预测到顺势作用元件的数量最多,达 44 个,其中光响应元件为 23 个、激素响应元件 16 个,CsWRKY33 可能在光响应和激素响应方面有重要的作用,并参与到多条代谢通路中。

2.6 大麻 WRKY 转录因子家族的基因同源性

通过 TBtools 的 Dual syteny plot for MCScanX 功能对大麻 WRKY 转录因子家族基因与其他植物 WRKY 家族进行共线性分析,分别选用模式植

物拟南芥及其他 3 种主要作物水稻、玉米、大豆进行比对,结果如图 5 所示。在大麻 WRKY 家族与拟南芥的比对中,有 25 对基因呈现共线性,水稻中 11 对,玉米中 8 对,大豆中共线基因较多,存在 77 对;不同的植物之间,基因存在相似性,在进化中可能来自统一祖先,因其种属分类等差异,共线性基因数量存在明显的差异。为了解进化过程中自然选择的作用,对 4 个比对结果中的共线基因对进行提取,置入 Simple ka/ks calculator 功能中进行 ka/ks 分析,如 $ka/ks < 1$,说明基因进行纯化选择,意图消除有害突变;如 $ka/ks = 1$,说明基因发生中性选择,如 $ka/ks > 1$,说明基因受到正向选择,新产生的突变基因在当前环境存在一定优势。值得一提的是,4 个物种共 121 对共线性基因,其中 82 对 ka/ks 均小于 1,证明进化过程中进行了纯化选择,削弱蛋白变异的幅度;剩余 39 对为负值,不符合 ka/ks 的取值范围,标定为 NaN。实际原因为同义突变较多、核苷酸序列差异大,存在丰富的同义突变使得氨基酸序列差异较小,共线性通过蛋白比对实现而非基因序列,因而所有类似蛋白的 ka/ks 比值被识别为错误结果,可理解为蛋白存在共线性,但对应基因序列的共线性较弱,无法计算 ka/ks 。

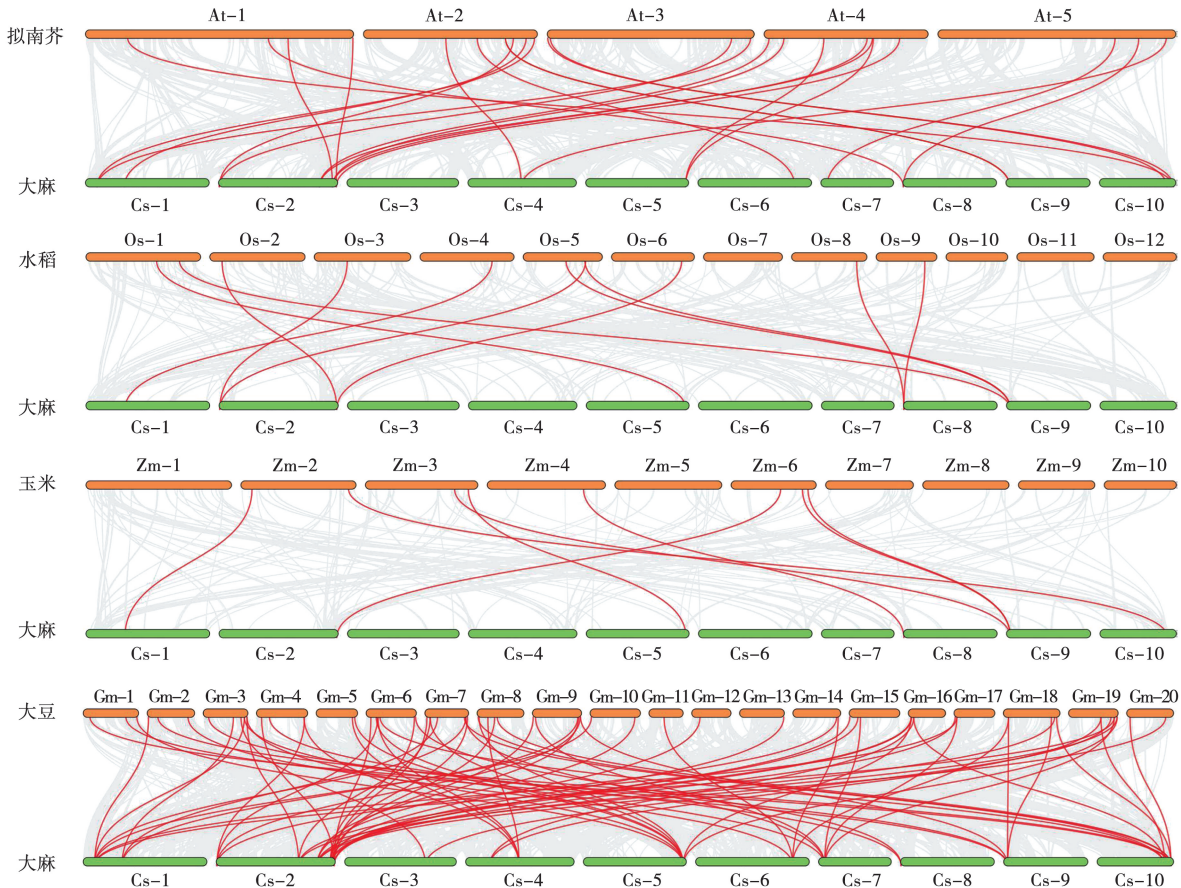


图 5 大麻与拟南芥、水稻、玉米和大豆的 WRKY 成员间的共线性关系

在大麻 WKRY 转录因子家族与 4 种高等植物的共线性基因对中, *CsWRKY19* 的共线性基因对最多, 为 13 对, 占比 10. 7%, *CsWRKY7* 和 *CsWRKY29* 分别为 11 对和 12 对, *CsWRKY7*/19/29 的祖先可能在植物染色体上产生了相对较多的片段重复。4 种高等植物中, 大豆与 *CsWRKYs* 比对出的共线性基因对最丰富, 为 77 对, 其中 Chr1 上有 17 对, Chr2 上有 21 对, 其他染色体上分布较少, 大麻 Chr1 和 Chr2 上的 WKRY 家族成员可能与大豆 WKRY 家族亲缘关系更近, 此外大麻 *CsWRKYs* 共有 17 个成员与大豆 WKRY 家族存在共线性, 两者均是双子叶植物, 与水稻、玉米等单子叶植物确实可能存在更近的亲缘关系。进一步对大麻 WKRY 家族的物

种内共线性进行分析, 如图 6 所示, Chr1/2/3/4/6/10 等染色体上均存在 WKRY 基因簇, 呈现共线性的 WKRY 成员, 多归属于同一亚族, 如 *CsWRKY2* 和 *CsWRKY29* 存在共线性但定位于不同染色体上, 两者均属于 Group III 亚族; *CsWRKY9* 和 *CsWRKY34* 存在共线性且定位于不同染色体上, 两者属于 Group II 亚族, 原因可能是在进化过程中发生了染色体片段复制事件。值得一提的是, 在物种间共线性的 *ka/ks* 分析中 *CsWRKY1* 和 *CsWRKY2* 的 *ka/ks* 值 > 1, 证实其进行了正向选择, 将可能有益的突变保留下来, 剩下的共线性基因对 *ka/ks* 多数 < 1, 即发生了纯化选择, 部分共线性基因对被定义为 NaN, 即序列差异较大。

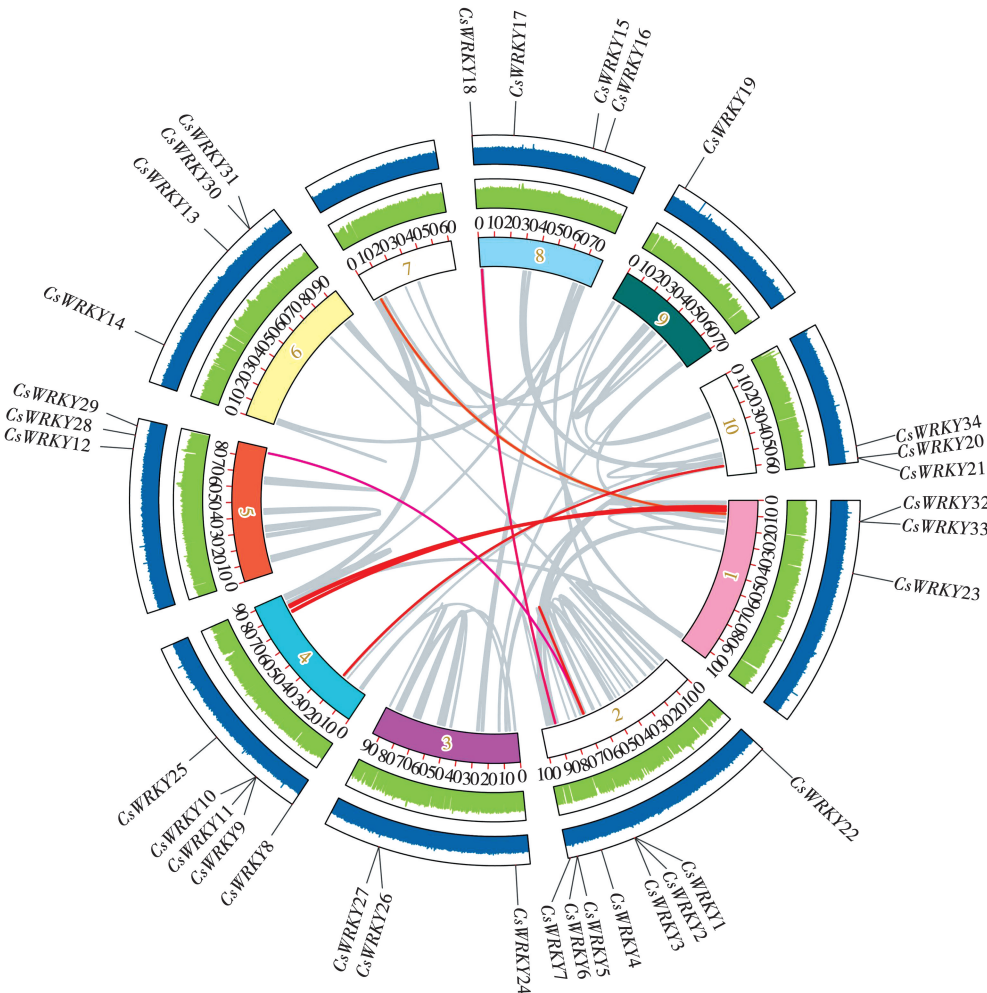


图 6 大麻物种内共线性关系

3 讨论

WRKY 转录因子是植物中重要的基因家族, 已证实参与到生物和非生物胁迫, 种子的形成、发

育、休眠和萌发, 叶片发育和衰老, 生物碱类、萜类、黄酮类等次生代谢物合成等重要生理进程。本研究在大麻中, 通过生物信息学分析鉴定到 34 个

WRKY 转录因子,可归类于 Group I、Group II、Group III 等 3 个亚族中,根据其他的分类方法,可将 Group II 亚族进一步分为 Group II a~Group II e 等 3 个小亚族。不同亚族间 WRKY 结构域、其他结构域、Motif 分布、内含子-外显子、启动子区域顺式作用元件等的数量和位置均存在差异,相同亚族间 WRKY 结构域、其他结构域、Motif 分布、内含子-外显子、启动子区域顺式作用元件等的数量和位置存在一定的相似性,同一亚族成员可能行使着近似但不完全相同的功能。

大麻 WRKY 家族的研究较少,Xin 等^[32]的研究中对大麻基因组进行分析,鉴定到 40 个 WRKY 基因,并构建了进化树,对 25 个 WRKY 成员在干旱、盐胁迫、镉胁迫等处理的表达量进行分析,证实大麻 WRKY 家族在非生物胁迫下响应中占据重要作用,40 个 WRKY 转录因子的基因结构、顺式作用元件、同源基因进化关系并未进一步展现。Wei 等^[33]选用籽麻作为试验材料,鉴定到 39 个 CasWRKYs,对 GA₃ 处理 CasWRKY 的表达情况进行了深入探究,证实籽用大麻的 WRKY 家族在纤维和茎的发育中具有重要作用。经过比对 Xin 等^[32]鉴定到的 40 个 CsWRKYs 中有 4 个成员与本研究定位到的基因显著相似,Wei 等^[33]鉴定到的 39 个 CasWRKYs 中有 2 个成员与本研究鉴定基因显著相似,而两者间存在 2 个相似鉴定基因。使用不同的高通量测序数据,定位得到的基因家族存在差异,通过比对能够得到数量较少的相似鉴定基因,不同测序数据中均能够鉴定得到的基因可能更具有真实性^[36]。

4 结论

本研究共鉴定出 34 个 WRKY 家族成员,对基因结构与分类、启动子区域顺式作用元件及同源基因进化关系进行了详细的分析,推测大麻 WRKY 家族各成员在植株生长过程中,可能扮演着不同的角色,具体生理功能有待继续研究发掘。此外,分析过程中剔除了 10 个无 WRKY 结构域的 WRKY Superfamily 家族成员,WRKY Superfamily 可能具有 WRKY 家族类似的功能,有待进一步分析鉴定。本研究为大麻 WRKY 家族成员的生物学功能研究奠定了数据基础。

参考文献:

[1] 郑其向,杨永红,罗欢,等. 不同程度干旱胁迫下火麻的生理形态变化及性别比例差异[J]. 中国麻业科学,2023,45(5): 206-214.

[2] 方兆平. 大麻纤维形态特征的鉴别及产品的开发利用[J]. 中国纤检,2010(7):82-84.

[3] 孙丹,孔令苹,孙颖,等. 大麻纤维制条工艺研究进展及思路[J]. 黑龙江纺织,2023(3):1-4.

[4] 董晓慧. 黑龙江省中南部地区纤用工业大麻雨露沤制技巧[J]. 园艺与种苗,2021,41(1):78-79.

[5] 齐丽颖,郭倩,赵明,等. 火麻仁蛋白及多肽提取工艺研究进展[J]. 高师理科学刊,2022,42(12):61-63,85.

[6] CERINO P, BUONERBA C, CANNAZZA G, et al. A review of hemp as food and nutritional supplement[J]. Cannabis and Cannabinoid Research, 2021, 6(1): 19-27.

[7] 王满生,杨晶,龚友才,等. 高钙型火麻仁植物蛋白饮料的研制[J]. 食品工业,2020,41(7):123-126.

[8] 王语聪,谢智鑫,杨文敏,等. 汉麻叶中大麻二酚的分离纯化及其抗抑郁功能[J]. 精细化工,2021,7(4):795-800,805.

[9] 李俊,朱雪雯,万会花,等. 大麻中大麻素类化学成分及其分析方法研究进展[J]. 中草药,2020,51(24):6414-6425.

[10] 王秋月,卢芳,刘树民. 大麻及大麻素药用价值的现代研究进展[J]. 中药药理与临床,2020,36(4):222-227.

[11] RUSHTON P J, SOMSSICH I E, RINGLER P, et al. WRKY transcription factors[J]. Trends in Plant Science, 2010,15(5):247-58.

[12] 倪辉,孙维红,丁乐,等. 香樟全基因组 WRKY 基因家族的鉴定与分析[J]. 植物科学学报,2022,40(4):513-523.

[13] EULGEM T, RUSHTON P J, ROBATZEK S, et al. The WRKY superfamily of plant transcription factors[J]. Trends in Plant Science, 2000, 5(5): 199-206.

[14] RUSHTON P J, BOKOWIEC M T, HAN S C, et al. Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae[J]. Plant Physiology, 2008, 147(1): 280-295.

[15] KHOSO M A, HUSSAIN A, RITONGA F N, et al. WRKY transcription factors (TFs): molecular switches to regulate drought, temperature, and salinity stresses in plants[J]. Frontiers in Plant Science, 2022, 13: 1039329.

[16] HE Y G, ZHU M H, LI Z H, et al. IPA1 negatively regulates early rice seedling development by interfering with starch metabolism via the GA and WRKY pathways[J]. International Journal of Molecular Sciences, 2021, 22(12): 6605.

[17] 陈乃钰,赵贺,蒋慧欣,等. 五种豆科植物 WRKY 基因家族全基因组鉴定及表达分析[J]. 农业大数据学报,2023(2): 16-26.

[18] 杜春芳,姚琳,孙璇,等. 甘蓝型油菜低温诱导的 WRKY 转录因子动态分析[J]. 分子植物育种,2021,19(10):3175-3184.

[19] SINGH D, DEBNATH P, SANE A P, et al. Tomato (*Solanum lycopersicum*) WRKY23 enhances salt and osmotic stress tolerance by modulating the ethylene and auxin pathways in transgenic *Arabidopsis*[J]. Plant Physiology and Biochemistry, 2023, 195: 330-340.

[20] GUPTA S, MISHRA V K, KUMARI S, et al. Deciphering genome-wide WRKY gene family of *Triticum aestivum* L. and their functional role in response to Abiotic stress[J]. Genes & Genomics, 2019, 41(1): 79-94.

[21] 刘微,蒋莉萍,池玉洁,等. 黄连 WRKY 基因家族鉴定及表达分析[J]. 世界科学技术-中医药现代化,2022,8(4): 1369-1380.

- [22] CHEN W, HUANG B R. Cytokinin or ethylene regulation of heat-induced leaf senescence involving transcriptional modulation of WRKY in perennial ryegrass[J]. *Physiologia Plantarum*, 2022, 174(5): e13766.
- [23] AMATO A, CAVALLINI E, ZENONI S, et al. A grapevine TTG2-like WRKY transcription factor is involved in regulating vacuolar transport and flavonoid biosynthesis[J]. *Frontiers in Plant Science*, 2017, 7: 1979.
- [24] 王彩云, 张晓东, 沈涛, 等. WRKY 在植物次生代谢物合成中的作用及研究进展[J]. *广东农业科学*, 2013, 40(20): 137-144.
- [25] WEI X S, ZHOU H L, XIE D Y, et al. Genome-wide association study in rice revealed a novel gene in determining plant height and stem development, by encoding a WRKY transcription factor[J]. *International Journal of Molecular Sciences*, 2021, 22(15): 8192.
- [26] SHI W Y, DU Y T, MA J, et al. The WRKY transcription factor *GmWRKY12* confers drought and salt tolerance in soybean[J]. *International Journal of Molecular Sciences*, 2018, 19(12): 4087.
- [27] WANG N N, XU S W, SUN Y L, et al. The cotton WRKY transcription factor (*GhWRKY33*) reduces transgenic *Arabidopsis* resistance to drought stress[J]. *Scientific Reports*, 2019, 9(1): 724.
- [28] 付家旭, 闫亚丽, 谢小文, 等. 干旱-复水处理下差异表达玉米 WRKY 基因的鉴定与分析[J]. *河南农业科学*, 2022, 51(8): 9-19.
- [29] 付乾堂, 余迪求. 拟南芥 *AtWRKY25*、*AtWRKY26* 和 *AtWRKY33* 在非生物胁迫条件下的表达分析[J]. *遗传*, 2010, 32(8): 848-856.
- [30] DUAN S W, WANG J J, GAO C H, et al. Functional characterization of a heterologously expressed *Brassica napus* WRKY41-1 transcription factor in regulating anthocyanin biosynthesis in *Arabidopsis thaliana* [J]. *Plant Science*, 2018, 268: 47-53.
- [31] 姜森, 于怡琳, 姜利, 等. 三七转录因子 *PnWRKY1* 对三七皂苷生物合成的影响[J]. *西北植物学报*, 2019, 39(3): 430-438.
- [32] XIN P F, GAO C S, CHENG C H, et al. Identification and characterization of hemp WRKY transcription factors in response to abiotic stresses[J]. *Biologia Plantarum*, 2016, 60(3): 489-495.
- [33] WEI H W, CHEN S Y, NIYTANGA S, et al. Genome-wide identification and expression analysis response to GA₃ stresses of WRKY gene family in seed hemp (*Cannabis sativa* L.)[J]. *Gene*, 2022, 822: 146290.
- [34] IRANBAKHSH A, ORAGHI ARDEBILI Z, MOLAEI H, et al. Cold plasma up-regulated expressions of WRKY1 transcription factor and genes involved in biosynthesis of cannabinoids in hemp (*Cannabis sativa* L.)[J]. *Plasma Chemistry and Plasma Processing*, 2020, 40(2): 527-537.
- [35] ADHIKARY D, KULKARNI M, EL-MEZAWY A, et al. Medical *Cannabis* and industrial hemp tissue culture: present status and future potential[J]. *Frontiers in Plant Science*, 2021, 12: 627240.
- [36] CHEN C J, CHEN H, ZHANG Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data [J]. *Molecular Plant*, 2020, 13(8): 1194-1202.

Genome-Wide Identification and Bioinformatics Analysis of the WRKY Gene Family in Industrial Hemp

ZHANG Yuanye, ZHANG Ligu, ZHANG Ming, FANG Yuyan, ZHENG Nan, YAN Bowei, SUI Yue

(Institute of Industrial Crops, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China)

Abstract: WRKY transcription factors are a class of genes with multiple biological functions, and the WRKY family in cannabis is rarely studied. We used *Arabidopsis* WRKY family protein sequences to compare with the whole genome of cannabis, and identified 34 WRKY family members, which were unequally distributed on 9 chromosomes, and could be divided into 7 subfamilies, including Group I, Group IIa—Group IIe, and Group III, among which Group I had the largest number of members, with 11 members. *CsWRKY15* contains the most domains, and *CsWRKY34* has the most cisacting elements in the promoter region, so they may have more biological functions. The collinearity analysis between the WRKY family of cannabis and *Arabidopsis thaliana*, rice, maize, soybean and cannabis itself showed that *CsWRKYs* had rich phylogenetic relationships. This course lays a data foundation for exploring the biological function of cannabis WRKY family.

Keywords: industrial hemp; WRKY; physicochemical properties of protein; homology analysis; bioinformatics analysis