

菊花脑 WRKY 基因家族鉴定与表达分析*

蒋莉萍¹, 徐宏峰¹, 程璐¹, 刘义飞^{2**}

(1. 武汉市第一医院 武汉 430022; 2. 湖北中医药大学药学院 武汉 430065)

摘要:目的 使用生物信息学方法鉴定和分析菊花脑 WRKY 基因家族成员,为后期研究 WRKY 转录因子调控菊花脑萜烯类物质生物合成提供理论基础。方法 利用 HMMER、BLAST、Pfam、NCBI 等在线工具对菊花脑基因组中的 WRKY 基因家族成员进行筛选和鉴定;ExPASy、WOLF PSORT 等在线工具进行理化性质分析;ClustalW、IQ-Tree、MEME 等软件进行系统进化、保守结构域分析;TBtools 软件进行染色体定位、串联复制、基因表达模式及可视化分析。结果 共鉴定出 110 个 *CnaWRKYs*,其 ORF 长度为 246–2367 bp,编码蛋白的氨基酸数量为 81–788 AA,相对分子质量为 9463.98–87062.38 Da,理论等电点为 4.63–10.15,93 个蛋白预测亚细胞定位于细胞核;系统进化分析将 *CnaWRKYs* 蛋白分为 3 个组,其中 II 组又分为 5 个亚组;蛋白结构域丰富,同一组 *CnaWRKYs* 蛋白结构域高度保守、保守基序相似;109 个菊花脑 WRKY 基因不均一地分布在 9 条染色体上,另外有 1 个分布在 Scaffold25 上;基因表达热图显示,108 个 *CnaWRKYs* 基因在整体水平上有表达,WRKY 基因家族成员在各个组织中的表达丰度存在明显差异。结论 系统分析了菊花脑 WRKY 基因家族成员,为进一步解析 *CnaWRKYs* 基因在菊花脑挥发性芳香化合物萜烯类物质生物合成的作用及其调控机制提供理论依据。

关键词:菊花脑 WRKY 转录因子 次生代谢 串联复制 表达分析

doi: 10.11842/wst.20230906002 中图分类号: R-058 文献标识码: A

菊花脑(*Chrysanthemum nankingense* Hand.-Mazz.)是菊科菊属植物,为药用野菊(*C. indicum* L.)的近缘种。菊花脑又名菊花菜,嫩茎叶具有降血压、生津开胃、清热解毒等功效,民间药食两用^[1]。菊花脑适应性极强,在江苏、贵州、湖南和湖北等省引种栽培并广泛分布。野菊倍性复杂(包括二倍体、四倍体、六倍体等),而菊花脑为二倍体物种,倍性简单故可作为菊属模式植物开展药用菊花的相关研究^[2]。菊花脑化学成分比较复杂,其天然活性成分主要为黄酮类化合物和挥发油中的萜类化合物^[3],其中挥发油的萜烯类物质不仅影响菊花脑的香气品质,而且还在抗病虫害等方面有重要的作用^[4–6]。

WRKY 转录因子(Transcription factors, TFs)是植物中最大的转录调控因子家族之一,其名字来源于 TFs 高度保守的 60 个氨基酸长的 WRKY 结构域,包含

一个在 N 端的高度保守的七肽氨基酸序列 WRKYGQK 和一个在 C 端锌指状基序 Cys(2)–His(2)或 Cys(2)–HisCys,并与靶基因启动子中的 TTGAC(C/T) W-box 顺式元件结合^[7–8]。根据 WRKY 蛋白结构域的数量和锌指样基序的类型,将 WRKY 蛋白分为 3 组。I 组的 WRKY 蛋白具有 2 个 WRKY 结构域,而 II 组和 III 组的 WRKY 蛋白只具有一个 WRKY 结构域。I 组和 II 组蛋白均具有相同的潜在锌配体模式 C–X_{4–5}–C–X_{22–23}–H–X–H^[9]。III 组蛋白的的锌指基序类型与 I 组和 II 组不同,为 C–X₇–C–X₂₃–H–X–C^[10]。II 组的 WRKY 蛋白根据初级氨基酸序列进一步分为 II a、II b、II c、II d 和 II e 五个亚组^[11–12]。

WRKYs 不仅在植物抗病、非生物胁迫反应、营养剥夺、衰老、种子和毛状体发育、胚胎发生以及额外的发育和激素控制过程等方面具有多种多样的生物学

收稿日期:2023–09–06

修回日期:2024–03–22

* 国家自然科学基金委员会面上项目(32270231):菊属植物 TS/CYP 基因簇的起源进化机制及功能研究,负责人:刘义飞。

** 通讯作者:刘义飞,研究员,博士研究生导师,主要研究方向:中药资源评价与组学研究。

功能,还在植物次生代谢产物生物合成过程中起着重要作用^[13-16]。在菊花脑中,萜烯是其最重要的挥发性化合物,而W-box是WRKY转录因子的同源结合位点,可以在大多数萜烯合成相关基因的启动子中被识别出来。然而,WRKY家族在菊花脑中萜烯合成中的作用却未被研究。本研究以菊花脑基因组数据为基础,通过生物信息学的方法,系统地分析了菊花脑WRKY基因家族,为后期研究菊花脑WRKY转录因子在菊花脑萜类生物合成途径中的调控机制奠定基础。

1 材料与方法

1.1 材料

本实验样品采自湖北中医药大学野菊种质资源保存圃,经植物标本比对鉴定为菊花脑。取舌状花、管状花、根、茎和叶5个不同的组织进行后续分析,生物学重复为每个部位3个单株重复。样品经液氮速冻后,-80℃冰箱保存备用。

1.2 菊花脑WRKY基因的筛选与鉴定

菊花脑基因组数据(CDS序列、PEP序列和GFF文件)和转录组数据来源于湖北中医药大学药学院中药资源评价和组学研究实验室数据库^[17],WRKY基因保守结构域相关文件PF03106.hmm从Pfam数据库(<http://pfam.xfam.org/>)下载获取。利用TBtools软件中的Simple HMM Search功能在菊花脑基因组蛋白质序列文件中搜索模型类似的序列,以NCBI网站UniProtKB/Swiss-Prot(swissprot)数据库中搜索已功能验证的WRKY基因序列集为响应(Query)序列,运用TBtools软件BlastP功能分析同源WRKY基因序列,即以已知功能的WRKY基因序列在全基因组蛋白数据中比对,并提取序列相似性高的序列。对以上两种方法分析得到的序列ID取交集,得到所有的候选WRKY基因。利用NCBI中的Batch CDD search(<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi/>)数据库进一步排除没有保守WRKY结构域的基因,最终获得全部的菊花脑WRKY基因家族成员。

根据基因组GFF注释文件得到WRKY基因CDS序列的长度。利用ExPASy(<https://web.expasy.org/protparam/>)在线工具预测菊花脑WRKY蛋白氨基酸序列的理化特性,包括编码的氨基酸数目、等电点和蛋白质相对分子量。使用WOLF PSORT(<https://www.genscript.com/wolf-psort.html>)在线工具预测菊花脑

WRKY蛋白的亚细胞定位。

1.3 菊花脑WRKY基因家族系统发育分析

基于拟南芥(*Arabidopsis thaliana*)和鉴定到的菊花脑WRKY氨基酸序列构建系统发育树^[7]。运用ClustalW软件进行多序列比对,通过IQ-Tree软件计算最佳模型VT+F+R6并用最大似然法(Maximum likelihood, ML)对拟南芥和菊花脑WRKY基因家族进行系统进化分析来确定菊花脑WRKY基因家族的分类,建树步长值(Bootstrap)重复次数为1000。

1.4 菊花脑WRKY基因家族保守基序和WRKY结构域分析

使用Multiple Em for Motif Elicitation(MEME, <https://meme-suite.org/meme/tools/meme>)在线网站预测菊花脑不同WRKYs蛋白的保守基序,设置参数为:保持重复数,任意;最大图案数,10;基序的最佳宽度≥20。分析前10个保守基序并下载xml文件和10个保守基序图。对不同组WRKY结构域60个保守氨基酸进行比较分析其N端的七肽氨基酸序列和C端的锌指状基序。

1.5 菊花脑WRKY基因家族染色体定位和串联复制分析

根据WRKY基因家族ID文件和菊花脑基因组GFF注释文件,利用TBtools软件中的Gene Location Visualize(Advanced)功能绘制CnaWRKYs基因在菊花脑染色体上的分布图。

利用TBtools软件中的Two Sequence Files、Text Merge for MCSanX、Quick Run MCSanX Wrapper和File Merge for MCSanX等功能分析WRKY基因家族的串联复制情况。

1.6 菊花脑WRKY基因家族的表达模式分析

原始转录组数据经Trimmomatic-0.36过滤低质量数据,过滤后数据经STAR_2.6.0a软件比对到基因组上。基于比对数据,转录本丰度基于HTSeq软件计算,并将原始计数转化为TPM(Transcripts Per Million)值,通过TBtools软件中的HeatMap功能可视化CnaWRKYs基因表达热图。

2 结果与分析

2.1 菊花脑WRKY基因家族成员鉴定与理化性质

通过构建隐马尔科夫模型,将PF03106.hmm功能域序列在菊花脑基因组数据库中比对;以NCBI中已

有的 WRKY 序列作为 Query 序列,对菊花脑蛋白质数据库进行 BlastP 分析;合并以上两种分析方法的结果,并经 NCBI 的 CDD-search 验证其保守结构域,最终确定了 110 个 *CnaWRKYs* 基因。菊花脑 *WRKY* 基因的开放阅读框长度为 246–2367 bp,其编码蛋白大小在 81–788 AA,蛋白相对分子质量为 9463.98–87 062.38 Da;理论等电点为 4.63–10.15,其中有 63 个酸性蛋白($pI < 7$),占 57.27%,47 个碱性蛋白,占 42.73%;不稳定指数

为 25.1–68.33,其中值小于 40 的为稳定蛋白;脂肪族氨基酸指数为 42.00–95.47;亲水性总平均值为 $-1.183--0.151$,均为亲水性蛋白;93 个蛋白亚细胞定位预测于细胞核,8 个蛋白亚细胞定位预测于细胞质,4 个蛋白亚细胞定位预测于叶绿体,还有 4 个蛋白分别位于高尔基体(*Cna0024632.1*)、过氧化物酶体(*Cna0005089.1*)、细胞骨架(*Cna0045100.1*)和质膜(*Cna0029351.1*)(表 1)。

表 1 菊花脑 WRKY 基因家族信息

基因名称	ORF 长度(bp)	分子量(Da)	理论等电点	不稳定指数	脂肪族氨基酸指数	亲水性总平均值	亚细胞定位
Cna0001853.1	1455	54 870.86	8.91	50.49	73.84	-0.602	细胞核
Cna0002807.1	1533	56 177.55	7.9	44.07	57.18	-0.704	细胞核
Cna0003991.1	546	20 591.87	9.18	29.26	58.67	-1.023	细胞核
Cna0004081.1	483	18 404.43	9.33	27.43	53.56	-1.183	细胞核
Cna0004545.1	798	30 397.06	6.54	40.02	71.66	-0.806	细胞核
Cna0004546.1	1020	38 157.26	5.79	45.47	68.67	-0.880	细胞核
Cna0004989.1	912	35 007.76	8.48	57.44	57.62	-1.104	细胞核
Cna0005087.1	1062	40 780.24	6.49	50.71	56.63	-0.928	细胞核
Cna0005089.1	612	22 979.96	9.44	40.02	72.41	-0.684	过氧化物酶体
Cna0006052.1	1461	55 032.40	8.11	37.90	95.47	-0.309	细胞质
Cna0006278.1	849	31 071.33	5.69	59.42	56.10	-0.666	细胞核
Cna0006756.1	879	32 685.45	6.79	55.69	59.73	-0.722	细胞核
Cna0006759.1	864	32 178.01	6.21	53.36	73.62	-0.679	细胞核
Cna0006762.1	990	36 436.16	6.50	41.93	82.67	-0.407	叶绿体
Cna0006764.1	1386	51 148.88	5.81	40.23	94.45	-0.168	细胞质
Cna0006766.1	684	25 505.99	9.11	46.91	79.82	-0.590	细胞核
Cna0006767.1	765	28 861.63	8.60	49.99	77.48	-0.602	细胞核
Cna0009806.1	1233	46 289.30	9.33	42.30	68.66	-0.775	细胞核
Cna0014622.1	1353	51 345.95	5.96	38.54	55.89	-1.003	细胞核
Cna0014889.1	651	24 813.83	8.27	66.08	68.10	-0.854	细胞核
Cna0014893.1	888	33 419.68	5.93	49.02	70.37	-0.507	叶绿体
Cna0014963.1	888	33 407.85	6.10	51.56	71.05	-0.511	细胞核
Cna0014964.1	891	33 363.57	5.87	46.95	68.82	-0.543	细胞核
Cna0014965.1	1008	38 345.07	5.23	50.99	76.21	-0.497	细胞核
Cna0014967.1	684	26 171.57	8.68	48.10	67.80	-0.821	细胞核
Cna0014969.1	666	25 271.69	5.68	64.63	71.45	-0.626	细胞核
Cna0014971.1	525	20 212.04	8.36	49.83	72.24	-0.618	细胞核
Cna0014980.1	966	36 628.91	6.01	38.75	57.98	-0.795	细胞核
Cna0014981.1	966	35 795.68	5.79	46.86	58.60	-0.842	细胞核
Cna0015027.1	966	35 790.75	5.66	45.98	61.34	-0.793	细胞核
Cna0015334.1	714	27 797.84	4.95	56.92	49.28	-1.178	细胞核
Cna0016808.1	921	34 144.58	6.70	42.92	49.31	-0.879	细胞核
Cna0017570.1	855	31 474.86	6.46	51.67	48.70	-0.844	细胞核
Cna0017979.1	1515	55 058.11	8.56	46.79	66.25	-0.680	细胞核
Cna0019642.1	1623	60 114.07	6.58	64.44	49.85	-0.898	细胞核

下转续表

续表

基因名称	ORF 长度(bp)	分子量(Da)	理论等电点	不稳定指数	脂肪族氨基酸指数	亲水性总平均值	亚细胞定位
Cna0020197.1	936	34 489.45	10.09	44.74	82.41	-0.531	细胞核
Cna0021713.1	1830	66 653.84	6.32	46.32	45.45	-0.957	细胞核
Cna0021969.1	615	22 457.29	9.18	45.84	58.33	-0.747	细胞核
Cna0022304.1	828	30 841.94	8.86	28.95	76.51	-0.551	细胞质
Cna0022795.1	1719	62 577.19	6.33	49.27	56.10	-0.800	细胞核
Cna0023942.1	939	34 819.28	5.69	58.38	50.93	-0.873	细胞核
Cna0024632.1	654	24 784.62	8.03	54.57	93.96	-0.151	高尔基体
Cna0025974.1	882	32 367.51	5.66	60.28	53.58	-0.903	细胞核
Cna0026021.1	1497	55 062.48	8.47	56.60	51.27	-0.971	细胞核
Cna0026236.1	591	22 184.41	6.75	44.69	55.10	-0.959	细胞核
Cna0029027.1	1551	56 200.54	6.43	41.52	65.68	-0.571	细胞核
Cna0029351.1	1926	73 056.46	9.22	35.75	93.01	-0.225	质膜
Cna0029825.1	903	33 768.13	5.10	63.01	43.27	-1.001	细胞核
Cna0029978.1	1209	44 524.86	5.89	44.82	52.86	-0.912	细胞核
Cna0030416.1	666	24 788.32	9.51	44.80	52.49	-0.582	细胞核
Cna0030603.1	978	36 559.31	6.18	57.27	45.91	-0.962	细胞核
Cna0031455.1	720	26 888.23	7.24	58.63	56.65	-0.824	细胞核
Cna0031722.1	1005	37 209.44	9.60	53.20	67.07	-0.654	细胞核
Cna0033252.1	993	36 696.38	7.67	53.57	73.58	-0.568	细胞核
Cna0033933.1	1458	54 195.93	8.39	46.76	42.00	-1.074	细胞核
Cna0034553.1	654	24 231.82	8.71	47.57	57.56	-0.888	细胞核
Cna0035010.1	246	9463.98	9.87	26.66	64.81	-0.82	细胞质
Cna0035078.1	2367	87 062.38	7.78	46.48	52.12	-0.829	细胞核
Cna0035215.1	621	23 571.74	7.60	53.15	65.29	-0.759	细胞核
Cna0036080.1	1971	71 060.33	6.14	46.74	58.52	-0.772	细胞核
Cna0036630.1	960	36 133.41	6.33	42.99	66.02	-0.696	细胞核
Cna0038675.1	870	33 117.36	8.20	40.92	70.80	-0.693	细胞核
Cna0041911.1	762	28 473.96	6.71	68.33	60.08	-0.695	细胞核
Cna0042672.1	825	30 683.27	6.40	31.90	69.34	-0.598	叶绿体
Cna0042720.1	678	25 162.33	4.84	47.32	49.82	-0.933	细胞核
Cna0044052.1	795	30 485.22	5.89	58.31	73.07	-0.626	细胞核
Cna0044053.1	966	36 622.02	6.01	48.31	60.44	-0.721	细胞核
Cna0044252.1	966	36 656.14	6.13	52.24	62.87	-0.702	细胞核
Cna0044968.1	1359	49 131.58	5.60	45.00	71.28	-0.629	细胞核
Cna0045100.1	1146	42 812.18	6.24	41.69	76.72	-0.594	细胞骨架
Cna0045103.1	1050	38 282.26	6.81	50.27	59.97	-0.869	细胞核
Cna0045824.1	1572	57 748.47	8.08	49.21	53.67	-0.958	细胞核
Cna0048080.1	699	26 251.76	4.78	29.68	77.76	-0.725	细胞质
Cna0048081.1	672	25 267.57	5.15	31.70	63.81	-0.867	细胞核
Cna0048202.1	297	11 281.87	9.71	25.10	57.65	-0.915	叶绿体
Cna0048835.1	810	30 851.65	7.61	44.08	71.08	-0.779	细胞核
Cna0048890.1	1788	66 727.02	5.97	37.29	58.47	-0.904	细胞核
Cna0049347.1	1032	37 560.45	6.42	39.10	76.73	-0.431	细胞核
Cna0050102.1	603	23 403.10	8.38	46.12	49.65	-0.918	细胞核
Cna0050954.1	1542	56 849.78	8.41	53.26	52.96	-0.924	细胞核

下转续表

续表

基因名称	ORF 长度(bp)	分子量(Da)	理论等电点	不稳定指数	脂肪族氨基酸指数	亲水性总平均值	亚细胞定位
Cna0050961.1	1308	48 548.75	6.95	48.27	57.98	-0.936	细胞核
Cna0051230.1	642	24 361.54	4.63	58.12	62.21	-1.096	细胞质
Cna0051318.1	903	32 690.05	9.27	52.83	69.87	-0.516	细胞核
Cna0053755.1	990	36 201.32	8.47	47.75	62.25	-0.823	细胞核
Cna0053785.1	990	36 187.29	8.47	46.31	62.25	-0.823	细胞核
Cna0056813.1	681	25 072.60	10.15	55.84	61.68	-0.689	细胞核
Cna0058018.1	741	27857.01	8.74	56.83	61.87	-0.935	细胞核
Cna0058102.1	639	23969.58	6.25	54.66	64.39	-0.831	细胞核
Cna0058217.1	1485	54 004.30	6.24	55.86	56.03	-0.937	细胞核
Cna0058873.1	1359	50 053.21	5.77	50.90	59.80	-0.891	细胞核
Cna0059942.1	1449	53 631.11	6.90	55.17	50.95	-0.894	细胞核
Cna0060212.1	1926	70 944.42	5.00	38.29	82.20	-0.301	细胞质
Cna0060291.1	1443	53 267.78	6.90	54.07	52.38	-0.853	细胞核
Cna0060943.1	915	34 384.02	9.70	48.99	65.07	-0.719	细胞核
Cna0061751.1	537	20 387.42	5.73	47.11	46.46	-0.956	细胞核
Cna0061973.1	681	24 897.14	10.06	60.90	55.58	-0.715	细胞核
Cna0062156.1	1329	49 271.13	6.29	54.05	61.92	-1.050	细胞核
Cna0062422.1	1332	49 385.23	6.29	53.95	61.78	-1.055	细胞核
Cna0065396.1	558	21 132.27	9.11	54.93	46.32	-1.082	细胞核
Cna0066232.1	1407	52 281.63	9.25	50.44	60.83	-0.86	细胞核
Cna0066233.1	855	31 746.54	8.93	47.31	60.07	-0.817	细胞核
Cna0066257.1	813	30 002.37	9.18	64.27	54.48	-0.839	细胞核
Cna0069987.1	798	29 787.28	5.38	57.37	65.85	-0.594	细胞核
Cna0069999.1	741	28 051.74	8.97	50.03	67.36	-0.704	细胞核
Cna0070253.1	945	35 630.30	5.79	38.88	73.22	-0.472	细胞质
Cna0072216.1	1008	37 739.79	5.68	41.92	57.97	-0.739	细胞核
Cna0072227.1	807	30 435.42	6.04	44.53	47.31	-0.906	细胞核
Cna0073026.1	723	27 350.51	6.51	39.09	70.25	-0.965	细胞核
Cna0073152.1	723	27 539.79	6.32	42.60	65.38	-0.987	细胞核
Cna0076894.1	993	36 338.46	8.47	47.64	62.06	-0.83	细胞核

2.2 菊花脑 WRKY 基因家族系统发育

为了进一步明确菊花脑 WRKY 基因家族的进化分类,本研究采用 58 条已有明确进化分类的拟南芥 WRKY 氨基酸序列和 110 条菊花脑 WRKY 氨基酸序列,进行系统聚类分析。结果表明菊花脑 WRKY 转录因子可以分为 3 个组,分别为 I 组(26 个)、II 组(56 个)和 III 组(28 个),其中 II 组又可分为 5 个不同的亚群(II a、II b、II c、II d 和 II e),与拟南芥系统发育结果一致(见图 1)。

2.3 菊花脑 WRKY 基因家族的 WRKY 结构域和保守基序

根据菊花脑 WRKY 成员的系统分类,进一步探究了其 WRKY 结构域和保守基序。利用 MEME 在线工

具对菊花脑 WRKY 氨基酸序列进行分析,选择前 10 个保守结构域进行展示(图 2 右)。WRKY 基因家族具有丰富的结构域,除了 12 个蛋白外,菊花脑其他 WRKY 蛋白均有 motif1 基序;motif2 基序通常出现在 motif1 基序后面,相对比较保守,这些保守功能结构域是其与底物结合的关键位置。motif3 基序只出现在 I 组中,motif6 基序只出现在 II a 和 II b 两个亚家族中,motif10 基序与 motif6 基序类似,但 I 组成员 Cna0050954.1 中也有 motif10。II d 和 II e 亚家族保守基序比较少,绝大多数以 motif1+motif2、motif1 或 motif2 的结构存在。接近一半的 III 组 WRKY 成员中有 motif5、motif8 和 motif9,是其保守基序。Motif4 多出现在 II c 亚家族中,且出现在 motif1 基序前端。I 组中的成员均含有 2 个 WRKY 蛋

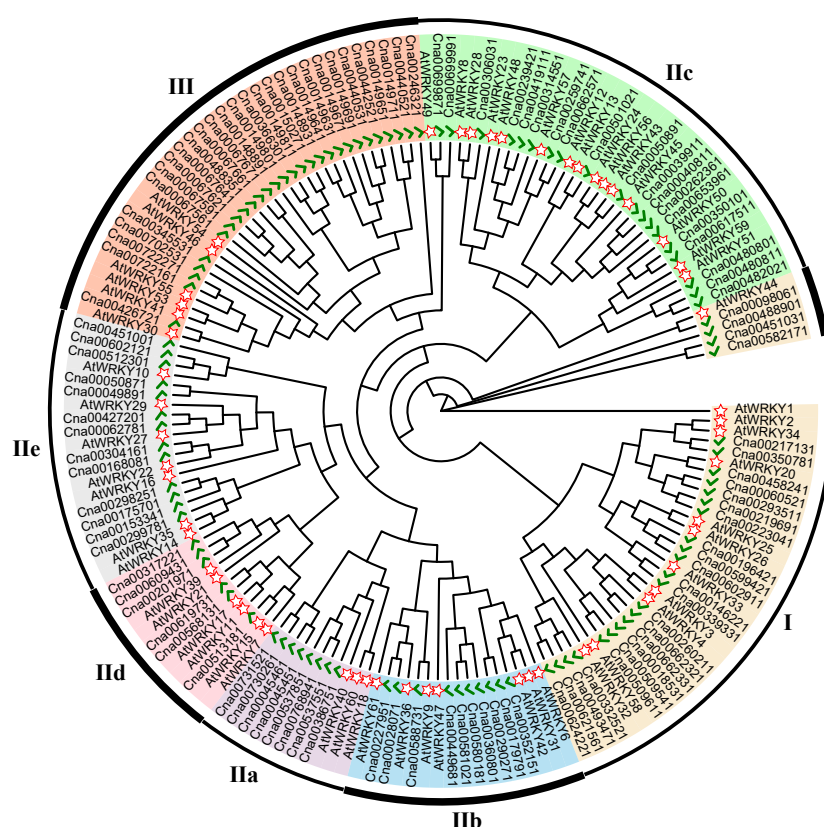


图1 拟南芥和菊花脑WRKY基因家族成员的系统发育树

白序列,其余各组中的成员只含有一个WRKY结构域。

菊花脑WRKY基因家族中WRKY结构域的七肽氨基酸序列WRKYGQK是高度保守的,有10个WRKY蛋白七肽氨基酸序列缺失,有少量蛋白出现个别氨基酸的突变,形成WLKYGQK、WKYGEQK、WKYGEK、WRKGYYK、WTKYGKK、WRKYGKK和WRKSYYK的七肽(或八肽)氨基酸序列类型。I组、II组中的CnaWRKYs蛋白的锌指结构类型为C-X₄₋₅-C-X₂₂₋₂₃-H-X-H,III组中的CnaWRKYs蛋白的锌指结构类型为C-X₇-C-X₂₃-H-X-C,每个组里也有个别WRKY蛋白锌指结构基序缺失(图2左)。

2.4 菊花脑WRKY基因家族染色体定位和基因重复分析

在110个基因中,109个CnaWRKYs基因分散在菊花脑的9条染色体上,另外1个CnaWRKY基因定位到了Scaffold25上。3号染色体上的CnaWRKY基因最多,为18个。在1-2、4-9号染色体上依次分布10、8、9、16、11、13、17、7个WRKY家族成员。I、III、IIa、IIb、IIc、IIe和IId各亚家族在染色体上分布也不一。为了进一步研究菊花脑WRKY基因的进化,对片段和

串联复制事件进行了研究,发现在菊花脑中WRKY基因家族中存在基因片段和串联复制事件,其中有8对串联复制的CnaWRKY基因对(红星标注),分别分布在1、2、3、6、7和8染色体上(图3)。

2.5 菊花脑WRKY基因家族的表达模式

利用菊花脑茎、叶、舌状花、管状花和根5个不同组织的基因表达数据,分析其时空表达情况,结果表明,108个CnaWRKY基因在整体水平上有表达,87个CnaWRKY基因在5个组织中均能表达,而2个CnaWRKY基因在5个组织中均没有表达。基因表达谱热图显示,WRKY基因家族成员各个组织中的表达丰度存在明显差异,例如Cna0070253.1、Cna0033933.1、Cna0015334.1、Cna0061751.1、Cna0019642.1和Cna0029027.1等基因在叶和茎组织中高度表达(log₁₀TMP≥2);Cna0058217.1、Cna0051318.1和Cna0070253.1等基因在舌状花和管状花组织中的表达量较高(log₁₀TMP≥1.6);Cna0053785.1、Cna0023942.1、Cna0076894.1、Cna0053755.1、Cna0004989.1和Cna0005087.1在根中高度表达(log₁₀TMP≥2)。菊花脑以头状花序入药,其活性物质主要在舌状花和管状花中

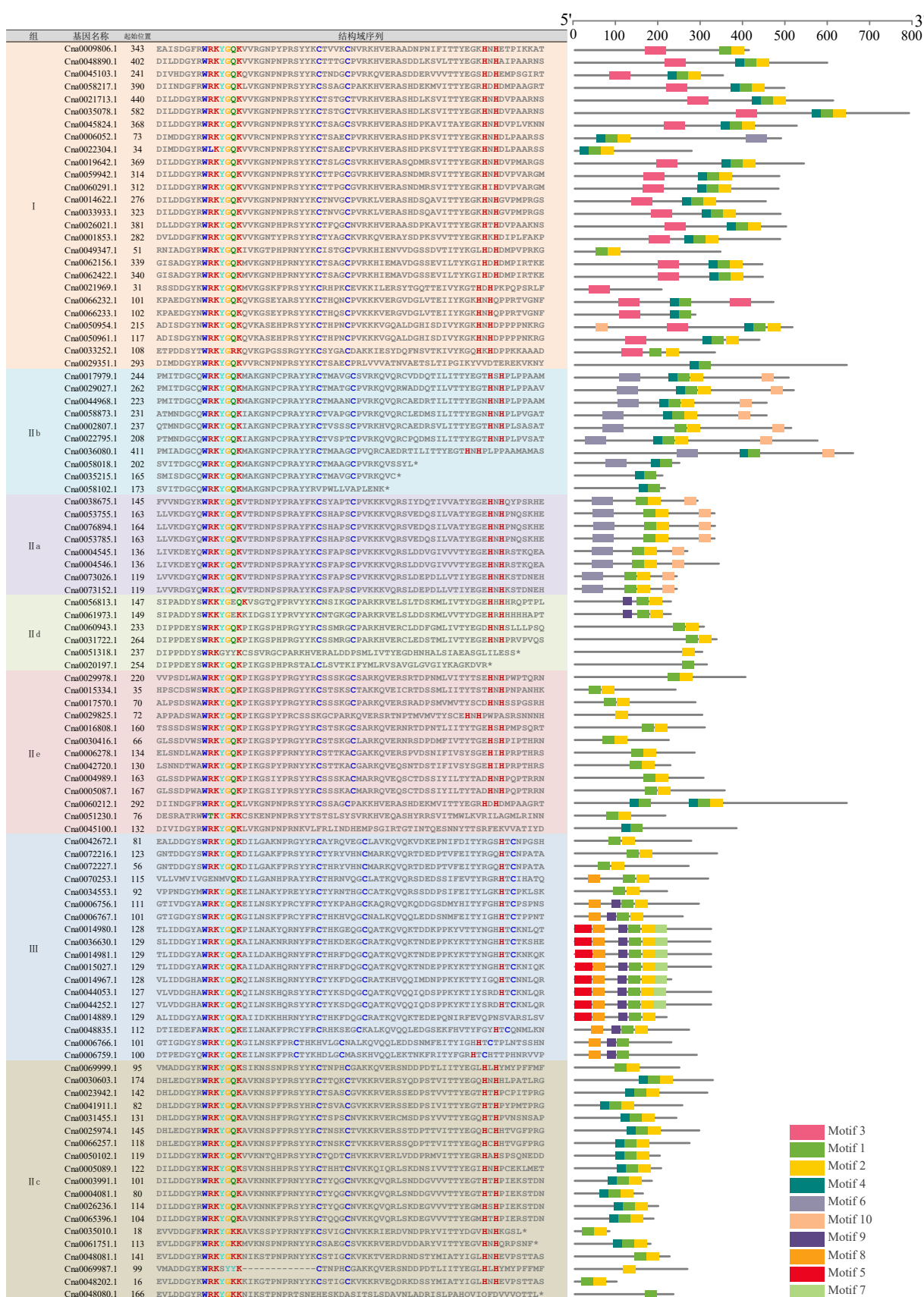


图2 菊花WRKY家族的WRKY结构域和保守基序

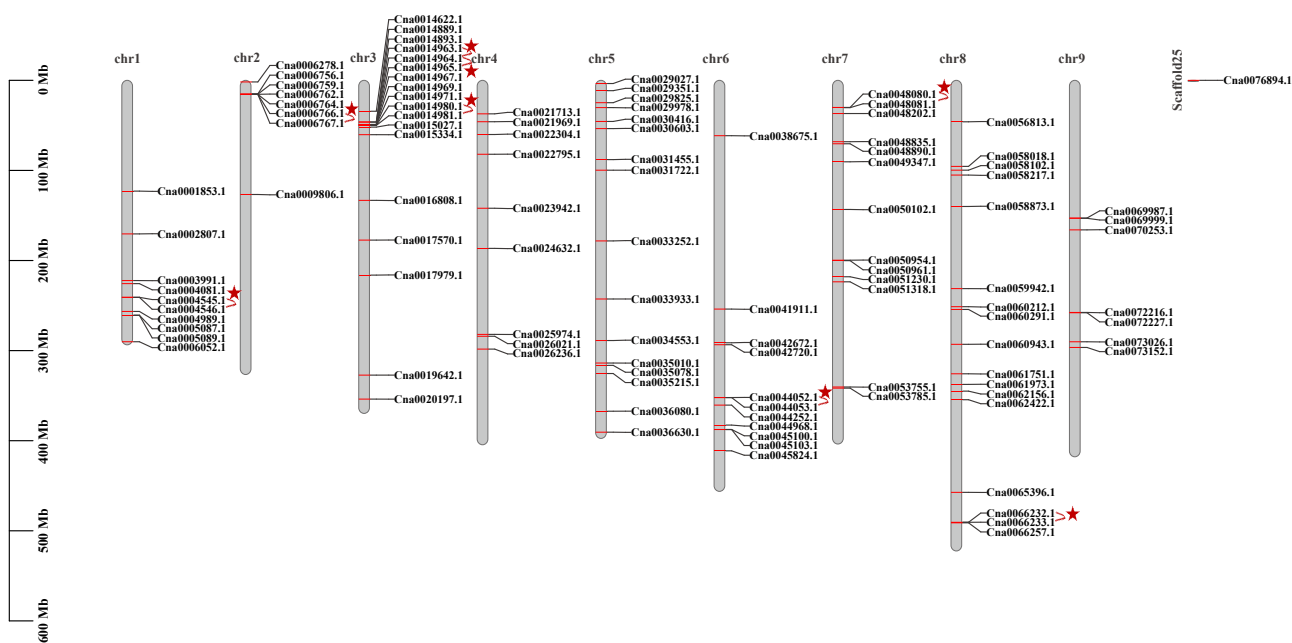


图3 菊花脑WRKY基因的染色体定位和串联复制

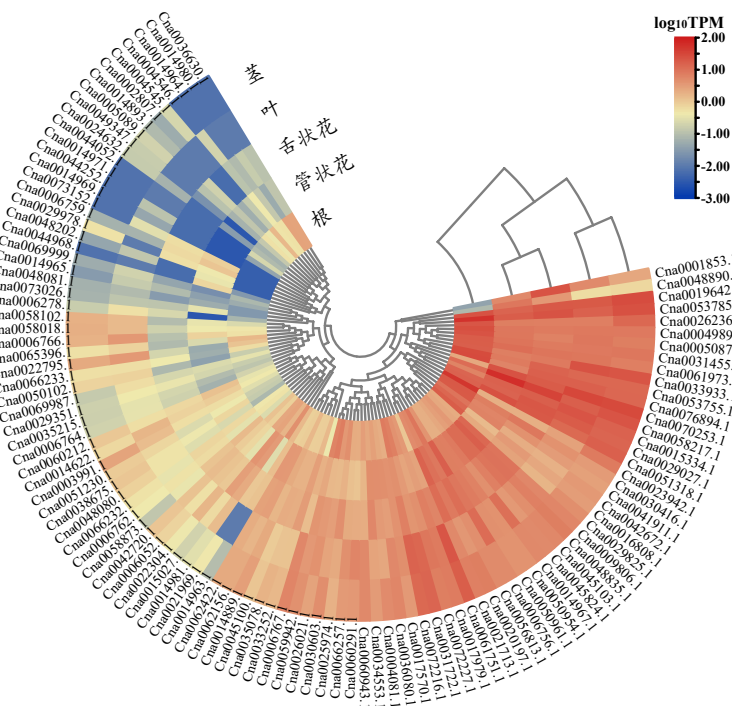


图4 菊花脑WRKY基因在不同组织中的表达谱

积累,故 *Cna0058217.1*、*Cna0051318.1* 和 *Cna0070253.1* 等在花组织中表达量的基因值得进一步挖掘其生物功能(见图4)。

3 讨论

WRKY 转录因子家族在植物中普遍存在,并广泛

存在于许多物种中。目前,随着全基因组测序的高速发展,使更多WRKY基因在植物中得到鉴定,例如在拟南芥、水稻、黄连和金银花等植物中分别有72、100、41、24和46个WRKY超家族成员^[18-21]。本研究基于菊花脑基因组数据,共鉴定到110个CnaWRKYs基因,并将其分为I、II和III组。

转录因子的功能与其定位之间存在着密切的关系^[22]。菊花脑 WRKY 蛋白主要定位于细胞核,有少部分定位于其他细胞器或质膜上。虽然转录调控是在细胞核内进行,但转录因子定位在细胞核中并不是一个绝对的结论^[23-24]。保守基序分析揭示了有关基因进化的有趣事实。I 组蛋白的 motif4 出现在含有 WRKY 结构域的 motif1 之前,II b 亚家族蛋白中也被发现,表明 IIb 亚家族的基因起源于 I 组基因 N 端 WRKY 结构域的缺失。WRKY 结构域的七肽序列 WKYRQK 在许多植物中相当保守,但也会有个别 WRKY 结构域突变,形成多种变体,例如在龙眼中有 3 个^[25]、水稻有 9 个^[26]、豆中有 4 个变体(WRKYGKK、WRKYGEK、WKYEDK 和 WKYCEDK^[27])。一般来说,变异的数字类型不同,但主要是 II b、II c 和 III 组的成员。本研究在菊花脑中观察到 11 个变体:5 个 WRKYGKK 变体属于 II c,除了 II a 和 II b 外,其他组均有变体存在。因此,我们的数据支持 WRKY 基因的进化理论,即 I 组是最古老的群体,II 组和 III 组都是从 I 组进化而来的。WRKYGQK 基序的变体可能通过结合特定序列而不是 W-box 元件来改变 DNA 与下游靶基因结合的特异性^[7]。这些氨基酸的变异可能使 WRKY 蛋白具有不同的生物学功能。有趣的是,通过对 *TPS* 基因启动子的全面分析,发现了大量的 TGAC 重复序列(W-box),这是 WRKY 转录因子的结合位点,这意味着 WRKY 和 *TPS* 基因之间存在潜在的调控关系。绝大多数菊花脑 WRKY 蛋白有锌簇结构域(有 10 个 WRKY 成员锌簇结构域不全),该结构域编码 WRKY GCM1 结构域蛋白,这些基因在植物发育和调控中的功能多样性^[28]。基因的分布表明 WRKYs 在不同物种中的基因结构多样性不同。因此,基因分布的不均匀是物种进化过程中产生的遗传变异的证据^[29]。串联重复是在植物进化过程中被 5 个或更少基因分隔的 100 kb 以内染色体区域的重复^[30],这被认为在基因家族的扩展中起着主导作用^[31]。由于 WRKY 家族在被子植物的各种生理过程中发挥着重要作用,因此在进化过程中可能会迅速扩大。据报道,最近的基因重复事件在水稻和拟南芥等许多作物中更为广泛地扩展了 WRKY 基因^[18]。在菊花脑 WRKY 基因分布研究中,也检测到一些串联和片段基因重复。值得注意的是,共有 8 对串

联重复的基因,其中 6 对来源于 III 组,这之前对拟南芥、柑橘的研究结果一致^[18,32]。另外,还有两对分别来源于 II a 和 II c,II 组的扩张与草莓的研究结果相似^[33]。研究表明,WRKY 基因家族 II 组和 III 组的串联重复事件是菊花脑 WRKY 基因家族快速扩增的主要原因。

WRKY 蛋白被认为参与了各种植物发育和生理过程的调控,包括次生代谢物的生物合成^[34-35]。WRKYs 单独与其他调控因子联合发挥作用,激活或抑制转录,从而调控参与萜类化合物的生物合成^[36]。番茄的 *SlWRKY73* 转录因子可以激活 *SITPS5*、*SITPS3* 和 *SITPS7* 等 3 个单萜合成酶基因启动子,从而调控番茄毛状体中萜烯的生物合成^[37]。在黄花蒿中,同样在毛状体中产生的抗疟药物青蒿素的生物合成受到 *AaWRKY1* 的调控^[38]。过表达 *AaWRKY1*,黄花蒿高产品种的青蒿素产量增加了一倍^[39]。*TcWRKY1* 参与了紫杉醇生物合成途径中的一个限速基因的表达的调控,为红豆杉细胞培养中增加紫杉醇积累提供了潜在的候选基因^[40]。棉花(陆地棉)中分离出的 *GaWRKY1* 过表达可高度激活倍半萜环化酶基因 *CAD1-A* 的启动子,从而发挥其在倍半萜生物合成调控中的作用^[34]。当茉莉酸甲酯诱导时,一些与防御化合物(如萜类和异黄酮)形成相关的 WRKY 基因被上调^[41]。同样,在玉米中,过表达 *CbWRKY24* 导致总皂苷含量提高,这是由于 MVA 通路关键基因的表达上调^[42]。玫瑰 *CrWRKY1* 基因在茉莉酸和乙烯诱导后优先在根中表达,在类固醇合成中与 *DXS* 和 *SLS* 基因以及调控因子 *CrMYC2* 和 *CrZCT* 相互作用^[43]。这些共同观察表明,对 WRKY 家族的研究可以为植物萜类的合成机制提供有价值的见解。识别所有参与特殊代谢的 WRKYs 对于解开连接天然产物合成的交织网络至关重要。此外,虽然已有研究表明 WRKY 在萜类合成中的调控作用,但对菊花脑中最重要的芳香物质单萜的调控机制知之甚少。基于菊花脑基因组和转录组数据,对菊花脑 WRKY 基因家族进行了鉴定,并对其理化性质、系统发育关系、结构域和保守基序、染色体定位和基因重复及基因表达模式展开了系统分析,这些研究结果为 *CnaWRKYs* 基因提供了全面的信息,有助于进一步研究 *CnaWRKYs* 在调控菊花脑挥发性芳香化合物合成中的作用及其调控网络。

参考文献

- 1 陈玉英, 孙怀志, 谭雪, 等. 保健野生蔬菜-菊花脑. 广西园艺, 2004, 15(2):47-48.
- 2 Song C, Liu Y, Song A, *et al.* The *Chrysanthemum nankingense* genome provides insights into the evolution and diversification of *Chrysanthemum* flowers and medicinal traits. *Mol Plant*, 2018, 11(12): 1482-1491.
- 3 王亚博, 都宏霞. 菊花脑活性成分及其提取技术研究进展. 广东化工, 2016, 43(21):109-110, 116.
- 4 Sharma E, Anand G, Kapoor R. Terpenoids in plant and arbuscular mycorrhiza-reinforced defence against herbivorous insects. *Ann Bot*, 2017, 119(5):791-801.
- 5 Langenheim J H. Higher plant terpenoids: a phytocentric overview of their ecological roles. *J Chem Ecol*, 1994, 20(6):1223-1280.
- 6 Boncan D A T, Tsang S S K, Li C, *et al.* Terpenes and terpenoids in plants: interactions with environment and insects. *Int J Mol Sci*, 2020, 21(19):7382.
- 7 Eulgem T, Rushton P J, Robatzek S, *et al.* The WRKY superfamily of plant transcription factors. *Trends Plant Sci*, 2000, 5(5):199-206.
- 8 Rushton P J, Torres J T, Parniske M, *et al.* 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.
- 9 Berg J M, Shi Y. The galvanization of biology: a growing appreciation for the roles of zinc. *Science*, 1996, 271(5252):1081-1085.
- 10 Li H, Xu Y, Xiao Y, *et al.* Expression and functional analysis of two genes encoding transcription factors, VpWRKY1 and VpWRKY2, isolated from Chinese wild *Vitis pseudoreticulata*. *Planta*, 2010, 232(6): 1325-1337.
- 11 Chen L, Song Y, Li S, *et al.* The role of WRKY transcription factors in plant abiotic stresses. *Biochim Biophys Acta*, 2012, 1819(2):120-128.
- 12 Phukan U J, Jeena G S, Shukla R K. WRKY transcription factors: molecular regulation and stress responses in plants. *Front Plant Sci*, 2016, 7:760.
- 13 Rushton P J, Somssich I E, Ringler P, *et al.* WRKY transcription factors. *Trends Plant Sci*, 2010, 15(5):247-258.
- 14 Song Y, Ai C R, Jing S J, *et al.* Research progress on functional analysis of rice *WRKY* genes. *Rice Sci*, 2010, 17(1):60-72.
- 15 Zhou X, Jiang Y, Yu D. WRKY22 transcription factor mediates dark-induced leaf senescence in *Arabidopsis*. *Mol Cells*, 2011, 31(4):303-313.
- 16 Wani S H, Anand S, Singh B, *et al.* WRKY transcription factors and plant defense responses: latest discoveries and future prospects. *Plant Cell Rep*, 2021, 40(7):1071-1085.
- 17 蒋莉萍, 周春苗, 钟浩天, 等. 菊花脑芳樟醇合酶基因 CnTPS1 的克隆与原核表达分析. 世界科学技术-中医药现代化, 2022, 24(4): 1391-1399.
- 18 Wu K L, Guo Z J, Wang H H, *et al.* The WRKY family of transcription factors in rice and *Arabidopsis* and their origins. *DNA Res*, 2005, 12(1):9-26.
- 19 刘微, 蒋莉萍, 池玉洁, 等. 黄连 WRKY 基因家族鉴定及表达分析. 世界科学技术-中医药现代化, 2022, 24(4):1369-1380.
- 20 常箫月, 王晨舒, 李维蛟, 等. 滇重楼 WRKY 基因家族的鉴定及密码子偏好性分析. 分子植物育种, 1-11 [2024-03-22]. <http://kns.cnki.net/kcms/detail/46.1068.S.20230705.0910.004.html>.
- 21 杨教童, 李鹏飞, 肖巧巧. 金银花 WRKY 基因家族的鉴定与功能分析. 西南大学学报(自然科学版), 2023, 45(7):87-97.
- 22 Millar A H, Carrie C, Pogson B, *et al.* Exploring the function-location nexus: using multiple lines of evidence in defining the subcellular location of plant proteins. *Plant Cell*, 2009, 21(6):1625-1631.
- 23 He G H, Xu J Y, Wang Y X, *et al.* Drought-responsive WRKY transcription factor genes TaWRKY1 and TaWRKY33 from wheat confer drought and/or heat resistance in *Arabidopsis*. *BMC Plant Biol*, 2016, 16(1):116.
- 24 Trofimov K, Ivanov R, Eutebach M, *et al.* Mobility and localization of the iron deficiency-induced transcription factor bHLH039 change in the presence of FIT. *Plant Direct*, 2019, 3(12):e00190.
- 25 Jue D, Sang X, Liu L, *et al.* Identification of *WRKY* gene family from *Dimocarpus longan* and its expression analysis during flower induction and abiotic stress responses. *Int J Mol Sci*, 2018, 19(8):2169.
- 26 Xie Z, Zhang Z L, Zou X, *et al.* Annotations and functional analyses of the rice *WRKY* gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant Physiol*, 2005, 137(1):176-189.
- 27 Wu J, Chen J, Wang L, *et al.* Genome-wide investigation of WRKY transcription factors involved in terminal drought stress response in common bean. *Front Plant Sci*, 2017, 8:380.
- 28 Babu M M, Iyer L M, Balaji S, *et al.* The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res*, 2006, 34(22):6505-6520.
- 29 Wu G A, Prochnik S, Jenkins J, *et al.* Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat Biotechnol*, 2014, 32(7):656-662.
- 30 Wang L, Guo K, Li Y, *et al.* Expression profiling and integrative analysis of the *CESA/CSL* superfamily in rice. *BMC Plant Biol*, 2010, 10:282.
- 31 Zhao P, Wang D, Wang R, *et al.* Genome-wide analysis of the potato Hsp20 gene family: identification, genomic organization and expression profiles in response to heat stress. *BMC Genomics*, 2018, 19(1):61.
- 32 Dai W S, Peng T, Wang M, *et al.* Genome-wide identification and comparative expression profiling of the WRKY transcription factor family in two *Citrus* species with different Candidatus *Liberibacter asiaticus* susceptibility. *BMC Plant Biol*, 2023, 23(1):159.
- 33 Wei W, Hu Y, Han Y T, *et al.* The WRKY transcription factors in the diploid woodland strawberry *Fragaria vesca*: identification and expression analysis under biotic and abiotic stresses. *Plant Physiol Biochem*, 2016, 105:129-144.
- 34 Xu Y H, Wang J W, Wang S, *et al.* Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-delta-cadinene synthase-A. *Plant Physiol*, 2004, 135(1):507-515.
- 35 Li C, Zha W, Li W, *et al.* Advances in the biosynthesis of terpenoids and their ecological functions in plant resistance. *Int J Mol Sci*, 2023, 24(14):11561.
- 36 Schluttenhofer C, Yuan L. Regulation of specialized metabolism by

- WRKY transcription factors. *Plant Physiol*, 2015, 167(2):295–306.
- 37 Spyropoulou E A, Haring M A, Schuurink R C. RNA sequencing on *Solanum lycopersicum* trichomes identifies transcription factors that activate terpene synthase promoters. *BMC Genomics*, 2014, 15(1):402.
- 38 Ma D, Pu G, Lei C, *et al.* Isolation and characterization of AaWRKY1, an *Artemisia annua* transcription factor that regulates the amorpho-4, 11-diene synthase gene, a key gene of artemisinin biosynthesis. *Plant Cell Physiol*, 2009, 50(12):2146–2161.
- 39 Han J, Wang H, Lundgren A, *et al.* Effects of overexpression of AaWRKY1 on artemisinin biosynthesis in transgenic *Artemisia annua* plants. *Phytochemistry*, 2014, 102:89–96.
- 40 Li S, Zhang P, Zhang M, *et al.* Functional analysis of a WRKY transcription factor involved in transcriptional activation of the DBAT gene in *Taxus chinensis*. *Plant Biol*, 2013, 15(1):19–26.
- 41 Naoumkina M A, He X, Dixon R A. Elicitor-induced transcription factors for metabolic reprogramming of secondary metabolism in *Medicago truncatula*. *BMC Plant Biol*, 2008, 8:132.
- 42 Sun W J, Zhan J Y, Zheng T R, *et al.* The jasmonate-responsive transcription factor *CbWRKY24* regulates terpenoid biosynthetic genes to promote saponin biosynthesis in *Conyza blinii* H. Lév. *J Genet*, 2018, 97(5):1379–1388.
- 43 Suttipanta N, Pattanaik S, Kulshrestha M, *et al.* The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Physiol*, 2011, 157(4):2081–2093.

Identification and Expression Analysis of WRKY Gene Family in *Chrysanthemum nankingense*

JIANG Liping¹, XU Hongfeng¹, CHENG Lu¹, LIU Yifei²

(1. Wuhan No.1 Hospital, Wuhan 430022, China; 2. College of Pharmacy, Hubei University of Chinese Medicine, Wuhan 430065, China)

Abstract: Objective Bioinformatics methods were used to identify and analyze members of the WRKY gene family in *Chrysanthemum nankingense*, which provided a theoretical basis for further studies on the regulation of WRKY transcription factors on the biosynthesis of terpenes in *C. nankingense*. Methods Online tools such as HMMER, BLAST, Pfam and NCBI were used to screen and identify the WRKY gene family members in the brain genome of *C. nankingense*. ExPASy, WOLF PSORT and other online tools for physical and chemical properties analysis; ClustalW, IQ-Tree, MEME and other software for system evolution, conservative domain analysis; TBtools software for chromosome localization, tandem replication, gene expression patterns and visual analysis. Results A total of 110 *CnaWRKYs* were identified, whose ORF length was 246–2367 bp, amino acid number was 81–788 AA, molecular weight was 9463.98–87 062.38 Da, theoretical isoelectric point was 4.63–10.15, and 93 *CnaWRKYs* were predicted to be located in the nucleus. *CnaWRKYs* protein was divided into 3 groups by phylogenetic analysis, and group II was divided into 5 subgroups. Protein domains are abundant, and the same group of *CnaWRKYs* protein domains are highly conserved with similar conserved motifs. There were 109 WRKY genes unevenly distributed on 9 chromosomes, and one *WRKY* gene was distributed on Scaffold25. The gene expression heat map showed that 108 *CnaWRKYs* genes were expressed at the whole level, and the expression abundance of WRKY gene family members in different tissues was significantly different. Conclusion The WRKY gene family members of *C. nankingense* were systematically analyzed to provide theoretical basis for further analysis of the role of *CnaWRKYs* gene in the biosynthesis of volatile aromatic compounds terpenes in *C. nankingense* and its regulatory mechanism.

Keywords: *Chrysanthemum nankingense*, WRKY transcription factor, Secondary metabolism, Tand-em replication, Expression analysis

(责任编辑: 李青)