

DOI: 10.11931/guihaia.gxzw202007039

乔刚, 李莉, 姜山. 小立碗藓 *WRKY* 基因家族生物信息学分析 [J]. 广西植物, 2022, 42(2): 267–276.QIAO G, LI L, JIANG S. Bioinformatics analysis of *WRKY* gene family in *Physcomitrella patens* [J]. *Guihaia*, 2022, 42(2): 267–276.

# 小立碗藓 *WRKY* 基因家族生物信息学分析

乔刚<sup>1</sup>, 李莉<sup>2</sup>, 姜山<sup>2\*</sup>

(1. 贵州师范大学 生命科学学院, 贵阳 550001; 2. 贵州师范大学 国际教育学院, 贵阳 550001)

**摘要:** *WRKY* 作为最先在植物中发现的转录因子, 在植物生长发育等过程中发挥重要作用。为了更好地研究小立碗藓 *WRKY* 蛋白的结构与功能, 该文以 Pfam 数据库中 *WRKY* 基因家族数据 (登录号为 PF03106) 为材料, 分析了小立碗藓 (*Physcomitrella patens*) *WRKY* 基因家族成员的理化性质、蛋白质的二级结构预测、染色体定位、内外显子分布及系统进化关系。结果表明: (1) 小立碗藓 *WRKY* 基因家族成员共有 38 个基因, 根据 *WRKY* 保守结构域个数和锌指结构类型分成 I、II 两大类, 不含 III 类 (锌指结构为 C2HC 型), 其中部分基因 *WRKY* 保守结构域发生变异。 (2) *WRKY* 蛋白氨基酸长度在 216~775 aa 之间、相对分子质量在 24.5~82.8 kDa 之间, 亚细胞定位显示 *WRKY* 家族成员蛋白质定位于细胞核中。 (3) *WRKY* 蛋白的二级结构以  $\alpha$ -螺旋、延伸链、 $\beta$ -转角、无规卷曲四种构成元件构成, 除 PpWRKY11 ( $\alpha$ -螺旋为主) 外, 其余无规卷曲占比高达 70%。 (4) 与拟南芥的系统进化关系表明, 植物在进化过程中 *WRKY* 家族成员的数目与进化方式发生改变, *WRKY* 基因家族成员外显子的个数为 3~7 个。 (5) 小立碗藓 *WRKY* 基因家族成员无规则分散于 21 条染色体上, 并未形成基因簇。该研究通过分析 *WRKY* 基因家族的基本结构与性质, 能为后续深入研究 *WRKY* 转录因子的功能奠定基础。

**关键词:** 小立碗藓, *WRKY* 转录因子, 基因家族分析, 生物信息

中图分类号: Q943 文献标识码: A 文章编号: 1000-3142(2022)02-0267-10

## Bioinformatics analysis of *WRKY* gene family in *Physcomitrella patens*

QIAO Gang<sup>1</sup>, LI Li<sup>2</sup>, JIANG Shan<sup>2\*</sup>

(1. School of Life Sciences, Guizhou Normal University, Guiyang 550001, China; 2. School of International Education, Guizhou Normal University, Guiyang 550001, China)

**Abstract:** As a transcription factor first found in plants, *WRKY* plays an important role in plant growth and development. However, it has not been studied in *Physcomitrella patens*. By using the *WRKY* gene family data (accession number is PF03106) in Pfam database, this paper studied the basic information of *WRKY* in *P. patens*, which included physicochemical properties, protein secondary structure prediction, chromosome localization, exon and intron distribution and phylogenetic relationship. The results were as follows: (1) The *WRKY* gene family in *P. patens* consisted of 38 members which were divided into two major categories—I and II, and the conserved domains of some *WRKY* genes

收稿日期: 2020-11-28

基金项目: 国家自然科学基金 (31260426, 31560508) [Supported by National Natural Science Foundation of China (31260426, 31560508)].

第一作者: 乔刚 (1992-), 硕士研究生, 主要从事植物分子生物学研究, (E-mail) 2335603452@qq.com。

\*通信作者: 姜山, 博士, 教授, 研究方向为植物病理学, (E-mail) kyosan200312@hotmail.com。

had mutated. (2) The amino acid length of WRKY protein was 216–775 aa and the relative molecular mass was 24.5–82.8 kDa; Subcellular location showed that the proteins of WRKY gene family were distributed in the nucleus. (3) The secondary structure of WRKY protein was composed of four constituent elements:  $\alpha$ -helix, extended strand,  $\beta$ -turn, and random coil; Except for PpWRKY11 ( $\alpha$ -helix dominated), the proportion of random coil was up to 70%. (4) The phylogenetic relationship with *Arabidopsis* showed that the number of the WRKY family members and the way of their revolution had changed in the process of plants' evolution, the number of exons of WRKY gene family members were 3–7. (5) The members of WRKY gene were randomly dispersed on 21 chromosomes without forming a gene cluster. This study analyzed the basic structure and properties of the WRKY gene family, and it was found that the WRKY gene family of *Physcomitrella patens* has evolutionary diversity and unique variability in conserved domains, laying the foundation for subsequent research.

**Key words:** *Physcomitrella patens*, WRKY transcription factor, gene family analysis, bioinformatics

目前,在植物中发现 60 多个转录因子家族 (Kaplan-Levy et al., 2012; Qin et al., 2014), 其中研究较多的有 MYB、NAC、WRKY、SBP、GARS 等。WRKY(transcription factor)作为最先在植物中发现的转录因子 (Wang et al., 2015), 它由 N-端的 WRKY 保守结构域 (WRKYGQK) 和 C-端的锌指结构组成 (Wang et al., 2018)。根据保守结构域的个数和锌指结构的类型, 可将 WRKY 转录因子分成三类: I 类, 含有 2 个 WRKY 结构域、锌指结构为 C2H2 型; II 类, 含有单个结构域, 锌指结构为 C2H2 (CX4-5-C-X22-23-H-X1-H), 由此将其分为五个亚型 (II a、II b、II c、II d、II e); III 类, 单个保守结构域、锌指为 C2HC (C-X7-C-X23-HX) 型 (Song et al., 2015; Jia et al., 2018)。但是, 在低等植物中不含有 III 类类型的 WRKY 转录因子 (苏琦等, 2007)。

小立碗藓 (*Physcomitrella patens*) 作为高等植物中的低等类群, 它的优势主要体现在以下几点: (1) 小立碗藓基因组测序结果已知 (<http://www.cosmoss.org/>) (Rensing et al., 2008), 大小为 511 Mb, 染色体共有 27 条 (Rensing et al., 2002); (2) 同源重组率高; (3) 生长周期短、易于扩繁 (Xu et al., 2009); (4) 细胞结构简单, 表型能直接观察。因此, 小立碗藓已成为研究植物基因结构与功能的一种理想植物 (蓝雨纯等, 2020)。

随着生物技术的发展, WRKY 转录因子不断被发现。目前, 已经鉴定出绿藻中有 3 个 WRKY 转录因子 (Rinerson et al., 2015)、卷柏 19 个 (Li et al., 2016)、山松 83 个 (Liu et al., 2009)、拟南芥 74 个 (Ulker et al., 2004)、水稻 97 个 (Xu et al., 2016)。同时, WRKY 作为转录因子已经被证明调控植物多重生理功能 (Eulgem et al., 2000)。有研究发现, WRKY 转录因子参与植物对干旱、高温、低温等非

生物胁迫的应答反应 (Chen et al., 2012) 以及植物对病原菌的抗性 (Sham et al., 2017), 如水稻 OsWRKY13 参与抵抗水稻稻瘟病菌 (*Magnaporthe grisea*) 的感染 (Schlutenhofer & Yuan, 2014), AtWRKY8 在盐胁迫下表达出现增加, 进而增强植株对盐胁迫的抗性 (Chen et al., 2013); WRKY 转录因子与植物的生长发育 (Chen et al., 2019) 和代谢产物生物的合成 (Zhang et al., 2018; Liu et al., 2019) 密切相关, 但在小立碗藓中, WRKY 是否参与防御反应及其作用机制却鲜有报道。随着生物技术、计算机技术与大数据的迅猛发展 (Pearson, 2001), 人们对于未知领域的研究也不断加强。生物信息学技术作为一门新兴的时尚学科, 结合计算机数据收集与分析, 不断探索未知基因的结构与功能, 使得预测结果逐渐具有说服力, 为科学研究做出了很大贡献, 这一技术广泛应用于蛋白分析、基因结构预测等众多领域。因此, 本文通过生物信息学分析研究 WRKY 基因家族成员的基本信息、保守结构域分析、染色体定位、内外显子分布等信息, 旨在为后续研究小立碗藓 WRKY 蛋白的结构与功能提供理论基础。

## 1 材料与amp;方法

### 1.1 材料

Bio-linux 操作系统; 虚拟机 virtual Box; 各种数据库, 即 Pfam (<http://pfam.xfam.org/>)、Ensembl Plants (Bolseret al., 2017) (<http://plants.ensembl.org/index.html>)、phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>); 在线构图工具及可视化软件, 即 Adobe illustrator、TBtools 等。

### 1.2 小立碗藓 WRKY 基因家族的确定

在 Pfam 数据库下载 WRKY 蛋白保守结构域

(\* .hmm) 文件,在 Ensembl Plants 数据库中下载小立碗藓基因组的 cds、cdna、pep、gff3 等序列作为备用文件。

在小立碗藓蛋白质数据库检索含有 *WRKY* (PF03106) 基因家族结构域的蛋白质序列(李昊阳等,2014),通过 E 值初筛,再用 SMART 网站(<http://smart.embl-heidelberg.de/>)进行手动筛选,最终结果用于后续研究。

### 1.3 小立碗藓 *WRKY* 基因家族成员的理化性质分析与亚细胞定位

由 ProtParam (<https://web.expasy.org/protparam/protparam-doc.html>) 网站预测家族成员的等电点(PI)、相对分子质量、氨基酸长度、蛋白质疏水性等性质;通过 PLOC (<http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/>)、Plant-mPLOC (植物细胞定位) 预测家族成员在细胞中的位置分布。

### 1.4 小立碗藓 *WRKY* 基因家族成员蛋白的二级结构分析

将确定的小立碗藓 *WRKY* 基因家族蛋白质的氨基酸序列通过在线网站 HNN Secondary Structure Prediction ([https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=npsa\\_sopma.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)) 预测其二级结构各构型的占比。

### 1.5 小立碗藓 *WRKY* 基因家族成员系统发育树的构建

利用 MEGA 7.0 软件进行进化树构建,程序选

用最大似然法,Bootstrap 值设定为 1 000,其他设定参数以系统默认值为准。

### 1.6 小立碗藓 *WRKY* 基因家族成员内外显子及保守结构域分析

由 MEME 在线网 (<http://gsds.cbi.pku.edu.cn/>),预测蛋白质的保守结构域的 motif 模型信息。通过 <http://gsds.cbi.pku.edu.cn/> 网站绘制小立碗藓 *WRKY* 基因家族成员的内外显子分布图,由 AI 软件合并绘图。

### 1.7 小立碗藓 *WRKY* 基因家族成员的染色体定位图

将 *WRKY* 蛋白所对应的氨基酸序列及成员在小立碗藓基因组中对应的名称文件、小立碗藓 27 条染色体的长度信息;由 [http://mg2c.iask.in/mg2c\\_v2.0/](http://mg2c.iask.in/mg2c_v2.0/) 在线网站绘制染色体定位图,通过 TBtools 工具进行可视化修饰。

## 2 结果与分析

### 2.1 小立碗藓 *WRKY* 基因家族成员的确定

通过对 38 个 *WRKY* 家族成员的保守结构域序列比对将 *WRKY* 家族成员主要分为以下几类: I 类(3 个基因)、II a(5 个基因)、II b(6 个基因)、II c(5 个基因)、II d(13 个基因)、II e(6 个基因)、不含第三类(C2HC 型),其中部分基因的保守结构域序列发生变异(表 1)。

表 1 小立碗藓 *WRKY* 保守结构域的变异

Table 1 Variation of conserved domain of *Physcomitrella patens* *WRKY*

基因名称 Gene name	<i>WRKY</i> 结构域序列 WRKY domain sequence	变异后保守结构域序列 Conserved domain sequence after mutation	分组 Groups
<i>PpWRKY7</i>	WRKYGQK	WKKYGNK	II a
<i>PpWRKY8</i>	WRKYGQK	WKKYGNK	II a
<i>PpWRKY14</i>	WRKYGQK	WKKYGNK	II a
<i>PpWRKY25</i>	WRKYGQK	WRKYGHK	II d
<i>PpWRKY30</i>	WRKYGQK	WRKYGQN	II b
<i>PpWRKY32</i>	WRKYGQK	WKKYGNK	II a
<i>PpWRKY35</i>	WRKYGQK	WKKYGNK	II a

### 2.2 小立碗藓 *WRKY* 基因家族成员理化性质分析及亚细胞定位

由表 2 可知,整个家族成员的氨基酸序列长度在 216~775 aa 之间,*PpWRKY25* 最小,*PpWRKY10* 最大。*WRKY* 蛋白氨基酸的理论等电点 PI 值在 5.10~9.79 之间,蛋白质的相对分子质量在 24.5~

82.8 kDa 之间,与氨基酸数量成正比。另外,通过蛋白质疏水性数值可以得到 *PpWRKY25*、*PpWRKY30*、*PpWRKY36* 既具有疏水性又有亲水性,其数值在 -0.5~+0.5 之间。负值越大说明亲水性越强,其他氨基酸都为亲水氨基酸。PLOC 亚细胞定位结果表明小立碗藓 *WRKY* 蛋白全部分布在细胞核中。

表 2 小立碗藓 WRKY 转录因子理化性质分析

Table 2 Analysis of physicochemical properties of WRKY transcription factors in *Physcomitrella patens*

基因名称 Gene name	等电点 PI	分子量 Molecular weight (kDa)	氨基酸长度 Length of amino acid (aa)	亚细胞定位 Subcellular localization	蛋白质疏水性 Grand average of hydropathicity (GRAVY)
<i>PpWRKY1</i>	9.71	43.1	395	Nucleus 细胞核	-0.693
<i>PpWRKY2</i>	6.41	82.0	765	Nucleus 细胞核	-0.601
<i>PpWRKY3</i>	8.79	75.4	705	Nucleus 细胞核	-0.970
<i>PpWRKY4</i>	9.76	43.2	396	Nucleus 细胞核	-0.655
<i>PpWRKY5</i>	9.10	74.2	684	Nucleus 细胞核	-1.052
<i>PpWRKY6</i>	6.62	81.3	761	Nucleus 细胞核	-0.560
<i>PpWRKY7</i>	6.08	37.7	343	Nucleus 细胞核	-1.047
<i>PpWRKY8</i>	7.68	40.3	365	Nucleus 细胞核	-0.936
<i>PpWRKY9</i>	6.71	60.6	558	Nucleus 细胞核	-0.648
<i>PpWRKY10</i>	8.21	82.8	775	Nucleus 细胞核	-0.751
<i>PpWRKY11</i>	9.79	42.8	385	Nucleus 细胞核	-0.719
<i>PpWRKY12</i>	5.66	55.8	507	Nucleus 细胞核	-0.845
<i>PpWRKY13</i>	5.89	41.1	372	Nucleus 细胞核	-0.991
<i>PpWRKY14</i>	8.31	65.2	589	Nucleus 细胞核	-0.759
<i>PpWRKY15</i>	9.75	43.2	395	Nucleus 细胞核	-0.643
<i>PpWRKY16</i>	5.69	71.2	642	Nucleus 细胞核	-0.840
<i>PpWRKY17</i>	9.72	42.4	385	Nucleus 细胞核	-0.829
<i>PpWRKY18</i>	8.55	48.7	437	Nucleus 细胞核	-0.688
<i>PpWRKY19</i>	8.45	78.9	749	Nucleus 细胞核	-0.717
<i>PpWRKY20</i>	7.60	55.7	504	Nucleus 细胞核	-0.820
<i>PpWRKY21</i>	6.06	59.3	533	Nucleus 细胞核	-0.754
<i>PpWRKY22</i>	5.76	56.6	511	Nucleus 细胞核	-0.809
<i>PpWRKY23</i>	6.46	81.7	758	Nucleus 细胞核	-0.813
<i>PpWRKY24</i>	5.27	63.0	579	Nucleus 细胞核	-0.637
<i>PpWRKY25</i>	7.15	24.5	216	Nucleus 细胞核	-0.295
<i>PpWRKY26</i>	7.37	80.5	740	Nucleus 细胞核	-0.694
<i>PpWRKY27</i>	5.10	52.8	480	Nucleus 细胞核	-0.752
<i>PpWRKY28</i>	5.62	62.8	576	Nucleus 细胞核	-0.789
<i>PpWRKY29</i>	5.40	64.0	584	Nucleus 细胞核	-0.806
<i>PpWRKY30</i>	8.01	44.3	409	Nucleus 细胞核	-0.463
<i>PpWRKY31</i>	5.33	57.2	525	Nucleus 细胞核	-0.737
<i>PpWRKY32</i>	5.94	59.9	532	Nucleus 细胞核	-0.955
<i>PpWRKY33</i>	7.79	61.2	566	Nucleus 细胞核	-0.720
<i>PpWRKY34</i>	5.99	63.2	588	Nucleus 细胞核	-0.571
<i>PpWRKY35</i>	5.67	48.1	435	Nucleus 细胞核	-0.571
<i>PpWRKY36</i>	8.71	52.7	491	Nucleus 细胞核	-0.454
<i>PpWRKY37</i>	8.23	79.5	748	Nucleus 细胞核	-0.512
<i>PpWRKY38</i>	5.96	79.3	723	Nucleus 细胞核	-0.867

### 2.3 小立碗藓 WRKY 基因家族成员二级结构预测

由表 3 可知,小立碗藓 WRKY 基因家族成员的蛋白质由四种结构元件构成( $\alpha$ -螺旋、延伸链、 $\beta$ -转角、无规卷曲),其中以无规卷曲为主要构成元

件,最高为 PpWRKY19,占比达 73.83%,其次是  $\alpha$ -螺旋;而 PpWRKY11 结果与家族成员不同,以  $\alpha$ -螺旋为主,占比高达 42.34%,无规卷曲为 39.74%。另外两种构成元件占比均较小,比较稳定。

## 2.4 小立碗藓 *WRKY* 基因家族成员系统进化关系分析

从图 1 小立碗藓 *WRKY* 基因家族成员的进化树分析可以看出, *WRKY* 基因家族成员在进化的过程中已经发生了改变, 38 个基因分布在不同的五个分支上(分别含有 I 型基因 3 个、II a 型基因 5 个、II b 型基因 6 个、II c 型基因 5 个、II d 型基因 13 个、II e 型 6 个)。从进化树还可以看出, 小立碗藓与拟南芥在进化上具有一定的差异性, 其中的 II a、II d 并未聚集在同一分支, 说明植物由低等向高等进化的过程中基因的进化方式发生了改变, 出现了分化。

## 2.5 小立碗藓 *WRKY* 基因家族成员内外显子分布及结构域分析

小立碗藓 *WRKY* 家族成员主要分布在几个不同的分支, 在进化上既具有差异性又具有相似性。由图 2:B 可知, 小立碗藓 *WRKY* 基因家族成员内外显子分布图, 外显子的个数为 3~7 个。其中, I 型有 4~6 个外显子、II a 与 II e 型外显子 4 个、II b 型含外显子 3~7 个、II c 型外显子 3 个、大多数 II d 型含有外显子 3 个。内含子的长度可以通过图下比例尺推断得到。由图 2:C 可知, 不同 *WRKY* 基因所对应的蛋白结构域共有 10 个 motif, 不同基因对应的 motif 也有所不同。由表 4 可知, 每个 motif 的宽度、E-value 值、对应 motif 的 Logo 图; 氨基酸宽度最大为 50, 最低仅有 21。这 38 个家族成员中约有 50% 具有两个保守结构域, 最多含有四个; 而 II a (*PpWRKY7*、*PpWRKY8*、*PpWRKY14*、*PpWRKY35*) 都只含有一个 motif 5 的保守结构域, 并未形成稳定的 motif 1-motif 3 的稳定结构。

## 2.6 小立碗藓 *WRKY* 基因家族成员的染色体分布

由图 3 可知, 通过分析最终将 38 个 *WRKY* 基因定位在小立碗藓 27 条染色体中的 21 条, I 型 (*PpWRKY3*、*PpWRKY5*、*PpWRKY19*) 分布于第 8、第 3、第 2 号染色体; II a 型 (*PpWRKY35*、*PpWRKY31*、*PpWRKY8*、*PpWRKY7*、*PpWRKY14*) 分布于第 14、第 4、第 27、第 6、第 5 号染色体; II b 型 (*PpWRKY36*、*PpWRKY30*、*PpWRKY2*、*PpWRKY6*、*PpWRKY37*、*PpWRKY26*) 分布于第 11、第 7、第 13、第 3、第 12、第 4 号染色体; II c 型 (*PpWRKY34*、*PpWRKY9*、*PpWRKY33*、*PpWRKY24*、*PpWRKY28*) 分布于第 20、第 24、第 23、第 3、第 13 号染色体; II d 型 (*PpWRKY29*、*PpWRKY22*、*PpWRKY25*、*PpWRKY38*、*PpWRKY16*、*PpWRKY10*、*PpWRKY23*、*PpWRKY21*、

表 3 小立碗藓 *WRKY* 蛋白质二级结构分析

Table 3 Secondary structure analysis of *WRKY* protein in *Physcomitrella patens*

基因名称 Gene name	二级结构 Secondary structure (%)			
	$\alpha$ -螺旋 $\alpha$ -helix	延伸链 Extended strand	$\beta$ -转角 $\beta$ -turn	无规卷曲 Random coil
<i>PpWRKY1</i>	26.33	12.91	6.08	54.68
<i>PpWRKY2</i>	32.03	13.59	2.61	51.76
<i>PpWRKY3</i>	14.61	9.50	3.97	71.91
<i>PpWRKY4</i>	21.97	15.91	6.06	56.06
<i>PpWRKY5</i>	16.96	8.63	4.24	70.18
<i>PpWRKY6</i>	29.57	13.53	3.55	53.55
<i>PpWRKY7</i>	15.74	13.70	6.71	63.85
<i>PpWRKY8</i>	22.74	15.34	7.67	54.25
<i>PpWRKY9</i>	24.37	8.96	3.41	63.26
<i>PpWRKY10</i>	19.35	9.81	6.97	63.87
<i>PpWRKY11</i>	42.34	11.34	6.49	39.74
<i>PpWRKY12</i>	31.76	8.88	6.90	52.47
<i>PpWRKY13</i>	33.39	10.19	5.49	50.94
<i>PpWRKY14</i>	26.32	10.53	3.90	59.25
<i>PpWRKY15</i>	25.06	14.94	8.35	51.65
<i>PpWRKY16</i>	28.19	12.15	6.70	52.96
<i>PpWRKY17</i>	25.45	12.47	4.94	51.74
<i>PpWRKY18</i>	24.49	10.76	6.18	58.58
<i>PpWRKY19</i>	12.68	9.35	4.14	73.83
<i>PpWRKY20</i>	33.73	8.13	4.76	53.37
<i>PpWRKY21</i>	27.77	10.51	5.44	56.29
<i>PpWRKY22</i>	19.37	9.59	6.65	64.38
<i>PpWRKY23</i>	18.07	8.05	5.41	68.47
<i>PpWRKY24</i>	17.10	12.61	7.77	62.52
<i>PpWRKY25</i>	43.06	7.41	1.85	47.69
<i>PpWRKY26</i>	27.43	12.70	2.97	56.98
<i>PpWRKY27</i>	32.29	8.96	5.42	53.33
<i>PpWRKY28</i>	21.18	11.81	4.69	62.33
<i>PpWRKY29</i>	27.10	9.42	7.53	55.14
<i>PpWRKY30</i>	27.63	15.65	6.60	50.12
<i>PpWRKY31</i>	20.00	9.33	3.24	67.43
<i>PpWRKY32</i>	29.89	11.28	4.51	54.32
<i>PpWRKY33</i>	23.85	8.83	4.42	62.90
<i>PpWRKY34</i>	27.55	9.01	4.42	59.01
<i>PpWRKY35</i>	23.91	9.89	5.06	61.15
<i>PpWRKY36</i>	32.59	13.65	6.31	47.45
<i>PpWRKY37</i>	26.60	12.70	3.07	57.62
<i>PpWRKY38</i>	27.25	6.64	5.67	60.44



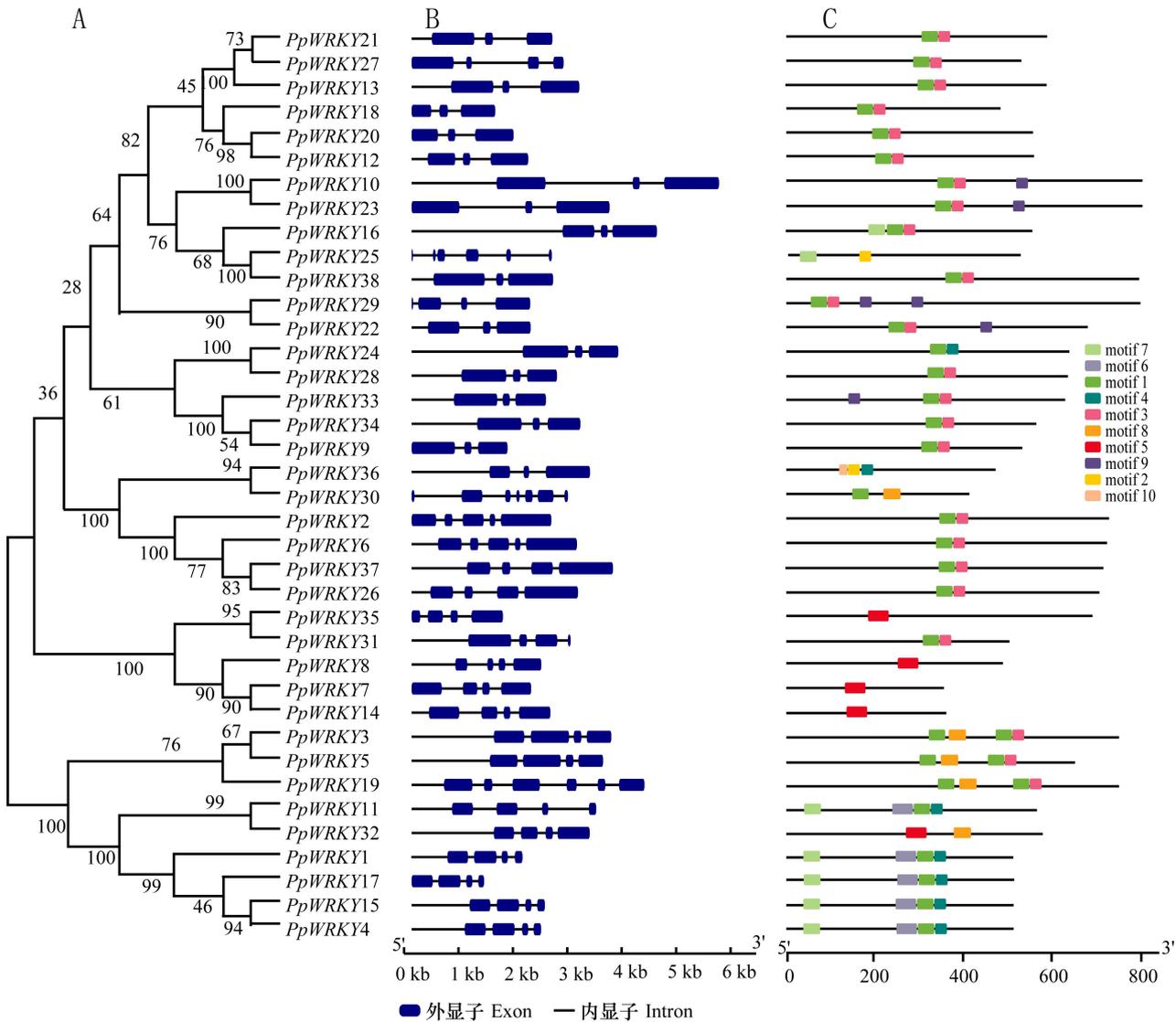


图 2 小立碗藓 *WRKY* 基因进化关系 (A) 和内外显子分布图 (B) 以及小立碗藓 *WRKY* 蛋白结构域的 motif 模型 (C)

Fig. 2 Evolutionary relationship (A) and distribution of exons and introns (B) of *Physcomitrella patens* *WRKY* gene family, and motif model of the *WRKY* protein domain of *P. patens* (C)

和锌指结构将小立碗藓 *WRKY* 家族分成两大类 (I 型 3 个、II a 型 5 个、II b 型 6 个、II c 型 5 个、II d 型 13 个、II e 型 6 个, 不含 III 型); 而在高等模式植物拟南芥中含 III 型基因 14 个 (Ulker et al., 2004)、水稻含有 47 个 (Ross et al., 2007), 且在高等植物中几乎所有的 III 类 *WRKY* 因子都参与生物胁迫反应, 而低等植物不含第 III 类 *WRKY* 因子 (苏琦等, 2007)。本实验室前期转录组数据显示, 小立碗藓在接种灰霉菌的过程中 *PpWRKY10* 出现表达量的上调, 说明 *WRKY* 转录因子参与了植物对

灰霉菌的防御反应 (实验数据未发表), 但关于具体的防御机制仍有待深入研究。进化分析表明 I 型被认为是 II 型与 III 型的原始祖先, II 型与 III 型是通过 I 型 C 末端或 N 末端 *WRKY* 结构域的变化或缺失演变而来 (Zhang et al., 2005), 说明植物由低等向高等进化的过程中对环境的适应能力不同最终可能导致进化的类型不同, 高等植物 *WRKY* 家族中的 III 型可能由 I 型在应对环境压力过程中产生, 表明 *WRKY* 基因家族在进化过程中具有多样性, 基因功能不断进化。

表 4 WRKY 转录因子的预测 motif 列表

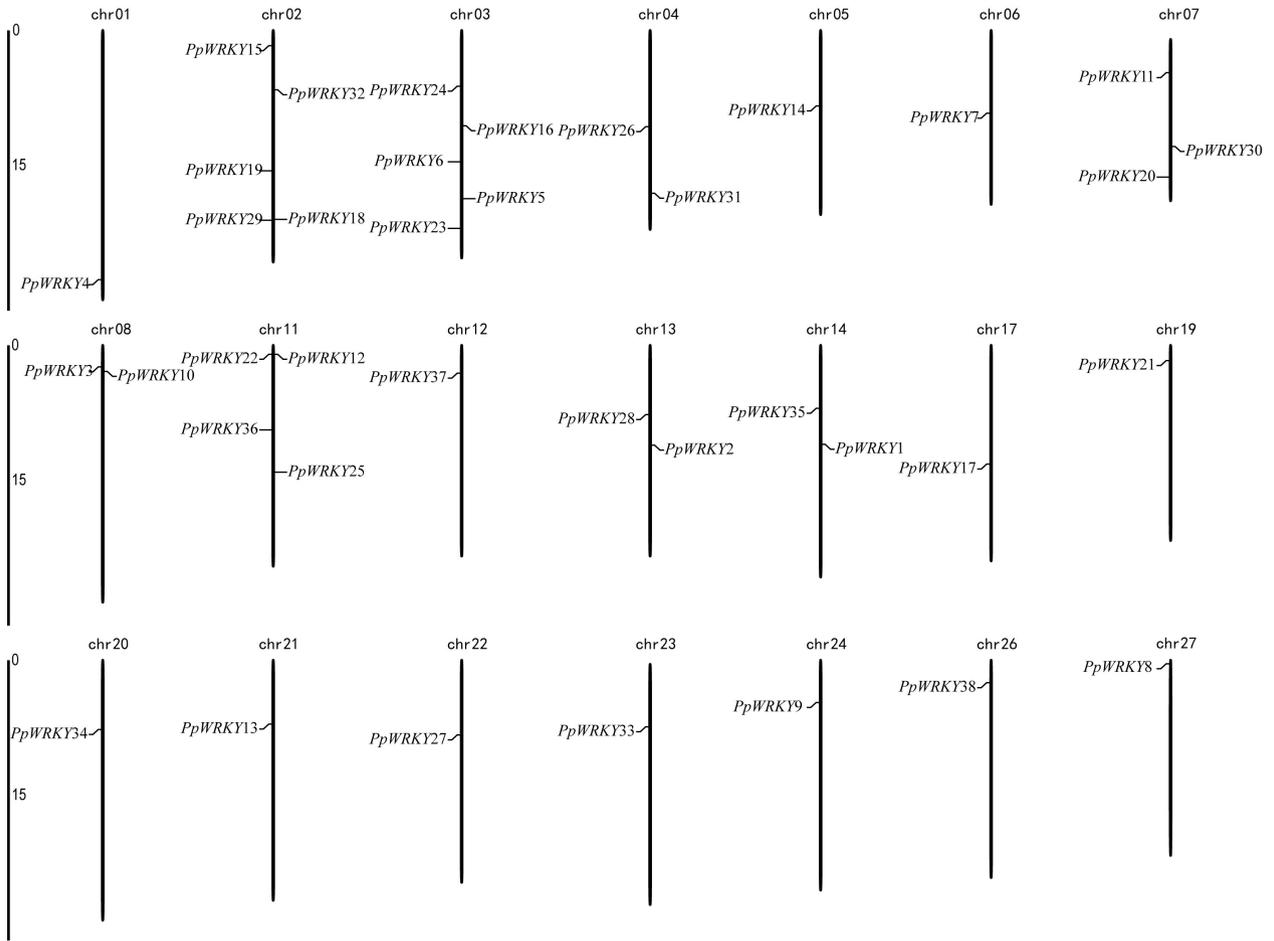
Table 4 Predictive motif list of WRKY transcription factors

基序名 Motif name	宽度 Width (aa)	E-value	Logo
motif 1	40	1.9e-1485	
motif 2	29	2.0e-1224	
motif 3	29	1.4e-1039	
motif 4	29	5.2e-969	
motif 5	50	2.0e-660	
motif 6	50	4.4e-592	
motif 7	41	3.3e-561	
motif 8	42	1.2e-349	
motif 9	29	4.8e-368	
motif 10	21	4.1e-286	

通过结构分析发现,小立碗藓部分 II 型 *WRKY* 基因保守结构域发生变异,主要涉及 WKYGNK、WRKYGHK、WRKYGQN 三种变异类型。张凡等 (2018) 研究表明,WRKY 家族保守结构域大都是 ‘Q’ 突变为 ‘E’ ‘K’ 或 ‘S’,这些变异会导致 WRKY 蛋白与 DNA 结合的活性减弱,并且不同变异类型对植株的功能影响也不相同。水稻 *OSWRKY45* 中 ‘Q’ 变异为 ‘K’ 后,其表达量在植株不同部位均出现了上调表达,并且在干旱胁迫下转基因的植株表现出更好的恢复力,过表达 *OSWRKY45* 植株的抗病与抗旱能力显著提高 (Qiu & Yu, 2009)。但是,在小立碗藓中出现了由 ‘Q’ 突变为 ‘H’ 的不同变异类型,结合实验室前期数据植株在接种灰霉菌后突变基因并未上调表达,可能突变基因并未参与生物胁迫反应,是否参与非生物胁迫目前尚未见有报道。本研究结果为后续深入研究其基因功能提供了方向。对 WRKY 家族保守结构域的 motif 模型分析发现,家族成员中出

现 motif 的缺失,并未全部形成 motif 1-motif 3 的稳定结构,解释了由单个保守结构域构成的 II a 型 WRKY 保守结构域部分发生突变的原因,表明物种可能在变异中不断获得进化。蛋白质的二级结构除 PpWRKY11 以  $\alpha$ -螺旋为主要构成元件外,其他基因均以无规卷曲 (占比达 70%) 为主要构成元件,且蛋白全分布于细胞核中。染色体定位显示,38 个基因分散分布于小立碗藓的 21 条染色体上,呈现明显的不规则分布;毛果杨中 86 个 WRKY 家族基因在染色体上也呈现不规则分布 (He et al., 2012)。

从进化关系图来看,WRKY 家族的聚类与分析一致,结构上 WRKY 蛋白保守程度较高,但也存在变异的情况,说明小立碗藓 WRKY 转录因子在进化过程中具有多样性。其中,结构域突变基因除了 Ppwrky25 没聚在一起外,其他均聚合在同一分支上,说明这些转录因子可能具有相似的功能,此现象同样出现在土豆 WRKY 转录因子中 (Liu et al., 2017)。

图 3 小立碗藓 *WRKY* 基因家族的染色体分布图Fig. 3 Chromosome distribution of *WRKY* gene family

*WRKY* 基因在植物生长发育及抗逆境胁迫中具有重要意义,一直是研究的热点问题,但在小立碗藓中 *WRKY* 基因的研究较少。本文在基因水平上对 *WRKY* 基因展开研究,为后续深入研究 *WRKY* 蛋白的结构与功能奠定了基础。

### 参考文献:

- BAI J, PENNILL LA, NING JC, et al., 2002. Diversity in nucleotide binding site-leucine-rich repeat genes in cereals [J]. *Cytogenet Genome Res*, 12(12): 1871–1884.
- BOLSER DM, RUQRTI B, ROBERT J, et al., 2017. Ensembl plants: Integrating tools for visualizing, mining, and analyzing plant genomics data [J]. *Methods Mol Biol*, 1533: 1–31.
- CHEN LG, ZHANG LP, LI DB, et al., 2013. WRKY8 transcription factor functions in the TMV-cg defense response

by mediating both abscisic acid and ethylene signaling in *Arabidopsis* [J]. *PNAS*, 110(21): E1963–E1971.

- CHEN X, CHEN RH, WANG YF, et al., 2019. Genome-Wide Identification of WRKY Transcription Factors in Chinese jujube (*Ziziphus jujuba* Mill.) and their involvement in fruit developing, ripening, and abiotic stress [J]. *Genes*, 10(5): 360.
- EULGEM T, RUSHTON PJ, ROBATZEK S, et al., 2000. The WRKY superfamily of plant transcription factors [J]. *Trends Plant Sci*, 5(5): 199–206.
- HE HS, DONG Q, SHAO YH, et al., 2012. Genome-wide survey and characterization of the WRKY gene family in *Populus trichocarpa* [J]. *Plant Cell Reports*, 31(7): 1199–1217.
- ISHIGURO S, NAKAMURA K, 1994. 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  $\beta$ -amylase from sweet potato [J]. *Mol Gen Genet*, 244(6): 563–571.
- JIA CH, WANG Z, ZHANG JB, et al., 2018. Cloning and expression analysis of eight WRKY transcription factors in

- bananas [J]. *Chin J Trop Crop*, 39(11): 87-93. [贾彩红, 王卓, 张建斌, 等, 2018. 香蕉中 8 个 WRKY 转录因子的克隆及表达分析 [J]. *热带作物学报*, 39(11): 87-93.]
- KAPLAN-LEVY RN, BREWER PB, QUON T, et al., 2012. The Trihelix family of transcription factors-light, stress and development [J]. *Trends Plant Sci*, 17(3): 163-171.
- LAN YC, HUANG B, WEI J, et al., 2020. Identification and bioinformatics analysis of the expansin gene family of *Physcomitrella patens* [J]. *Guihaia*, 40(6): 854-863. [蓝雨纯, 黄彬, 韦娇, 等, 2020. 小立碗藓扩展蛋白基因家族的鉴定与生物信息学分析 [J]. *广西植物*, 40(6): 854-863.]
- LIU Y, CAO T, CHEN JW, 2007. Advances on the study of the moss *Physcomitrella patens*, a potential model plant [J]. *Guihaia*, 27(1): 90-94. [刘艳, 曹同, 陈静文, 2007. 有前景的模式植物小立碗藓的研究新进展 [J]. *广西植物*, 27(1): 90-94.]
- LIU JJ, EKRAMODDOULLAH AKJ, 2009. Identification and characterization of the WRKY transcription factor family in *Pinus monticola* [J]. *Genome*, 52(1): 77-88.
- LIU QN, LIU Y, XIN ZZ, et al., 2017. Genome-wide identification and characterization of the WRKY gene family in potato (*Solanum tuberosum*) [J]. *Biochem Syst Ecol*, 71: 212-218.
- LIU Y, YANG TY, LIN ZK, et al., 2019. A WRKY transcription factor PbrWRKY53 from *Pyrus betulaefolia* is involved in drought tolerance and AsA accumulation [J]. *Plant Biotechnol*, 17(9): 1770-1787.
- LI HY, SHI Y, DING YN, et al., 2014. Bioinformatics analysis of expansin gene family in poplar genome [J]. *J Beijing For Univ*, 36(2): 59-67. [李昊阳, 施杨, 丁亚娜, 等, 2014. 杨树扩展蛋白基因家族的生物信息学分析 [J]. *北京林业大学学报*, 36(2): 59-67.]
- LI MY, XU ZS, TIAN C, et al., 2016. Genomic identification of WRKY transcription factors in carrot (*Daucus carota*) and analysis of evolution and homologous groups for plants [J]. *Sci Rep*, 6: 23101.
- PEARSONNR, 2001. Training for bioinformatics and computational biology [J]. *Bioinformatics*, 17(9): 761-762.
- QIN Y, MA X, YU GH, et al., 2014. Evolutionary history of trihelix family and their functional diversification [J]. *DNA Res*, 21(5): 499-510.
- QIU YP, YU DQ, 2009. Over-expression of the stress-induced OsWRKY45 enhances disease resistance and droughttolerance in *Arabidopsis* [J]. *Environ Exp Bot*, 65(1): 35-47.
- RENSING SA, ROMBAUTS S, PEER YVD, et al., 2002. Moss transcriptome and beyond [J]. *Trends Plant Sci*, 7(12): 535-538.
- ROSS CA, LIU Y, SHEN QJ, 2007. The WRKY gene family in rice (*Oryza sativa*) [J]. *J Integr Plant Biol*, 49(6): 827-842.
- RINERSON CI, RABARA RC, TRIPATHIP, et al., 2015. The evolution of WRKY transcription factors [J]. *BMC Plant Biol*, 15(1): 66.
- SCHUTTENHOFER C, YUAN L, 2014. Regulation of specialized metabolism by WRKY transcription factors [J]. *Plant Physiol*, 167(2): 295-306.
- SU Q, SHANG YH, DU MY, et al., 2007. Progress on plant WRKY transcription factor [J]. *Chin Agric Sci Bull*, 23(5): 94-98. [苏琦, 尚宇航, 杜密英, 等, 2007. 植物 WRKY 转录因子研究进展 [J]. *中国农学通报*, 23(5): 94-98.]
- SONG H, SUN WH, YANG GF, et al., 2018. WRKY transcription factors in legumes [J]. *BMC Plant Biol*, 18(1): 243.
- ULKER B, SOSICH IEJ, 2004. WRKY transcription factors: from DNA binding towards biological function [J]. *Curr Opin Plant Biol*, 7(5): 491-498.
- WANG YY, FENG L, ZHU YX, et al., 2015. Comparative genomic analysis of the WRKY III gene family in populus, grape, arabidopsis and rice [J]. *Biol Direct*, 10: 28.
- WANG CT, RU JN, LIU YW, et al., 2018. Maize WRKY transcription factor *ZmWRKY106* confers drought and heat tolerance in transgenic plants [J]. *Int J Mol Sci*, 19(10): 2-15.
- XU ZY, ZHANG DD, HU J, et al., 2009. Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom [J]. *BMC Bioinformatics*, 10 (Suppl. 11): 1471-1486.
- XU HJ, WATANABE KA, ZHANG LY, et al., 2016. WRKY transcription factor genes in wild rice *Oryza nivara* [J]. *DNA Res*, 23(4): 311-323.
- ZHAO H, ZHAO XG, HE YK, et al., 2004. *Physcomitrella patens*, a potential model system in plant molecular biology [J]. *Chin Bull Bot*, 21(2): 129-138. [赵奂, 赵晓刚, 何奕昆, 等, 2004. 植物分子生物学研究极具前景的模式系统—小立碗藓 [J]. *植物学报*, 21(2): 129-138.]
- ZHANG F, YIN J, GUO YQ, et al., 2018. Research advances on WRKYtranscription factors [J]. *Biotechnol Bull*, 34(1): 40-48. [张凡, 尹俊龙, 郭英琪, 等, 2018. WRKY 转录因子的研究进展 [J]. *生物技术通报*, 34(1): 40-48.]
- ZHANG M, CHEN Y, NIE L, et al., 2018. Transcriptome-wide identification and screening of WRKY factors involved in the regulation of taxol biosynthesis in *Taxus chinensis* [J]. *Sci Reports*, 8(1): 5197.
- ZHANG YJ, WANG LJ, 2005. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants [J]. *BMC Evol Biol*, 5(1): 1.