

竹叶花椒 *TPS* 基因家族全基因组分析

任妙珍^{1,2}, 董凯麟^{1,2}, 张 剑², 陈之端², 苏俊霞²
(1. 山西师范大学生命科学学院, 太原 030092;
2. 中国科学院植物研究所系统与进化植物学国家重点实验室, 北京 100093)

摘 要: 为了探究香料植物 *TPS* 基因家族的潜在功能, 本研究以香料植物竹叶花椒(*Zanthoxylum armatum* DC.) 为材料, 利用生物信息学方法对其萜类合成酶基因家族的理化性质、基因重复类型、亚细胞定位、进化关系、染色体定位与共线性、基因结构和基因表达进行了系统地分析。我们在竹叶花椒全基因组中共鉴定到 53 个 *TPS* 基因家族成员, 它们编码蛋白氨基酸长度为 173~859 aa, 分子量为 20. 21~98. 44 kDa, 等电点范围在 4. 87~9. 10, 主要定位于细胞质和叶绿体当中。进化分析显示, *ZaTPSs* 可以划分为 6 个亚家族, *TPS-a* 和 *TPS-b* 亚家族的成员最多。其中 48 个成员都包含 *TPS* 家族的保守基序 motif1。染色体定位与共线性分析显示, 53 个成员在 11 条染色体上不均匀分布, 存在 4 处串联重复, 共线性分析证明竹叶花椒与其同科物种甜橙[*Citrus sinensis* (L.) Osbeck] 的 *TPS* 基因有更近的基因进化关系。基因表达差异分析显示, *ZaTPSs* 有一定的组织表达特异性, 在幼花中表达量最高, 其次是果皮。这些研究结果表明 *TPS* 基因家族在竹叶花椒果皮挥发油的合成和防御等方面发挥了重要作用。

关键词: 竹叶花椒; 萜类合成酶; 基因家族; 生物信息学

中图分类号: Q941+. 2 **文献标识码:** A **DOI:** 10. 19907/j. 0490-6756. 2022. 046001

Genome-wide analysis of *TPS* gene family of *Zanthoxylum armatum*

REN Miao-Zhen^{1,2}, DONG Kai-Lin^{1,2}, ZHANG Jian², CHEN Zhi-Duan², SU Jun-Xia²
(1. School of Life Sciences, Shanxi Normal University, Taiyuan 030092, China;
2. State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany,
The Chinese Academy of Sciences, Beijing 100093, China)

Abstract: In order to explore the potential function of the *TPS* gene family in perfume plant, a comprehensive investigation of the *TPS* gene family was conducted in the well-known perfume plant *Zanthoxylum armatum*, and the basic physical and chemical properties, gene duplication type, subcellular localization, evolutionary relationship, chromosome localization and collinear analysis, gene structure and gene expression pattern were analyzed. A total of 53 *ZaTPS* genes were identified in the whole genome of *Z. armatum*. The amino acid length of their proteins ranges from 173 to 859 aa, and the molecular weight ranges from 20. 21 to 98. 44 kDa. The isoelectric point ranges from 4. 87 to 9. 10, and the subcellular localizations are mainly in the cytoplasm and chloroplast. Phylogenetic analysis showed that *ZaTPSs* could be divided into six subfamilies, and the members of *TPS-a* and *TPS-b* subfamilies were

收稿日期: 2021-11-19
基金项目: 国家自然科学基金(31300180, 42071067)
作者简介: 任妙珍(1997—), 女, 山西运城人, 硕士研究生, 研究方向为植物基因家族进化。E-mail: renmiaozen @ibcas. ac. cn
通讯作者: 苏俊霞。E-mail: sjx_xia@163. com

the majority ones. There are 48 *ZaTPSs* containing motif1 that are conserved in the *TPS* proteins. Chromosomal localization and collinear analysis showed that 53 *ZaTPSs* were unevenly distributed on 11 chromosomes, and there were four pairs of tandem repeats, and the collinear analysis revealed that *TPS* genes in *Z. armatum* and *Citrus sinensis* from the same family have closer evolutionary relationships. The gene differential expression analysis based on transcriptome datasets showed that *ZaTPSs* had tissue-specific expression patterns, with the highest expression level was found in young flowers, followed by husks. These results indicate that *ZaTPSs* played an important role in the synthesis of husks volatile oil and defense.

Keywords: *Zanthoxylum armatum*; Terpene synthase; Gene family; Bioinformatics

1 引言

竹叶花椒(*Zanthoxylum armatum* DC.)属于芸香科(Rutaceae)花椒属(*Zanthoxylum* L.),俗称“藤椒”,是一种落叶小乔木或灌木,枝具宽而扁锐的刺,全株有浓烈辛香味,主要分布在我国秦岭以南的广大西南区域^[1]. 由于其清香浓郁的特殊风味,竹叶花椒的干燥成熟果皮,常用作风味调节剂和防腐剂,是中国美食中不可或缺的一种调料. 它也是一种重要的药用植物,根、茎、叶、果实等部位都可入药,可以辅助治疗胃痛、牙痛、肠内蠕虫及发热、风湿等;其提取的化合物具有消炎、杀菌和抗氧化的作用^[2-4]. 竹叶花椒不仅是我国传统的香料、油料和药用树种,也是重要的生态经济树种,深受广大人民的喜爱. 其中含有丰富的挥发油,主要为小分子醇类或萜类物质^[5].

特化次生代谢物的出现,提高了植物适应环境的能力^[6]. 在植物组织中,挥发性有机物(volatile organic compounds, VOCs)的合成具有重要的生物学功能,包括抵抗病原菌,寄生生物和食草动物等. 根据其来源,VOCs 又可分为萜类、苯类芳烃和脂肪酸衍生物,其中萜类占据了大多数^[7]. 萜类是天然化合物中规模最大,结构最多样的一类次生代谢物^[8]. 萜类化合物源自甲羟戊酸(MVA pathway)途径和甲基赤藓糖磷酸酯途径(MEP pathway)产生的化合物和衍生物,分子骨架基于异戊二烯或其异构体二甲基烯丙基焦磷酸 C_5 单元,包含倍半萜、单萜、二萜、三萜等几种类型^[9]. 萜类化合物在植物生理和生化中扮演了重要的角色,例如光合作用、电子传递、发育调控等^[10]. 根据植物中萜类化合物的作用可将其分为两大类:初级代谢产物和次级代谢产物,例如赤霉素、脱落酸、类胡萝卜素、甾醇等为重要的初级代谢物,可调节细胞伸长

和植物生长,参与光合作用以及控制膜流动性等;植物中大多数萜类是次生代谢物,如青蒿素,紫杉醇,棉酚等,在植物与环境相互作用中起着关键作用,并且具有重要的药用价值. 萜类合成依赖于萜类合成酶(Terpene synthase, TPS),萜类合成酶的多样性造就了萜类的多样性,根据生成的产物不同,可将其分为单萜合成酶、倍半萜合成酶、二萜合成酶等. 根据植物 *TPS* 基因家族的系统发育关系,前人将该家族划分为 8 个亚家族:TPS-a, TPS-b, TPS-c, TPS-d(裸子植物特有), TPS-e, TPS-f, TPS-g, 和 TPS-h(卷柏属特有)^[11]. 为了适应特殊的生态位,例如吸引传粉者,散播种子,对抗病原菌,防止食草动物的啃食等,大部分植物的萜类物质会发生谱系特异进化,在各个亚家族中发生不同程度的扩张和变异^[10,12,13]. 萜类合成酶 *TPS* 基因家族在多种植物中都被鉴定,例如在拟南芥 [*Arabidopsis thaliana* (L.) Heynh.] 中鉴定到 32 条,在玉米(*Zea mays* L.)中鉴定到 30 条^[14],在水稻(*Oryza sativa* L.)中鉴定到 34 条^[11],在陆地棉(*Gossypium hirsutum* L.)中鉴定出 41 条^[15]. 然而,关于有名的香料植物竹叶花椒萜类合成酶基因家族的成员鉴定与功能研究却还没有相关报道. 近年来随着基因组测序技术的成熟,竹叶花椒基因组测序也已经完成^[16]. 因此,本研究开展了竹叶花椒 *TPS* 基因家族的鉴定与表达分析,利用生物信息学方法深入解析了竹叶花椒 *TPS* 基因家族的进化和表达模式,为进一步探索竹叶花椒 *TPS* 基因的功能提供了参考,为深入探究竹叶花椒挥发油的合成调控提供重要的目标基因.

2 材料与方法

2.1 材料

竹叶花椒的全基因组数据,相关蛋白序列和

GFF 文件从 Figshare (https://figshare.com/articles/dataset/The_chromosome-level_genome_assembly_of_Sichuan_pepper/14400884/1) 网站下载获取。拟南芥 (*A. thaliana*) 与甜橙 [*Citrus sinensis* (L.) Osbeck] 的蛋白序列分别来自于 TAIR10 (<https://www.arabidopsis.org/>) 和 Phytozome13 (https://phytozome-next.jgi.doe.gov/info/Csinensis_v1_1) 网站。

2.2 方法

2.2.1 基因家族成员的鉴定和序列分析 根据 TPS 蛋白包含保守 N 端结构域 Terpene_synth (PF01397) 和 C 端结构域 Terpene_synth_C (PF03936) 的特征, 从 Pfam34.0 网站 (<http://pfam.xfam.org/>)^[17] 下载这两个隐马尔科夫模型文件。分别用 PF01397 和 PF03936 隐马尔科夫模型文件对竹叶花椒蛋白数据进行 hmmsearch, 使用默认阈值 10.0, 可以在初始搜索时覆盖所有可能的 TPS 基因, 最终得到两个序列集, 对这两个数据集合并去冗余, 得到 hmmsearch 的数据集 (1)。为了使数据集更完善, 使用前人已鉴定的拟南芥 TPS 蛋白序列^[18], 进行 blastp 搜索竹叶花椒的蛋白库, 获得竹叶花椒同源序列数据集 (2)。将数据集 (1) 和 (2) 合并去冗余, 并在 Pfam34.0 网站 (<http://pfam.xfam.org/>)^[17] 对合并得到的数据集进行结构域核查, 获得最终的竹叶花椒 TPS 基因列表。甜橙中的 TPS 基因序列采用同样方法获得。等电点、分子量等理化性质分析用 perl 脚本计算完成。亚细胞定位使用 WoLF PSORT 网站 (<https://wolfpsort.hgc.jp/>) 进行预测。

2.2.2 系统发育分析 在 MAFFT v7 在线网站 (<https://mafft.cbrc.jp/alignment/software/>)^[19] 对竹叶花椒、甜橙、拟南芥和三条经过功能验证的胡椒木的 TPS 蛋白序列进行多序列比对, 比对好的序列在 IQ-tree 在线网站 (<http://iqtree.cibiv.univie.ac.at/>)^[20] 上进行系统发育树构建, Bootstrap 值设置为 1000。竹叶花椒 TPS 基因类别划分根据 2011 年 Chen 的系统发育分类标准^[11]。

2.2.3 基因结构及蛋白保守基序分析 使用在线网站 MEME (<http://meme-suite.org/tools/meme>)^[21] 对 ZaTPSs 保守基序进行分析, 预测值设置为 20 个; 利用竹叶花椒基因组的 GFF 文件和蛋白序列, 采用 TBTOOLS^[22] 中的 Gene Structure

View 功能对竹叶花椒 TPS 基因结构和保守基序进行可视化绘图。

2.2.4 染色体定位及基因重复类型分析 染色体定位使用以下几个文件: MCScanX 中的 duplicate_gene_classifier 程序生成的串联重复对文件, 染色体序列长度文件, GFF 文件以及竹叶花椒的 TPS 基因 ID 列表, 运行脚本生成 Mapchart 所需要的配置文件, 在 Mapchart 软件中进行绘制和美化。竹叶花椒基因重复类型分析采用 PGDD (<http://chibba.agtec.uga.edu/duplication/>) 的方法进行本地化分析。从基因组注释文件中提取 ZaTPSs 基因家族成员的染色体定位信息和相关基因对信息, 而后使用 MCScanX 鉴定 ZaTPSs 基因的重复类型。

2.2.5 全基因组共线性分析 为了揭示竹叶花椒与其近缘种甜橙及模式植物拟南芥的 TPS 家族同源基因的共线性关系, 比对三者基因组, 进行全基因组共线性分析, 使用 JCVI 画图, 竹叶花椒、甜橙和拟南芥的 TPS 基因在共线性块中的连线将会高亮显示。

2.2.6 基因表达模式分析 使用 Hisat2 软件分别将竹叶花椒的顶芽、幼叶、成熟叶片、叶柄、幼花、种子、果皮、茎和皮刺 9 个不同组织的转录组数据 (<https://www.ncbi.nlm.nih.gov/sra/?term=Zanthoxylum+armatum>), 比对到参考基因组上^[16], 使用 featurecounts R 包计算 TPM 值^[23], 分析竹叶花椒 TPS 家族基因在不同组织中的表达模式, 使用 TBTOOLS^[22] 软件中的 Heatmap 功能完成可视化。

3 结果与分析

3.1 基因鉴定及理化性质分析

在竹叶花椒基因组中鉴定出 53 个 TPS 基因。根据染色体位置信息分别命名, 见表 1。对 ZaTPS 蛋白理化性质的分析结果显示, ZaTPS 蛋白长度在 173~859 aa 之间, 分子量在 20.21~98.44 kD 之间, 等电点范围在 4.87~9.10, 其中有 52 个成员偏酸性, 只有 1 个成员偏碱性, 说明 ZaTPSs 蛋白在酸性环境下更容易沉淀。亚细胞定位结果显示 ZaTPSs 大部分分布在叶绿体 (49.06%) 和细胞质 (39.62%), 少部分位于细胞核 (9.43%) 和线粒体 (1.89%) 中。

表 1 竹叶花椒 *TPS* 基因家族成员信息

Tab. 1 Information of *TPS* gene family members from *Z. armatum*

基因编号	基因号	蛋白长度/aa	分子量/kDa	理论等电点/pI	亚细胞定位预测 ¹	重复基因类型 ²
<i>ZaTPS1</i>	Zardc00213	499	57.97	5.00	cyto	TRD
<i>ZaTPS2</i>	Zardc02158	602	69.73	5.86	cyto	WGD
<i>ZaTPS3</i>	Zardc02232	563	65.22	5.21	nucl	DSD
<i>ZaTPS4</i>	Zardc02558	626	72.17	9.10	chlo	PD
<i>ZaTPS5</i>	Zardc02562	204	23.62	6.03	cyto	PD
<i>ZaTPS6</i>	Zardc02564	607	70.12	5.86	chlo	PD
<i>ZaTPS7</i>	Zardc02591	607	70.10	5.86	chlo	DSD
<i>ZaTPS8</i>	Zardc04707	663	76.11	5.18	cyto	WGD
<i>ZaTPS9</i>	Zardc06748	551	63.81	5.23	cyto	WGD
<i>ZaTPS10</i>	Zardc16053	571	66.35	4.88	chlo	WGD
<i>ZaTPS11</i>	Zardc18199	859	98.44	5.97	cyto	DSD
<i>ZaTPS12</i>	Zardc20465	173	20.21	5.78	cyto	TRD
<i>ZaTPS13</i>	Zardc20496	524	61.03	5.17	cyto	WGD
<i>ZaTPS14</i>	Zardc20508	406	47.48	5.22	cyto	PD
<i>ZaTPS15</i>	Zardc20510	567	65.51	5.78	cyto	PD
<i>ZaTPS16</i>	Zardc21140	460	52.91	5.24	chlo	SL
<i>ZaTPS17</i>	Zardc21145	409	47.26	5.09	cyto	WGD
<i>ZaTPS18</i>	Zardc21149	558	64.59	6.05	cyto	TD
<i>ZaTPS19</i>	Zardc21150	402	46.61	5.32	chlo	TD
<i>ZaTPS20</i>	Zardc21169	546	63.02	5.10	cyto	WGD
<i>ZaTPS21</i>	Zardc21170	546	63.05	4.99	cyto	TD
<i>ZaTPS22</i>	Zardc21171	520	60.01	5.07	cyto	TD
<i>ZaTPS23</i>	Zardc21195	557	64.02	5.03	chlo	TD
<i>ZaTPS24</i>	Zardc21201	585	67.21	5.80	mito	WGD
<i>ZaTPS25</i>	Zardc22320	826	94.80	6.11	chlo	WGD
<i>ZaTPS26</i>	Zardc22924	569	66.14	6.56	chlo	TD
<i>ZaTPS27</i>	Zardc22925	481	55.74	5.40	chlo	TD
<i>ZaTPS28</i>	Zardc23195	593	69.02	6.13	cyto	WGD
<i>ZaTPS29</i>	Zardc23197	508	59.79	6.47	cyto	SL
<i>ZaTPS30</i>	Zardc23653	610	70.71	5.35	chlo	DSD
<i>ZaTPS31</i>	Zardc23655	613	70.76	5.65	chlo	DSD
<i>ZaTPS32</i>	Zardc23656	600	69.42	5.88	chlo	TRD
<i>ZaTPS33</i>	Zardc23770	611	71.10	5.07	chlo	DSD
<i>ZaTPS34</i>	Zardc23788	606	70.25	5.25	chlo	PD

(续表 1)

基因编号	基因号	蛋白长度/aa	分子量/kDa	理论等电点/pI	亚细胞定位预测 ¹	重复基因类型 ²
<i>ZaTPS35</i>	Zardc23793	610	70.61	5.35	chlo	PD
<i>ZaTPS36</i>	Zardc23795	600	69.67	6.11	chlo	WGD
<i>ZaTPS37</i>	Zardc35342	576	65.79	4.98	nucl	PD
<i>ZaTPS38</i>	Zardc35346	588	66.82	4.98	nucl	TD
<i>ZaTPS39</i>	Zardc35347	588	66.82	4.98	nucl	TD
<i>ZaTPS40</i>	Zardc35348	587	66.75	4.98	chlo	TD
<i>ZaTPS41</i>	Zardc35840	762	86.86	5.52	chlo	TRD
<i>ZaTPS42</i>	Zardc42695	561	64.32	6.05	cyto	TRD
<i>ZaTPS43</i>	Zardc42834	704	81.14	6.15	nucl	TD
<i>ZaTPS44</i>	Zardc43153	602	70.26	6.11	chlo	TRD
<i>ZaTPS45</i>	Zardc44524	611	70.72	4.95	chlo	DSD
<i>ZaTPS46</i>	Zardc44530	600	69.71	6.28	chlo	TRD
<i>ZaTPS47</i>	Zardc46571	522	60.04	6.21	chlo	DSD
<i>ZaTPS48</i>	Zardc46763	506	58.31	6.08	cyto	TRD
<i>ZaTPS49</i>	Zardc47396	596	69.63	6.56	chlo	TRD
<i>ZaTPS50</i>	Zardc48693	435	50.51	5.36	cyto	TRD
<i>ZaTPS51</i>	Zardc50072	532	61.10	4.87	chlo	TRD
<i>ZaTPS52</i>	Zardc51844	614	71.85	6.22	chlo	TRD
<i>ZaTPS53</i>	Zardc52568	377	43.88	4.88	cyto	TRD

注:1 亚细胞定位预测中,cyto 表示细胞质,chlo 表示叶绿体,nucl 表示细胞核,cyto_nucl 表示细胞质和细胞核,mito 表示线粒体;2 基因重复类型中,DSD 表示散在重复,TD 表示串联重复,TRD 表示转座重复,PD 表示近端重复,WGD 表示全基因组复制,SL 表示单拷贝。

3.2 基因家族系统发育分析

使用 32 个拟南芥、28 个甜橙和 53 个竹叶花椒 *TPS* 蛋白序列构建系统发育进化树,见图 1。三个物种的 *TPS* 基因聚为 6 个分支,分别对应 *TPS* 基因的 6 个亚家族:TPS-a,TPS-b,TPS-c,TPS-e,TPS-f 和 TPS-g,在系统树上每个亚家族的聚类关系都得到了很好的支持。在被子植物中,TPS-a 和 TPS-b 是倍半萜和单萜的主要合成酶。在竹叶花椒中,TPS-a 和 TPS-b 亚家族成员较多,分别为 20 条和 24 条,TPS-e 和 TPS-f 亚家族中成员最少,只有 1 条。与之相似,甜橙的 TPS-a 和 TPS-b 亚家族成员也较多,分别为 15 条和 9 条,TPS-c 和 TPS-e 亚家族成员较少,均为 1 条,TPS-f 亚家族则一条也没有。

3.3 基因结构及蛋白保守基序分析

利用 MEME 在线工具对 *ZaTPS* 蛋白序列进

行分析,并得到 motif 保守基序,根据含有保守基序的序列数目和其显著性,依次命名为 motif 1 至 20,见图 2。其中 motif 1(E-value 为 7.5e-945)有 *TPS* 最典型的保守结构域,“富含天冬氨酸残基的 DDxxD”,该残基被认为对金属离子起结合作用,从而在引导底物催化时发挥重要作用^[24],48 个 *ZaTPSs* 都含有 motif1。Motif 12 包含 TPS-a 和 TPS-b 的保守基序“RRX8W”。包含 Motif 20 的序列条数最少,只有 9 条,都为 TPS-b 亚家族成员。整体来看,TPS-a 和 TPS-b 亚家族包含的保守基序较多,TPS-g 亚家族保守基序较为一致,猜测 TPS-g 在竹叶花椒中比较保守。通过竹叶花椒全基因组 GFF 注释文件和 *TPS* 基因家族 CDS 序列信息,统计外显子和内含子,分析发现 53 个 *ZaTPSs* 均正常编码,但各成员外显子个数存在差异,*ZaTPS25* 外显子数目最多,达到 15 个。

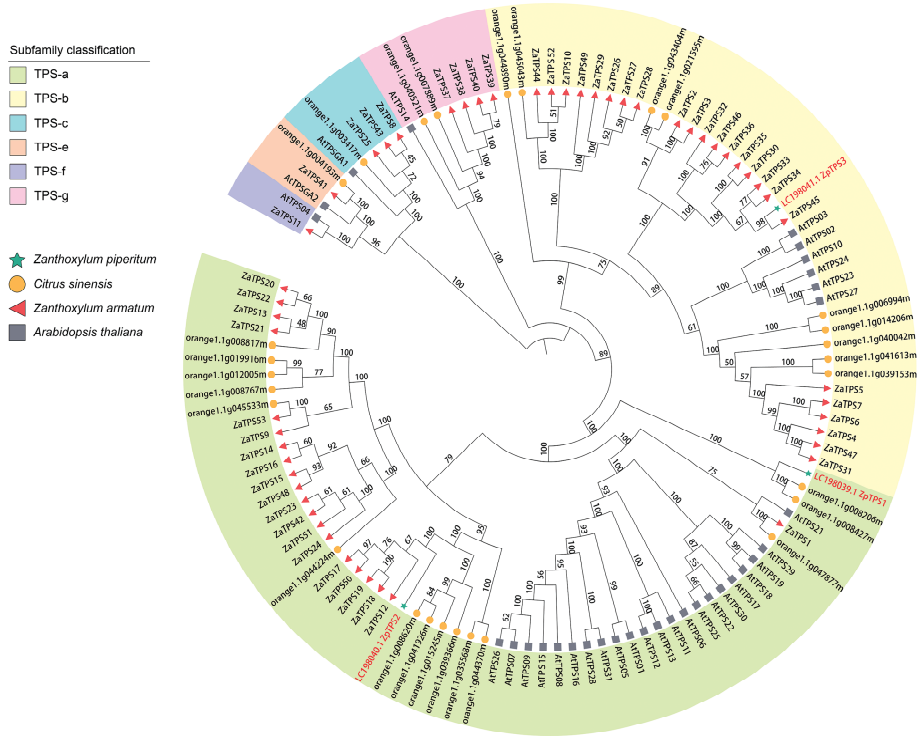


图 1 竹叶花椒、甜橙和拟南芥 TPS 蛋白的系统发育树

红色三角形表示竹叶花椒 TPS 基因,橙色圆形表示甜橙 TPS 基因,灰色矩形表示拟南芥 TPS 基因,绿色五角星表示胡椒木 TPS 基因

Fig. 1 The phylogenetic tree of TPS proteins from *Z. armatum*, *C. sinensis* and *A. thaliana*

The red triangle represents the TPS genes of *Z. armatum*, the orange circle represents the TPS genes of *C. sinensis*, the gray rectangle represents the TPS genes of *A. thaliana*, and the green pentagram represents the TPS genes of *Z. piperitum*

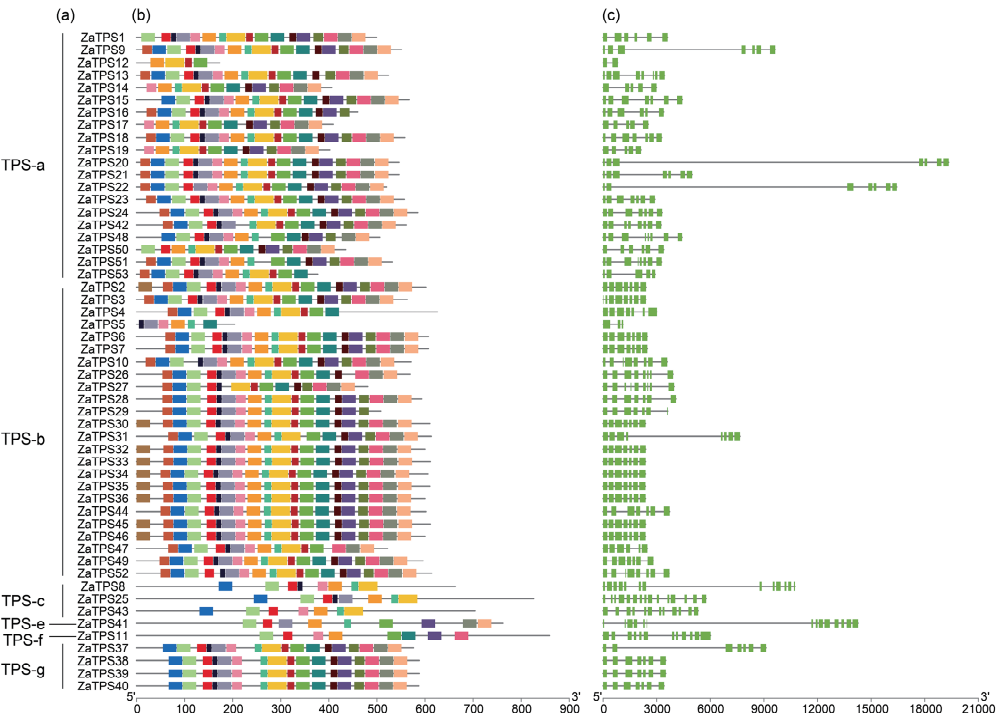


图 2 竹叶花椒 *ZaTPSs* 保守基序分布及外显子、内含子分布

(a) 竹叶花椒 53 个 TPS 基因的分类; (b) *ZaTPS* 蛋白的保守基序分布, 20 个 motif 用不同颜色矩形框表示; (c) *ZaTPS* 基因的外显子、内含子分布。绿色矩形框表示蛋白的编码区(即 CDS), 中间横断的黑色线条表示内含子区域。通过底部比例尺可推断各区域长度

Fig. 2 The conserved motif distribution and exon-intron distribution of *ZaTPSs* in *Z. armatum*

(a) The classification of 53 *ZaTPSs* in *Z. armatum*; (b) conserved motif distribution of *ZaTPS* proteins, 20 motifs are represented by colorful rectangular boxes; (c) exon and intron distribution of *ZaTPS* genes. The green rectangular boxes indicate the coding region of protein (CDS), and the black lines in the middle of CDS indicate the intron regions. The length of each region can be inferred from the bottom scale

3.4 染色体定位及基因重复类型

染色体定位分析结果显示竹叶花椒有 44 条 *ZaTPS*s 不均匀地分布在 11 条染色体上, 结果见图 3, 剩余的 9 个基因分布在未挂载到染色体上的片段序列中. 其中 14 号染色体含有的 *ZaTPS* 基因最多, 且主要为串联重复, 呈小型基因簇. 16 号, 24 号染色体上都分别发生了不同程度的串联重复. 除了标红的几个串联重复基因簇外, 还有一些基因分布距离较近, 猜测是由比较古老的串联重

复形成. *TPS* 基因分布最少的为 3 号、4 号、10 号、12 号、15 号、25 号染色体, 仅含有 1 个 *ZaTPS* 基因. 同时, 我们对 *ZaTPS* 基因重复类型的分析结果显示, 转座重复 (TRD) (占 24. 53%), 串联重复 (TD) (占 20. 75%) 和全基因组复制 (WGD) (占 20. 75%) 是竹叶花椒 *TPS* 基因家族扩张的主要的驱动因素. 与甜橙物种分化后, 竹叶花椒发生了一次特异的 WGD 事件^[16], 这与 WGD 对竹叶花椒 *TPS* 基因特异扩张的贡献相吻合.

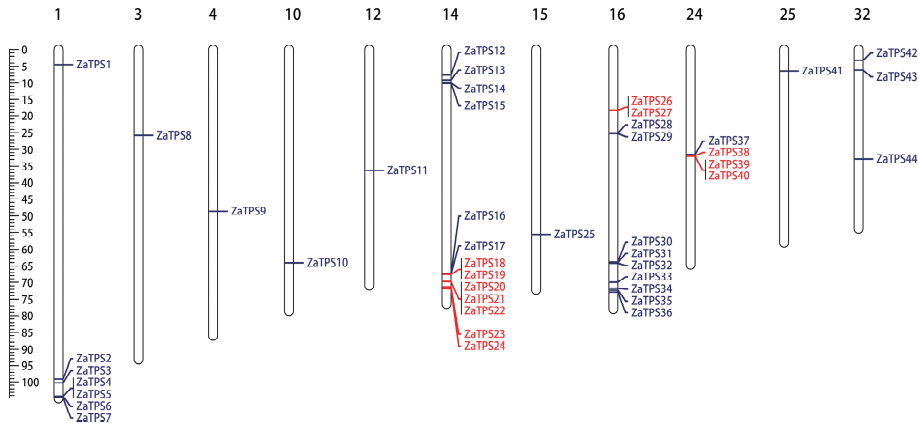


图 3 竹叶花椒 *TPS* 基因染色体定位
(红色标注表示基因间有串联重复)
Fig. 3 Chromosome localization map of *TPS* genes in *Z. armatum*
(red marks in the figure indicate tandem duplication gene pairs)

3.5 竹叶花椒、拟南芥及甜橙 *TPS* 基因全基因组共线性分析

我们对竹叶花椒、拟南芥及甜橙 *TPS* 基因进行了全基因组共线性分析, 见图 4. 竹叶花椒有 1 条染色体上的 *TPS* 基因与拟南芥的 *TPS* 基因存在共线性关系, 有 2 条染色体与甜橙的 *TPS* 基因存

在共线关系. 与拟南芥相比, 竹叶花椒与甜橙具有更近的基因进化关系, 共线基因对为 14 对, 而与拟南芥仅有 1 对. 竹叶花椒与甜橙的 *TPS* 共线性基因对和竹叶花椒与拟南芥的 *TPS* 共线性基因对中, 没有共享的 *ZaTPS* 基因.

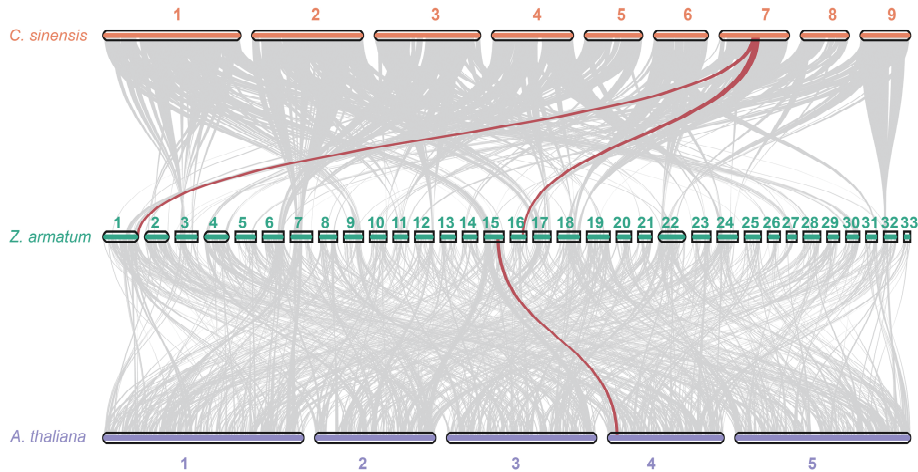


图 4 竹叶花椒与甜橙及拟南芥 *TPS* 基因共线性分析
(灰色线条代表物种之间所有存在共线性的基因对, 红色高亮线条代表竹叶花椒 *TPS* 基因与对应物种的共线性基因对)
Fig. 4 Syntenic relationship of *TPS* genes in *Z. armatum*, *C. sinensis* and *A. thaliana*
(gray lines represent all syntenic gene pairs between two species, and red lines represent syntenic gene pairs between *ZaTPS*s and *TPS*s in other species)

3.6 TPS 基因在竹叶花椒中的表达模式分析

利用竹叶花椒转录组数据分析 *ZaTPSs* 在竹叶花椒顶芽、幼叶、成熟叶片、叶柄、幼花、种子、茎、皮刺和果皮中的表达量, 见图 5。总体来看, *ZaTPSs* 在竹叶花椒的各个部位都有表达。相比于其他亚家族, TPS-a($\log_2$ TPM $\approx$ 11.86)和 TPS-b($\log_2$ TPM $\approx$ 8.42)亚家族平均表达量大于其他亚家族。从各个组织的表达量来看, *ZaTPSs* 在种子、成熟叶片和皮刺中表达量很低或不表达, 但皮刺中 *ZaTPS51* 却表达异常高($\log_2$ TPM $\approx$ 11.85), 忽略此异常值后, *ZaTPSs* 在幼花组织中平均表达量最突出, 尤其是 TPS-a($\log_2$ TPM $\approx$ 3.30), TPS-b($\log_2$ TPM $\approx$ 2.89)和 TPS-g($\log_2$ TPM $\approx$ 4.97)亚家族成员, 其中 TPS-b 亚家族的 *ZaTPS45* 和

ZaTPS36 在幼花中的表达量最多。TPS-g 亚家族成员在种子、茎和皮刺中不表达, 但在幼花中表达总量较明显。除幼花组织外, *ZaTPSs* 在果皮中总体表达量也比较高, 最突出的是 *ZaTPS23* ($\log_2$ TPM $\approx$ 8.79), 在系统发育树上与其聚为一支的 *ZaTPS24*、*ZaTPS42* 和 *ZaTPS51* 在果皮中也表达较高。一些基因在幼叶中有表达但在成熟叶中不表达, 如 *ZaTPS4*、*ZaTPS6*、*ZaTPS7* 等, 而一些基因在幼叶和成熟叶中都不表达, 如 *ZaTPS14*、*ZaTPS44*、*ZaTPS52* 等。 *ZaTPS5*、*ZaTPS10*、*ZaTPS44* 只在茎中有微量表达。

4 讨 论

TPS 基因家族广泛分布于各种生物体内, 参与多种生长发育过程, 发挥调节功能。目前, 已在多种植物中鉴定到了大量的 TPS 基因家族成员, 例如, 在花椒 (*Z. bungeanum*) 中, 鉴定得到 70 条 TPS 蛋白, TPS-a 和 TPS-b 亚家族中成员最多, 分别为 23 条和 31 条^[25], 这与我们在竹叶花椒中确定的 TPS 基因数目大体相近。最近一项基于竹叶花椒转录组数据的研究中, 调查了果实中萜类合成通路的关键基因, 筛选出 28 个表达显著上调的关键基因, 主要为以下几个基因家族: TPS-like, GGPPS-like, GCPE-like 和 DXS-like, TPS-like 中主要报道了 5 个 TPS-a 和 TPS-b 亚家族成员, 但目前并没有针对竹叶花椒 TPS 基因家族做全面系统性的研究和报道^[26,27]。鉴于竹叶花椒全基因组数据的发表, 本研究开展了竹叶花椒 TPS 基因家族的鉴定与表达分析, 为进一步了解 *ZaTPSs* 基因的功能提供了参考。

基于竹叶花椒全基因组数据, 共有 53 条 TPS 基因被鉴定, 其中 TPS-a 和 TPS-b 亚家族成员较多, 分别为 20 条和 24 条, TPS-e 和 TPS-f 中成员最少, 只有 1 条, 同科植物甜橙在 TPS-a 和 TPS-b 亚家族中成员也较多, 分别为 23 条和 16 条, 这一现象符合大多数被子植物中 TPS 家族分布的规律^[14]。通过统计竹叶花椒不同亚家族中的基因重复类型情况, 我们发现 TPS-a 亚家族成员中 35% 来自于转座重复, 其次是串联重复(25%)和全基因组复制(25%), 并且该亚家族中转座重复来源的基因表达平均值最高。在 TPS-b 亚家族中, 分散重复来源的基因占据了 29.17%, 其次是近端重复(20.83%)和转座重复(20.83%), 然而在该亚家族中全基因组复制来源的基因表达平均值最高, 其次

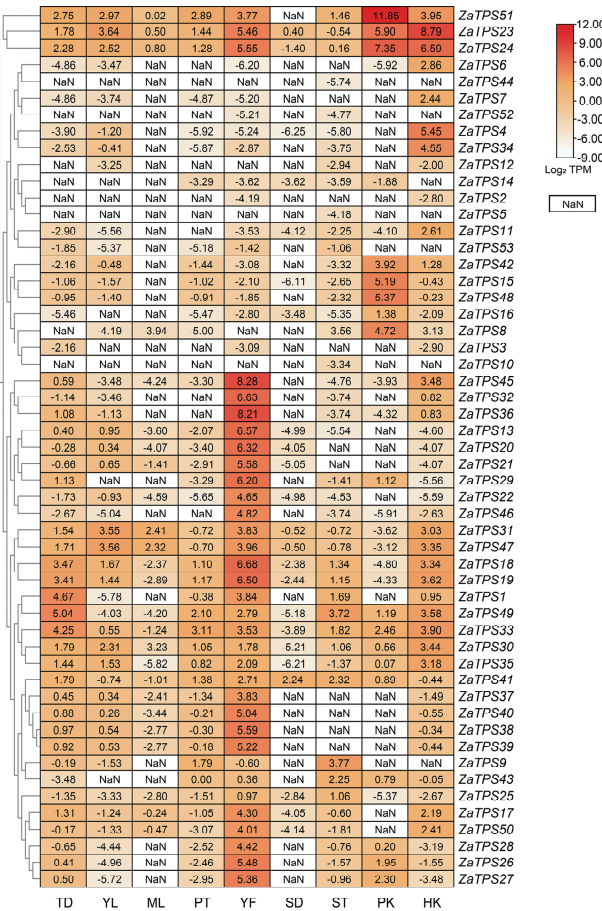


图 5 竹叶花椒 TPS 基因在不同组织中的表达热图(TD, 顶芽; YL, 幼叶; ML, 成熟叶; PT, 叶柄; YF, 幼花; SD, 种子; ST, 茎; PK, 皮刺; HK, 果皮), 白色表示没有表达, 颜色越深表示表达值越高

Fig. 5 Expression heatmap of the TPS genes in different tissues of *Z. armatum* (TD, terminal bud; YL, young leaf; ML, mature leaf; PT, petiole; YF, young flower; SD, seed; ST, stem; PK, prickle; HK, husk). The white rectangles represent no expression, and darker color represents higher expression.

是散在重复。总体来说,转座重复(24.53%),串联重复(20.75%)和全基因组复制(20.75%)是竹叶花椒 TPS 基因家族扩张的主要来源。转座重复是一种通过 DNA 或 RNA 复制机制,将基因或基因片段从原来的位置转移到一个新的染色体位置的基因重复类型^[28]。DNA 转座子,例如高粱(sorghum)的 CACTA 转座元件,能够将基因或基因片段移动到一个新的染色体位置,而基于 RNA 的转座重复通常是指反转录转座,能够从信使 RNA 中的多个外显子中产生单外显子进行反转录,反转录产生的基因将会转座到一个新的染色体^[29]。结合竹叶花椒基因组中含有大量转座元件的结构特征,猜测 TPS 基因家族扩张受到该结构特征的影响。竹叶花椒与甜橙物种分化后,在大约 26.6 个百万年前发生过一次物种特异的全基因组复制事件,TPS 基因家族的扩张也受到了该事件的促进^[16]。

萜类生成的前体是异戊烯焦磷酸(IPP)与其异构体二甲基烯丙基二磷酸(DMAPP),二者分别由发生在细胞质中的甲羟戊酸途径(MVA pathway)和质体中的甲基赤藓糖磷酸酯途径(MEP pathway)生成^[9]。通过二者的变换结合,在萜类合成酶的作用下,生成一系列中间化合物:牻牛儿基焦磷酸(GPP),橙花基焦磷酸(NPP),法尼基焦磷酸(FPP)和牻牛儿牻牛儿基焦磷酸(GGPP)等,最终生成单萜、倍半萜、二萜等^[30,31]。根据 TPS 基因家族的亚细胞定位分布模式,倍半萜合酶一般定位于细胞质中,单萜合酶都位于质体中^[11,32]。亚细胞定位结果显示 *ZaTPSs* 大部分定位在细胞质和叶绿体,且 TPS-a 亚家族蛋白大部分分布于细胞质中,TPS-b 亚家族蛋白大部分分布于叶绿体中。众所周知,TPS-a 亚家族基因成员编码倍半萜合酶,TPS-b 亚家族基因成员编码单萜合酶。这与大部分 TPS-a 分布在细胞质中,大部分 TPS-b 分布于叶绿体中的事实相契合。有研究表明,代谢物质合成的基因在染色体的基因排布上并不是随意分散,而是彼此相邻,形成基因簇。萜类合成酶基因形成基因簇有助于基因组合的共遗传和共表达,减少单一通路基因在重组后丢失的可能性^[33]。染色体定位分析显示,16 号染色体和 24 号染色体上均有串联重复,且 16 和 24 号染色体上分布的全为 TPS-b 亚家族的成员,猜测这加强了基因的共表达。通过观察竹叶花椒 TPS 基因在不同组织中的表达分析发现,幼花中基因表达最为明显,主要集中于 TPS-a,TPS-b 和 TPS-g 三个亚家族。在拟南芥

TPS 基因功能研究中发现,花器官中检测到的萜类化合物数量最多,与两个功能相关:吸引短距离内的小型昆虫进行传粉,另外可能与防御机制有关^[34]。玉米中,已有实验证明 *TPS23* 能够调控石竹烯的释放,来吸引食草动物的天敌,间接达到驱赶食草动物的目的^[35]。此外,在棉花中也发现能产生一系列具有杀虫特性的萜类化合物,例如众所周知的棉酚,半棉酚,杀食夜蛾素等^[36]。这说明萜类作为一种具有重要生物学意义的挥发性有机化合物及其合成相关的 TPS 基因家族,很有可能在竹叶花椒吸引传粉者、防止食草动物的侵食、防御病原菌等方面起到重要作用,从而来适应特殊的生态位。单萜和倍半萜为小分子萜类,多具有挥发性,在植物体内常以挥发油的形式存在,*ZaTPSs* 基因在果皮中也呈现较高的表达量,猜测与果皮合成、释放挥发性物质有关。竹叶花椒同属植物胡椒木中已有研究验证,*ZpTPS1* 和 *ZpTPS2* 能够催化 FPP 生成倍半萜(C15),*ZpTPS3* 催化 GPP 生成单萜(C10)。将这三条 *ZpTPSs* 蛋白序列加入系统发育分析后发现,*ZaTPS12*、*ZaTPS17-19* 和 *ZaTPS50* 与 *ZpTPS2* 聚为一支,得到了较高的支持率,猜测这几个 *ZaTPSs* 基因也发挥与 *ZpTPS2* 同样的功能。*ZpTPS3* 在植物体内催化形成单萜化合物,在系统进化树上与 *ZaTPS2*、*ZaTPS3*、*ZaTPS30*、*ZaTPS32-36* 和 *ZaTPS45*、*ZaTPS46* 聚为一支,这些基因在幼花和果皮中都有表达,但在幼叶和成熟叶中表达有所差别,在成熟叶中不表达,推测这些基因在幼花和果实中功能较为保守,参与幼花和果实单萜类化合物的合成,在叶中不会持续表达。*ZaTPS23* 在果皮中表达量最高,在系统发育树上与其聚为一支的 *ZaTPS24*、*ZaTPS42* 和 *ZaTPS51* 在果皮中也表达较高,推测这些成员都参与了果皮中挥发油的合成,后续的功能验证实验可通过基因敲除或过表达实验来观察果皮中挥发油含量的变化^[37]。

综上所述,基于全基因组数据,本研究全面地鉴定了竹叶花椒中 TPS 基因家族成员,并对其分类、进化和表达模式进行了深入分析,为后续功能验证实验提供了重要的理论依据,为探究竹叶花椒的生态适应性提供了备选基因,同时为竹叶花椒挥发油应用的研究提供了理论指导。

参考文献:

- [1] Zhang D, Thomas G H, David J M. RUTACEAE.

- Flora of China [M]. Beijing & St. Louis: Science Press & Missouri Botanical Garden Press, 2008: 51.
- [2] Bhatt V, Kumar N, Sharma U, *et al.* Comprehensive metabolic profiling of *Zanthoxylum armatum* and *Zanthoxylum acanthopodium* leaves, bark, flowers and fruits using ultra high performance liquid chromatography [J]. Sep Sci Plus, 2018, 1: 311.
- [3] Byun-mckay A, Godard K A, Toudefallah M, *et al.* Wound-induced terpene synthase gene expression in sitka spruce that exhibit resistance or susceptibility to attack by the white pine weevil [J]. Plant Physiol, 2006, 140: 1009.
- [4] 哈立洋, 杨斌, 尹可欢, 等. 竹叶花椒挥发油提取工艺优化及镇痛抗炎活性研究[J]. 中药药理与临床, 2021, 37: 127.
- [5] 杨森, 何晗, 易达, 等. GC/MS/SIM 法测定比较川产竹叶花椒挥发油中 12 种挥发性成分[J]. 食品与发酵工业, 2021, 1: 10.
- [6] Waters E. Molecular adaptation and the origin of land plants [J]. Mol Phylogenet Evol, 2003, 29: 456.
- [7] Holopainen J K, Gershenzon J. Multiple stress factors and the emission of plant VOCs [J]. Trends Plant Sci, 2010, 15: 176.
- [8] Tholl D. Biosynthesis and biological functions of terpenoids in plants [J]. Adv Biochem Eng Biotechnol, 2015, 148: 63.
- [9] Mcgarvey D J, Croteau R. Terpenoid metabolism [J]. The Plant Cell, 1995, 7: 1015.
- [10] Pichersky E, Raguso R A. Why do plants produce so many terpenoid compounds? [J]. New Phytol, 2018, 220: 692.
- [11] Chen F, Tholl D, Bohlmann J, *et al.* The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom; terpene synthase family [J]. Plant J, 2011, 66: 212.
- [12] Gershenzon J, Dudareva N. The function of terpene natural products in the natural world [J]. Nat Chem Biol, 2007, 3: 408.
- [13] Irmisch S, Jiang Y F, Chen F, *et al.* Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar (*Populus trichocarpa*) [J]. BMC Plant Biol, 2014, 14: 270.
- [14] Jiang S Y, Jin J J, Sarojam R, *et al.* A comprehensive survey on the terpene synthase gene family provides new insight into its evolutionary patterns [J]. Genome Biol Evol, 2019, 11: 2078.
- [15] Huang X Z, Xiao Y T, Köllner T G, *et al.* The terpene synthase gene family in *Gossypium hirsutum* harbors a linalool synthase GhTPS12 implicated in direct defence responses against herbivores [J]. Plant Cell Environ, 2018, 41: 261.
- [16] Wang M C, Tong S F, Ma T, *et al.* Chromosome - level genome assembly of sichuan pepper provides insights into apomixis, drought tolerance, and alkaloid biosynthesis [J]. Mol Ecol Resour, 2021, 21: 2533.
- [17] Mistry J, Chuguransky S, Williams L, *et al.* Pfam: the protein families database in 2021 [J]. Nucl Acid Res, 2021, 49: D412.
- [18] Aubourg S, Lecharny A, Bohlmann J. Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana* [J]. Mol Genet Genomics, 2002, 267: 730.
- [19] Katoh K, Rozewicki J, Yamada K D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization [J]. Brief Bioinform, 2019, 20: 1160.
- [20] Trifinopoulos J, Nguyen L-T, Von haeseler A, *et al.* W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis [J]. Nucl Acid Res, 2016, 44: W232.
- [21] Bailey T L, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymer [J]. Proc Int Conf Intell Syst Mol Biol, 1994, 2: 28.
- [22] Chen C J, Chen H, Zhang Y, *et al.* TBtools: an integrative toolkit developed for interactive analyses of big biological data [J]. Mol Plant, 2020, 13: 1194.
- [23] Liao Y, Smyth G K, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features [J]. Bioinformatics, 2014, 30: 923.
- [24] Degenhardt J, Köllner T G, Gershenzon J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants [J]. Phytochemistry, 2009, 70: 1621.
- [25] Feng S J, Liu Z S, Cheng J, *et al.* *Zanthoxylum*-specific whole genome duplication and recent activity of transposable elements in the highly repetitive paleotetraploid *Z. bungeanum* genome [J]. Hortic Res, 2021, 8: 205.

[26] Hui W K, Zhao F Y, Wang J Y, *et al.* De novo transcriptome assembly for the five major organs of *Zanthoxylum armatum* and the identification of genes involved in terpenoid compound and fatty acid metabolism [J]. BMC Genomics, 2020, 21: 81.

[27] Hui W K, Wang J Y, Ma L X, *et al.* Identification of key genes in the biosynthesis pathways related to terpenoids, alkaloids and flavonoids in fruits of *Zanthoxylum armatum* [J]. Sci Hortic, 2021, 290: 110523.

[28] Wang Y P, Wang X Y, Paterson A H. Genome and gene duplications and gene expression divergence: a view from plants: gene duplication and expression divergence [J]. Ann N Y Acad Sci, 2012, 1256: 1.

[29] Kaessmann H, Vinckenbosch N, Long M. RNA-based gene duplication: mechanistic and evolutionary insights [J]. Nat Rev Genet, 2009, 10: 19.

[30] Li M Y, Feng K, Hou X L, *et al.* The genome sequence of celery (*Apium graveolens* L.), an important leaf vegetable crop rich in apigenin in the Apiaceae family [J]. Hortic Res, 2020, 7: 9.

[31] Sacchettini J C, Poulter C D. Creating isoprenoid diversity [J]. Science, 1997, 277: 1788.

[32] Gao F Z, Liu B F, Li M, *et al.* Identification and characterization of terpene synthase genes accounting for volatile terpene emissions in flowers of *Freesia x Hybrida* [J]. J Exp Bot, 2018, 69: 4249.

[33] Nützmann H W, Huang A C, Osbourn A. Plant metabolic clusters-from genetics to genomics [J]. New Phytol, 2016, 211: 771.

[34] Tholl D, Lee S. Terpene specialized metabolism in *Arabidopsis thaliana* [J]. Arabidopsis Book, 2011, 9: e0143.

[35] Capra E, Colombi C, De poli P, *et al.* Protein profiling and tps23 induction in different maize lines in response to methyl jasmonate treatment and *Diabrotica virgifera* infestation [J]. J Plant Physiol, 2015, 175: 68.

[36] Hagenbucher S, Olson D M, Ruberson J R, *et al.* Resistance mechanisms against arthropod herbivores in cotton and their interactions with natural enemies [J]. Crit Rev Plant Sci, 2013, 32: 458.

[37] Fujita Y, Koeduka T, Aida M, *et al.* Biosynthesis of volatile terpenes that accumulate in the secretory cavities of young leaves of Japanese Pepper (*Zanthoxylum piperitum*): isolation and functional characterization of monoterpene and sesquiterpene synthase genes [J]. Plant Biotechnol, 2017, 34: 17.

引用本文格式:

中 文: 任妙珍, 董凯麟, 张剑, 等. 竹叶花椒 *TPS* 基因家族全基因组分析[J]. 四川大学学报: 自然科学版, 2022, 59: 046001.

英 文: Ren M Z, Dong K L, Zhang J, *et al.* Genome-wide analysis of *TPS* gene family in *Zanthoxylum armatum* [J]. J Sichuan Univ: Nat Sci Ed, 2022, 59: 046001.