

## 大麻仁基原植物**bZIP**基因家族鉴定及其调控油脂代谢的功能初探

怀浩<sup>1,2</sup>, 宁康<sup>2</sup>, 侯聪<sup>2</sup>, 杨树明<sup>3</sup>, 汪鋈植<sup>1</sup>, 陈士林<sup>2</sup>, 董林林<sup>2\*</sup>

(1. 三峡大学生物与制药学院, 湖北宜昌 443002; 2. 中国中医科学院中药研究所, 北京 100700;  
3. 云南大麻产业投资有限公司, 云南昆明 650217)

**摘要:** *bZIP* (basic leucine zipper) 基因家族是真核生物中最大的转录因子家族之一, 其成员在逆境响应、次级代谢、植物生长、种子发育等方面发挥着重要作用。为探究大麻仁基原植物大麻 (*Cannabis sativa* L.) *bZIP* (*CsbZIP*) 基因的生物学功能, 本研究基于大麻全基因组和转录组数据, 利用生物信息学方法对 *CsbZIP* 基因家族进行系统性研究。结果表明, 在大麻中鉴定到 55 个 *CsbZIP* 基因家族成员 (*CsbZIP1*~*CsbZIP55*), 分布在 10 条染色体上, 属于 12 个亚家族, 同一亚家族成员之间的基因结构和蛋白基序分布相似。片段重复是 *CsbZIP* 基因家族扩张的主要因素。顺式作用元件分析表明 73 个油脂合成基因的启动子区含有 G-box 或 A-box 元件, qRT-PCR 实验表明 7 个 *CsbZIP* 基因和 7 个油脂合成基因在大麻仁中的相对表达量较高。相关性分析表明 7 个 *CsbZIP* 基因和 7 个油脂合成基因之间存在显著的正相关关系。本研究揭示了 *CsbZIP* 基因的结构特征、进化方式和表达模式, 为进一步研究 *CsbZIP* 基因对大麻仁油脂代谢的调控提供了重要线索。

**关键词:** 大麻仁; 药用大麻; *bZIP* 基因家族; 油脂代谢; 表达模式

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2022)08-2528-15

## The identification of *bZIP* gene family in *Cannabis sativa* L. and its preliminary research of the function in regulation of lipid metabolism

HUAI Hao<sup>1,2</sup>, NING Kang<sup>2</sup>, HOU Cong<sup>2</sup>, YANG Shu-ming<sup>3</sup>, WANG Jun-zhi<sup>1</sup>,  
CHEN Shi-lin<sup>2</sup>, DONG Lin-lin<sup>2\*</sup>

(1. College of Biological & Pharmaceutical Sciences, China Three Gorges University, Yichang 443002, China;  
2. Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China;  
3. Yunnan Hemp Industrial Investment Co. Ltd., Kunming 650217, China)

**Abstract:** The *bZIP* (basic leucine zipper) gene family is one of the largest transcription factor families in eukaryotes, and its members play important roles in stress response, secondary metabolism, plant growth, seed development and other aspects. To investigate the biological functions of the *bZIP* (*CsbZIP*) gene in *Cannabis sativa* L., we systematically investigated the *CsbZIP* gene family using bioinformatics methods based on the whole-genome and transcriptome data. The results showed that 55 *CsbZIP* gene family members (*CsbZIP1*-*CsbZIP55*) were identified and distributed on 10 chromosomes, belonging to 12 subfamilies. The gene structure and protein motif distribution of the same subfamily members were similar. Segment repeats were the main reasons for the expansion of *CsbZIP* gene family. *Cis*-elements analysis showed that the promoter regions of 73 lipid synthesis genes contained G-box or A-box element. qRT-PCR showed that the relative expression levels of 7 *CsbZIP* genes and 7 lipid synthesis genes were relatively high in hemp seed. 7 *CsbZIP* genes had a significant positive correlation

收稿日期: 2022-04-13; 修回日期: 2022-05-05.

基金项目: 海南省院士创新平台科研项目 (SQ2021PTZ0052); 国家重点研发计划资助: 中药多组学方法创新及新品种选育研究 (2019YFC1711100); 高品质工业大麻品种培育及开发研究 (H2021038).

\*通讯作者 E-mail: lldong@icmm.ac.cn

DOI: 10.16438/j.0513-4870.2022-0440

with 7 lipid synthesis genes. This study revealed the structural features, evolutionary patterns and expression patterns of *CsbZIP*, providing important clues for further study on the regulation of *CsbZIP* on oil metabolism of hemp seed.

**Key words:** hemp seed; non-psychoactive medicinal cannabis; *bZIP* gene family; lipid metabolism; expression profile

亮氨酸拉链 (basic leucine zipper, bZIP) 是广泛分布于真核生物中的一类可编码转录因子的超基因家族<sup>[1]</sup>。植物 bZIP 转录因子含有两个功能域: 一个基本区域和一个相邻的亮氨酸拉链, 基本区域在真核生物中非常保守, 是 DNA 结合域, 一般由 18 个氨基酸残基组成, 具有相同的氨基酸序列模型 (N-X7-R/K-X9)<sup>[2]</sup>。亮氨酸拉链区由重复氨基酸序列 L-X6-L-X6-L 组成, 该区域变化较大, 亮氨酸位于七肽序列的第 7 个氨基酸位置, 可以被异亮氨酸、缬氨酸、苯丙氨酸或蛋氨酸取代<sup>[3,4]</sup>。植物 bZIP 转录因子偏好 ACGT 核心序列, 如 A-box (TACGTA)、C-box (GACGTC)、G-box (CACGTG)<sup>[5]</sup>。bZIP 蛋白一般通过亮氨酸拉链形成二聚体来发挥作用, 依赖基本区域与 DNA 结合, 从而实现对基因的灵活调控<sup>[2]</sup>。

bZIP 转录因子广泛参与植物的生长发育和胁迫响应等多种生命活动<sup>[6]</sup>。例如, 种子萌发<sup>[7]</sup>、花器官的发育<sup>[8]</sup>、胁迫应答<sup>[9]</sup>、次生代谢<sup>[10]</sup>。在拟南芥中, *AtbZIP29* 可以调控根尖分生组织细胞中细胞壁形成有关基因的表达, 从而影响根的生长发育<sup>[11]</sup>。在番茄中过表达 *SlAREB* 基因可以提高植株对水分和盐胁迫的耐受性<sup>[12]</sup>。bZIP 转录因子可调控植物的次生代谢, 研究表明过表达 *SmbZIP1* 基因可抑制丹参酮的积累, 而利用基因编辑技术敲除 *SmbZIP1* 基因可显著提高丹参酮含量<sup>[13]</sup>。bZIP 转录因子在植物生长发育的调控、胁迫应答、次生代谢和信号介导中具有重要作用。

大麻 (*Cannabis sativa* L.) 属于大麻科 (Cannabaceae) 大麻属 (*Cannabis*), 一年生草本植物, 起源于亚欧大陆, 有着悠久的种植历史, 在世界各地广泛种植<sup>[14]</sup>。宁康等<sup>[15]</sup>将四氢大麻酚 (THC) 质量分数小于 0.3% 且大麻二酚 (CBD) 含量高 (原则上应大于 2.0%) 的大麻定义为非精神活性药用大麻。火麻仁, 即大麻的种子, 是一种良好的药食同源材料, 《神农本草经》对火麻仁的描述为“麻子, 味甘, 性平。主补中益气, 肥健不老”, 火麻仁归脾、胃、大肠经, 具有润肠通便、缓解月经紊乱、通淋活血等功效<sup>[16,17]</sup>, 此外火麻仁还被开发成火麻油、饮料和高蛋白营养品等<sup>[18]</sup>。火麻仁富含油脂、蛋白质、多种微量元素以及黄酮和酚类物质, 其中油脂含量占 27.04%~37.67%, 而油脂中含量最多的是不饱

和脂肪酸, 占比高达 77.55%~81.21%<sup>[17]</sup>。油脂是火麻仁中的一类重要的物质, 其中富含的不饱和脂肪酸具有重要的营养价值和药理作用<sup>[19]</sup>。现代临床研究表明, 火麻仁富含的不饱和脂肪酸具有降血糖血脂、抗炎镇痛、改善肠道微环境、减轻肠道炎症反应等功效<sup>[20]</sup>。然而, 目前关于火麻仁油脂合成的分子机制尚不清楚, 因此, 探究火麻仁油脂合成调控机制的分子基础, 挖掘油脂合成关键基因和转录调控因子, 对于培育高油脂含量的火麻仁基原植物具有重要意义。

*bZIP* 基因家族的鉴定和系统性研究在许多植物中均有报道, 例如拟南芥<sup>[4]</sup>、水稻<sup>[21]</sup>、葡萄<sup>[22]</sup>、玉米<sup>[23]</sup>、油菜<sup>[24]</sup>、高粱<sup>[25]</sup>和大豆<sup>[26]</sup>等。bZIP 转录因子在调控油脂合成过程中有着重要的作用<sup>[2]</sup>。在拟南芥中, *AtbZIP67* 转录因子能够结合 *FAD3* 基因启动子区的 G-box 作用元件, 增强 *FAD3* 基因的表达, 从而提高了拟南芥种子中  $\omega$ -3 多不饱和脂肪酸的含量<sup>[27]</sup>。*GmbZIP123* 基因可参与大豆种子脂质积累的调控, 研究发现, 在转基因拟南芥中过表达 *GmbZIP123* 可提高拟南芥种子的脂质含量<sup>[28]</sup>。bZIP 基因在大麻中的研究还未曾报道过, 且 *CsbZIP* 基因调控火麻仁油脂代谢的相关研究亦未见报道。

鉴于 *bZIP* 基因家族在调控油脂代谢中的重要作用, 以及该基因家族在大麻中尚未鉴定且功能缺少系统性研究。因此, 本研究基于大麻全基因组数据, 对大麻 *bZIP* 基因家族进行了系统性研究, 包括家族成员鉴定、序列特征分析、系统发育分析、基因结构及保守基序 (motif) 分析、共线性关系分析、蛋白互作网络分析及表达模式分析。此外, 基于转录组数据, 构建了火麻仁油脂生物合成通路, 分析了油脂合成基因的表达水平, 并对 *CsbZIP* 基因和油脂合成基因进行了相关性分析, 为进一步研究 *CsbZIP* 基因在油脂代谢中的作用提供基础。

## 材料与方法

**材料** 本研究所用的大麻品种为非精神活性药用大麻, 即 THC 质量分数小于 0.3% 且 CBD 含量高的大麻品种<sup>[15]</sup>。大麻全基因组数据和全基因组注释文件均来自 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/>), 大麻 AA126 品种植株的根、茎、叶、雌花、雄花和种子及 5

个不同的发育时期 (5 个时期依据种子的发育状态来定义, S1: 第一阶段, 顶端分生组织, 雌花未出现; S2: 第二阶段, 此时雌花已出现, 柱头呈现白色; S3: 第三阶段, 此时授粉完成, 柱头呈现橙色; S4: 第四阶段, 此时种子呈现绿色尚未成熟; S5: 第五阶段, 此时种子成熟且呈现出棕褐色) 的 RNA-seq 数据由本课题组提供<sup>[16]</sup>。拟南芥、葡萄、水稻及玉米的全基因组、全基因组注释文件及蛋白质文件均来自 Ensembl Plants 数据库 (<https://plants.ensembl.org/index.html>)。

**CsbZIP 基因的鉴定及蛋白序列特征分析** 使用 TBtools 软件来提取拟南芥 bZIP 蛋白序列用作查询, 对大麻蛋白质数据库进行 BLASTP 搜索 (value 为  $1 \times 10^{-5}$ )。从 PFAM 数据库 (<http://pfam.xfam.org/>) 中下载 bZIP 蛋白的 Hidden Markov 模型 (PF00170), 基于大麻蛋白质数据库, 通过 HMMER 3.0 软件搜索具有该结构域的蛋白 (value 为  $1 \times 10^{-5}$ )。合并去重后将剩余序列提交到 NCBI-CDD (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) 和 PFAM (<http://pfam.xfam.org/>) 数据库中来验证 bZIP 蛋白的保守域。依据它们的染色体位置命名, 使用 ExPASy (<https://web.expasy.org/prot-param/>) 分析蛋白质的理化性质, 利用 CELLO v2.5 (<http://cello.life.nctu.edu.tw/>) 进行亚细胞定位分析。

**CsbZIP 基因的系统发育分析** 采用 ClustalW 工具对已鉴定到的 55 条 CsbZIP 蛋白序列和 77 条拟南芥 bZIP 蛋白序列 (AtbZIP53 由于序列差异太大被排除) 进行多序列比对<sup>[4]</sup>, 然后利用 MGEA7.0 软件构建系统发育树, 采用 NJ 邻接法, bootstrap 值设置为 1 000。CsbZIP 蛋白的分类方法参考拟南芥之前的研究<sup>[4]</sup>。

**CsbZIP 基因的结构及保守基序分析** 基于大麻全基因组注释信息, 使用 Gene Structure Display Server (GSDS2.0) (<http://gsds.gao-lab.org/index.php>) 来展示基因的结构。使用 MEME 在线工具 (<http://meme-suite.org/>)

分析 CsbZIP 蛋白的保守基序。最大 motif 数设置为 10。通过 TBtools 软件可视化, 利用 NCBI-CDD 数据库分析 motif 的注释信息。

**CsbZIP 基因的染色体定位、基因复制事件和共线性分析** 根据 CsbZIP 基因在大麻基因组中的位置信息, 用 MapChart 软件绘制 CsbZIP 基因在染色体上的位置。利用多重共线性扫描工具 (multiple collinearity scan toolkit, McScanx) 分析基因复制事件, 使用 TBtools 软件自带的 Dual Systemy Plotter 工具分析大麻和其他 4 个物种 (拟南芥、葡萄、水稻、玉米) 的 bZIP 基因之间的共线性关系。

**CsbZIP 基因的互作网络分析** 基于拟南芥和大麻之间的 bZIP 基因的同源关系, 使用 STRING 数据库 (<https://string-db.org/>) 研究 CsbZIP 基因的互作网络。利用 Cytoscape 软件来展示预测的互作网络。

**基因的表达分析** 利用本课题组前期自测的转录组数据来分析 CsbZIP 基因和油脂合成基因的表达模式, 基于 FPKM 值, 利用在线云平台工具 (<https://www.omicstudio.cn/index>) 绘制热图, 进行可视化分析。

**油脂合成基因的启动子区顺式作用元件分析** 利用 TBtools 工具提取 86 个油脂合成基因的启动子区序列 (上游 2 kb), 然后提交到 PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 数据库来预测顺式作用元件。利用 TBtools 工具进行可视化。

**植物 RNA 的提取和 qRT-PCR** 采用快速 RNA 提取试剂盒 (北京华越洋生物科技有限公司), 按照生产厂家的说明书来提取总 RNA。使用 KR118 反转录试剂盒 (天根生化科技 (北京) 有限公司) 通过一步法合成 cDNA。采用 Primer 5.0 软件设计引物 (表 1), 以 *Actin* 基因作为内参基因, 引物合成由擎科生物科技有限公司完成。使用 StarLighter SYBR Green qPCR Mix 试剂盒 (北京启衡星生物科技有限公司) 进行 qRT-PCR 实验。

**Table 1** Details of primers for qRT-PCR in this study

Gene	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
<i>CsbZIP12</i>	GTTGGCTGATTGGCACA	GCTTCCCTTTGCTACTCCTAA
<i>CsbZIP13</i>	GAGAAACGGGGCTTACACG	GGAACAGCCAACCCCATAG
<i>CsbZIP28</i>	CTCCCTGATTCTGCTGCCA	TGGGTCCTTCTTGCTCCG
<i>CsbZIP41</i>	CATTTCTGCTGCTGCGGT	TCAACATTGTCCAACGCC
<i>CsbZIP42</i>	TCAAGTCACTGGAGAGGGC	TCTAAATCGCTCAAGTATGCC
<i>CsbZIP43</i>	GAGTCCTGCGGTCCAATC	CGAGTCTCCACCACCGAA
<i>CsbZIP44</i>	GTTGGGTTTCGGTGGTG	CCATCAACGCCAAGACCA
<i>CsHAD-1</i>	CTCACTCAACCGACCCACAA	ACAGAAACGGAAACCGATGG
<i>CsENR-1</i>	AGTGGACAGTTCAGGAAGTTG	CGTGTTGACTCTAACTCGGTG
<i>CsLACS-4</i>	TTTGGTGGTGCTCGTATTG	AAGAGAGCGAAGAAGCCG
<i>CsGPAT-3</i>	AAGAAAGGCGAATACACGAA	CACTGACACTCCAACCCCG
<i>CsDGAT-1</i>	TCGTCATCATTCCCATCG	GCCAGTAAGGTCAGCAAGT
<i>CsPDAT-3</i>	TGCTCGGTCTTCAAACGC	CAAACCTCACCCTACACG
<i>CsPDAT-4</i>	CCAAAAGTTGATGGCACGA	TAAACCGAGTCTTCCCCC
<i>CsActin</i>	TGATGAGTCAGGTCCATCCA	GCCTCTCTCAAAGCAGCAG

反应程序为: 95 °C酶激活 5 min, 95 °C变性 30 s, 60 °C退火 30 s, 72 °C延伸 1 s, 循环次数为 40。基因的相对表达量采用 $2^{-\Delta\Delta Ct}$ 方法计算。

**相关性分析** 基于**CsbZIP**基因和油脂合成基因之间的表达水平, 利用R语言进行Spearman相关性分析并可视化。

## 结果与分析

### 1 **CsbZIP**基因的鉴定及理化性质分析

通过HMMER和BLASTP比对方法分别鉴定到了92个和129个**CsbZIP**基因, 合并去除重复后使用NCBI-CDD和Pfam数据库来验证保守域, 最终鉴定到55个**CsbZIP**基因, 依据它们在染色体上的位置进行命名(**CsbZIP1**~**CsbZIP55**)。蛋白序列特征分析表明, 55个**CsbZIP**基因的编码蛋白的长度范围在133~745个氨基酸之间, 平均长度为348个氨基酸。蛋白质分子质量介于15.41~80.39 kDa, 平均蛋白质分子质量为38.38 kDa; 理论等电点介于4.62~10.26, 平均等电点为7.25, GRAVY均为负值, 表明这些**CsbZIP**蛋白均为亲水性蛋白。亚细胞定位预测结果表明只有**CsbZIP41**蛋白定位于叶绿体中, 其他的**CsbZIP**蛋白均定位于细

胞核内(表2)。

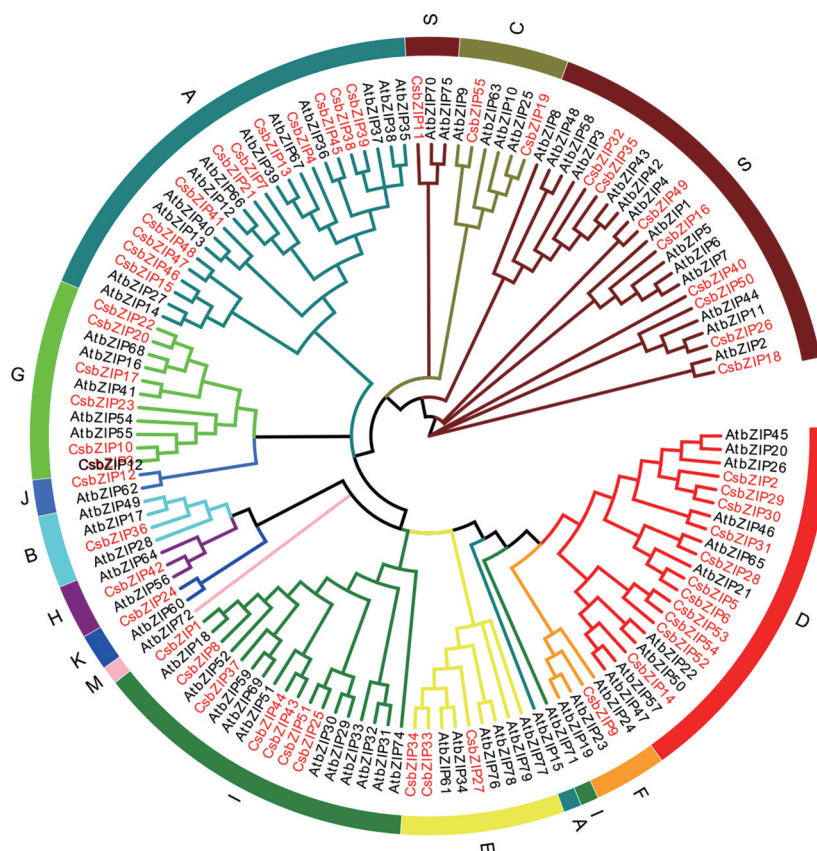
### 2 **CsbZIP**基因的系统发育分析

利用**CsbZIP**基因的氨基酸序列和拟南芥**bZIP**基因进行多序列比对, 分析**CsbZIP**基因家族的进化关系(图1)。系统发育树表明**CsbZIP**基因可分成12个亚家族, 这12个亚家族分别是A、B、C、D、E、F、G、H、I、J、K和S, 分别含有12、1、2、11、3、1、6、1、7、1、1和9个**CsbZIP**基因, 其中最大的3个亚家族分别是A、D和S。

### 3 **CsbZIP**基因结构和保守基序分析

通过分析**CsbZIP**基因的内含子和外显子分布, 进一步研究其结构特征。系统发育分析表明同一亚家族内的基因通常具有相似的内含子和外显子结构(图2)。这些**CsbZIP**基因的外显子数量范围在1~12之间, 有10个**CsbZIP**基因(18.18%)只含有一个外显子, 其中绝大多数是S亚家族成员。D亚家族中的成员含有外显子的数量范围在8~12之间; 亚家族G含有6~12个外显子; 亚家族A含有3~6个外显子。这些结构特征可能与它们在基因组中的功能有关。

在55个**CsbZIP**蛋白中鉴定到了10个保守的motif结构, 同一亚家族蛋白成员之间的motif分布相似(图2)。Motif1普遍存在于所有**CsbZIP**蛋白中, 代表着高



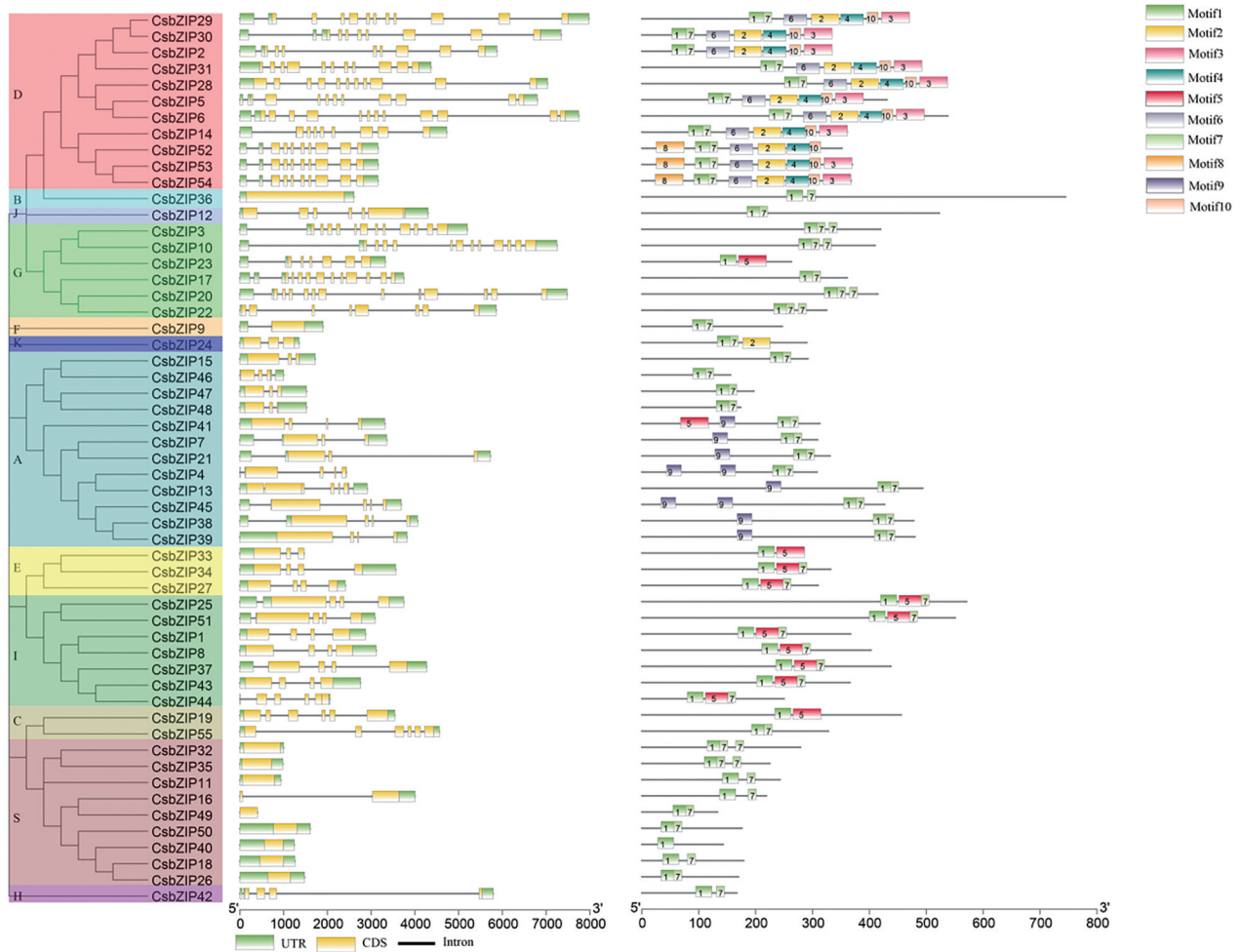
**Figure 1** Phylogenetic tree of **bZIP** genes in *Cannabis sativa* and *Arabidopsis thaliana* was constructed using MEGA7.0 based on the NJ method; bootstrap was 1 000 replicates

**Table 2** Basic information of *CsbZIP* genes identified in *Cannabis sativa*

Gene name	Gene ID	CHR	Chromosome location	Gene length/bp	ORF length/bp	Deduced protein				Subcellular location
						Size(aa)	MW/kDa	pI	GRAVY	
<i>CsbZIP1</i>	XP_030490468.1	1	131184-133529	2 345	1 100	367	40.88	6.07	-0.95	Nucleus
<i>CsbZIP2</i>	XP_030490768.1	1	45266477-45271510	5 033	991	332	37.17	8.61	-0.55	Nucleus
<i>CsbZIP3</i>	XP_030489233.1	1	79796624-79799733	3 109	1 251	420	45.65	8.38	-0.85	Nucleus
<i>CsbZIP4</i>	XP_030506327.1	2	889819-892115	2 296	923	308	35.81	6.81	-1.00	Nucleus
<i>CsbZIP5</i>	XP_030503171.1	2	932250-938488	6 238	1 286	431	48.60	7.81	-0.74	Nucleus
<i>CsbZIP6</i>	XP_030503170.1	2	932250-939218	6 968	1 605	538	60.21	6.72	-0.78	Nucleus
<i>CsbZIP7</i>	XP_030503792.1	2	14777800-14779739	1 939	927	309	34.19	6.72	-0.68	Nucleus
<i>CsbZIP8</i>	XP_030505964.1	2	87269479-87271912	2 433	1 208	403	43.53	6.07	-0.81	Nucleus
<i>CsbZIP9</i>	XP_030495829.1	3	482152-482895	743	743	247	27.70	5.65	-0.79	Nucleus
<i>CsbZIP10</i>	XP_030493871.1	3	4062813-4066744	3 931	1 221	410	44.47	6.44	-0.83	Nucleus
<i>CsbZIP11</i>	XP_030495583.1	3	22692505-22693236	731	731	243	26.98	8.69	-0.77	Nucleus
<i>CsbZIP12</i>	XP_030494140.1	3	47006072-47009777	3 705	1 565	523	57.75	8.35	-0.79	Nucleus
<i>CsbZIP13</i>	XP_030492683.1	3	80289965-80292288	2 323	1 479	494	53.50	6.82	-0.92	Nucleus
<i>CsbZIP14</i>	XP_030498547.1	4	66699237-66702286	3 049	1 078	361	40.83	6.48	-0.46	Nucleus
<i>CsbZIP15</i>	XP_030499953.1	4	79302204-79303389	1 185	876	292	32.11	9.84	-0.86	Nucleus
<i>CsbZIP16</i>	XP_030498081.1	4	80330869-80334502	3 633	658	219	25.16	8.58	-0.86	Nucleus
<i>CsbZIP17</i>	XP_030499008.1	4	83685986-83688483	2 497	1 075	361	37.80	5.85	-0.86	Nucleus
<i>CsbZIP18</i>	XP_030498421.1	4	84898096-84898635	539	539	179	20.51	6.29	-0.81	Nucleus
<i>CsbZIP19</i>	XP_030498505.1	4	91804778-91808068	3 290	1 365	456	49.05	6.25	-0.73	Nucleus
<i>CsbZIP20</i>	XP_030501220.1	5	33851498-33857736	6 238	1 231	415	43.96	6.02	-0.88	Nucleus
<i>CsbZIP21</i>	XP_030501737.1	5	66297971-66302307	4 336	993	331	36.75	9.13	-0.79	Nucleus
<i>CsbZIP22</i>	XP_030501753.1	5	66567617-66573052	5 435	970	325	34.29	6.99	-0.87	Nucleus
<i>CsbZIP23</i>	XP_030500765.1	5	84545743-84547629	1 886	786	263	28.81	5.17	-0.94	Nucleus
<i>CsbZIP24</i>	XP_030510819.1	6	31989647-31990811	1 164	870	290	32.79	4.62	-0.65	Nucleus
<i>CsbZIP25</i>	XP_030510247.1	6	72721569-72724262	2 693	1 712	571	62.24	6.16	-0.91	Nucleus
<i>CsbZIP26</i>	XP_030510623.1	6	74719303-74719815	512	512	170	18.89	6.51	-0.62	Nucleus
<i>CsbZIP27</i>	XP_030479688.1	7	1599399-1601449	2 050	929	310	34.12	5.33	-0.76	Nucleus
<i>CsbZIP28</i>	XP_030481936.1	8	26913865-26920330	6 465	1 600	536	59.97	7.15	-0.71	Nucleus
<i>CsbZIP29</i>	XP_030482272.1	8	31132156-31138924	6 768	1 396	468	51.33	7.81	-0.55	Nucleus
<i>CsbZIP30</i>	XP_030482275.1	8	31134053-31138924	4 871	991	332	37.05	8.85	-0.63	Nucleus
<i>CsbZIP31</i>	XP_030481818.1	8	35602476-35606107	3 631	1 462	490	54.54	6.1	-0.50	Nucleus
<i>CsbZIP32</i>	XP_030481671.1	8	38542363-38543202	839	839	279	32.27	5.66	-1.12	Nucleus
<i>CsbZIP33</i>	XP_030482255.1	8	51565650-51566793	1 143	828	276	30.21	5.82	-0.72	Nucleus
<i>CsbZIP34</i>	XP_030482254.1	8	51565650-51568127	2 477	995	332	36.66	6.01	-0.82	Nucleus
<i>CsbZIP35</i>	XP_030482052.1	8	55335000-55335677	677	677	225	26.03	5.65	-0.86	Nucleus
<i>CsbZIP36</i>	XP_030482193.1	8	56575957-56578194	2 237	2 237	745	80.39	6.51	-0.55	Nucleus
<i>CsbZIP37</i>	XP_030508159.1	9	50617099-50620262	3 163	1 313	438	47.83	6.09	-0.86	Nucleus
<i>CsbZIP38</i>	XP_030486255.1	X	3101301-3104000	2 699	1 433	478	51.18	9.47	-0.63	Nucleus
<i>CsbZIP39</i>	XP_030486266.1	X	3263820-3266559	2 739	1 439	480	51.30	9.47	-0.62	Nucleus
<i>CsbZIP40</i>	XP_030489067.1	X	9161244-9161675	431	431	143	16.03	9.09	-0.60	Nucleus
<i>CsbZIP41</i>	XP_030493109.1	X	28546514-28546594	2 518	938	313	34.12	8.54	-0.55	Chloroplast
<i>CsbZIP42</i>	XP_030495242.1	X	43111516-43116940	5 424	500	167	18.30	9.33	-1.12	Nucleus
<i>CsbZIP43</i>	XP_030497414.1	X	59123877-59125886	2 009	1 097	366	40.03	8.94	-0.84	Nucleus
<i>CsbZIP44</i>	XP_030507823.1	X	59583168-59585199	2 031	747	250	28.22	10.26	-0.91	Nucleus
<i>CsbZIP45</i>	XP_030499089.1	X	75289063-75291690	2 627	1 280	427	45.87	9.63	-0.61	Nucleus
<i>CsbZIP46</i>	XP_030507426.1	X	82004861-82005544	683	468	156	17.81	9.37	-0.95	Nucleus
<i>CsbZIP47</i>	XP_030508845.1	X	82637160-82638001	841	591	197	22.27	9.76	-0.82	Nucleus
<i>CsbZIP48</i>	XP_030508849.1	X	82637160-82637919	759	522	174	19.62	9.12	-0.77	Nucleus
<i>CsbZIP49</i>	XP_030477746.1	X	87682842-87683243	401	401	133	15.41	7.76	-0.75	Nucleus
<i>CsbZIP50</i>	XP_030484774.1	X	102894348-102894878	530	530	176	20.26	6.19	-1.01	Nucleus
<i>CsbZIP51</i>	XP_030485273.1	X	103452793-103455182	2 389	1 652	551	59.85	6.26	-0.83	Nucleus
<i>CsbZIP52</i>	XP_030485546.1	X	104179989-104182252	2 263	1 050	352	39.77	5.38	-0.61	Nucleus
<i>CsbZIP53</i>	XP_030485505.1	X	104179989-104182292	2 303	1 104	370	42.29	6.13	-0.59	Nucleus
<i>CsbZIP54</i>	XP_030485530.1	X	104180194-104182292	2 098	1 099	368	42.10	6.13	-0.59	Nucleus
<i>CsbZIP55</i>	XP_030485618.1	X	104347422-104351748	4 326	981	328	36.25	5.95	-0.75	Nucleus

度保守的亮氨酸拉链结构域 (leucine zipper domain)。Motif2 只存在于亚家族 D 和亚家族 K 中；motif3、motif4、motif6 和 motif10 只存在于亚家族 D 中。亚家族 E 和 I 的每个成员均含有 motif5, *CsbZIP23* 蛋白 (亚

家族 G)、*CsbZIP41* 蛋白 (亚家族 A) 和 *CsbZIP19* 蛋白 (亚家族 C) 也都含有 motif5; 除了 *CsbZIP17* 蛋白 (亚家族 G), *CsbZIP33* 蛋白 (亚家族 E) 和 *CsbZIP19* 蛋白 (亚家族 C) 不含有 motif7, 其他的所有 *CsbZIP* 蛋白均含有



**Figure 2** Phylogenetic relationship, gene structure and conserved motifs of *bZIP* genes in *Cannabis sativa*

motif7; motif8 只存在于亚家族 D 的这 3 个成员中 (CsbZIP52、CsbZIP53 和 CsbZIP54); motif9 只存在于亚家族 A 的这 8 个成员中 (CsbZIP41、CsbZIP7、CsbZIP21、CsbZIP4、CsbZIP13、CsbZIP45、CsbZIP38 和 CsbZIP39)。系统发育分析结果与内含子/外显子分布和 motif 分布一致, 表明同一亚家族 *CsbZIP* 基因的结构差异较低, 进化关系密切, 保守程度较高。

#### 4 *CsbZIP* 基因的染色体定位、基因复制事件和共线性分析

55 个 *CsbZIP* 基因不均匀地分布在大麻的 10 条染色体上 (图 3)。其中 X 染色体上的基因数量最多, 有 18 个, 9 号染色体上最少, 只有 1 个。染色体的长度和染色体上基因的数量之间不存在正相关关系。

在 55 个 *CsbZIP* 基因中鉴定到 6 对片段重复事件, 涉及 9 个 *CsbZIP* 基因, *CsbZIP38* (亚家族 A) 和 *rna-XM030630405.1* (由于基因组数据不完整, 所以该基因在基因组注释信息中没有被成功注释, 因此这一对片段重复基因在图 4 中没有展示出来); *CsbZIP50* (亚家族

S) 和 *CsbZIP18* (亚家族 S); *CsbZIP50* (亚家族 S) 和 *CsbZIP26* (亚家族 S); *CsbZIP1* (亚家族 I) 和 *CsbZIP8* (亚家族 I); *CsbZIP18* (亚家族 S) 和 *CsbZIP26* (亚家族 S); *CsbZIP32* (亚家族 S) 和 *CsbZIP35* (亚家族 S)。结果表明, 形成一个基因复制事件的一对基因来自同一亚家族, 其中大部分都是亚家族 S 中的基因 (图 4)。然而, *CsbZIP* 基因家族中不存在串联重复事件。这些结果表明片段重复事件是 *CsbZIP* 基因家族扩张的主要驱动因素。

为了探究 *bZIP* 基因在不同物种中的进化关系, 对大麻和 4 种有代表性的植物进行了基因共线性分析 (图 5), 包括 2 种单子叶植物 (水稻和玉米) 和 2 种双子叶植物 (拟南芥和葡萄)。大麻和葡萄之间存在 36 个 *bZIP* 基因对, 其次分别是拟南芥 (18 对)、水稻 (10 对)、玉米 (5 对)。结果表明, 与单子叶植物相比, *CsbZIP* 的共线基因对更多的出现在双子叶植物中。此外, *CsbZIP1* 和 *CsbZIP40* 在大麻和其他 4 种植物之间均存在共线对, 表明在双子叶植物和单子叶植物分化之前, 这些同

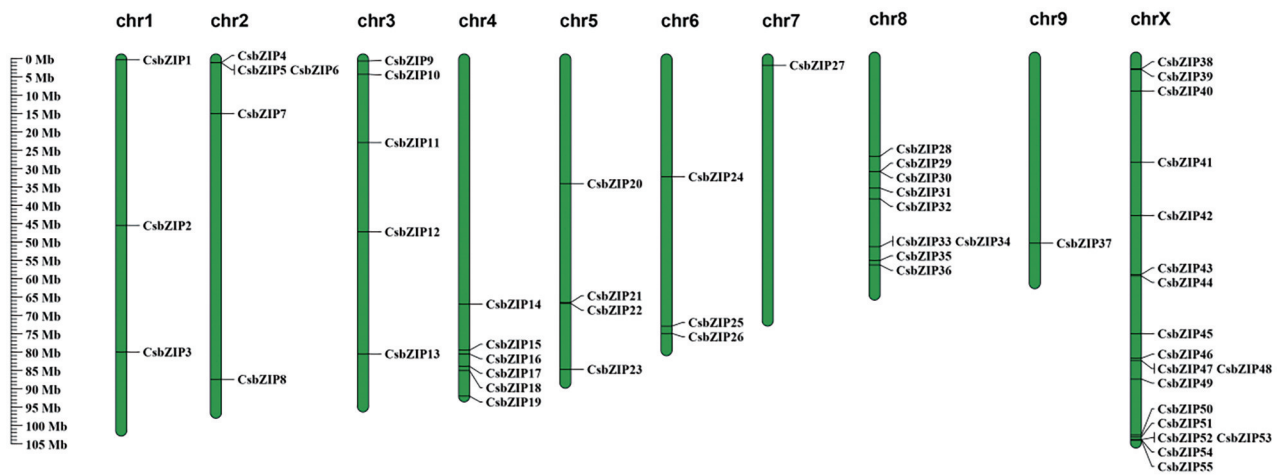


Figure 3 Chromosomal locations of *CsbZIP* genes on *Cannabis sativa* chromosomes

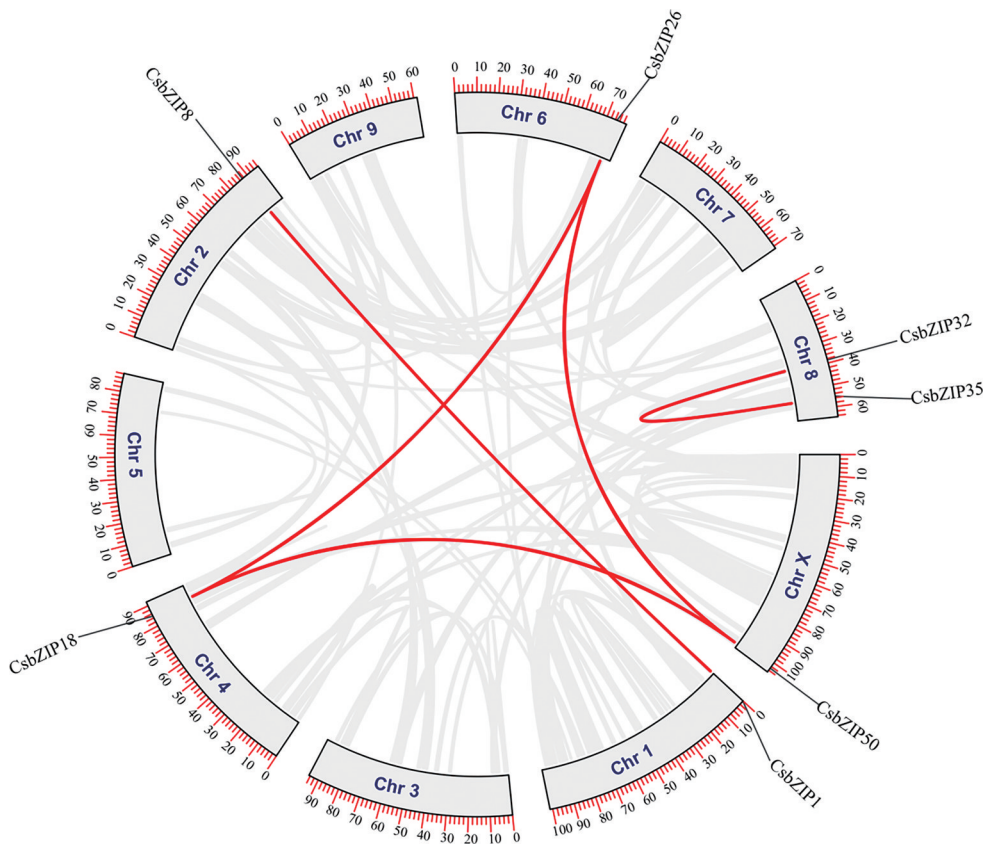


Figure 4 Collinearity analysis of the *bZIP* gene family in *Cannabis sativa*. The gray lines indicate syntenic blocks in the *Cannabis sativa* genome, while the red lines between chromosomes indicate segmental duplicated gene pairs

源基因对可能已经存在。然而,有6个 *CsbZIP* 基因 (*CsbZIP14*、*CsbZIP18*、*CsbZIP25*、*CsbZIP31*、*CsbZIP37* 和 *CsbZIP50*) 的共线对只存在于拟南芥和葡萄中,而在水稻和玉米中不存在。结果表明,这些特殊的共线基因对可能是在双子叶植物和单子叶植物分化后形成的。

### 5 *CsbZIP* 基因的互作网络分析

为了进一步研究 *CsbZIP* 基因的生物学功能,基于

*CsbZIP* 基因与拟南芥之间的同源关系构建了 *CsbZIP* 基因的蛋白互作网络 (图6)。发现了26个与拟南芥同源的 *CsbZIP* 基因和141个与之相互作用的基因。这些互作基因包括一些重要的转录调控因子,如 *bZIP*、*WRKY* 和 *MYB*; 一些调控植物生长发育的基因,如 *SNRK*、*ROXY*、*CRP*、*IRE* 和 *SPA*。还有一些参与植物信号转导的基因,如 *ABI1*、*KEG* 和 *AFP* 等。结果表明



**Figure 5** Synteny analysis of *bZIP* genes between *Cannabis sativa* and four representative plant species. Gray lines in the background indicate the collinear blocks within *Cannabis sativa* and other plant genomes, while the red lines highlight the syntenic *bZIP* gene pairs. The specie names with the prefixes 'C. sativa', 'A. thaliana', 'V. vinifera', 'O. sativa' and 'Z. mays' indicate *Cannabis sativa*, *Arabidopsis thaliana*, *Vitis vinifera*, *Oryza sativa*, and *Zea mays*, respectively

*CsbZIP* 基因在大麻中的作用非常重要, 广泛参与各种生命活动过程。

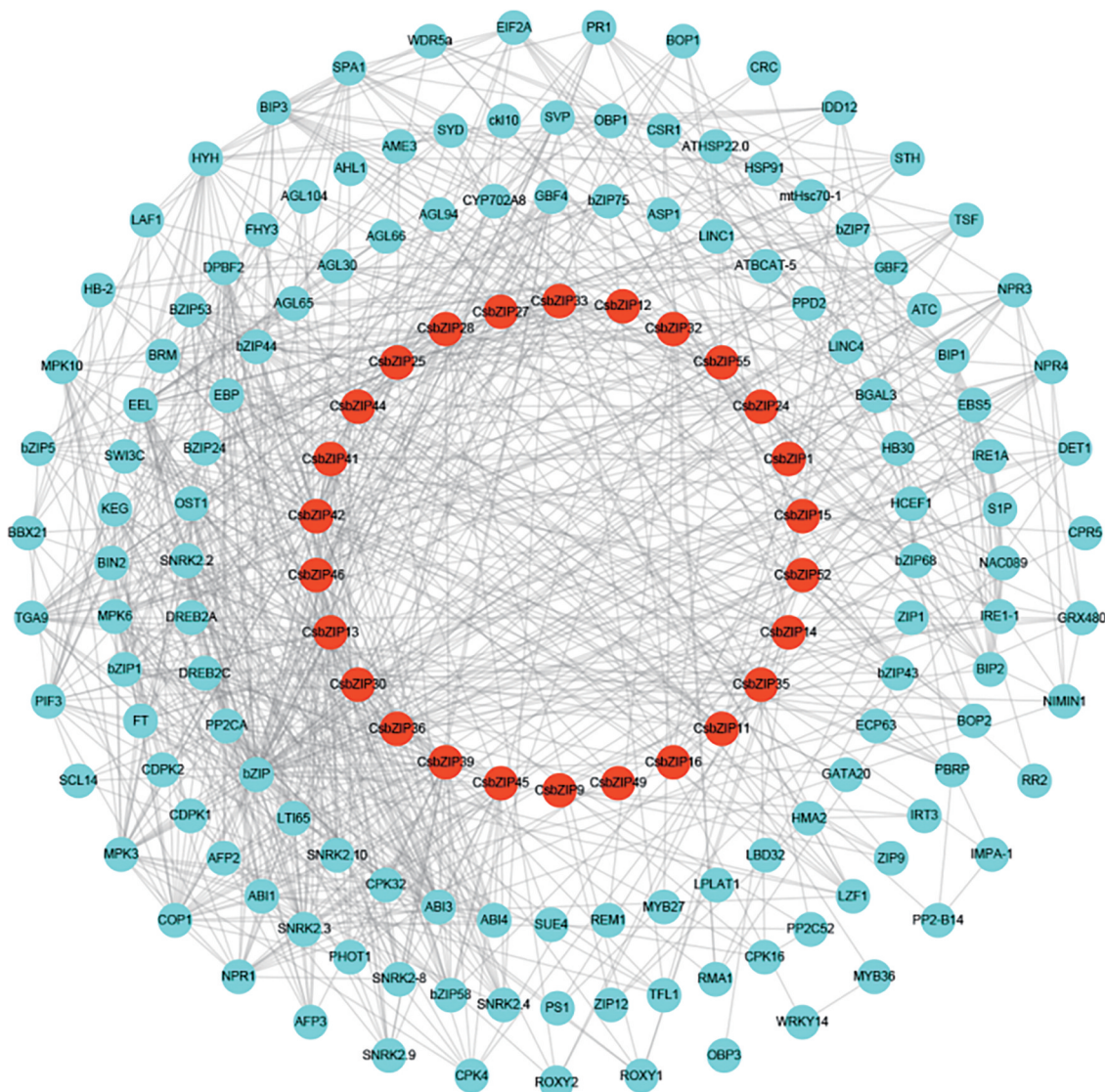
### 6 *CsbZIP* 基因的表达分析

为了研究 *CsbZIP* 基因的表达模式, 基于大麻的不同组织和不同生长时期的转录组数据绘制了热图。结果表明, 有些 *CsbZIP* 基因的表达呈现出明显的组织特异性, 如 *CsbZIP4*、*CsbZIP11* 和 *CsbZIP46* 只在雌花中高表达。*CsbZIP12* 和 *CsbZIP13* 只在种子中高表达, *CsbZIP24* 和 *CsbZIP40* 只在根中高表达。有些基因在不同的组织器官中均有表达, 如 *CsbZIP28* 在雄花和种子中的表达水平均较高, *CsbZIP14* 和 *CsbZIP52* 在根和茎中的表达水平较高, *CsbZIP45* 在根、茎、叶和花(雌花和雄花)中都有较高的表达水平。有些基因在所有的组织器官内不表达或表达水平极低, 如 *CsbZIP16*、*CsbZIP27*、*CsbZIP46*、*CsbZIP48* 和 *CsbZIP49* (图 7A)。

在不同的发育时期中, *CsbZIP* 基因的表达情况也类似。有些基因的表达水平表现出明显的时期特异性,

如 *CsbZIP13* 和 *CsbZIP26* 只在 S4 时期有较高的表达水平; *CsbZIP15* 和 *CsbZIP33* 只在 S1 时期高表达。有些基因在不同的时期均有较高的表达水平, 如 *CsbZIP4* 在 S3 和 S4 时期的表达水平较高; *CsbZIP18* 在 S3、S4、S5 时期的表达水平较高; *CsbZIP24*、*CsbZIP30*、*CsbZIP36* 和 *CsbZIP41* 在 5 个时期都有较高的表达水平。有些基因在所有时期内几乎不表达或表达水平极低, 如 *CsbZIP5*、*CsbZIP11* 和 *CsbZIP16* (图 7B)。表达模式分析结果表明 *CsbZIP* 基因成员在大麻的不同组织器官和不同发育时期中的表达均存在差异性, 说明 *CsbZIP* 基因之间已出现功能分化。

依据基因在转录组中的表达水平, 以根为对照组, 以种子为实验组, 从中挑选了 7 个有代表性的基因 (*CsbZIP12*、*CsbZIP13*、*CsbZIP28*、*CsbZIP41*、*CsbZIP42*、*CsbZIP43* 和 *CsbZIP44*), 利用 qRT-PCR 实验研究基因的表达水平。结果表明, 这 7 个 *CsbZIP* 基因在种子中的表达水平较高 (图 7C), 表明这些基因可能在种子的生



**Figure 6** Predicted protein-protein interaction networks of CsbZIP transcription factors (TFs) with other proteins in *Cannabis sativa* using STRING tool. The red circles represent CsbZIP TFs, and the cyan circle represent proteins that interact with CsbZIP TFs

长发育中有着重要的调控作用。此外, qRT-PCR 实验结果与转录组数据保持一致, 表明转录组数据是可靠的。

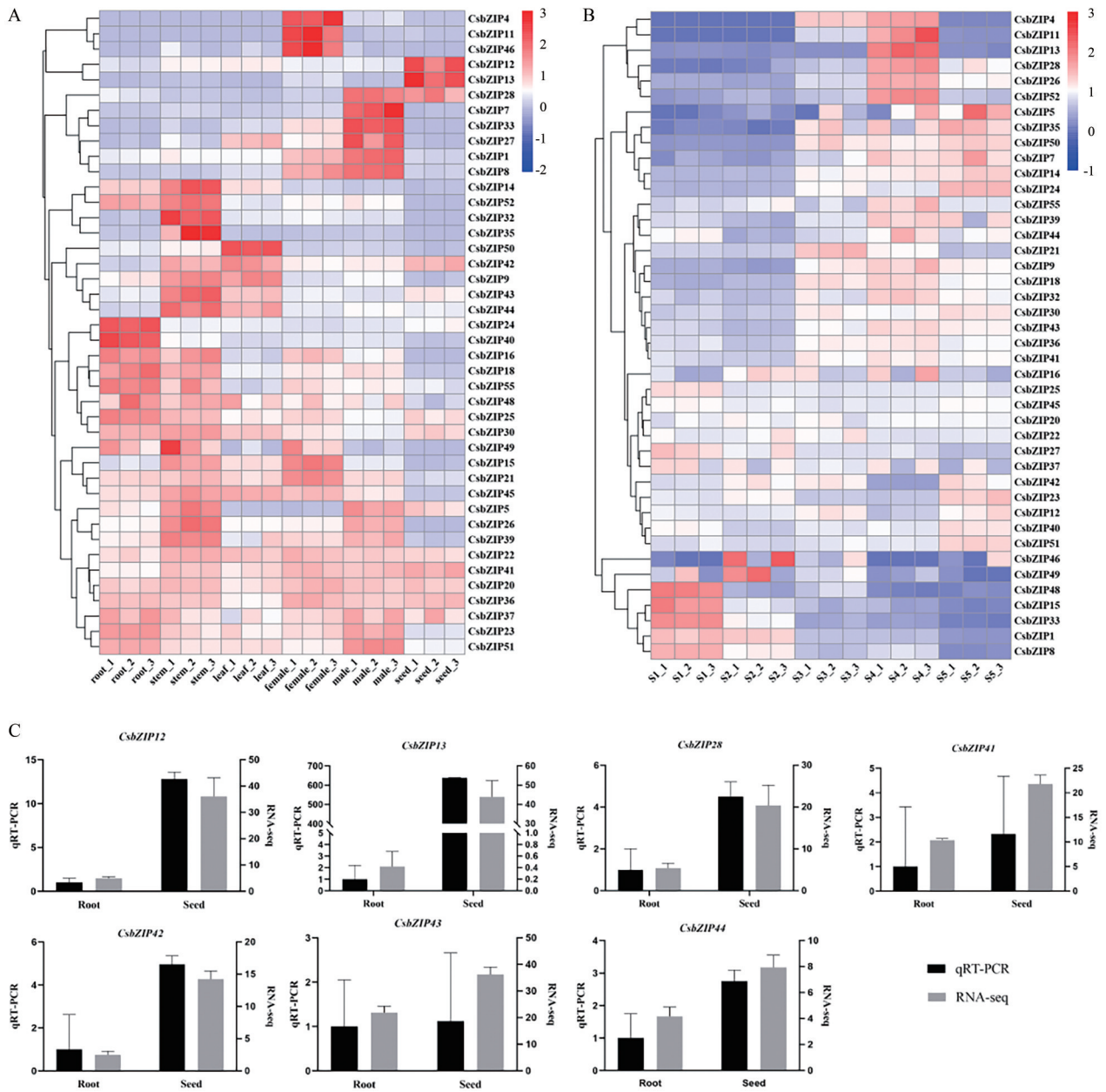
### 7 油脂合成基因的表达分析

基于已有的研究<sup>[29]</sup>, 构建了大麻仁油脂合成通路(图 8A), 在转录组数据中共鉴定到 86 个油脂合成基因, 对这些油脂合成基因进行启动子区顺式作用元件分析, 结果表明, 有 73 个油脂合成基因的启动子区含有 G-box 或 A-box 作用元件(图 8B), 因此, 对这 73 个油脂合成基因的表达水平进行分析。转录组数据表明, 有些油脂合成基因的表达存在组织特异性, 如 *KAS II-1*、*LACS-7*、*LACS15*、*LACS16* 和 *LACS17* 只在雄花中高表达; *SAD-8* 只在雌花中高表达; *FAD3-1* 在根中的表达水平显著高于其他组织器官, *FAD3-2* 在叶中的表达水平显著高于其他组织器官。有些基因在不同的组

织器官内均有表达, 如 *PDH-2*、*PDH-3*、*PDH-4*、*PDH-5* 和 *ACC-1*; *FATA-3* 在叶、雌花、雄花内高表达, 而在根、茎和种子内几乎不表达。有些基因在所有组织器官内不表达或表达水平极低, 如 *FATB-2*、*PDAT-9*、*FAD2-4*、*FAD2-9* 和 *FAD2-II*(图 9A)。

在不同的发育时期中, 油脂合成基因的表达情况类似。有些基因只在特定的时期表达, 如 *LACS-13* 和 *FAD2-II* 只在 S4 时期表达, 而在其他时期不表达或表达水平极低。有些基因在不同的时期均有表达, 如 *PDH-2*、*PDH-3*、*PDH-4* 和 *PDH-5* 在 5 个发育时期中均有表达。有些基因在所有时期中几乎不表达或表达水平极低, 如 *FATA-1*、*FATA-2*、*FATB-2*、*LACS-7*、*PDAT-7*、*PDAT-9* 和 *FAD2-9*(图 9B)。

基于油脂合成基因的表达水平以及启动子区的顺



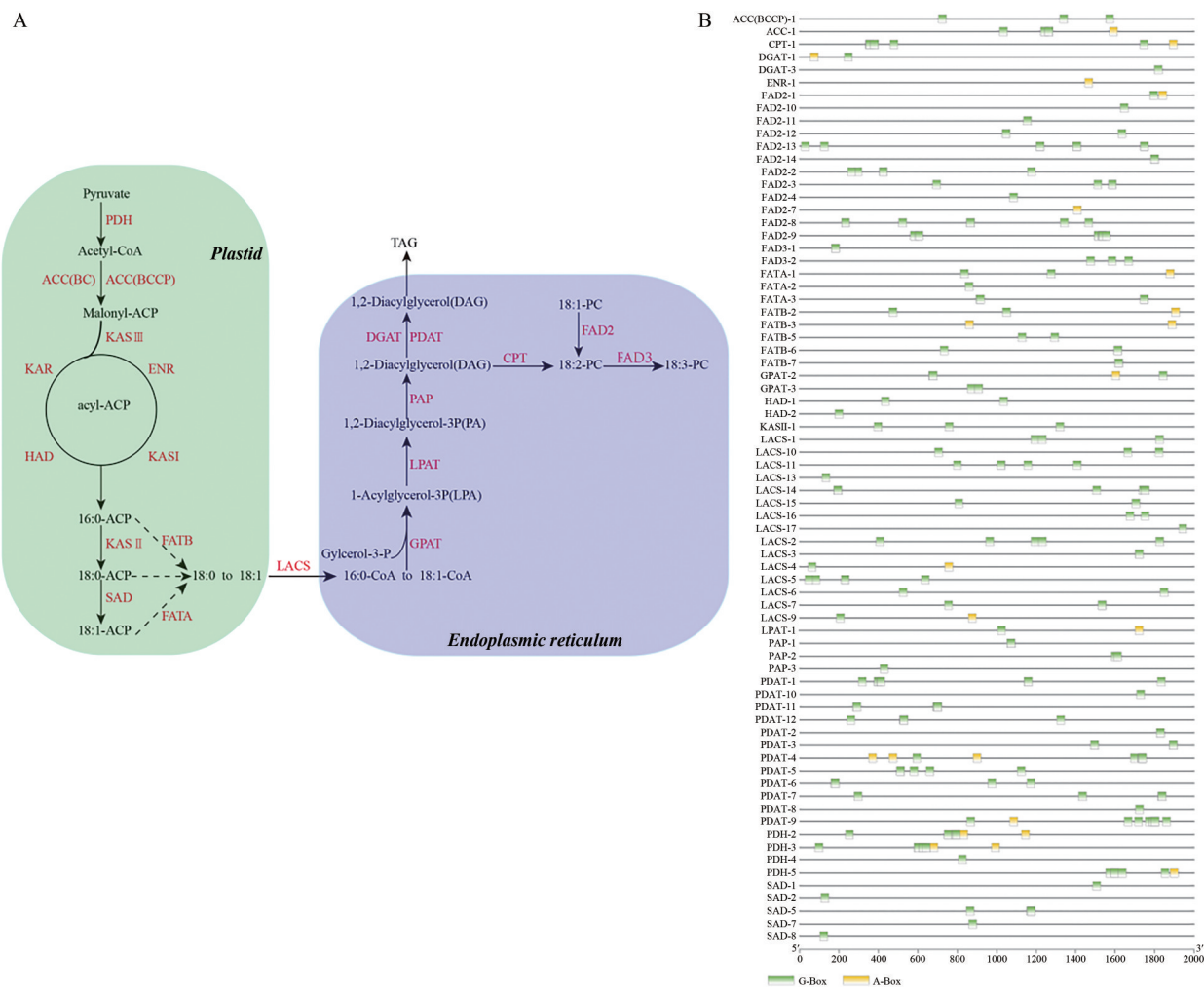
**Figure 7** (A) Expression profiles of *CsbZIP* genes in different tissues including root, stem, leaves, female flower, male flower and seed. (B) Expression profiles of *CsbZIP* genes in different stages, S1: Phase I; S2: Phase II; S3: Phase III; S4: Phase IV; S5: Phase V. (C) The expression levels of *CsbZIP* genes were verified by qRT-PCR

式作用元件预测结果, 筛选了7个油脂合成基因 (*HAD-1*、*ENR-1*、*LACS-4*、*GPAT-3*、*DGAT-1*、*PDAT-3* 和 *PDAT-4*), 以根为对照组, 种子为实验组, 使用qRT-PCR实验研究基因的表达水平(图9C)。结果表明, 这7个油脂合成基因在种子中有较高的表达水平, 表明这些油脂合成基因的表达可能会影响火麻仁的生长发育。

### 8 *CsbZIP*基因与油脂合成基因之间的相关性分析

基于转录组数据, 对7个*CsbZIP*基因和7个油脂合成基因进行相关性分析(图10), 结果显示, *CsbZIP12*与*GPAT-3*/*PDAT-3*之间存在显著正相关关系; *CsbZIP13*与

*GPAT-3*/*PDAT-4*之间存在显著正相关关系; *CsbZIP28*与*LACS-4*/*DGAT-1*/*PDAT-3*之间存在显著正相关关系; *CsbZIP41*与*HAD-1*之间存在显著正相关关系; *CsbZIP42*与*GPAT-3*之间存在显著正相关关系; *CsbZIP43*与*LACS-4*/*DGAT-1*/*PDAT-3*/*PDAT-4*之间存在显著正相关关系; *CsbZIP44*与*HAD-1*/*ENR-1*/*LACS-4*/*DGAT-1*/*PDAT-4*之间存在显著正相关关系。结果表明*CsbZIP*基因可能对油脂合成基因的表达具有调控作用, 从而影响火麻仁中油脂的积累。



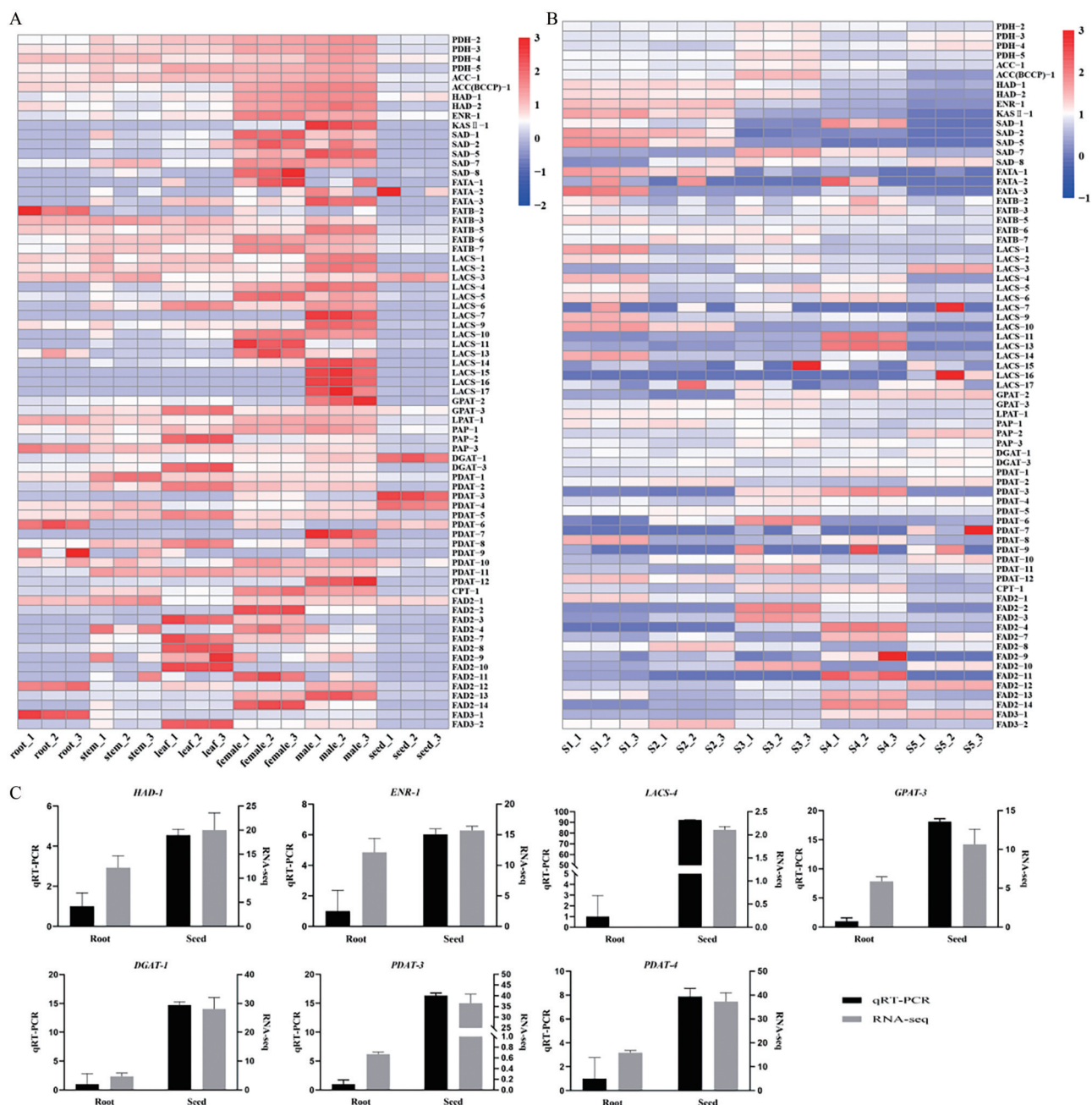
**Figure 8** (A) The lipid biosynthesis pathway in *Cannabis sativa*. Enzymes in this pathway are marked in red words. PDH: Pyruvate dehydrogenase; ACC(BC): Biotin carboxylase subunit of heteromeric acetyl-CoA carboxylase (ACCase); ACC (BCCP): Biotin carboxyl carrier protein of heteromeric ACCase; KAS: Ketoacyl-ACP synthase; KAR: 3-Ketobutyryl-ACP reductase; HAD: 3-Hydroxybutyryl-ACP dehydratase; ENR: Enoyl-ACP reductase; SAD: Stearoyl-ACP desaturase; FATA: Acyl-ACP thioesterase A; FATB: Acyl-ACP thioesterase B; LACS: Long-chain acyl-CoA synthetase; GPAT: Glycerol-3-phosphate acyltransferase; LPAT: Lysophosphatidic acid acyltransferase; PAP: Phosphatidic acid phosphatase; DGAT: Diacylglycerol acyltransferase; PDAT: Phospholipid: diacylglycerol acyltransferase; CPT: Diacylglycerol cholinephosphotransferase; FAD2:  $\nu$ -6 desaturase; FAD3:  $\nu$ -3 desaturase; TAG: Triacylglycerol. (B) The prediction of upstream *cis*-elements of lipid synthesis genes

### 讨论

本研究在大麻中共鉴定到 55 个 *CsbZIP* 基因, 利用生物信息学方法对 55 个 *CsbZIP* 基因进行了系统性研究, 结果表明 55 个 *CsbZIP* 基因的编码蛋白长度范围介于 133~745 个氨基酸之间; 蛋白质分子质量介于 15.41~80.39 kDa。55 个 *CsbZIP* 基因可分成 12 个亚家族, 同一亚家族成员之间的基因结构分布相似。片段重复事件是 *CsbZIP* 基因家族扩张的主要驱动因素。表达模式分析结果表明 7 个 *CsbZIP* 基因 (*CsbZIP12*、*CsbZIP13*、*CsbZIP28*、*CsbZIP41*、*CsbZIP42*、*CsbZIP43* 和 *CsbZIP44*) 和 7 个油脂合成通路基因 (*HAD-1*、*ENR-1*、

*LACS-4*、*GPAT-3*、*DGAT-1*、*PDAT-3* 和 *PDAT-4*) 在火麻仁中的相对表达水平较高。顺式作用元件分析表明上述 7 个油脂合成基因的启动子区含有 A-box 或 G-box 作用元件。相关性分析表明上述 7 个 *CsbZIP* 基因和 7 个油脂合成基因之间的表达水平存在显著的正相关关系。结果表明, *CsbZIP* 转录因子可能对火麻仁油脂代谢具有重要的调控作用。

*bZIP* 基因家族是植物中最大的转录因子家族之一, 可参与多种生物学过程, 包括植物生长发育的调控、生物和非生物胁迫响应、次级代谢物质的生物合成<sup>[30]</sup>。*bZIP* 基因在一些物种中已经被报道过, 如拟南芥<sup>[4]</sup>、水

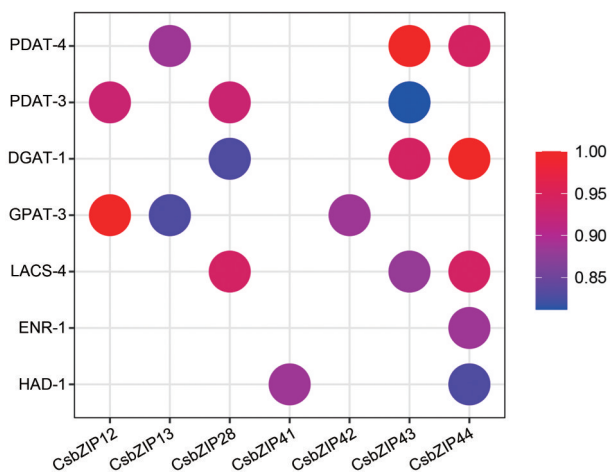


**Figure 9** (A) Expression profiles of lipid synthesis genes in different tissues including root, stem, leaves, female flower, male flower and seed. (B) Expression profiles of lipid synthesis genes in different stages, S1: Phase I; S2: Phase II; S3: Phase III; S4: Phase IV; S5: Phase V. (C) The expression levels of lipid synthesis genes were verified by qRT-PCR

稻<sup>[21]</sup>、葡萄<sup>[22]</sup>、玉米<sup>[23]</sup>等。相比较于拟南芥、水稻和玉米，大麻中的*bZIP*基因的数量较少，只有55个，与葡萄中*bZIP*基因的数量一致，物种间的共线性关系也表明，*CsbZIP*基因与葡萄*bZIP*基因之间的共线对的数量最多，表明大麻与葡萄之间的亲缘关系更密切。亚细胞定位预测结果显示除了*CsbZIP41*蛋白定位在叶绿体上，其他的家族成员都定位在细胞核中，表明*CsbZIP*转录因子主要在细胞核内发挥作用，而*CsbZIP41*转录因子则可能在叶绿体中发挥特殊的作用，这与油桐中

的报道类似<sup>[31]</sup>。

系统发育分析结果表明，55个*CsbZIP*基因可分成12个亚家族。拟南芥的78个*bZIP*基因被分成13个亚家族<sup>[4]</sup>，与拟南芥相比较，大麻缺少M亚家族。研究表明，葡萄*bZIP*基因可分成10个亚家族，缺少K、M和S亚家族<sup>[22]</sup>。芝麻*bZIP*基因可分成9个亚家族，缺少E、J、K和M亚家族<sup>[32]</sup>。由此可见，在植物*bZIP*基因亚家族中，存在基因缺失事件，*CsbZIP*基因的亚家族缺失可能是由于该家族成员在进化过程中被丢失。



**Figure 10** Correlation analysis of *CsbZIP* genes with lipid synthesis genes. The different colors represent the correlation coefficient

基因结构分析表明, *CsbZIP* 基因含有 1~12 个外显子, 其中 18% 的基因只含有一个外显子, 它们绝大多数分布在 S 亚家族中, 研究表明, S 亚家族的成员通常与 C 亚家族的成员互作形成二聚体, 在植物的能量代谢调控过程中发挥着关键的作用<sup>[4]</sup>。例如, 在拟南芥中, *bZIP63* 转录因子可与 S 亚家族的其他成员结合形成二聚体结构, 促进支链氨基酸的分解代谢, 提供 ATP 以维持植物在逆境中的生存<sup>[33,34]</sup>。D 亚家族的成员含有 8~12 个外显子, 大多数 D 亚家族成员可介导水杨酸、茉莉酸等激素信号通路来提高植物在生物胁迫下的耐受性<sup>[4]</sup>。结果表明, 不同亚家族成员之间的基因结构不同, 可能有着不同的生物学功能。同一亚家族中的内含子/外显子结构和 motif 的排列分布相似, 并倾向于执行相同的生物学功能。例如, 同属于 H 亚家族的拟南芥 *AtHY5* 蛋白和水稻 *OsbZIP48* 蛋白分别在双子叶和单子叶系统的光形态建成中发挥重要作用<sup>[35]</sup>。值得关注的是, 有些 motif 只存在于特定的亚家族中, 如 *Motif2* 只存在于亚家族 D 和亚家族 K 中, 鉴于 *bZIP* 转录因子的功能多样性, 有些 motif 可能具备一些特殊的功能, 需要进一步深入研究。

基因复制事件是遗传系统进化的主要驱动因素之一, 其中串联重复和片段重复是植物基因家族扩张的两个主要因素<sup>[36,37]</sup>。本次研究鉴定到 6 对片段重复基因, 占 *CsbZIP* 基因的 14.5%。但是, 没有一对 *CsbZIP* 基因被鉴定为串联重复, 表明在 *CsbZIP* 基因家族的进化和扩张过程中, 片段重复事件起着重要作用, 这与葡萄、芝麻和水稻中的研究一致<sup>[21,22,32]</sup>。物种间共线性分析结果表明, 大麻与双子叶植物之间的共线性基因对的数量显著多于单子叶植物, 表明在物种进化过程中,

亲缘关系相对密切的物种, 有更多的共线基因对, 例如大麻和葡萄。此外, 有 6 个 *CsbZIP* 基因 (*CsbZIP14*、*CsbZIP18*、*CsbZIP25*、*CsbZIP31*、*CsbZIP37* 和 *CsbZIP50*) 的共线对只存在于双子叶中, 而在单子叶中不存在, 说明这些特殊的共线基因对可能出现在双子叶植物和单子叶植物分化之后。

*bZIP* 基因家族参与植物的生长发育过程, 包括花的发育和种子的成熟<sup>[3,8]</sup>。为了探究 *CsbZIP* 基因的生物学功能, 对 *CsbZIP* 基因在大麻的不同组织器官和不同发育阶段中的表达模式进行分析, 研究表明, 虽然大多数 *CsbZIP* 基因在所有组织中均有表达, 但仍有一些 *CsbZIP* 基因在不同组织中的表达存在显著差异, 这与水稻、玉米和葡萄中的研究一致<sup>[21-23]</sup>。例如, 在大麻中, *CsbZIP4* 基因只在雌花中特异性表达, *CsbZIP13* 基因只在种子中特异性表达, 表明它们可能在大麻特定的组织器官发育中发挥关键作用。

基于拟南芥同源基因, 构建了 *CsbZIP* 转录因子的蛋白互作网络。在 *CsbZIP* 基因家族蛋白网络互作分析中, *CsbZIP12* 基因与 *AT1G19490.1* 基因互为同源基因, 研究表明 *AT1G19490.1* 基因的编码蛋白是 *AtbZIP62* 转录因子, 该转录因子在拟南芥中可与其他 *bZIP* 转录因子互作, 调控拟南芥对干旱胁迫的响应<sup>[38]</sup>。基于同源基因的功能相似性, 推测 *CsbZIP12* 基因可能在大麻参与逆境胁迫响应中发挥作用。*CsbZIP13* 基因与 *ABI5* 基因互为同源基因, 在拟南芥中, *ABI5* 基因是种子发育、萌发和幼苗生长所必需的<sup>[39]</sup>。蛋白互作网络图显示 *ABI3* 蛋白、*SNRK2.2* 蛋白、*AFP2* 蛋白等多种蛋白都能与 *ABI5* 蛋白发生相互作用。转录组表达谱聚类分析表明, *CsbZIP13* 基因不仅在大麻的种子中高表达, 并且在火麻仁成熟发育期高表达。因此, 基于 *CsbZIP13* 基因与 *ABI5* 基因之间的同源关系, 推测 *CsbZIP13* 基因在大麻中可能与多种蛋白相互作用共同调控火麻仁的生长发育。

研究表明, 在植物中, *HAD*、*ENR*、*LACS*、*GPAT*、*DGAT* 和 *PDAT* 基因的表达会影响油脂的积累<sup>[40,41]</sup>。在本次研究中, 转录组数据及 qRT-PCR 实验结果均表明 *HAD-1*、*ENR-1*、*LACS-4*、*GPAT-3*、*DGAT-1*、*PDAT-3* 和 *PDAT-4* 在火麻仁中的表达水平较高。顺式作用元件分析结果表明, 这 7 个油脂合成基因的启动子区存在 A-box 或 G-box 作用元件, *bZIP* 转录因子可以通过 A-box 或 G-box 作用元件与基因结合, 从而调控基因的表达<sup>[42]</sup>。相关性分析结果表明, 7 个 *CsbZIP* 基因 (*CsbZIP12*、*CsbZIP13*、*CsbZIP28*、*CsbZIP41*、*CsbZIP42*、*CsbZIP43* 和 *CsbZIP44*) 的表达水平与 7 个油脂合成基因 (*HAD-1*、*ENR-1*、*LACS-4*、*GPAT-3*、*DGAT-1*、*PDAT-3*

和*PDAT-4*)的表达水平之间存在显著的正相关。这些结果表明,在火麻仁中,*CsbZIP*转录因子可能通过调控油脂合成代谢基因的表达,从而影响火麻仁中油脂含量的积累。

本研究利用生物信息学方法对*CsbZIP*基因家族进行了系统性研究,解析了*CsbZIP*基因的结构特征、进化方式和表达模式。此外,基于转录组数据,构建了火麻仁油脂代谢通路,预测了火麻仁油脂生物合成基因启动子区的顺式作用元件,并对油脂合成基因和*CsbZIP*基因的表达水平进行了相关性分析,初步筛选出了与火麻仁油脂代谢相关的*CsbZIP*转录因子。为后续深入研究*CsbZIP*转录因子在火麻仁油脂代谢调控机制中的作用提供了重要的参考。

**作者贡献:** 怀浩负责文章撰写及数据分析;董林林、宁康负责实验设计及论文修改;侯聪、杨树明负责数据分析和实验材料的收集;汪鋈植、陈士林指导文章撰写并提出修改意见;董林林负责论文设计及项目开展。

**利益冲突:** 所有作者均声明不存在利益冲突。

## References

- [1] Zhao K, Chen S, Yao WJ, et al. Genome-wide analysis and expression profile of the *bZIP* gene family in poplar [J]. *BMC Plant Biol*, 2021, 21: 122.
- [2] Rong SY, Wu ZY, Cheng ZZ, et al. Genome-wide identification, evolutionary patterns, and expression analysis of *bZIP* gene family in Olive (*Olea europaea* L.) [J]. *Genes (Basel)*, 2020, 11: 510.
- [3] Jakoby M, Weisshaar B, Dröge-Laser W, et al. bZIP transcription factors in *Arabidopsis* [J]. *Trends Plant Sci*, 2002, 7: 106-111.
- [4] Dröge-Laser W, Snoek BL, Snel B, et al. The *Arabidopsis* bZIP transcription factor family-an update [J]. *Curr Opin Plant Biol*, 2018, 45: 36-49.
- [5] Li HY, Li LX, ShangGuan GD, et al. Genome-wide identification and expression analysis of *bZIP* gene family in *Carthamus tinctorius* L [J]. *Sci Rep*, 2020, 10: 15521.
- [6] Li ST, Lauri A, Ziemann M, et al. Nuclear activity of ROXY1, a glutaredoxin interacting with TGA factors, is required for petal development in *Arabidopsis thaliana* [J]. *Plant Cell*, 2009, 21: 429-441.
- [7] Yang Y, Gao SQ, Tang YM, et al. Advance of bZIP transcription factors in plants [J]. *J Trit Crops (麦类作物学报)*, 2009, 29: 730-737.
- [8] Abe M, Kobayashi Y, Yamamoto S, et al. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex [J]. *Science*, 2005, 309: 1052-1056.
- [9] Joo H, Lim CW, Lee SC. A pepper ring-type E3 ligase, CaASRF1, plays a positive role in drought tolerance via modulation of CaAIBZ1 stability [J]. *Plant J*, 2019, 98: 5-18.
- [10] Zhang YQ, Zheng S, Liu ZJ, et al. Both HY5 and HYH are necessary regulators for low temperature-induced anthocyanin accumulation in *Arabidopsis* seedlings [J]. *J Plant Physiol*, 2011, 168: 367-374.
- [11] Van Leene J, Blomme J, Kulkarni SR, et al. Functional characterization of the *Arabidopsis* transcription factor bZIP29 reveals its role in leaf and root development [J]. *J Exp Bot*, 2016, 67: 5825-5840.
- [12] Hsieh TH, Li CW, Su RC, et al. A tomato bZIP transcription factor, SIAREB, is involved in water deficit and salt stress response [J]. *Planta*, 2010, 231: 1459-1473.
- [13] Deng CP, Shi M, Fu R, et al. ABA-responsive transcription factor bZIP1 is involved in modulating biosynthesis of phenolic acids and tanshinones in *Salvia miltiorrhiza* [J]. *J Exp Bot*, 2020, 71: 5948-5962.
- [14] Wan ZT, Lu M, Wu SS, et al. Identification and expression analysis of the MIKC-type MADS-box gene family in *Cannabis sativa* L [J]. *Acta Pharm Sin (药理学报)*, 2021, 56: 3173-3183.
- [15] Ning K, Dong LL, Li MZ, et al. Application and development of non-psychoactive medicinal cannabis [J]. *Chin J Exp Tradit Med Form (中国实验方剂学杂志)*, 2020, 26: 228-240.
- [16] Huai H, Dong LL, Ning K, et al. Genome-wide identification of the *Hsp20* gene family in *Cannabis sativa* hemp seed and its expression profile [J]. *Acta Pharm Sin (药理学报)*, 2022, 57: 1203-1215.
- [17] Su FF, Yang G, Zhang J, et al. Research on protein extraction method of comprehensive value of hemp seed [J]. *J Chin Cereals Oils Assoc (中国粮油学报)*, 2022. <https://kns.cnki.net/kcms/detail/11.2864.TS.20220111.1948.004.html>.
- [18] Wei F, Tu DP, Wang LP. Research progress in edible development and pharmacological action of hemp seed [J]. *Chin J Gerontol (中国老年学杂志)*, 2015, 35: 3486-3488.
- [19] Zhang HW, Zhang WJ, Zhang GF, et al. Research progress on the pharmacological effects of hemp seed based on the compatibility of traditional Chinese medicine [J]. *Chin J Hosp Pharm (中国医院药学杂志)*, 2022, 42: 659-664.
- [20] Li GL, Guo CY, Wu SH, et al. Regulatory effects of Cannabis Fructus oil on gut microecology in *D-galactose*-induced aging mice [J]. *Chin Tradit Herb Drugs (中草药)*, 2020, 51: 3509-3517.
- [21] Nijhawan A, Jain M, Tyagi AK, et al. Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice [J]. *Plant Physiol*, 2008, 146: 333-350.
- [22] Liu JY, Chen NN, Chen F, et al. Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (*Vitis vinifera*) [J]. *BMC Genomics*, 2014, 15: 281.
- [23] Wei KF, Chen J, Wang YM, et al. Genome-wide analysis of bZIP-encoding genes in maize [J]. *DNA Res*, 2012, 19: 463-476.
- [24] Zhou Y, Xu DX, Jia LD, et al. Genome-wide identification and structural analysis of bZIP transcription factor genes in *Brassica*

- napus* [J]. Genes (Basel), 2017, 8: 288.
- [25] Wang JZ, Zhou JX, Zhang BL, et al. Genome-wide expansion and expression divergence of the basic leucine zipper transcription factors in higher plants with an emphasis on sorghum [J]. J Integr Plant Biol, 2011, 53: 212-231.
- [26] Liao Y, Zou HF, Wei W, et al. Soybean *GmbZIP44*, *GmbZIP62* and *GmbZIP78* genes function as negative regulator of ABA signaling and confer salt and freezing tolerance in transgenic *Arabidopsis* [J]. Planta, 2008, 228: 225-240.
- [27] Mendes A, Kelly AA, Erp H, et al. bZIP67 regulates the omega-3 fatty acid content of *Arabidopsis* seed oil by activating *fatty acid desaturase3* [J]. Plant Cell, 2013, 25: 3104-3116.
- [28] Song QX, Li QT, Liu YF, et al. Soybean *GmbZIP123* gene enhances lipid content in the seeds of transgenic *Arabidopsis* plants [J]. J Exp Bot, 2013, 64: 4329-4341.
- [29] Ding MZ, Lou HQ, Chen WC, et al. Comparative transcriptome analysis of the genes involved in lipid biosynthesis pathway and regulation of oil body formation in *Torreya grandis* kernels [J]. Ind Crops Prod, 2020, 145: 112051.
- [30] Li H, Zheng XY, Wang RF, et al. Biology function of bZIP transcription factors and their application prospect in quality improvement of Chinese medicinal materials [J]. Acad J Shanghai Univ Tradit Chin Med (上海中医药大学学报), 2021, 35: 1-7.
- [31] Liang JH, Zhao QY. Identification and analysis of bZIP transcription factor family regulating oil metabolism in *Vernicia fordii* [J]. Plant Physiol J (植物生理学报), 2021, 57: 1135-1150.
- [32] Wang YY, Zhang YJ, Zhou R, et al. Identification and characterization of the bZIP transcription factor family and its expression in response to abiotic stresses in sesame [J]. PLoS One, 2018, 13: e200850.
- [33] Mair A, Pedrotti L, Wurzinger B, et al. SnRK1-triggered switch of bZIP63 dimerization mediates the low-energy response in plants [J]. Elife, 2015, 4: e200850.
- [34] Pedrotti L, Weiste C, Nägele T, et al. Snf1-RELATED KINASE1-controlled C/S<sub>1</sub>-bZIP signaling activates alternative mitochondrial metabolic pathways to ensure plant survival in extended darkness [J]. Plant Cell, 2018, 30: 495-509.
- [35] Burman N, Bhatnagar A, Khurana JP. OsbZIP48, a HY5 transcription factor ortholog, exerts pleiotropic effects in light-regulated development [J]. Plant Physiol, 2018, 176: 1262-1285.
- [36] Jeffäres DC, Penkett CJ, Bahler J. Rapidly regulated genes are intron poor [J]. Trends Genet, 2008, 24: 375-378.
- [37] Vision TJ, Brown DG, Tanksley SD. The origins of genomic duplications in *Arabidopsis* [J]. Science, 2000, 290: 2114-2117.
- [38] Rolly NK, Imran QM, Shahid M, et al. Drought-induced *AtbZIP62* transcription factor regulates drought stress response in *Arabidopsis* [J]. Plant Physiol Biochem, 2020, 156: 384-395.
- [39] Ju L, Jing YX, Shi PT, et al. JAZ proteins modulate seed germination through interaction with ABI5 in bread wheat and *Arabidopsis* [J]. New Phytol, 2019, 223: 246-260.
- [40] Huang RM, Zhou Y, Zhang JP, et al. Transcriptome analysis of Walnut (*Juglans regia* L.) embryos reveals key developmental stages and genes involved in lipid biosynthesis and polyunsaturated fatty acid metabolism [J]. J Agric Food Chem, 2021, 69: 377-396.
- [41] Li CQ, Hu LZ, Que BB, et al. Expression profiles of genes involved in fatty acid and lipid biosynthesis in developing seeds of *Paeonia ostii* [J]. Gen Genomics, 2021, 43: 885-896.
- [42] Herath V, Verchot J. Insight into the *bZIP* gene family in *Solanum tuberosum*: genome and transcriptome analysis to understand the roles of gene diversification in spatiotemporal gene expression and function [J]. Int J Mol Sci, 2020, 22: 253.