

翅果油树油体蛋白基因家族鉴定及其在种子发育中的表达模式分析

王海燕¹, 张馨之¹, 蒋晋豫¹, 赵雅², 刘玉林^{2*}

(1. 陕西省林业科学院, 西安 710002; 2. 西北农林科技大学 林学院, 陕西 杨凌 712100)

摘要: 油体蛋白 (Oleosin) 是油体表面丰度最高的蛋白, 调控油脂储存。目前, 木本油料植物翅果油树中 Oleosin 家族的功能及其在油脂合成积累等关键生物学过程中的作用尚未明晰。因此, 该研究采用生物信息学方法鉴定翅果油树油体蛋白, 并对该基因家族成员理化性质、进化关系、系统发育、保守基序、顺式作用元件结构特征、密码子偏好性和基因表达模式进行了系统的分析, 结果表明: (1) 从翅果油树的基因组中鉴定出 8 个 *Oleosin* 基因, 根据其在染色体上的位置命名为 *EmoOLE1-EmoOLE8*。(2) 系统进化树、保守基序和基因结构分析显示, 8 个 *EmoOLE* 基因被划分为 3 个亚家族。(3) *EmoOLE* 基因除核心启动子元件 (TATA-box 和 CAAT-box) 外, 共鉴定出 22 类功能各异的调控元件, 依据其生物学功能划分为光响应、植物激素响应、生长发育调控以及非生物胁迫响应 4 大类。

(4) 密码子偏好性分析显示, 该基因家族密码子偏好以 A/T 碱基结尾, 4 个密码子无偏好性, 27 个有偏好性。(5) 基于翅果油树 4 个发育阶段的种子转录组及种子中三酰甘油、二酰甘油含量, 表达谱分析显示 *EmoOLE* 在种子成熟期表达上调, qRT-PCR 证实了这一结果; 相关性分析结果表明, 7 个 *EmoOLE* 基因 (*EmoOLE1*、*EmoOLE3*、*EmoOLE4*、*EmoOLE5*、*EmoOLE6*、*EmoOLE7* 和 *EmoOLE8*) 的表达水平与三酰甘油或二酰甘油合成积累显著正相关。综上, 该研究结果为深入解析翅果油树 *Oleosin* 基因家族的功能及其高含油量分子育种提供了理论基础。

关键词: 翅果油树, *Oleosin* 基因家族, 系统进化, 密码子偏好性, 表达模式分析

中图分类号: Q943.2

文献标识码: A

文章编号:

Identification of Oleosin gene family in *Elaeagnus mollis* and analysis of its expression patterns during seed development

WANG Haiyan¹, ZHANG Xinzhi¹, JIANG Jinyu¹, ZHAO Ya², LIU Yulin^{2*}

(1. Shaanxi Academy of Forestry Sciences, Xi'an 710002, China; 2. College of Forestry, Northwest A & F University, Yangling 712100, Shannxi, China)

Abstract: Oleosin is the most abundant structural protein localized on the surface of oil bodies and serves a critical function in regulating lipid storage. Currently, the functions of the Oleosin gene family in the woody oilseed plant *Elaeagnus mollis*, particularly its roles in regulating key biological processes such as lipid biosynthesis and accumulation, remain unclear. Therefore, this study employed bioinformatics methods to identify *Oleosin* genes in *E. mollis* and systematically analyzed the physicochemical properties, protein structures, conserved motifs, phylogenetic relationships, cis-regulatory element structures, codon usage bias, and gene expression patterns of the gene family members. The results demonstrated that: (1) A total of 8 *Oleosin* genes were identified from the *E. mollis* genome and were designated *EmoOLE1* to *EmoOLE8* according

基金项目: 陕西省林业科技创新计划专项 (SXLK2024-0211)。

第一作者: 王海燕 (1973—), 本科, 高级工程师, 研究方向为林业育种和产业化发展等, (E-mail) 1158891663@qq.com。

***通信作者:** 刘玉林, 博士, 副教授, 研究方向为经济林种质发掘与利用, (E-mail) liuyulin@nwfufu.edu.cn。

to their chromosomal order. (2) Phylogenetic analysis, together with conserved motif and gene structure examination, classified the 8 *EmoOLE* genes into three distinct subfamilies. (3) In addition to core promoter elements such as TATA-box and CAAT-box, a total of 22 types of cis-regulatory elements were identified in the promoters of *EmoOLE* genes. These elements were categorized into four major functional modules: light response, phytohormone response, growth and development, and abiotic stress response. (4) Analysis of codon usage bias indicated that the *EmoOLE* gene family exhibits a strong preference for A/T-ended codons. Among them, 4 codons exhibited no significant bias, while 27 codons exhibited clear usage bias. (5) Based on the transcriptome data of the seeds at 4 developmental stages and the contents of triacylglycerol (TAG) and diacylglycerol (DAG), which demonstrated that *EmoOLE* expression was upregulated during seed maturation, and this result was also confirmed by qRT-PCR. Furthermore, correlation analysis revealed that 7 *EmoOLE* genes (*EmoOLE1*, *EmoOLE3*, *EmoOLE4*, *EmoOLE5*, *EmoOLE6*, *EmoOLE7* and *EmoOLE8*) displayed significant positive correlations with TAG or DAG accumulation, suggesting their crucial roles in lipid storage in *E. mollis* seeds. In conclusion, these findings establish a theoretical foundation for further functional studies and support molecular breeding initiatives aimed at enhancing oil content in this economically important woody oil species.

Keywords: *Elaeagnus mollis*, Oleosin gene family, phylogenetic analysis, codon usage bias, expression pattern analysis

翅果油树 (*Elaeagnus mollis*) 为胡颓子科 (Elaeagnaceae) 胡颓子属 (*Elaeagnus*) 多年生木本油料植物, 是经第四纪冰川作用后残存下来的中国特有古老孑遗植物, 具有独特的生态适应性与资源价值。翅果油树根系发达, 耐旱、耐瘠薄能力较强, 可自主固氮, 在生长条件较为恶劣的黄土高原地区能够正常生长结实, 具有重要的生态价值。此外, 其种子粗脂肪含量达 46.2% (不饱和脂肪酸占 90% 左右), 出油率为 30% 以上。翅果籽油的理化性质与二级芝麻油、花生油相近, 可直接食用。此外, 其富含多种营养组分, 包括维生素 E、必需氨基酸和植物甾醇等。其中, 维生素 E 含量达到 88–128mg/100g, 约为山茶油的 7~9 倍、橄榄油的 4~6 倍、棕榈油的 3~4 倍、毛椰子油的 2~3 倍。基于其独特的营养成分和良好的食用特性, 翅果籽油具有很高的经济价值与开发潜力 (Liang et al., 2015; Guo et al., 2019; Wang et al., 2019)。

植物油以三酰甘油 (triacylglycerol, TAG) 为主要储存形式, 并特异性地在油体中组装。油体是植物储存脂质的关键细胞器, 其稳定性和油脂代谢效率直接影响种子含油量及品质。它广泛分布于植物的种子、果实、花粉及多种营养器官中 (Murphy, 2012; Huang, 2018)。油体生物合成起始于质体中脂肪酸 (Fatty acid, FA) 的合成。随后, 游离脂肪酸转运至内质网 (endoplasmic reticulum, ER), 在 ER 膜上完成 TAG 的生物合成 (范世航等, 2021)。新合成的 TAG 分子在 ER 双层膜间积累并向外膨出, 形成前体油体, 最终被其特征性的 3 类膜蛋白——油体蛋白 (Oleosin)、油体钙蛋白 (Calciosin) 和油体固醇蛋白 (Steroleosin) 包裹后释放。其中, Oleosin 家族蛋白在成熟油体表面占据主导地位 (约 90%), 是决定油体形态结构和功能的核心成分 (Shao et al., 2019; Scholz et al., 2022)。

Oleosin 作为油体膜结构蛋白的核心组分, 不仅通过 N 端、C 端亲水结构域及中央疏水结构域维持油体结构稳定, 还参与调控脂质代谢、种子萌发及逆境响应等生物学过程。目前, Oleosin 基因家族已在模式植物拟南芥 (*Arabidopsis thaliana*) 和一些木本油料树种, 如油桐 (*Vernicia fordii*) 和核桃 (*Juglans regia*) 等中得到系统鉴定与分析 (Cao et al., 2014; Shimada et al., 2018; Hashemipetroudi & Ahmadi, 2024)。基于系统发育分析, 油体蛋白的进化轨迹可追溯至绿藻向种子植物的适应性分化过程, 其蛋白家族被划分为 5 大谱系: 原始型 (primitive, P)、通用型 (universal, U)、种子低分子量型 (seed low molecular weight, SL)、种子高分子量型 (seed high molecular weight, SH) 及花药绒毡层特异型 (tapetum, T) (Huang & Huang, 2015)。其中, P 谱系作为早期演化分支主要存在于绿藻中, 可能为 U 谱系的进化前体, 而 SL、SH 和 T 谱系则进一步由 U 谱系分化形成。值得注意的是, T 谱系表现出严格的科属特异性, 目前仅在十字花科 (Brassicaceae) 中报道其功能活性 (Fang et al., 2014)。研究表明, 不同物种间 Oleosin 基因家族成员数量及亚型组成存在显著性差异, 且其表达模式与功能可能因组织特异性或代谢需求而发生差异。因此, 系统解析油料植物中 Oleosin 基因家族的成员特征及其调控网络, 对于揭示物种特异性油脂储存机制、筛选关键功能基因具有重要理论意义, 同时可为作物油脂合成能力的遗传改良提供靶点资源。

本研究基于翅果油树全基因组数据，在系统鉴定 *Oleosin* 基因家族成员的基础上，结合不同发育阶段翅果油树种子的转录组数据和 TAG、二酰甘油 (diacylglycerol, DAG) 含量，拟探讨以下问题：（1）翅果油树 *Oleosin* 基因家族的理化特征、系统发育、基因结构、启动子顺式元件以及其密码子使用偏好性等；（2）翅果油树 *Oleosin* 基因家族在种子不同发育阶段中的表达特征及其与 TAG 和 DAG 含量的相关性。研究结果旨在深化对本本植物油脂储存调控网络的理解，为翅果油树高含油量分子育种提供分子基础。

1 材料与方法

1.1 数据获取

翅果油树基因组数据以及基因注释文件获取自己公开发表的文献 (Li et al., 2024)。从 Pfam (<https://www.ebi.ac.uk/interpro/entry/pfam/#table>) 数据库网站下载 *Oleosin* 隐马尔可夫模型，登录号为 PF01277。拟南芥基因组数据下载自 TAIR (<https://www.arabidopsis.org>)，从国家基因库核酸序列归档系统 (<https://ftp.cnbg.org/pub/CNSA/data2/CNP0001846/CNS0383685/CNA0022752/>) 下载胡颓子科沙棘属 (*Hippophae*) 的沙棘 (*Hippophae rhamnoides*) 基因组数据用于构建进化树以及进行共线性分析。

1.2 *Oleosin* 家族成员的鉴定及理化性质

基于隐马尔可夫模型 (Hidden Markov Model, HMM) 的蛋白保守结构域特征，利用 hmmsearch v3.3.2 (Potter et al., 2018) 软件对翅果油树、拟南芥及沙棘的全基因组蛋白序列进行跨物种同源基因检索。为排除随机匹配，保留远缘同源基因，将 E-value 阈值设为 $1e^{-5}$ ，其余参数默认。基于检索获得的蛋白序列借助 Pfam, SMART (<http://smart.embl-heidelberg.de/#>) 和 CDD (<https://www.ncbi.nlm.nih.gov/cdd/>) 数据库进行结构域交叉验证，仅保留同时具有典型 *Oleosin* 保守结构域的成员。基于 Expasy (http://web.expasy.org/compute_pi/) 对翅果油树 *Oleosin* 蛋白的理化性质 (包括氨基酸残基数、分子量等) 进行批量预测。亚细胞定位使用 WoLF PSORT (<https://wolfpsort.hgc.jp/>) 和 Cell-PLoc 2.0 在线工具进行预测 (Chou & Shen, 2010)。

1.3 系统发育树构建

基于 MEGA11.0.13 (Koichiro et al., 2021) 软件对翅果油树、拟南芥和沙棘的 *Oleosin* 蛋白保守结构域序列进行多序列比对，采用最大似然法 (maximum likelihood, ML) 构建系统进化树，Bootstrap 重复为 1 000 次，其他参数默认。根据聚类分析结果以及拟南芥的亚家族分类对翅果油树 *Oleosin* 基因进行亚家族划分，利用 iTOLv6.7.6 (<https://itol.embl.de/>) 在线工具对进化树进行美化。

1.4 染色体定位和共线性分析

使用 MCscanX 获得翅果油树的串联重复对文件，结合翅果油树的基因注释文件，利用 TBtools v2.326 软件可视化翅果油树 *Oleosin* 基因在染色体上的位置 (Wang et al., 2012; Chen et al., 2023)。通过 MCscanX 比对获得翅果油树、拟南芥和沙棘的 *Oleosin* 基因的共线性关系。

1.5 保守基序、基因结构及顺式作用元件分析

基于 MEME (<http://meme-suite.org/tools/meme>) 在线软件对翅果油树 *Oleosin* 家族成员进行保守基序分析，利用 TBtools v2.326 构建基因结构及基序图谱，并提取各基因翻译起始位点上游 2 000 bp 启动子序列，通过 PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 数据库对启动子区顺式作用元件进行系统性扫描与功能注释。

1.6 密码子偏好性分析

利用 CodonW1.4.2 软件分析翅果油树 *Oleosin* 家族的编码序列文件，获取密码子的使用特性参数，包括 T3s、C3s、A3s、G3s 和 GC3s 等。利用生信云平台 (<http://112.86.217.82:9919/#/tool/alltool/detail/261>) 中 CDS 序列密码子不同位置 GC 含量计算工具获得翅果油树 *Oleosin* 基因不同位置上的 G/C 含量以及 3 个位置的 G/C (GC) 总含量和同义密码子相对使用度。依据以上数据，计算出第 1、第 2 位的 G/C 含量 (GC12)，再利用 Origin 2024 软件绘制箱型图。

1.7 表达模式及相关性分析

使用本研究前期获得的 4 个不同发育时期的翅果油树种子：花后 30 d(days after flower, DAF30)、60d(DAF60)、90 d(DAF90)、120 d(DAF120)的转录组和脂质组数据，每个时期 3 次重复(Li et al., 2024)。基于 FPKM (Fragments per kilobase of exon model per million mapped fragments) 定量分析翅果油树种子四个发育阶段中 *Oleosin* 基因的表达动态，通过 TBtools 平台构建基因表达热图。使用 Origin2024 对翅果油树种子 4 个关键发育阶段中 EmoOLE 家族基因的表达模式与 DAG 和 TAG 总含量的动态变化进行相关性分析。

1.8 qRT-PCR 验证分析

利用 qRT-PCR 验证 *EmoOLE3*、*EmoOLE4* 和 *EmoOLE8* 基因的表达特征。种子样品采自西北农林科技大学灌木园。选用 *EmoActin* 基因作为内参基因标准化基因的表达水平。特异性引物使用 Primer Premier 6.0 设计，利用 TBtools 软件再次评估引物特异性，引物序列见表 1。采用 $2^{-\Delta\Delta CT}$ 方法计算 3 个生物重复的基因的相对表达水平。RNA 提取、反应体系及 PCR 程序均与前期研究中的方法保持一致（Gao et al., 2024）。

表 1 qRT-qPCR 引物
Table 1 Primers for the qRT-PCR

基因名称 Gene name	序列 (5'→3') Sequences (5'→3')	
<i>EmoActin</i>	F: ACCGAGGCACCTTTGAACCC	R: ACCGCCTGGATAGCAACATAC
<i>EmoOLE3</i>	F: TTCCTGTTGGTGGGTCGC	R: CAGCAAGGCCGACAGTGA
<i>EmoOLE4</i>	F: GTCTTGCTGTCACCGCCT	R: GTTGATCCGGCACCGACA
<i>EmoOLE8</i>	F: TACTGGCCACACCACTGC	R: ACGCCACAACCACCAGAG

2 结果与分析

2.1 翅果油树 Oleosin 家族基因鉴定及蛋白理化性质分析

全基因筛选结果表明，翅果油树 *Oleosin* 基因家族包含 8 个成员，根据基因在染色体上的顺序将其命名为 *EmoOLE1-EmoOLE8*。EmoOLE 基因家族成员氨基酸大小为 142~183 aa，分子量最小为 14.85 kD，最大为 19.80 kD，等电点为 4.80~9.63。根据不稳定系数可知，EmoOLE 基因家族成员包含 5 个稳定蛋白以及 3 个不稳定蛋白。亲水性平均系数表明，87.5%的 EmoOLE 蛋白为疏水性蛋白，仅 EmoOLE1 蛋白为亲水性蛋白。此外，亚细胞定位预测结果显示翅果油树 *Oleosin* 家族主要分布在细胞膜（cell membrane, CE）上。此外，EmoOLE3 和 EmoOLE8 还分别在叶绿体（chloroplast, CH）和高尔基体（golgi apparatus, GA）上有所分布（表 2）。

表 2 翅果油树 Oleosin 家族成员的理化性质

Table 2 Physicochemical properties of Oleosin in *E. mollis*

基因名称 Gene name	基因 ID Gene ID	氨基酸数目 Amino acid number (aa)	分子量 Molecular weight (kD)	等电点 Isoelectric point	不稳定系数 Instability index	亲水性平均系数 Grand average of hydropathicity	亚细胞定位 Subcellular locations
<i>EmoOLE1</i>	Emo00344	183	19.80	8.92	40.44	-0.126	CE
<i>EmoOLE2</i>	Emo05413	166	18.01	9.56	30.13	0.328	CE
<i>EmoOLE3</i>	Emo06059	174	18.46	9.52	43.37	0.160	CE/CH
<i>EmoOLE4</i>	Emo12300	166	17.82	9.26	44.26	0.146	CE
<i>EmoOLE5</i>	Emo23025	175	18.91	8.93	30.15	0.045	CE
<i>EmoOLE6</i>	Emo23961	178	18.79	4.80	29.68	0.035	CE
<i>EmoOLE7</i>	Emo28186	142	14.85	9.63	30.87	0.447	CE
<i>EmoOLE8</i>	Emo35319	149	15.94	9.41	19.73	0.429	CE/GA

2.2 翅果油树 Oleosin 家族的进化关系分析

为阐明 Oleosin 基因家族的进化保守性，本研究采用跨物种比较基因组学策略，基于隐马尔可夫模型的蛋白结构域筛选方法，利用 HMMER 对翅果油树、拟南芥和沙棘进行基因家族鉴定，分别鉴定出 8、17 和 7 个 Oleosin 基因家族成员。为进一步解析其亲缘关系，基于最大似然法构建系统发育树。根据系统进化分析可将其划分为 4 个亚家族，分别命名为 Group I-Group IV，分别代表 SL、U、SH 和 T 谱系，其中 Group IV 亚家族即 T 谱系仅包含拟南芥。从进化分析上看，翅果油树与沙棘的 Oleosin 基因家族成员紧密聚集在一起（图 1），表明二者 Oleosin 基因家族的进化过程中存在较高的保守性。

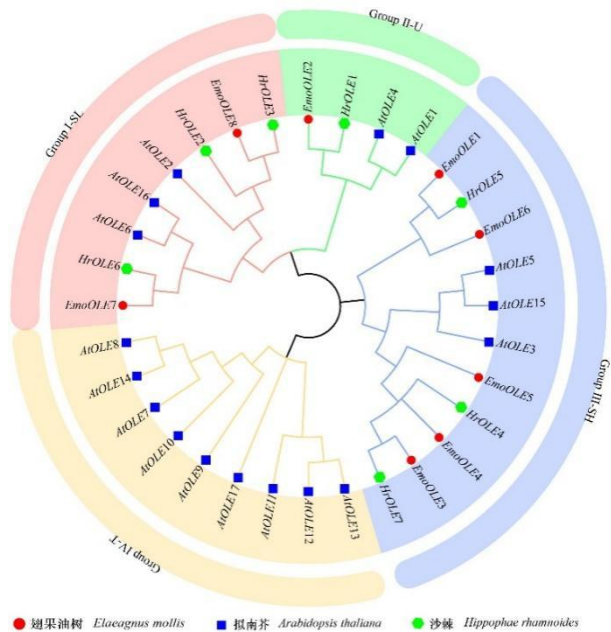


图 1 Oleosin 基因家族系统进化树

Fig. 1 Phylogenetic tree of Oleosin gene family

2.3 翅果油树 Oleosin 家族的系统发育、保守基序、基因结构及顺式作用元件分析

构建 EmoOLE 基因家族的进化树，结果表明 8 个 EmoOLE 基因可分为 3 个亚家族（I、II、III）。其中，III 亚家族中的数量最多，有 5 个，II 亚族中的数量最少，只有 1 个（图 2:A）。保守基序分析表明，所有的 EmoOLE 蛋白都含有多种保守基序，其中 Motif 1、Motif 2 和 Motif 3 是 8 个基因所共有的保守基序（图 2:B）。EmoOLE 基因内含子的数量均只有 1 个，EmoOLE2 只有 1 个外显子，并且其在 5’和 3’两端都没有 UTRs，其余 7 个基因均包含 3 个外显子（图 2:C）。



图 2 EmoOLEs 的系统发育、保守基序和基因结构

Fig. 2 Phylogenetic, conserved motif and gene structure of EmoOLEs

为解析 *EmoOLE* 基因的转录调控机制，对其起始密码子上游 2 000 bp 区域的启动子序列进行系统性顺式作用元件分析。除核心启动子元件（TATA-box 和 CAAT-box）外，共鉴定出 22 类功能各异的调控元件，依据其生物学功能划分为光响应、植物激素响应、生长发育调控以及非生物胁迫响应 4 大类。进一步分析表明，激素响应模块包含 5 种特异性元件，分别参与赤霉素、水杨酸、茉莉酸甲酯、生长素及脱落酸的信号转导途径，5 种激素响应元件的数量存在差异，其中脱落酸响应元件占比最多（图 3）。

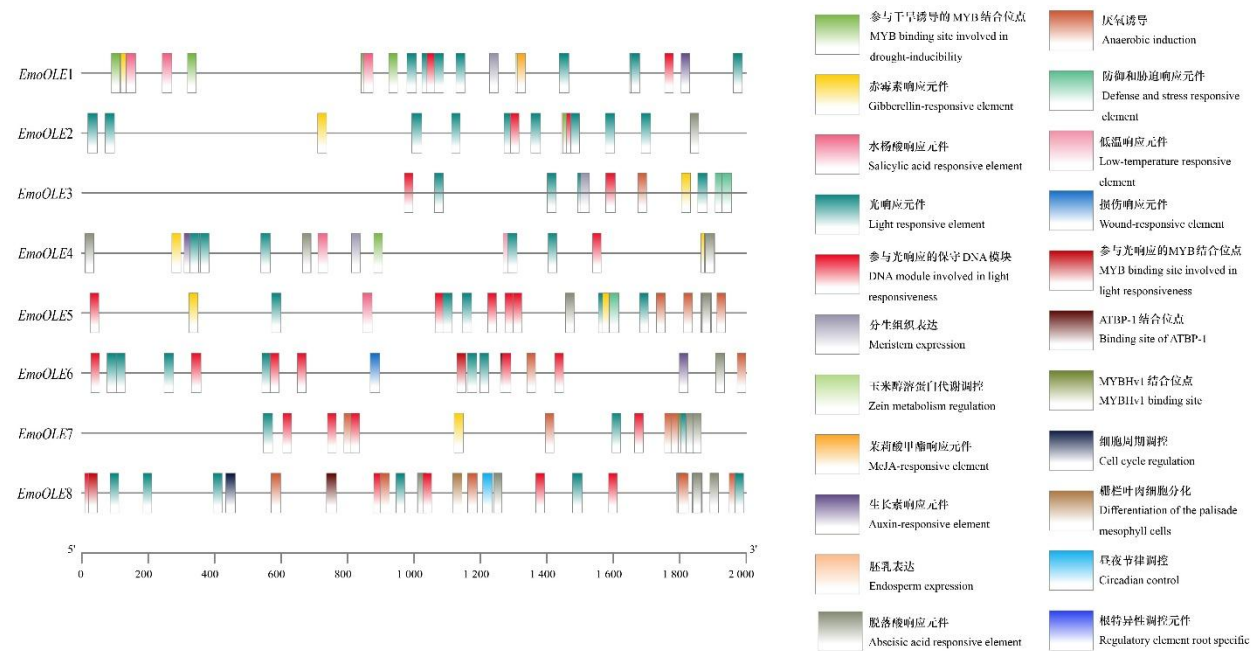


图 3 *EmoOLEs* 的启动子分析
Fig. 3 Promoter regions analysis of *EmoOLEs*

2.4 染色体定位与共线性分析

基于 TBtools 对 *EmoOLE* 基因家族成员进行染色体物理定位分析。结果显示其 8 个成员非均匀分布于翅果油树基因组的 6 条染色体上，第 2 和第 8 号染色体上鉴定到 2 个 *EmoOLE* 基因，第 1、第 4、第 10 以及第 14 号染色体上各分布 1 个 *EmoOLE* 基因。此外，在 5 个基因中观察到片段复制（segmental duplication）或全基因组复制（whole genome duplication, WGD）事件（图 4）。

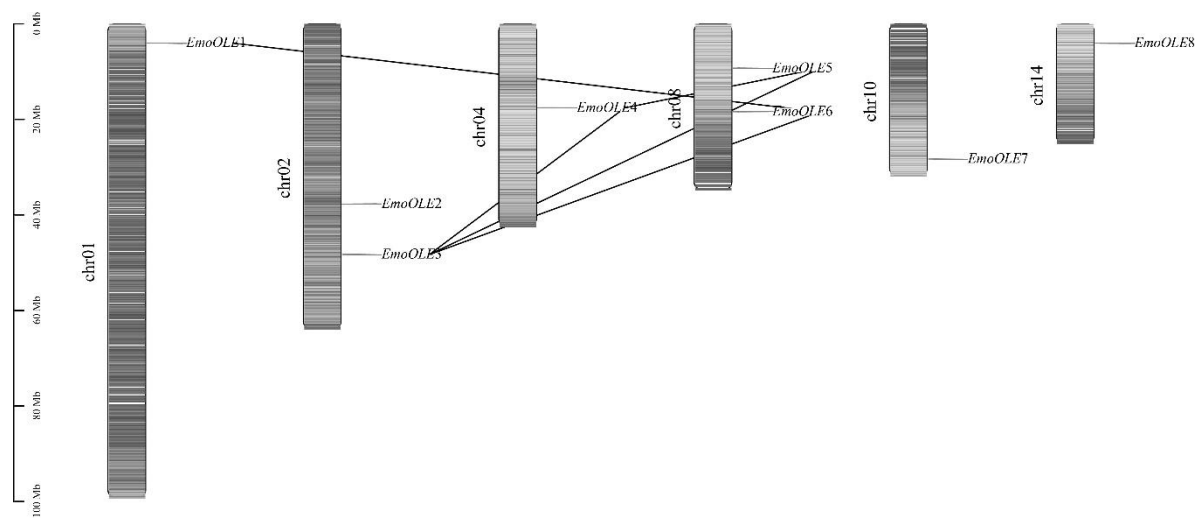


图 4 *EmoOLEs* 染色体定位和基因复制
Fig. 4 Chromosomal localization and gene duplication of *EmoOLEs*

为研究翅果油树 *Oleosin* 基因家族的基因重复事件,使用 MCScanX 软件分析了翅果油树 *Oleosin* 基因与拟南芥、沙棘的共线性关系。结果显示,翅果油树与拟南芥共有 14 对 *Oleosin* 同源基因对,与沙棘共有 15 对 *Oleosin* 同源基因对,进一步分析显示,跨物种同源基因对呈现“一对一”保守型及“一对多”扩张型两种互作模式(图 5)。

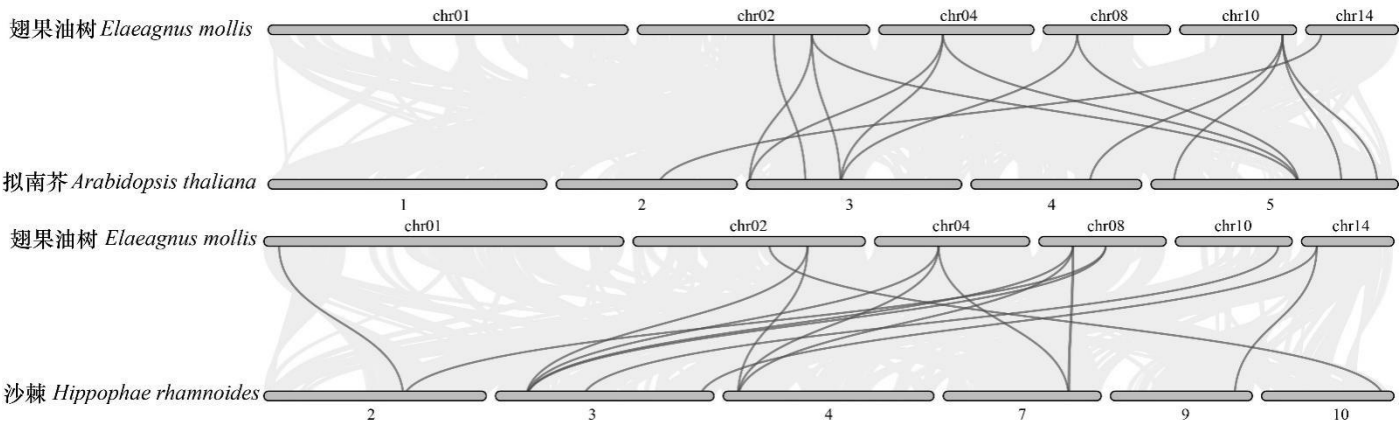


图 5 翅果油树与其他物种 *Oleosin* 基因的共线性分析

Fig. 5 Collinearity analysis of *Oleosin* between *Elaeagnus mollis* and other species

2.5 *EmoOLE* 基因密码子使用偏好性分析

EmoOLE 基因碱基组成分析结果表明,翅果油树 *EmoOLE* 基因家族密码子使用偏好以 A/T 而不是 G/C 碱基结尾,且密码子第 1 和第 2 位的 G/C 碱基含量高于第 3 位。同时,翅果油树 *Oleosin* 基因家族密码子的第 3 位 C3s 的值较高,这也使得其 G/C 碱基含量相对较高(图 6)。

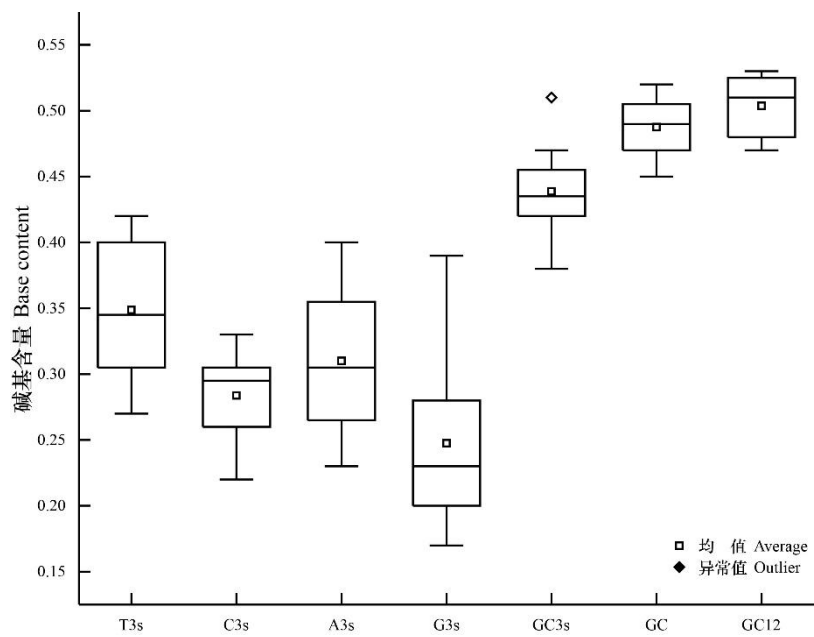


图 6 *EmoOLEs* 碱基组成

Fig. 6 Base composition of *EmoOLEs*

为解析 *EmoOLE* 基因的翻译效率调控特征,本研究通过同义密码子相对使用频率 (relative synonymous codon usage, RSCU) 量化其密码子偏好性。结果显示,在终止密码子外的所有编码子中,27 个密码子呈现显著使用偏好 (RSCU > 1),其中 5 个高偏好密码子 (RSCU > 1.5) 包括 GCU (Ala)、CUC (Leu)、CCA (Pro)、AGG (Arg) 及 UCA (Ser),而脯氨酸对应密码子 CCA 的偏好性尤为突出 (RSCU=2.09),提示其在蛋白疏水结构域形成中可能具有关键作用(表 3)。除终止密码子外,UGC、UGU、AUG 和 UGG 的 RSCU 值等于 1,说明 *EmoOLE* 基因中有 4 个密码子没有偏好性。

表 3 翅果油树 *Oleosin* 基因同义密码子相对使用度

Table 3 RSCU of *Oleosin* gene in *Elaeagnus mollis*

氨基酸	密码子	数目	相对密码子使用度	氨基酸Amino	密码子	数目	相对密码子使用度
Amino acid	Codon	Number	RSCU	acid	Codon	Number	RSCU
丙氨酸 Ala	GCA	34	0.96	脯氨酸 Pro	CCA	34	2.09
	GCC	32	0.91		CCC	8	0.49
	GCG	16	0.45		CCG	12	0.74
	GCU	59	1.67		CCU	11	0.68
半胱氨酸 Cys	UGC	1	1.00	谷氨酰胺 Gln	CAA	53	1.33
	UGU	1	1.00		CAG	27	0.68
天冬氨酸 Asp	GAC	14	0.51	精氨酸 Arg	AGA	12	1.41
	GAU	41	1.49		AGG	15	1.76
谷氨酸 Glu	GAA	22	1.42	丝氨酸 Ser	CGA	6	0.71
	GAG	9	0.58		CGC	5	0.59
苯丙氨酸 Phe	UUC	35	1.30		CGG	2	0.24
	UUU	19	0.70		CGU	11	1.29
甘氨酸 Gly	GGA	19	0.63	苏氨酸 Thr	AGC	11	0.71
	GGC	34	1.12		AGU	15	0.97
	GGG	31	1.02		UCA	29	1.87
	GGU	37	1.22		UCC	7	0.45
组氨酸 His	CAC	13	0.93	缬氨酸 Val	UCG	16	1.03
	CAU	15	1.07		UCU	15	0.97
异亮氨酸 Ile	AUA	25	0.82		ACA	36	1.26
	AUC	32	1.05		ACC	35	1.23
	AUU	34	1.12		ACG	8	0.28
赖氨酸 Lys	AAA	23	0.84	亮氨酸 Leu	ACU	35	1.23
	AAG	32	1.16		GUA	11	0.45
亮氨酸 Leu	CUA	16	0.65		GUC	20	0.82
	CUC	42	1.70		GUG	30	1.24
	CUG	22	0.89	GUU	36	1.48	
	CUU	28	1.14	色氨酸 Trp	UGG	9	1.00
	UUA	17	0.69		酪氨酸 Tyr	UAC	17
	UUG	23	0.93	UAU		26	1.21
蛋氨酸 Met	AUG	21	1.00	*	UAA	3	1.13
天冬氨酸 Asn	AAC	19	1.12	*	UAG	2	0.75
	AAU	15	0.88	*	UGA	3	1.13

注：加粗字体的密码子为具有偏好性的密码子；*：终止密码子。

Note: the codons in bold are those with codon usage bias; *: termination codon.

2.6 *EmoOLE* 基因在不同发育阶段种子中的表达模式分析及 qRT-PCR 验证

为探究 *EmoOLE* 基因在翅果油树种子油脂积累过程中的生物学功能，本研究结合翅果油树种子自身发育特点与油脂含量变化，利用 4 个不同发育阶段的翅果油树种子的转录组数据进行表达模式分析（图 7）。整体上，*EmoOLE* 基因家族在翅果油树种子发育阶段的表达量呈先升高后降低的趋势。其中，*EmoOLE3*、*EmoOLE4*、*EmoOLE5* 和 *EmoOLE6* 在翅果油树种子 DAF90 阶段表达量最高，与翅果油树种子油脂快速积累期相对应，到 DAF120 阶段表达量下降；*EmoOLE1* 和 *EmoOLE7* 在 DAF90 和 DAF120 阶段表达量均相对较高，并未出现明显的下降趋势；而 *EmoOLE8* 的表达量始终为上升趋势，在 DAF120 时期的表达量达到最高。为验证转录组结果的可靠性，利用 qRT-PCR 验证 *EmoOLE3*、*EmoOLE4* 和 *EmoOLE8* 的表达量变化规律；结果表明，qRT-PCR 和转录组展现的基因表达规律趋势具有较好的一致性（图 8）。

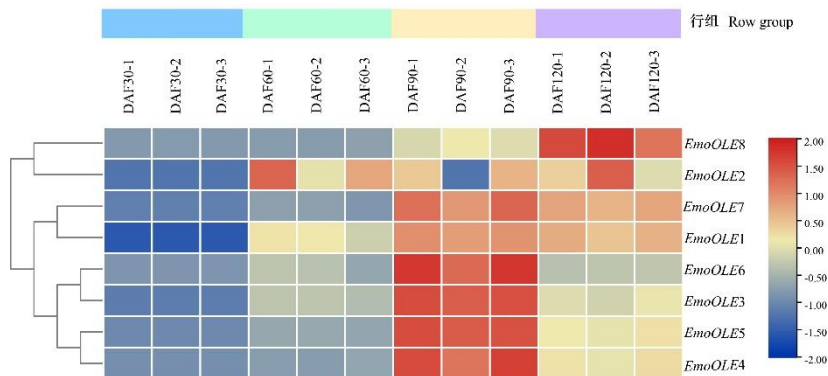


图 7 翅果油树 *Oleosin* 基因在不同发育阶段种子中的表达特征

Fig. 7 Expression profile of *EmoOLE* genes in seeds at different developmental stages

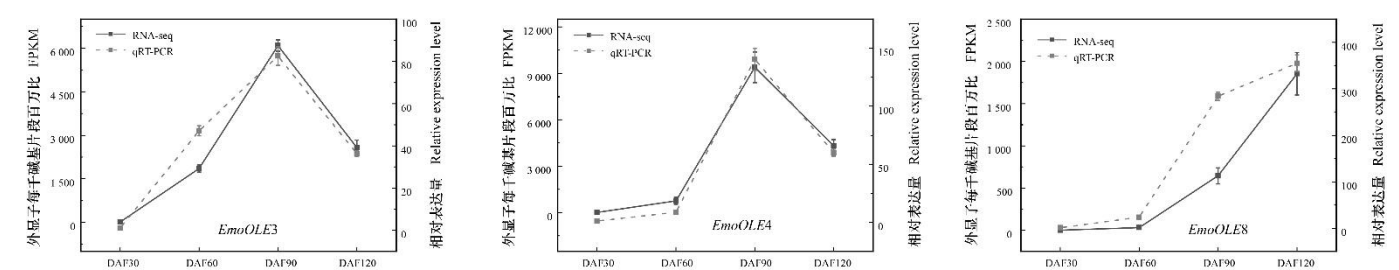


图 8 *EmoOLE* 基因表达量的 qRT-PCR 验证

Figure 8 qRT-PCR analysis of *EmoOLE* genes

为深入解析 *EmoOLE* 基因在脂质代谢中的调控特征，通过 Spearman 相关性分析对翅果油树种子 4 个发育阶段中 *EmoOLE* 家族基因的表达模式与 TAG 和 DAG 总含量的动态变化进行相关性分析。如图 9 所示，6 个 *EmoOLE* 基因（*EmoOLE1*、*EmoOLE3*、*EmoOLE4*、*EmoOLE5*、*EmoOLE6* 和 *EmoOLE7*）的表达水平与 TAG 总含量呈显著正相关，表明这些基因可能协同调控 TAG 的生物合成与贮藏。与之相反，*EmoOLE8* 的表达量仅与 DAG 总含量显著正相关。上述关联模式的显著差异，揭示了 *EmoOLE* 基因家族成员在脂质代谢网络中可能具有功能分化。

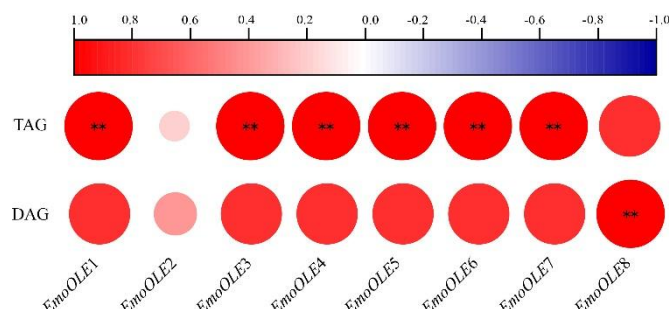


图 9 翅果油树 *Oleosin* 基因表达量与 TAG、DAG 总含量相关性分析

Fig. 9 Correlation analysis between *Oleosin* gene expression and total TAG、DAG content in *Elaeagnus mollis*

3 讨论与结论

Oleosin 基因家族作为调控植物脂质储存的关键因子，在低等绿藻到高等开花植物的进化谱系中均广泛存在（赵浩强等，2022）。该家族通过精确调控油体的生物发生、动态降解及其形态稳定性，在植物油脂储存过程中发挥不可替代的生物学功能（Shao et al., 2019）。尽管历经数百万年演化，不同物种的 Oleosin 蛋白在功能上呈现出适应性分化，但其核心编码序列仍保持着显著的结构保守性。尤为突出的是其疏水核心结构域中特有的“脯氨酸结”基序——这一油体特异性结构特征不仅构成 Oleosin 蛋白的功能核心，更成为其分子鉴定的标志性特征（Huang, 2018）。

本研究基于 Oleosin 保守结构域，最终鉴定出 8 个翅果油树 *Oleosin* 基因，在数量上，少于拟南芥 17 个、油桐 44 个和大豆（*Glycine max*）15 个，与沙棘 7 个相接近（Cao et al., 2014; Shimada et al., 2018; Zhang et al., 2019）。这种数量差异反映了不同植物物种在进化历程中面临的选择压力和基因组进化事件的不同。对于油料作物而言，种子含油量是一个至关重要的经济性状。在大豆中，*GmOLEO1* 基因被证明在驯化过程中受到了强烈的人工选择。过表达该基因不仅能增加油体数量，还能提高籽粒的油分积累，尤其是不饱和脂肪酸的含量（Zhang et al., 2019）。油桐的 *Oleosin* 基因家族显著扩张很可能也与其历史上特定的近期全基因组复制事件有关，这些事件被人工选择所利用，以增强油脂储存能力。相较于大豆、油桐等油料植物，翅果油树、沙棘作为野生资源，其进化更多受自然选择支配且可能缺乏近期的全基因组复制事件，因而限制了 *Oleosin* 基因家族的过度扩张，导致二者的基因家族规模保持在一个相对保守和自然的状态。从进化谱系角度分析，翅果油树与大部分物种一样，*EmoOLE* 基因家族成员主要在 U、SL 和 SH 亚家族中有分布。*Oleosin* 基因家族的系统进化分析揭示其成员在绿藻至种子植物中可分为 5 大亚家族：P、U、SL、SH 及 T 类。其中，P 亚家族作为系统发育的原始类群，特异性分布于绿藻中，被认为是 *Oleosin* 基因的祖先型；T 亚家族则呈现显著的科级特异性，仅存在于十字花科物种中（Huang & Huang, 2015; Huang, 2018）。这种进化格局提示，U/SL/SH 亚家族成员可能在植物界油体形成过程中承担核心生物学功能，且三类蛋白可能通过形成功能复合体协同调控油体动态平衡（Miquel et al., 2014）。

在油料作物中，*Oleosin* 基因的表达水平与油脂积累量呈显著正相关关系（Jia et al., 2022; Zou et al., 2022）。*EmoOLE* 基因表达模式分析表明，翅果油树 *EmoOLE* 基因在种子发育中呈现出动态且分化的表达模式，6 个基因（*EmoOLE1*、*EmoOLE3*、*EmoOLE4*、*EmoOLE5*、*EmoOLE6* 和 *EmoOLE7*）在 DAF90 阶段（对应油脂快速积累期）高表达。这一结果与油体形成的关键时期相契合，即 *Oleosin* 基因的表达水平越高，植物储存油脂的潜力越大（陈晨等，2025）。然而，不同基因的表达峰值存在细微差异，如 *EmoOLE2* 和 *EmoOLE8* 在 DAF60 或 DAF120 阶段高表达，可能反映家族成员在时序调控中的功能多样性和互补性。此外，本研究主要基于生物信息学预测，未来可利用过表达、基因敲除或基因编辑等多维度验证 *EmoOLE* 在油脂含量和油体形态调控中的具体作用。

本研究通过分析 *EmoOLE* 基因家族在翅果油树种子发育阶段的表达模式，加深了对 *EmoOLE* 基因家族功能的认识，也为进一步探索翅果油树种子油脂合成与积累的分子机制提供理论基础。在全基因组范围内共鉴定出 8 个 *Oleosin* 基因，不均匀分布在 6 条染色体上，被划分为 3 个亚族。翅果油树 *Oleosin* 基因家族密码子偏好以 A/T 碱基结尾。基因表达谱和关联分析表明，*EmoOLE* 基因表达与 TAG 或 DAG 积累显著正相关。研究结果为进一步解析 *EmoOLE* 基因家族功能奠定了基础。

参考文献：

- CAO H, ZHANG L, TAN X, et al., 2014. Identification, classification and differential expression of *oleosin* genes in tung tree (*Vernicia fordii*) [J]. PLoS ONE, 9(2): e88409.
- CHEN C, WU Y, LI J, et al., 2023. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining [J]. Molecular Plant, 16(11): 1733-1742.
- CHOU KC, SHEN H B, 2010. Plant-mPLOC: a top-down strategy to augment the power for predicting plant protein subcellular localization [J]. PLoS ONE, 5(6): e11335.

- FAN SH, LIU N, HUA W, 2021. Research advances in the biosynthesis and regulation of lipid in oil crops [J]. Chinese Journal of Oil Crop Sciences, 43(3): 361-375. [范世航, 刘念, 华玮, 2021. 油料作物油脂合成调控研究进展 [J]. 中国油料作物学报, 43(3): 361-375.]
- FANG Y, ZHU RL, MISHLER BD, 2014. Evolution of oleosin in land plants [J]. PLoS ONE, 9(8): e103806.
- FENG XX, LI J, CHEN QQ, et al., 2016. Amino acid composition and nutritional evaluation of proteins extracted from *Elaeagnus mollis* Diels. seed kernels [J]. Food Science, 37(22): 160-165.
- GAO WL, NIE JB, YAO J, et al., 2024. Genomic survey and expression analysis of cellulose synthase superfamily and COBRA-like gene family in *Zanthoxylum bungeanum* stipule thorns [J]. Physiology and Molecular Biology of Plants, 30(3): 369-382.
- GUO C, QIAO J, ZHANG S, et al., 2019. *Elaeagnus mollis* oil attenuates non-alcoholic fatty disease in high-fat diet induced obese mice via modifying the expression of lipid metabolism-related genes [J]. Journal of Oleo Science, 68(9): 893-908.
- HASHEMIPETROUDI SH, AHMADI F, 2024. Identification and expression analysis of Oleosin gene family in walnut (*Juglans regia* L.) [J]. Journal of Plant Molecular Breeding, 12(1): 150-165.
- HUANG AH, 2018. Plant lipid droplets and their associated proteins: potential for rapid advances [J]. Plant Physiology, 176(3): 1894-1918.
- HUANG MD, HUANG AH, 2015. Bioinformatics reveal five lineages of oleosins and the mechanism of lineage evolution related to structure/function from green algae to seed plants [J]. Plant Physiology, 169(1): 453-470.
- JIA Y, YAO M, HE X, et al., 2022. Transcriptome and regional association analyses reveal the effects of oleosin genes on the accumulation of oil content in *Brassica napus* [J]. Plants, 11(22): 3140.
- KOICHIRO T, GLEN S, SUDHIR K, 2021. MEGA11: molecular evolutionary genetics analysis version 11. [J]. Molecular biology and evolution, 38(7): 3022-3027.
- LI CL, ZHANG XZ, GAO WL, et al., 2024. The chromosome-level *Elaeagnus mollis* genome and transcriptomes provide insights into genome evolution, glycerolipid and vitamin E biosynthesis in seeds [J]. International Journal of Biological Macromolecules, 281(P1): 136273.
- LIANG SH, YANG RN, DONG CW, et al., 2015. Physicochemical properties and fatty acid profiles of *Elaeagnus mollis* Diels. nut oils [J]. Journal of Oleo Science, 64(12): 1267-1272.
- MIQUEL M, TRIGUI G, D'ANDREA S, et al., 2014. Specialization of oleosins in oil body dynamics during seed development in Arabidopsis seeds [J]. Plant Physiology, 164(4): 1866-1878.
- MURPHY DJ, 2012. The dynamic roles of intracellular lipid droplets: from archaea to mammals [J]. Protoplasma, 249(3): 541-585.
- POTTER SC, LUCIANI A, EDDY SR, et al., 2018. HMMER web server: 2018 update [J]. Nucleic Acids Research, 46(1): 200-204.
- SCHOLZ P, CHAPMAN KD, MULLEN RT, 2022. Finding new friends and revisiting old ones—how plant lipid droplets connect with other subcellular structures [J]. New Phytologist, 236(3): 833-838.
- SHAO Q, LIU X, SU T, et al., 2019. New insights into the role of seed oil body proteins in metabolism and plant development [J]. Frontiers in Plant Science, 10: 1568.
- SHIMADA TL, HAYASHI M, HARA-NISHIMURA I, 2018. Membrane dynamics and multiple functions of oil bodies in seeds and leaves [J]. Plant Physiology, 176(1): 199-207.
- WANG CX, DUAN ZH, FAN LP, et al., 2019. Supercritical CO₂ fluid extraction of *Elaeagnus mollis* Diels. seed oil and its antioxidant ability [J]. Molecules, 24(5): 911-922.
- WANG Y, TANG H, DEBARRY JD, et al., 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity [J]. Nucleic Acids Research, 40(7): 49.
- ZHANG D, ZHANG H, HU Z, et al., 2019. Artificial selection on *GmOLEO1* contributes to the increase in seed oil during soybean domestication [J]. PLoS Genetics, 15(7): e1008267.

- ZHAO HQ, WANG XF, GAO SP, 2022. Progress on the functional role of oleosin gene family in plants [J]. *Hereditas*, 44(12): 1128-1140. [赵浩强, 王小斐, 高少培, 2022. 植物油体蛋白基因家族研究进展 [J]. *遗传*, 44(12): 1128-1140.]
- ZOU Z, ZHAO YG, ZHANG L, 2022. Genomic insights into lineage-specific evolution of the oleosin family in Euphorbiaceae [J]. *BMC Genomics*, 23: 178.