

# 油莎豆 5-烯醇式丙酮酰莽草酸-3-磷酸合酶基因 *CeEPSPS* 的克隆与分析

邹智<sup>1\*</sup>, 肖艳华<sup>1,2</sup>, 张丽<sup>1,2</sup>, 赵永国<sup>1,3\*</sup>

1. 中国热带农业科学院三亚研究院/中国热带农业科学院热带生物技术研究所/海南省南繁生物安全与分子育种重点实验室, 海南海口 571101; 2. 中南民族大学生命科学院/武陵山区特色资源植物种质保护与利用湖北省重点实验室, 湖北武汉 430074; 3. 广东石油化工学院生物与食品工程学院, 广东茂名 525000

**摘要:** 起源于非洲和地中海沿岸的油莎豆以适应性广、产油量高而成为我国的新型油料作物。为促进该物种的开发与利用, 研究基于基因组和转录组数据对油莎豆及代表性单子叶植物的 *EPSPS* 基因进行系统鉴定。结果表明: 与大多数植物类似, 油莎豆仅含有 1 个 *EPSPS* 基因, 其包含 7 个内含子, 命名为 *CeEPSPS*; RT-PCR 分离到的 *CeEPSPS* 编码区序列为 1584 bp, 预测编码 514 个氨基酸, 其 N 端的前 70 个残基为叶绿体信号肽, 77~508 位残基为高度保守的 EPSP 合酶结构域 (PF00275); 与 EPSP 合酶结构域相比, 信号肽在不同植物中变异较大; *CeEPSPS* 成熟蛋白的理论分子量为 47.32 kDa, 等电点为 5.49, 总平均疏水指数为 0.069, 脂肪族指数为 93.76, 不稳定系数为 31.73, 与其他物种相近, 均属于稳定的疏水型酸性蛋白; 进化分析支持油莎豆划归为禾本目莎草科。序列比对和基因组测序分析结果显示, 56 份油莎豆种质不存在已报道的草甘膦抗性变异。qRT-PCR 分析结果显示, *CeEPSPS* 主要在成熟叶片和块茎中表达, 其表达水平显著高于芽、幼嫩叶片、衰老叶片和芽茎。此外, 本研究还构建了 *CeEPSPS* 的植物过表达载体, 这为下一步的草甘膦抗性育种奠定了坚实的基础。

**关键词:** 油莎豆; 莎草科; 油料作物; 5-烯醇式丙酮酰莽草酸-3-磷酸合酶; 草甘膦

中图分类号: S565.9 文献标识码: A

## Cloning and Characterization of *CeEPSPS*, a Gene Encoding 5-Enolpyruvylshikimate-3-Phosphate Synthase from Tigernut (*Cyperus esculentus* L.)

ZOU Zhi<sup>1\*</sup>, XIAO Yanhua<sup>1,2</sup>, ZHANG Li<sup>1,2</sup>, ZHAO Yongguo<sup>1,3\*</sup>

1. Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences / Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences / Hainan Key Laboratory for Biosafety Monitoring and Molecular Breeding in Off-Season Reproduction Regions, Haikou, Hainan 571101, China; 2. College of Life Science, South-Central University for Nationalities / Hubei Provincial Key Laboratory for Protection and Application of Special Plants in Wuling Area of China, Wuhan, Hubei 430074, China; 3. Guangdong University of Petrochemical Technology, Maoming, Guangdong 525000, China

**Abstract:** Tigernut (*Cyperus esculentus* L.) is a perennial herb that is native to Africa and the Mediterranean coast. Due to the shortage of edible vegetable oil and biodiesel in China, tigernut has been emerging as a promising oil crop for its wide adaptability and high oil production per unit. To promote the development and utilization of this special crop, *EPSPS* (5-enolpyruvylshikimate-3-phosphate synthase) genes, which encode a key enzyme targeted by herbicide glyphosate in the shikimate pathway for the synthesis of aromatic amino acids, were identified from tigernut as well as

收稿日期 2022-04-13; 修回日期 2022-05-16

基金项目 海南省自然科学基金项目 (No. 320RC705); 国家自然科学基金项目 (No. 31971688); 中国热带农业科学院基本科研业务费专项资金 (No. 1630052017011)。

作者简介 邹智 (1982—), 男, 硕士, 研究员, 研究方向: 基因组进化与油脂调控。\*通信作者 (Corresponding author): 邹智 (ZOU Zhi), E-mail: zouzhi2008@126.com; 赵永国 (ZHAO Yongguo), E-mail: zhaoyongguo@gdpu.edu.cn。

several representative monocots based on genome and transcriptome data available. Comparative analyses revealed that (a) tigernut harbors a single *EPSPS* gene (denoted *CeEPSPS*) with seven introns as observed in most plant species; (b) its complete coding sequence of 1584 bp was further isolated using RT-PCR; (c) among 514 amino acids encoded by *CeEPSPS*, the first 70 residues from the N-terminal were characterized as the chloroplast signal peptide, whereas the peptide from sites 77 to 508 belongs to the highly conserved EPSP\_ synthase domain (under the Pfam accession number of PF00275); (d) compared with the EPSP\_ synthase domain, the chloroplast signal peptide is relatively more variable; and (e) the maturation protein of *CeEPSPS* was predicted to possess the theoretical molecular weight (Mw) of 47.32 kDa, the isoelectric point (pI) of 5.49, the grand average of hydropathicity (GRAVY) of 0.069, the aliphatic index (AI) of 93.76, and the instability index (II) of 31.73, which is similar to other reported EPSPSs and could be categorized as a type of hydrophilic, acidic, and stable proteins; (f) phylogenetic analysis using deduced EPSPS proteins supports that tigernut is a Cyperaceae plant within Poales. Further sequence alignment and genome resequencing analyses revealed that no target site resistance mutations to glyphosate were found in all 56 germplasms investigated in this study. Gene expression analysis using qRT-PCR showed that *CeEPSPS* was dominantly expressed in mature leaves and tubers, which were significantly higher than that in buds, young leaves, senescent leaves, and rhizomes. Additionally, a plant vector of over-expressing *CeEPSPS* was also constructed, which would lay a solid foundation for the following molecular breeding for glyphosate resistant in tigernut and other species.

**Keywords:** *Cyperus esculentus* L.; Cyperaceae; oil crop; EPSPS; glyphosate

**DOI:** 10.3969/j.issn.1000-2561.2023.01.004

草甘膦是现代农业生产上应用最广的一类广谱除草剂,其作用靶标为莽草酸途径的 5-烯醇式丙酮酰莽草酸-3-磷酸合酶(5-enolpyruvylshikimate-3-phosphate synthase, EPSPS, EC 2.5.1.19)。草甘膦可竞争性地抑制 EPSPS 而使芳香氨基酸合成受阻,最终导致植物黄化死亡<sup>[1]</sup>。EPSPS 基因于 1983 年首次在鼠伤寒沙门氏菌中得到克隆,后发现其广泛存在于各类微生物和植物中<sup>[2]</sup>。根据进化关系及其对草甘膦的敏感性,微生物源 EPSPS 可分为两大类, I 型对草甘膦敏感, II 型则具有较高的耐受能力,其典型代表是根癌农杆菌 *CP4-EPSPS*, 已被广泛用于作物除草剂抗性育种<sup>[1, 3]</sup>。植物源 EPSPS 按敏感性可归为 I 型,然而,草甘膦的长期施用致使越来越多的杂草 (>30 种) 出现抗性变异,其抗性机制主要表现为 EPSPS 基因扩张和单/多位点突变,主要对应于拟南芥 EPSPS 成熟蛋白的 Thr<sub>102</sub>、Ala<sub>103</sub> 和 Pro<sub>106</sub><sup>[2, 4]</sup>。

油莎豆 (*Cyperus esculentus* L.) 隶属于禾本目莎草科,是一种起源于非洲和地中海沿岸的多年生草本,但在生产上作为一年生作物栽培<sup>[5]</sup>。油莎豆地下结豆(即块茎)、地上长草,其地上部分全为叶片,叶长 80~150 cm,可产鲜草  $3 \times 10^4$  kg/hm<sup>2</sup> 以上,是牛、羊、兔等食草动物的优质饲料;匍匐茎起源的块茎富含淀粉(25%~45%)、油脂(24%~35%)和糖(15%~30%),以及丰富的蛋白质(5%~10%)、膳食纤维(8%~10%)、维生素 C/E(8~14 mg/100 g)和矿物质(如钾、磷、钠、

钙、镁),综合利用价值高,可开发成可口食品、饮料、保健食用油、糖、酱油、醋、酒和饲料等。作为一种新型的油料作物,油莎豆具有产量高(产鲜豆 1.2 万~3 万 kg/hm<sup>2</sup>,折合油脂>100 kg)、适应性广、抗逆性强、适合机械化等特点,这便于其在不挤占现有耕地的情况下利用沙化边际土地增加我国的食用油和饲料原料供给,提高人民的生活质量,并减少对国外大豆的依赖程度,进而服务国家的战略需求<sup>[6]</sup>。然而,随着劳动力成本的日益攀升,草害成为制约油莎豆大面积推广的重要因素,因此培育耐除草剂的新品种具有积极的现实意义。本研究对油莎豆的 EPSPS 基因进行克隆,并在此基础上重点比较分析其序列特征、遗传变异、进化及表达特性,以期为今后的开发与利用奠定基础。

## 1 材料与amp;方法

### 1.1 材料

1.1.1 植物材料 供试材料为‘热研 1 号’及另外 55 份油莎豆种质<sup>[7]</sup>。用于 DNA 提取的幼嫩叶片直接采集于中国热带农业科学院热带生物技术研究所文昌试验基地,而用于 RNA 提取的‘热研 1 号’参照邹智等<sup>[8]</sup>的方法,在采样前 10 d 移栽至海口温室,然后单独采集幼嫩叶片、成熟叶片、衰老叶片、芽、芽茎(即分化产生芽的匍匐茎)和块茎等组织样本。

1.1.2 菌株及载体 感受态大肠杆菌 DH5 $\alpha$  和植物

表达载体 *pCAMBIA1301* 由本实验室制备和保存。

1.1.3 主要试剂 引物合成和常规 DNA 测序委托生工生物工程（上海）股份有限公司完成；Omega 快速质粒小提试剂盒、Omega DNA 纯化回收试剂盒、诺唯赞 ClonExpress II One Step Cloning Kit (C112-02) 试剂盒、宝生物高保真 DNA 聚合酶、普洛麦格 DNase I 及各类分析纯生

化试剂和实验耗材均购自相应的试剂公司。

## 1.2 方法

1.2.1 *EPSPS* 基因的鉴定 从 Araport11 (表 1) 下载拟南芥的 *AtEPSPS* (AT2G45300) 和 *AtEPSPS2* (AT1G48860) 基因, 并以其蛋白序列作为种子 tBLASTn 搜索代表性植物的基因组 (表 1) 和油莎豆的全长转录组<sup>[8]</sup>。

表 1 生物信息学分析所用数据库、软件及其网址

Tab. 1 Databases, software and related websites used for bioinformatics analysis

数据库/软件 Database/software	网址 Website	分析内容 Content for analysis
Araport 11	<a href="https://www.arabidopsis.org/">https://www.arabidopsis.org/</a>	拟南芥 <i>EPSPS</i> 基因下载
Phytozome v12	<a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a>	基因组数据下载
NCBI	<a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a>	基因组数据下载
NCBI-BLAST	<a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi">https://blast.ncbi.nlm.nih.gov/Blast.cgi</a>	核苷酸和蛋白序列比对
ORFfinder	<a href="https://www.ncbi.nlm.nih.gov/orffinder/">https://www.ncbi.nlm.nih.gov/orffinder/</a>	开放读码框查找
ProtParam	<a href="https://web.expasy.org/protparam/">https://web.expasy.org/protparam/</a>	蛋白理化特性分析
ProtScale	<a href="http://web.expasy.org/cgi-bin/protscale/">http://web.expasy.org/cgi-bin/protscale/</a>	亲水/疏水性预测
ChloroP 1.1	<a href="http://www.cbs.dtu.dk/services/ChloroP/">http://www.cbs.dtu.dk/services/ChloroP/</a>	亚细胞定位预测
TMHMM 2.0	<a href="http://www.cbs.dtu.dk/services/TMHMM">http://www.cbs.dtu.dk/services/TMHMM</a>	跨膜结构预测
SMART	<a href="http://smart.embl-heidelberg.de/">http://smart.embl-heidelberg.de/</a>	保守结构域预测
InterPro	<a href="http://www.ebi.ac.uk/interpro/search/">http://www.ebi.ac.uk/interpro/search/</a>	生物学功能预测
MEGA 6.0	<a href="https://mega6.software.informer.com/">https://mega6.software.informer.com/</a>	多序列比对和进化树的构建

1.2.2 总 RNA 提取与 cDNA 第一链的合成 不同组织的总 RNA 采用天根植物多糖多酚 RNA 提取试剂盒单独提取, 经纯度、浓度和完整性检测合格后, 用普洛麦格 DNase I 清除残存的 DNA, 并用赛默飞反转录试剂盒合成 cDNA 第一链。

1.2.3 基因克隆 基于转录组获得的 *CeEPSPS* 转录本序列, 采用 Primer Premier 5.0 设计引物对 *CeEPSPSF/R* (GTCACCTCCACTGCATTTCTT/ACTTTGATACAAGCAATATGC) 和 *CeEPSPSHF/R* (CACGGGGGACTCTTGACCATGGCGCAAGCGACCATGGCC/CTGGTCACCTGTAAATCACACCTAATATTTAGCAAACTTCTG), 下划线为可与 *pCAMBIA1301* 重组的同源臂), 分别用于基因克隆和植物表达载体的构建。基因克隆参照邹智等<sup>[9]</sup>的方法, PCR 产物用 1% 琼脂糖凝胶电泳检测; 为进一步完成植物过表达载体的构建, 以稀释 100 倍的第一轮 PCR 产物作为模板, 利用 *CeEPSPSHF/R* 进行第二轮 PCR, 产物电泳检测后切胶回收目的条带; 同时, 用 *Nco* I 和 *Pml* I 于 37°C 酶切 *pCAMBIA1301*

12 h, 产物经 0.8% 琼脂糖凝胶电泳检测后回收目的条带; 然后, 将上述线性化载体和 PCR 胶回收产物按 1:2 比例混合, 并用 *Exnase* II 于 37°C 催化 2 h, 构建载体 *pCAMBIA1301-CeEPSPS*。重组质粒转化 DH5 $\alpha$  感受态细胞后, 参照邹智等<sup>[9]</sup>的方法进行后续的菌落 PCR、测序及质粒的提取。

1.2.4 生物信息学分析 参照肖艳华等<sup>[7]</sup>的方法, 分析 *EPSPS* 成熟蛋白的理论分子量 (Mw)、等电点 (pI)、总平均疏水指数 (GRAVY)、脂肪族指数 (AI)、不稳定系数 (II) 等, 所用软件及其网址详见表 1。

1.2.5 基因表达分析 荧光定量分析具体参照邹智等<sup>[10]</sup>的方法, 以 *18S rRNA* 作为内参, 引物 *18SF/R* (TGTGATGCCCTTAGATGTTCTGG/GACGTAGTCAACGCGAGCTGA) 和 *CeEPSPSFq/Rq* (CGAGTAGGGGTGCTGTTGTCA/TGCGATTGAGAGGGATTAG), 每样品 3 次生物学重复; 用  $2^{\Delta\Delta C_T}$  法和 SPSS 软件分别进行基因相对表达分析和方差分析。

## 2 结果与分析

### 2.1 *CeEPSPS* 基因的克隆

通过搜索团队前期获得的油莎豆全长转录组文库<sup>[8]</sup>共得到 2 条同源转录本，其长度为 1896、2144 bp，分别包含 1545、1344 bp 的开放读码框（ORF）。前者的序列如图 1 所示，其编码的蛋白序列和长度与其他植物类似，将该基因命名为 *CeEPSPS*；后者则缺失了正常的 C 端。将序列比对到本地的油莎豆基因组，结果发现 *CeEPSPS* 含有 7 个内含子，而上述较长的转录本为第 6 内含子滞留的可变剪接形式。

为克隆 *CeEPSPS* 基因，在基因的转录非翻译区（UTR）设计引物对 *CeEPSPSF/R*，以‘热研 1 号’反转录的 cDNA 作为模板，首轮 PCR 成功扩增得到一条约 1853 bp 的特异条带；然后以上述 PCR 产物作为模板，*CeEPSPSHF/R* 作为引物，第二轮 PCR 扩增得到一条约 1584 bp 的目标条带，将该条带切胶回收，再通过同源重组法克隆到 *pCAMBIA1301* 上。测序结果显示，分离到的 CDS 与全长转录本的相应区域完全一致，并与 NCBI 中释放的 4 条序列（登录号为 KT757586.1、KT757587.1、KM052384.1 和 KM052385.1）的一致性分别为 100.00%、98.30%、98.50%和 98.60%。

```

1 gtcacctccactgcatcttctcacctaccaccatctgctctctcttcttacttctgacttaccacctccacttccacactaccaacc
91 ccaattgagagtccaattcccaccaaatctgacttcttggaccaggaaccctcacaacagaaacacaacaacaccagctgtcaccgat
1 M A Q A T M A K G L E P R I Y V S G S N L S R S R A Q A V P
181 cgATGGCGCAAGCGACCATGGCCAAAGGGGCTGAGCCAGGATCTACGTCTCCGGCTCAAACCTCTCCAGATCTCGCGCCAGGCGCTCC
31 I A S S L R F F K S G S L R C R N G G S A S S R G A V V R A
271 CCATCGCTTCTTCGCTGCGATTTTCAAATCGGGGTCGCTGAGGTGCAGGAATGGGGGCTCGGCGTCGAGTAGGGGTGCTGTGTGTCAGGG
61 S V A A A P E K P S T V P E I V L Q P I K E I S G T I N L P
361 CTTCTGTCGCGCGCTGCTCCCGAGAAACCGTCTACCGTGCCGGAGATTGTGCTCCAGCCCATCAAGGAGATTTCTGGGACCACTCAACTGC
91 G S K S L S N R T L L L A A L S E G T T V V D N L L S S D D
451 CTGGATCTAAATCCCTCTCCAATCGCACGCTTCTCCTTGGCCGACTCTCTGAGGGAACAACCTGTGTGGATAACCTGTTAAGCAGCGACG
121 I H Y M L E A L K T L G L S V E V D K E A K R A V V A G S G
541 ACATACACTACATGCTCGAGGCTTTGAAAACGCTTGGCCCTTCTGTAGAGGTGGACAAAGAGGCTAAAAGAGCAGTCGTTGCGGGGCTG
151 G Q F P A G K E S K E V E L F L G N A G T A M R P L T A A L
631 GCGGTCACTCCAGCTGGTAAAGAATCAAAGGAGGTGGAACCTTCTTAGGTAATGCTGGCACTGCCCCCTTTGACTGCTGCAT
181 T V A G G H G S Y V L D G V P R M R E R P I G D L V V G L K
721 TGACTGTCGCTGGTGGACATGGAAGTTATGTGCTTGTATGGAGTCCACGCATGAGGGAAGACCAATAGGTGATTTGGTGTGGGCTTAA
211 Q L G A D V D C F L G T N C P P V R V R G I G G L P G G K V
811 AACAACTGGCGCTGATGTGGATTGTTTTTGGGTACCAACTGCCACCACTTCGTGTAGAGGCATTGGCGCCCTTCTGGTGGAAAGG
241 K L S G S I S S Q Y L S A L L M A G P L A L G D I E I E I I
901 TTAAGCTGTCGGGTTCTATTAGTAGTCAGTACTTGAGTGCATTGCTCATGGCTGGCCCACTTGCCCTAGGAGATATCGAAATAGAATCA
271 D K L I S V P Y V E M T I K L M E R F G V S V E H T D S W D
991 TCGATAAATAATATCTGTCCTTACGTGGAATGACCATAAAATTAATGGAACGCTTTGGTGTCACTGTTGAGCATACTGACAGCTGGG
301 R F F I K G G Q K Y R S P G K A F V E G D A S S A S Y F L A
1081 ATCGCTTCTTCATTAAGGTTGGTCAAGTACAGGTCACTGGAAAGGCTTTCGTAGAAGGTGATGCATCCAGTCCGAGTTATTTCTTGG
331 G A A I T G G T V T V E G C G T A S L Q G D V K F A E V L E
1171 CTGGTGTGCTATAACTGGTGTACAGTCACTGTAGAAGGCTGTGGTACAGCCAGCTTACAGGGAGATGTGAAATTTGCTGAACTACTTG
361 K M G A K V T W T E N S V T V T G P P R G S S F R E K H L K
1261 AGAAAATGGGAGCAAAGTTACATGGACTGAGAATAGTGTAACTGTCACTGGTCCACCACGTTGGTTCATTTCAGAGAGAAACACTTGA
391 A V D V N M N K M P D V A M T L A V V A L F A D G P T A I R
1351 AAGCTGTTGATGTGAACATGAACAAGATGCCAGATGTTGCCATGACCTTAGCAGTCGTCGCACCTTTTGTGATGGCCCTACTGCTATCA
421 D V A S S W R V K E T E R M V A I C T E L R K L G A T V E E G
1441 GAGATGGCTTTGGAGAGTGAAGGAGACTGAGAGAATGGTTGCCACTTTGCCACTGAAGAAAGTTGGTGGTCAACCGGTAAGAGAAG
451 P D Y C I I T P P E K L N I A A I D T Y D D H R M A M T F S
1531 GGCCGACTATTGCATCATCACTCCACCTGAGAAGCTAAACATAGCAGCAATAGACACCTATGATGACCATAGAATGGCCATGACATTTT
481 L A A C A E V P V T I N D P G C T R K T F P D Y F D V L Q K
1621 CTCTAGCTGCATGTGCTGAAGTGCCGGTAACTATAAATGACCCCTGGTGTGCACACGCAAAACCTTCCCGGATTACTTCGATGTGCTCAGA
511 F A K Y *
1711 AGTTTGCTAAATATTAGcaaaattgaggagttctttttccctaattgtaaccaagaatttaaactaggaggaaaattatattgtgtctc
1801 tatgagtaataaattataggaaggtgaattttgcatattgcttgtatcaagtgttatcagattggcaataatattgttcatatttgag

```

大写碱基字母表示基因编码区，小写碱基字母表示 5'UTR 和 3'UTR，单下划线表示基因克隆引物 *CeEPSPSF/R*，双下划线表示同源重组引物 *CeEPSPSHF/R* 的基因特异性序列。

Uppercase letters represent sequences of the coding region, lowercase letters represent sequences of 5'UTR and 3'UTR, the primer pair *CeEPSPSF/R* for gene isolation is marked with the single underline, and partial sequences of the primer pair *CeEPSPSHF/R* are marked with the double underline.

图 1 *CeEPSPS* 的全长 cDNA 及其推测编码的氨基酸

Fig. 1 Full-length cDNA sequence of *CeEPSPS* and its deduced coding amino acids

### 2.2 代表性单子叶植物中 *EPSPS* 基因的鉴定

为揭示基因在单子叶植物中的进化特征，研究基于康藏蒿草 (*Carex littledalei*) 等植物的全基

因组序列对 *EPSPS* 基因进行系统鉴定。结果显示，*EPSPS* 在研究的 13 种代表性单子叶植物中多以单拷贝的形式存在，约占 69.23%；含有 2 个拷贝的

分别为椰枣 (*Phoenix dactylifera*)、油棕 (*Elaeis guineensis*)、菠萝 (*Ananas comosus*) 和香蕉 (*Musa acuminata*)，其在核酸和蛋白水平的一致性分别为 99.40%/99.40%、90.20%/89.70%、100.00%/100.00%、88.40%/90.90%，接近或略高于拟南芥的 89.60%/89.30%；拟南芥和 15 种单子叶植物中的 *EPSPS*

基因均含有 7 个内含子 (表 2)。此外，从 NCBI 中还搜索到 4 条香附子 (*C. rotundus*) 的 *EPSPS* 完整或部分编码区序列，即 KT757585.1 (1545 bp)、KM052381.1 (1530 bp)、KM052382.1 (1474 bp) 和 KM052383.1 (1477 bp)，其与 *CeEPSPS* 的一致性分别为 97.20%、97.50%、97.40%和 97.10%。

表 2 拟南芥和 15 种单子叶植物中 *EPSPS* 基因的结构及其编码蛋白的理化特性

Tab. 2 Gene structures and protein physicochemical properties of *EPSPS* genes in arabidopsis and 15 monocots

基因 Gene	内含子数量 Number of introns	氨基酸数量 Number of amino acids	信号肽长度 Length of signal peptide	分子量 Mw/kDa	等电点 pI	总平均疏 水指数 GRAVY	脂肪族 指数 AI	不稳定 系数 II
<i>AtEPSPS</i>	7	520	76	47.55	5.15	0.011	93.76	32.79
<i>AtEPSPS2</i>	7	521	77	47.47	5.22	0.032	94.21	32.64
<i>PeEPSPS</i>	7	527	83	47.27	5.22	0.077	94.41	35.46
<i>PdEPSPS1</i>	7	522	78	47.40	5.46	0.047	92.68	31.44
<i>PdEPSPS2</i>	7	522	78	47.31	5.56	0.054	92.68	30.86
<i>EgEPSPS1</i>	7	522	78	47.32	5.30	0.080	94.41	32.18
<i>EgEPSPS2</i>	7	522	78	47.37	5.48	0.057	93.54	29.90
<i>SpEPSPS</i>	7	456	13	46.86	5.13	0.111	95.96	30.47
<i>AcEPSPS1</i>	7	520	76	47.22	5.30	0.038	91.78	32.54
<i>AcEPSPS2</i>	7	520	76	47.22	5.30	0.038	91.78	32.54
<i>MaEPSPS1</i>	7	518	74	47.28	5.58	0.040	91.80	30.87
<i>MaEPSPS2</i>	7	518	74	47.31	5.49	0.041	92.25	28.38
<i>OsEPSPS</i>	7	515	71	47.41	5.51	0.064	94.86	31.20
<i>HvEPSPS</i>	7	503	59	47.30	5.15	0.094	94.84	30.31
<i>BdEPSPS</i>	7	510	66	47.43	5.38	0.065	93.96	29.94
<i>SiEPSPS</i>	7	511	67	47.29	5.52	0.049	93.78	29.68
<i>SbEPSPS</i>	7	506	63	47.21	5.40	0.087	95.08	30.24
<i>ZmEPSPS</i>	7	506	63	47.23	5.31	0.072	94.20	31.21
<i>ClEPSPS</i>	7	516	72	47.17	5.21	0.065	93.31	32.03
<i>CrEPSPS</i>	-	514	70	47.26	5.31	0.082	94.19	31.26
<i>CeEPSPS</i>	7	514	70	47.32	5.49	0.069	93.76	31.73

### 2.3 *CeEPSPS* 基因的生物信息学分析

2.3.1 *CeEPSPS* 基因编码蛋白的理化特性、亚细胞定位及保守结构域分析 如图 1 所示，*CeEPSPS* 基因编码区的 GC 含量为 49.64%，预测编码 514 个氨基酸，其中含量较高的为 Ala (10.30%)、Gly (9.70%)、Val (9.10%)、Leu (8.90%) 和 Ser (8.20%)；强酸性、强碱性、极性和疏水氨基酸分别占 10.70%、11.09%、22.18% 和 37.16%。亚细胞定位及与拟南芥中的 *AtEPSPS* 蛋白比对分析显示，*CeEPSPS* 属于叶绿体定位蛋白，其 N 端的前 70 个残基为信号肽，长度与所分析的多数物种相近；值得注意的是，紫萍 (*Spirodela polyrhiza*) 中的 *SpEPSPS* 蛋白的信号

肽仅有 13 个残基，由碱基插入与缺失造成，其是否足以介导叶绿体定位还有待研究。*EPSPS* 成熟蛋白的理论分子量、等电点、总平均疏水指数、脂肪族指数、不稳定系数在所分析的物种中均非常接近，暗示其均为稳定的疏水型酸性蛋白 (表 3)。TMHMM 和 SMART 分析显示，*CeEPSPS* 无跨膜螺旋，其 77~508 位为 EPSP 合酶结构域 (PF00275, *E* 值 2.7e-150) (图 2)；InterPro 分析显示，其主要参与芳香族氨基酸的生物合成 (GO:0009073)，具有 3-磷酸莽草酸-1-羧基乙烯基转移酶活性 (GO:0003866)、甲基以外的芳烃基转移酶活性 (GO:0016765) 以及催化活性 (GO:0003824) 等分子功能。

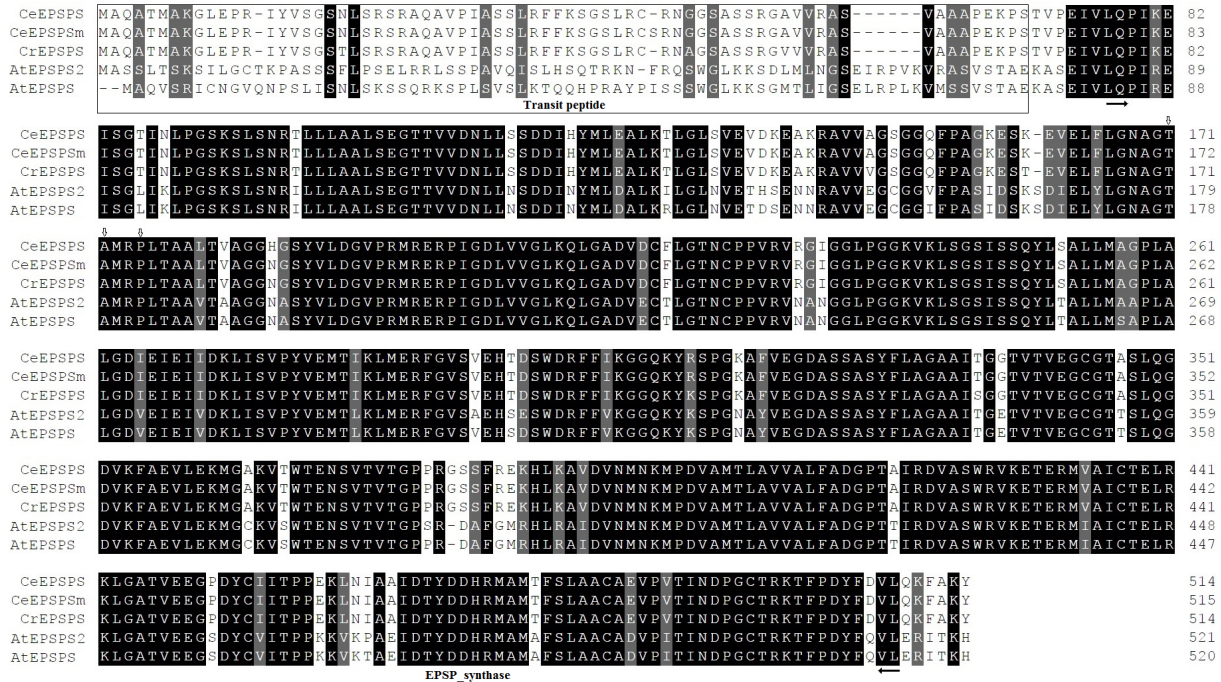


图 2 *CeEPSPS* 蛋白的多序列比对和序列特征

Fig. 2 Multiple sequence alignment and sequence features of *CeEPSPS*

2.3.2 SNP 分析 如图 2 所示，可引起草甘膦抗性改变的 Thr<sub>102</sub>、Ala<sub>103</sub> 和 Pro<sub>106</sub> 在所比对的蛋白间高度保守，暗示‘热研 1 号’的 *CeEPSPS* 蛋白可能对除草剂敏感。为摸清团队所收集的其他种质是否存在抗性突变，我们利用前期获得的基因

组重测序数据（即将 56 份材料的叶片 DNA 等量混合后测序）<sup>[7]</sup>进行多态性位点鉴定，共计获得 20 个单核苷酸多态性（SNP）位点（表 3）；经深入分析显示，它们仅造成 3 个氨基酸的变异，即对应于成熟蛋白的 His<sub>116</sub>Asn、Leu<sub>139</sub>Val 和 Asn<sub>153</sub>Asp。

表 3 SNP 分布情况

Tab. 3 Distribution of SNPs identified in this study

位点 Site	30	153	210	354	537	556	603	625	639	667	810	867	1008	1218	1224	1419	1464	1515	1524	
<i>CeEPSPS</i>	G	G	T	C	A	C	A	T	C	A	C	T	T	A	C	C	T	G	C	T
SNP	A	A	C	T	G	A	G	C	T	G	T	C	C	G	T	T	C	A	T	A

2.3.3 进化分析 为揭示 EPSPS 的进化特征及油莎豆的分类学地位，研究构建了包括拟南芥 EPSPS 在内 21 个蛋白的无根进化树。如图 3 所示，*AtEPSPS* 和 *AtEPSPS2* 作为唯一的双子叶植物来源蛋白位于进化树的基部，而单子叶植物来源蛋白较好地聚为 7 支，其中，同为莎草属的 *CeEPSPS* 和 *CrEPSPS* 聚在一起，并与同为莎草科的 *CIEPSPS* 形成一个分支；同为禾本目但隶属于禾本科的 *OsEPSPS*、*HvEPSPS*、*BdEPSPS*、*SiEPSPS*、*SbEPSPS* 和 *ZmEPSPS* 聚为莎草科的姊妹支；棕榈目棕榈科的 *PdEPSPS1/2*、*EgEPSPS1/2* 聚为一支；禾本目凤梨科的 *AcEPSPS1/2*、芭蕉目芭蕉科

的 *MaEPSPS1/2*、微子目兰科的 *PeEPSPS* 和天南星目浮萍科的 *SpEPSPS* 则形成另外的 5 个分支。由于来自同一物种的旁系同源蛋白都聚在一起，并具有很高的序列相似性（89.3%~100.0%），暗示其可能通过物种特异性的基因重复产生。

### 2.4 *CeEPSPS* 基因的表达特性分析

为揭示 *CeEPSPS* 的表达特性，并为草甘膦的适时施用提供参考，研究采用 qRT-PCR 技术检测基因在叶片（分幼嫩、成熟和衰老 3 个时期）、芽、芽茎和块茎等主要组织中的表达水平。如图 4 所示，*CeEPSPS* 在成熟叶片和块茎中的表达丰度最高，显著高于芽、幼嫩叶片、衰老叶片和芽茎；

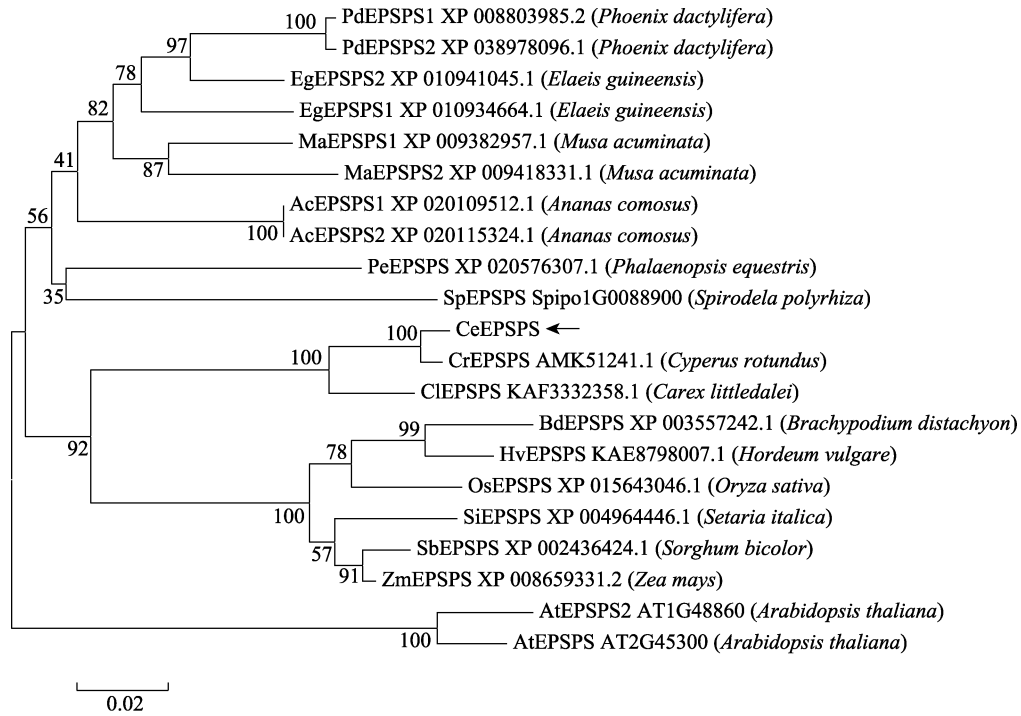
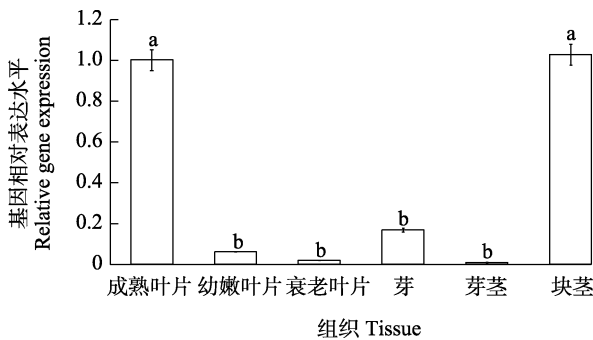


图 3 EPSPS 蛋白的进化分析

Fig. 3 Phylogenetic analysis of EPSPS proteins



不同小写字母表示差异显著 ( $P < 0.05$ )。

Different lowercase letters indicate significant difference ( $P < 0.05$ ).

图 4 *CeEPSPS* 在不同组织中的表达模式Fig. 4 Expression patterns of *CeEPSPS* in various tissues

从叶片的发育过程来看, 基因呈现先增后减的趋势, 其在成熟叶片中的表达水平显著高于幼嫩和衰老叶片, 而幼嫩和衰老叶片之间差异不显著。

### 3 讨论

作为莽草酸途径的关键酶, EPSPS 催化烯醇式丙酮酸和莽草酸-3-磷酸形成 EPSP, 继而参与酪氨酸、苯丙氨酸、色氨酸等芳香族氨基酸的生物合成, 以及香豆素、绿原酸、类黄酮等次生代谢物的合成。同时, EPSPS 又是广谱型除草剂草甘膦的作用靶标<sup>[2]</sup>。虽然绝大多数植物的 EPSPS 对草甘膦高度敏感, 但高选择压引起的基因扩张

及关键位点的突变赋予了部分群体或生态型的抗药性, 这些位点包括 Thr<sub>102</sub>、Ala<sub>103</sub> 和 Pro<sub>106</sub><sup>[2,4,11-12]</sup>。Pro<sub>106</sub> 突变是抗性获得的常见形式, 至今已在 10 种以上的杂草中发现, 如 Pro<sub>106</sub>Ser/Thr/Ala/Leu, 虽然此类突变多赋予低水平的抗性 (<10 倍), 但 Pro<sub>106</sub>Ser 突变在黑麦草、长芒苋和 *Leptochloa virgata* 等植物中也发现了高水平的抗性 (>10 倍); Thr<sub>102</sub>Ile 突变一般赋予高水平的抗性, 但 Thr<sub>102</sub>Ser 突变仅赋予低水平的抗性; 与单突变相比, 多位点突变如 Thr<sub>102</sub>Ile+Pro<sub>106</sub>Ser/Ala/Thr (TIPS/A/T)、Thr<sub>102</sub>Ile+Ala<sub>103</sub>Val+Pro<sub>106</sub>Ser 可赋予更高水平的草甘膦抗性<sup>[4, 13-16]</sup>。其中, 编码 TIPS 突变的 EPSPS 基因已成功用于玉米 (*Zea mays*) 除草剂抗性育种<sup>[17]</sup>。在油莎豆中, 至今还未见有关于草甘膦抗性变异的报道。

本研究利用基因组和转录组数据对油莎豆及另外 13 种代表性单子叶植物的 EPSPS 基因进行了鉴定, 并在此基础上完成了 *CeEPSPS* 编码区序列的分离及植物过表达载体的构建。这些单子叶植物包括莎草科的康藏嵩草, 禾本科的水稻 (*Oryza sativa*)、大麦 (*Hordeum vulgare*)、二穗短柄草 (*Brachypodium distachyon*)、谷子 (*Setaria italica*)、高粱 (*Sorghum bicolor*) 和玉米, 凤梨科的菠萝, 棕榈科的椰枣和油棕, 芭蕉科的香蕉,

兰科的小兰屿蝴蝶兰, 以及浮萍科的紫萍。结果显示, 与拟南芥一样, 这些植物的 *EPSPS* 基因均含有 7 个内含子; 多数植物以单拷贝的形式存在, 而椰枣、油棕、菠萝和香蕉中各自存在 1 对重复基因, 主要通过物种特异性的基因重复产生 (由于菠萝 *AcEPSPS1* 和 *AcEPSPS12* 的编码区序列完全一样, 不排除是基因组错误组装所致); 进化树依科将所分析的蛋白分成 9 支, 其中, 属于双子叶的拟南芥位于基部, *CeEPSPS* 与 *CrEPSPS* 的亲缘关系最近, 并与 *CIEPSPS* 形成莎草科分支; 莎草科分支与禾本科分子形成姊妹支, 支持油莎豆划归为禾本目<sup>[7, 18-19]</sup>。

研究克隆到的 *CeEPSPS* 编码 514 个氨基酸, 其序列长度与其他物种相近, 并含有叶绿体定位的信号肽和高度保守的 *EPSPS* 合酶结构域。*CeEPSPS* 成熟蛋白的分子量为 47.32 kDa、pI 为 5.49、GRAVY 为 0.069、AI 为 93.76、II 为 31.73, 与其他物种中同源蛋白的理化特性基本一致, 即分别为 46.86~47.55 kDa、5.13~5.58、0.011~0.111、91.78~95.96、28.38~35.46, 暗示其为稳定的疏水型酸性蛋白。进一步的序列比对证实 *CeEPSPS* 不存在已知的草甘膦抗性突变。育种实践中也发现团队收集的 56 份油莎豆种质均对草甘膦高度敏感, 研究利用前期获得的基因组重测序数据<sup>[8]</sup>进一步对相关 SNP 进行了鉴定, 虽然获得了 20 个 SNP, 但它们均与已知的抗药性变异无关, 这表明现有群体不存在抗药性突变。从表达模式来看, *CeEPSPS* 倾向于在成熟叶片和块茎中表达, 这与其活跃的次生代谢一致。

综上, 研究完成了包括油莎豆在内的代表性单子叶植物 *EPSPS* 基因的鉴定、*CeEPSPS* 编码区的分离与植物过表达载体的构建及序列特征、遗传变异、进化和表达特性分析, 较好地揭示了油莎豆的分类学地位, 初步证实团队前期收集的 56 份油莎豆种质不存在草甘膦抗性变异, 这为油莎豆的开发与合理栽培提供了理论支撑, 并为通过利用过表达或基因编辑技术创制油莎豆新种质奠定了坚实的基础。

## 参考文献

- [1] DUKE S O, POWLES S B. Glyphosate: a once-in-a-century herbicide[J]. Pest Management Science, 2008, 64(4): 319-325.
- [2] SAMMONS R D, GAINES T A. Glyphosate resistance: state of knowledge[J]. Pest Management Science, 2014, 70(9): 1367-1377.
- [3] 张玉池, 王晓蕾, 徐文蓉, 刘琪, 相世刚, 戴伟民, 强胜, 宋小玲. 国内外抗除草剂基因专利的分析[J]. 杂草学报, 2017, 35(2): 1-22.  
ZHANG Y C, WANG X L, XU W R, LIU Q, XIANG S G, DAI W M, QIANG S, SONG X L. Analysis on the patents of herbicide resistance gene at home and abroad[J]. Journal of Weed Science, 2017, 35(2): 1-22. (in Chinese)
- [4] HEAP I. Herbicide resistant weeds[J]. Integrated Pest Management Reviews, 2013, 3: 281-301.
- [5] DE CASTRO O, GARGIULO R, DEL GUACCHIO E, CAPUTO P, DE LUCA P. A molecular survey concerning the origin of *Cyperus esculentus* (Cyperaceae, Poales): two sides of the same coin (weed vs. crop)[J]. Annals of Botany, 2015, 115(5): 733-745.
- [6] 张学昆. 我国油莎豆产业研发进展报告[J]. 中国农村科技, 2019(287): 67-69.  
ZHANG X K. Research progress of the tigernut industry in China[J]. China Rural Science Technology, 2019(287): 67-69. (in Chinese)
- [7] 肖艳华, 邹智, 赵永国, 郭安平, 张丽. 油莎豆乙酰乳酸合酶基因 *CeALS* 的克隆与分析[J]. 生物技术通报, 2022, 38(4): 12-21.  
XIAO Y H, ZOU Z, ZHAO Y G, GUO A P, ZHANG L. Molecular cloning and characterization of an acetolactate synthase gene (*CeALS*) from tigernut (*Cyperus esculentus* L.)[J]. Biotechnology Bulletin, 2022, 38(4): 12-21. (in Chinese)
- [8] 邹智, 赵永国, 张丽, 孔华, 郭运玲, 郭安平. 基于单分子实时测序的油莎豆全长转录组分析[J]. 中国油料作物学报, 2021, 42(2): 229-235.  
ZOU Z, ZHAO Y G, ZHANG L, KONG H, GUO Y L, GUO A P. Single-molecule real-time (SMRT)-based full-length transcriptome analysis of tigernut (*Cyperus esculentus* L.)[J]. Chinese Journal of Oil Crop Sciences, 2021, 43(2): 229-235. (in Chinese)
- [9] 邹智, 杨礼富. 巴西橡胶树铁硫簇支架蛋白 cDNA 的克隆与分析[J]. 热带作物学报, 2010, 31(10): 1752-1756.  
ZOU Z, YANG L F. Cloning and analysis of a cDNA encoding an iron-sulfur cluster scaffold protein ISU1 from *Hevea brasiliensis* Müll. Arg[J]. Chinese Journal of Tropical Crops, 2010, 31(10): 1752-1756. (in Chinese)
- [10] 邹智, 郭运玲, 孔华. 橡胶树叶片衰老相关基因 *HbLEA3* 的克隆与表达分析[J]. 西南林业大学学报(自然科学版),

- 2021, 41(4): 42-48.
- ZOU Z, GUO Y L, KONG H. Cloning and expression analysis of *HbLEA3*, a gene associated with leaf senescence in rubber tree (*Hevea brasiliensis* Muell. Arg.)[J]. Journal of Southwest Forestry University (Natural Science), 2021, 41(4): 42-48. (in Chinese)
- [11] JUGULAM M, NIEHUES K, GODAR A S, KOO D H, DANILOVA T, FRIEBE B, SEHGAL S, VARANASI V K, WIERSMA A, WESTRA P, STAHLMAN P W, GILL B S. Tandem amplification of a chromosomal segment harboring 5-enolpyruvylshikimate-3-phosphate synthase locus confers glyphosate resistance in *Kochia scoparia*[J]. Plant Physiology, 2014, 166(3): 1200-1207.
- [12] CHEN J, JIANG C, HUANG H, WEI S, HUANG Z, WANG H, ZHAO D, ZHANG C. Characterization of *Eleusine indica* with gene mutation or amplification in EPSPS to glyphosate[J]. Pesticide Biochemistry and Physiology, 2017, 143: 201-206.
- [13] YU Q, JALALUDIN A, HAN H, CHEN M, SAMMONS R D, POWLES S B. Evolution of a double amino acid substitution in the 5-enolpyruvylshikimate-3-phosphate synthase in *Eleusine indica* conferring high-level glyphosate resistance[J]. Plant Physiology, 2015, 167(4): 1440-1447.
- [14] SAMMONS R D, YOU J, QI Y, FLASINSKI S, KAVANAUGH C, WASHAM J, OSTRANDER E, WANG D, HECK G. Evaluation of glyphosate resistance in *Arabidopsis thaliana* expressing an altered target site EPSPS [J]. Pest Management Science, 2018, 74(5): 1174-1183.
- [15] PEROTTI V E, LARRAN A S, PALMIERI V E, MARTINATTO A K, ALVAREZ C E, TUESCA D, PERMINGEAT H R. A novel triple amino acid substitution in the EPSPS found in a high-level glyphosate-resistant *Amaranthus hybridus* population from argentina[J]. Pest Management Science, 2019, 75(5): 1242-1251.
- [16] TAKANO H K, FERNANDES V N, ADEGAS F S, OLIVEIRA R S, WESTRA P, GAINES T A, DAYAN F E. A novel TIPT double mutation in EPSPS conferring glyphosate resistance in tetraploid *Bidens subalternans*[J]. Pest Management Science, 2020, 76(1): 95-102.
- [17] SPENCER M, MUMM R, GWYN J J, MCELROY D, STEPHENS M A. Glyphosate resistant maize lines: EP0975778 B8[P]. 2005-04-21.
- [18] LARRIDON I, REYNDERS M, HUYGH W. Affinities in *C<sub>3</sub>* *Cyperus* lineages (Cyperaceae) revealed using molecular phylogenetic data and carbon isotope analysis[J]. Botanical Journal of the Linnean Society, 2011, 167(1): 19-46.
- [19] ZOU Z, ZHAO Y G, ZHANG L, XIAO Y H, GUO A P. Analysis of *Cyperus esculentus* *SMP* family genes reveals lineage-specific evolution and seed desiccation-like transcript accumulation during tuber maturation[J]. Industrial Crops and Products, 2022, 187: 115382.