



吴立仁,张冬雪,唐贵,等.玉米 SEs 基因家族生物信息学分析[J].黑龙江农业科学,2022(7):14-19.

# 玉米 SEs 基因家族生物信息学分析

吴立仁<sup>1</sup>,张冬雪<sup>2</sup>,唐 贵<sup>2</sup>,隋冬华<sup>2</sup>,武新娟<sup>2</sup>,高佳缘<sup>2</sup>,王 腾<sup>2</sup>,王璐瑶<sup>2</sup>

(1. 黑龙江省农业科学院 经济作物研究所,黑龙江 哈尔滨 150086; 2. 黑龙江省农业科学院 乡村振兴科技研究所,黑龙江 哈尔滨 150023)

**摘要:**为获悉玉米 SEs 基因家族的功能和系统进化关系,利用生物信息学方法对玉米 SEs 基因家族进行鉴定,同时分析其家族成员理化性质、蛋白质二级结构、进化关系、基因结构、保守结构、氨基酸序列、染色体及启动子顺式作用元件。结果表明,在玉米中共鉴定到 10 个 SEs 基因家族成员,编码蛋白质的氨基酸序列长度为 320~534 aa,分子量大小为 33.85~56.94 kDa,等电点为 6.57~9.15;亚细胞定位预测结果表明大部分玉米 SEs 蛋白定位在细胞质和原生质;系统进化树将玉米 SEs 基因家族分为 3 支;基因结构的保守基序分析说明玉米 SEs 基因家族存在一定的差异性;10 个玉米 SEs 基因家族成员主要分布在 Chr1、Chr5 和 Chr9 上;氨基酸多序列比对分析表明 10 个玉米 SEs 基因家族成员存在一定的结构相似性;蛋白序列中存在 10 种 Motif;启动子顺式作用元件分析表明玉米 SEs 基因主要受光响应、茉莉酸信号响应调控。

**关键词:**玉米;鲨烯环氧酶;生物信息学

鲨烯环氧酶(Squalene Epoxidase, SE)又称鲨烯单加氧酶,属于单加氧酶。在酵母、真菌及动植物体内广泛存在<sup>[1]</sup>,在哺乳动物和酵母中 SEs 基因序列已经被注释<sup>[2]</sup>。鲨烯环氧酶可以催化鲨烯 C=C 双键发生反应,生成单氧态氧化鲨烯,而单氧态鲨烯在紫外线或强光作用下生成 2,3-氧化鲨烯,而 2,3-氧化鲨烯又经过糖基化、羟基化和环化等修饰作用最终生成不同的三萜类物质,如三萜皂苷、甾醇药用活性物质。三萜类物质合成通路研究发现,2,3-氧化鲨烯是三萜类物质合成的前体,故鲨烯环氧酶是三萜类化合物生物合成通路中的关键酶和限速酶<sup>[3]</sup>,亦是氧化还原反应过程所需要的重要结合位点。不同物种中 SEs 基因拷贝数量不同,在植物中通常含有 2 个及以上数量的 SEs 基因<sup>[4]</sup>,目前,鲨烯环氧酶已经在拟南芥<sup>[5]</sup>、人参<sup>[6]</sup>、三七<sup>[7]</sup>和蒺藜苜蓿<sup>[8]</sup>中被鉴定并克隆得到功能验证。

玉米(*Zea mays* L.)是世界重要的粮食作物之一,对玉米的研究现多集中在玉米遗传育种及栽培等方面,对玉米药用功能方面的研究相对较

少。因此本研究以玉米全基因组数据为基础,对 SEs 基因家族进行鉴定,并对其进行生物信息学分析,为解析玉米中 SEs 基因奠定基础,同时为挖掘玉米药用生物活性成分提供理论参考。

## 1 材料与方法

### 1.1 玉米 SEs 基因家族的鉴定

在玉米基因组数据库 Maize GDB(<http://www.maize.gdb.org>)下载玉米基因组,在 Pfam 数据库(<http://pfam.xfam.org>)下载 SE 结构域的隐马尔可夫模型(PF08491),并利用 HMMER 3.0 程序在玉米基因组数据库中对 SE 结构域进行搜索,以 E 值为 0.001 对结构域进行筛选。在 NCBI-CDD(<https://www.ncbi.nlm.nih.gov/cdd>)、SMART(<http://smart.embl-heidelberg.de/>)数据库进一步确认筛选到的基因的保守结构域。通过 TAIR(<http://arabidopsis.org>)下载拟南芥 SEs 基因家族核酸序列。

### 1.2 玉米 SEs 基因家族的生物信息学分析

利用 ExPASy(<https://www.expasy.org>)对 SEs 基因家族成员进行理化性质分析,利用 CELLO(<http://cello.life.nctu.edu.tw/>)进行亚细胞定位预测,利用 SOPMA SECONDARY STRUCTURE PREDICTION METHOD([https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_auto-mat.pl?page=npsa\\_sopma.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_auto-mat.pl?page=npsa_sopma.html))在线工具对 SEs 蛋白进行二级结构预测。

利用玉米 SEs 氨基酸序列和拟南芥 SEs 氨

收稿日期:2022-04-06

基金项目:黑龙江省农业科学院“农业科技创新跨越工程”玉米科技创新专项(HNK2019CX03)。

第一作者:吴立仁(1965—),男,硕士,高级农艺师,从事特色作物种质资源创新及综合利用研究。E-mail:15045578999@163.com。

通信作者:唐贵(1977—),男,硕士,副研究员,从事玉米育种和高产栽培技术研究。E-mail:hailuntanggui@163.com。

基酸序列进行系统进化树构建,将所有 SEs 氨基酸序列导入 MEGA 中,以 Cluster W 对蛋白序列进行多重比对,通过邻近法(Neighbor joining)构建系统进化树。利用在线工具 GSDS(<https://itol.embl.de/>)对基因结构进行分析,利用在线工具 MEME(<http://meme-suite.org/>)和 TBtools 软件对保守结构域进行分析。根据玉米 SEs 的位置信息,利用在线工具 MapGene2Chromosome2.0 ([http://mg2c.iask.in/mg2c\\_v2.0/](http://mg2c.iask.in/mg2c_v2.0/))对基因进行染色体定位分析。利用 DNAMAN 对玉米 SEs 蛋白序列进行氨基酸多序列比对分析。利用在线数据库 PlantCARE 分析玉米 SEs 基因起始密码子 ATG 上游 2.0 kb 启动子序列的顺式作用元件。

## 2 结果与分析

### 2.1 玉米 SEs 基因家族的鉴定及理化性质分析

根据 SE 结构域隐马尔可夫模型(PF08491),通过 HMMER 筛选及 NCBI-CDD、SMART 保守结构域进一步确定玉米中共有 10 个 SEs 基因家族成员,并根据其染色体位置将其命名为 *ZmSE1* ~ *ZmSE10*(表 1)。对得到的玉米所有 SEs 基因家族成员进行氨基酸长度、分子量、等电点、亲水性、

脂肪族氨基酸指数、亚细胞定位分析,结果表面氨基酸序列长度在 320~534 aa,编码蛋白质长度最长的是 *ZmSE10*,由 534 个氨基酸组成,编码蛋白质最短的是 *ZmSE8*,仅由 320 个氨基酸组成;分子量大小为 33.85~56.94 kDa;等电点 pI 为 6.57~9.15,多数蛋白质等电点大于 7,说明玉米 SEs 基因家族成员富含碱性氨基酸较多;不稳定系数为 38.77~45.50;脂肪族氨基酸指数为 85.50~100.51;亲水性为-0.164~0.159;亚细胞定位表明大多数 ZmSEs 蛋白定位在细胞质和原生质上,少部分定位在细胞内膜上,这也说明 ZmSEs 蛋白在不同的微环境下具有不同的功能,有的 ZmSEs 蛋白定位在不同位置,这也说明 ZmSEs 蛋白可能具有双重功能(表 1)。通过对 10 个 ZmSEs 蛋白进行二级结构预测表明,ZmSEs 蛋白的二级结构以 $\alpha$ -螺旋和无规则卷曲结构为主, $\beta$ 转角和扩展链结构所占比例较小。*ZmSE1*~*ZmSE9* 蛋白的无规则卷曲所占比例较大,为 39.32%~45.08%,其次是 $\alpha$ -螺旋结构,为 33.44%~40.16%;而 *ZmSE10* 蛋白的 $\alpha$ -螺旋所占比例最大,其次是无规则卷曲结构,所占比例分别为 39.89%和 37.27%(表 2)。

表 1 玉米 SEs 基因家族成员理化性质分析

基因名称	基因编号	染色体	编码氨基酸长度/aa	分子量/kDa	等电点 pI	不稳定系数	脂肪族氨基酸指数	亲水性	亚细胞定位
<i>ZmSE1</i>	Zm00001eb009510_P001	Chr1	529	56.84	9.14	45.50	99.58	0.139	内膜
<i>ZmSE2</i>	Zm00001eb009510_P002	Chr1	530	56.94	9.14	44.99	99.94	0.147	内膜
<i>ZmSE3</i>	Zm00001eb211070_P001	Chr5	457	50.26	7.72	41.09	88.12	-0.132	细胞质原生质
<i>ZmSE4</i>	Zm00001eb211070_P002	Chr5	522	56.81	8.05	42.40	86.53	-0.126	原生质
<i>ZmSE5</i>	Zm00001eb211070_P003	Chr5	381	41.89	6.57	49.19	87.77	-0.152	细胞质
<i>ZmSE6</i>	Zm00001eb211070_P004	Chr5	522	56.81	8.05	42.40	86.53	-0.126	原生质
<i>ZmSE7</i>	Zm00001eb211070_P005	Chr5	398	43.85	6.60	38.77	85.50	-0.164	细胞质
<i>ZmSE8</i>	Zm00001eb399370_P001	Chr9	320	33.85	8.58	43.15	99.97	0.068	原生质细胞质
<i>ZmSE9</i>	Zm00001eb399370_P002	Chr9	333	35.32	8.33	43.77	98.98	0.090	原生质
<i>ZmSE10</i>	Zm00001eb399370_P003	Chr9	534	56.89	9.15	44.03	100.51	0.159	内膜

### 2.2 玉米 SEs 基因家族的系统进化树分析

为分析玉米 SEs 基因在植物中的进化模式,对玉米中鉴定到的 10 个 SEs 蛋白的氨基酸序列和拟南芥中的 SE 蛋白序列构建系统进化树如图 1 所示,该基因家族可以被分为三大支,其中 *ZmSE5* 和 *ZmSE6* 聚为一支,*ZmSE3*、

*ZmSE4* 与 *ZmSE7* 聚为一支,剩余玉米中 SEs 蛋白与拟南芥 *AtSE1*、*AtSE2* 和 *AtSE3* 被聚为一支。玉米中 *ZmSE3*、*ZmSE4*、*ZmSE5*、*ZmSE6*、*ZmSE7* 这 5 个 SEs 基因未与拟南芥 *AtSEs* 聚在一支,这些基因具有的功能和作用有待进一步发掘。

表 2 玉米 SEs 蛋白二级结构预测分析

蛋白质名称	基因编号	$\alpha$ -螺旋/ %	$\beta$ -转角/ %	扩展链结构/ %	无规则卷曲/ %
ZmSE1	Zm00001eb009510_P001	37.43	6.43	16.82	39.32
ZmSE2	Zm00001eb009510_P002	37.36	6.79	16.42	39.43
ZmSE3	Zm00001eb211070_P001	37.20	3.94	13.79	45.08
ZmSE4	Zm00001eb211070_P002	38.51	3.83	14.18	43.49
ZmSE5	Zm00001eb211070_P003	40.16	3.94	15.22	40.68
ZmSE6	Zm00001eb211070_P004	38.51	3.83	14.18	43.49
ZmSE7	Zm00001eb211070_P005	35.43	3.77	17.34	43.47
ZmSE8	Zm00001eb399370_P001	33.44	6.88	20.31	39.38
ZmSE9	Zm00001eb399370_P002	35.44	5.41	17.72	41.44
ZmSE10	Zm00001eb399370_P003	39.89	6.37	16.48	37.27

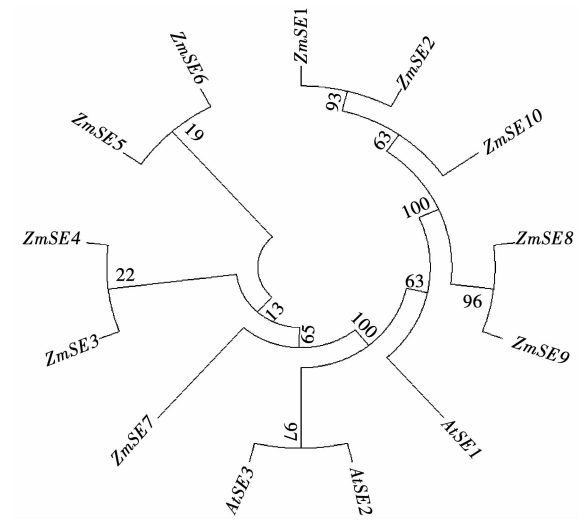


图 1 玉米 SEs 基因家族系统进化树

2.3 玉米 SEs 基因结构与 Motif 保守基序分析

利用玉米 SEs 基因家族 gff 文件,对其进行基因结构分析的结果如图 2 所示,10 个 SEs 基因

随机分布在玉米基因组中。ZmSE1、ZmSE2、ZmSE6、ZmSE7、ZmSE10 基因仅在 5' 端和 3' 端含有非编码区,剩余 ZmSEs 基因含有的非编码区均为 2 个以上。玉米中 ZmSEs 基因含有 CDS 的个数为 5~10 个,含有的内含子个数为 4~10 个。

利用在线软件 MEME 鉴定玉米 SEs 基因家族保守基序和 Motif 的结果如图 3 和图 4 所示,玉米 SEs 基因家族包括 10 个保守基序,所有基因中含有的保守基序个数有所差异,同时含有的保守基序位置也不相同,保守基序 1,2,4,7 在所有基因中全部存在,保守基序 10 更靠近 5' 端,保守基序 6 和 8 更靠近 3' 端。其中 ZmSE1 和 ZmSE2,ZmSE8 和 ZmSE9 两组含有的保守基序和位置大致相同,这一结果进一步支持了系统进化树的结果。Motif1~Motif10 由 32~50 不等的磷酸位点组成,其中 Motif3 共有 50 个保守磷酸位点,由 18 种氨基酸组成。

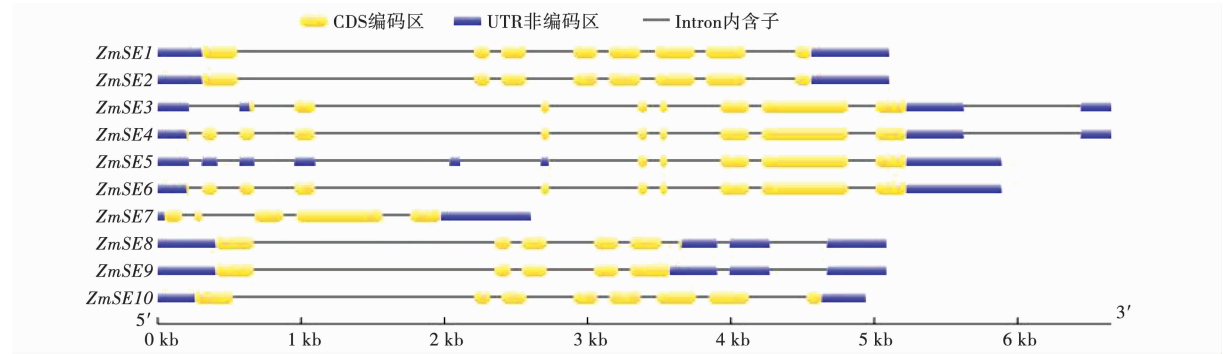


图 2 玉米 SEs 基因家族基因结构图



图 3 玉米 SEs 基因家族保守基序

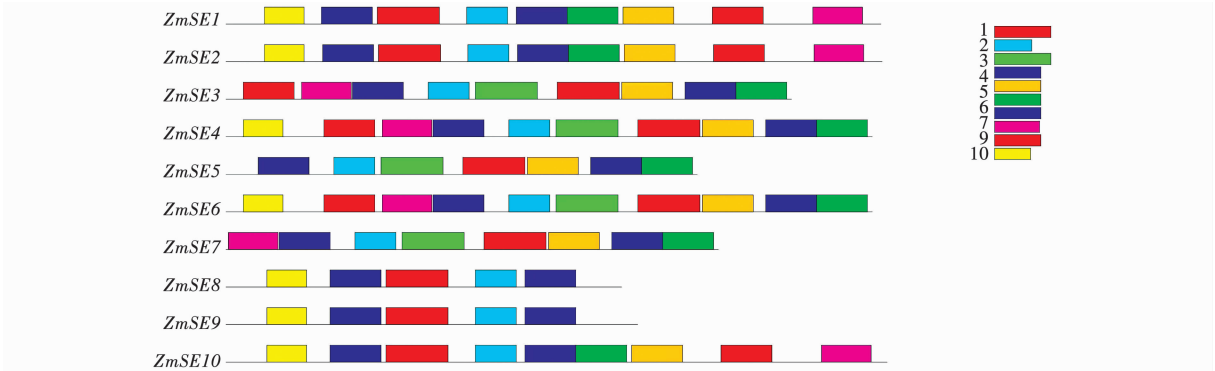


图 4 玉米 SEs 基因家族 Motif 分布图

2.4 玉米 SEs 基因家族染色体定位

为了探讨玉米 SEs 基因家族在染色体上的分布,对玉米 SEs 进行染色体定位分析,玉米 SEs 家族成员不均匀地分布在玉米染色体上。10 条

ZmSEs 基因主要分布在 Chr1、Chr5 和 Chr9 三条染色体上,其中 Chr1 上分布有 ZmSE1、ZmSE2,在 Chr5 上端分布着 ZmSE3~ZmSE7 共 5 个基因,ZmSE8、ZmSE9 和 ZmSE10 分布于 Chr9 上(图 5)。

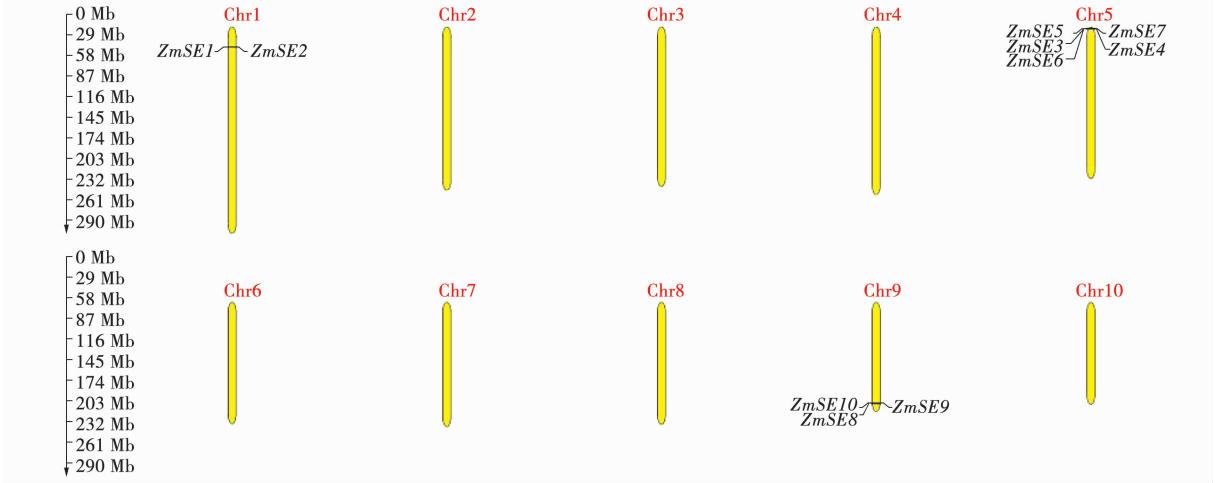


图 5 玉米 SEs 基因家族染色体定位

2.5 玉米 SEs 基因家族序列对比分析

利用 DNAMAN 对玉米 SEs 氨基酸序列进行多重比对的结果显示,10 个玉米 SEs 氨基酸序

列之间相似性极高,并对 10 个玉米 SEs 氨基酸序列进行保守位点预测,发现 10 个玉米 SEs 氨基酸序列均存在多个相同保守位点(图 6)。

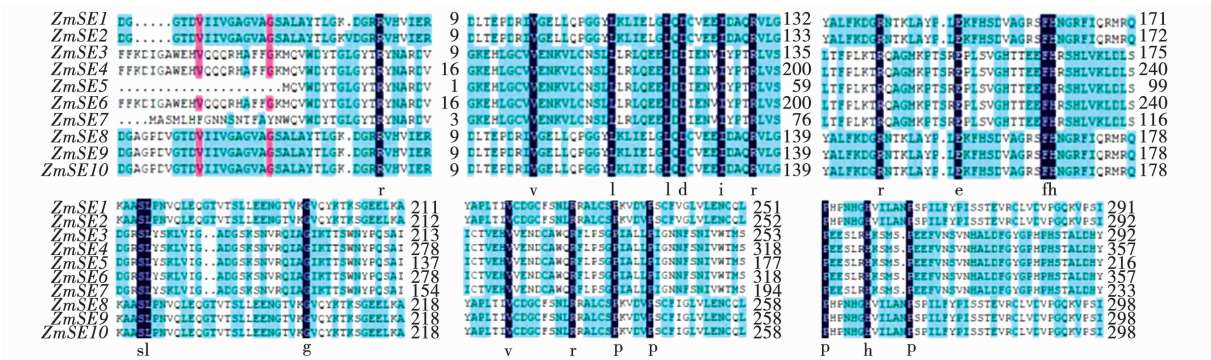


图6 玉米SEs基因家族序列对比图

2.6 玉米SEs基因家族启动子顺式作用元件分析

对玉米SEs基因顺式作用元件的分析结果如图7所示,玉米SEs基因的启动子区域均分布有顺式作用元件,且分布规律及数量差别较大。

其中ZmSE8、ZmSE9和ZmSE10基因启动子区域含有较多光响应、茉莉酸信号响应和脱落酸元件。玉米中所有SEs基因对光合茉莉酸信号响应元件均多于其他元件,说明玉米中SEs受到光调控和茉莉酸调控的影响较大。

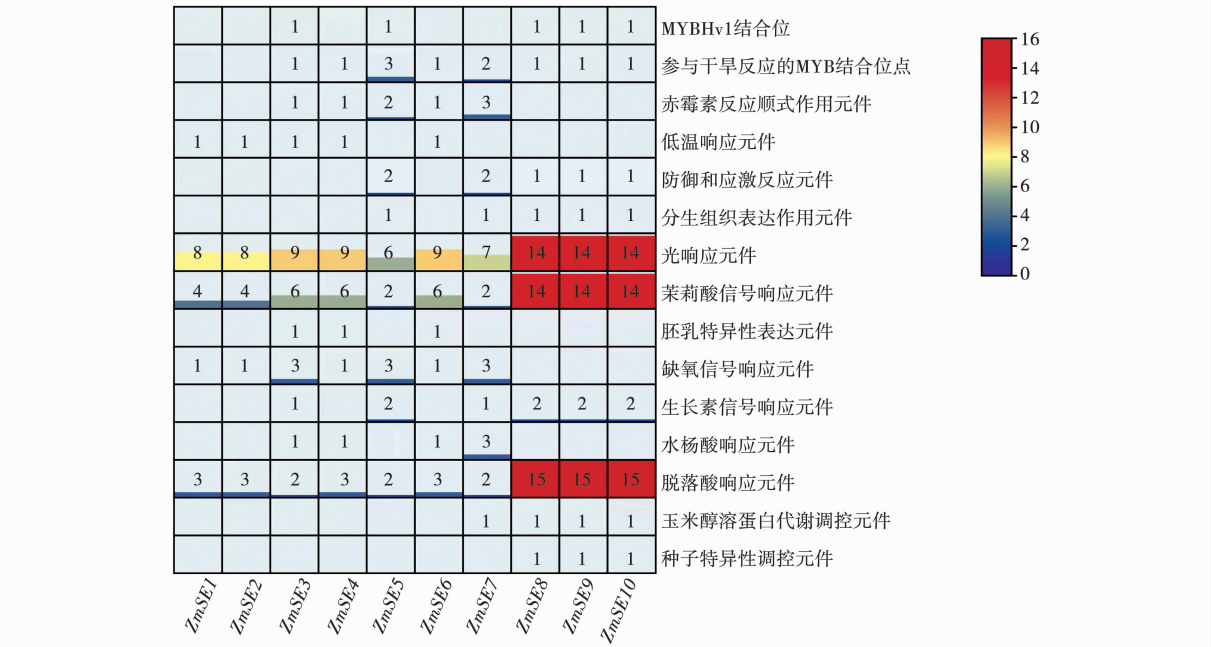


图7 玉米SEs基因启动子顺式作用元件数量分析图

3 讨论

鲨烯环氧酶催化植物甾醇和三萜皂苷等药用生物合成过程的第一个氧化步骤,因此被认为是限速酶。植物甾醇具有重要的药理学特性,可以降低胆固醇含量和抗肿瘤<sup>[9]</sup>,三萜皂苷也具有广泛的生物活性,在作为天然药物上具有重要的经济意义。刘艳<sup>[10]</sup>对雷公藤中鲨烯环氧酶的研究发现,TwSE1、TwSE2基因的表达量经茉莉酸甲酯诱导后呈先上升后下降的变化趋势。龙月红等<sup>[11]</sup>通过对刺五加中鲨烯环氧酶的表达分析研究得出EsSE1较EsSE2基因与刺五加三萜皂苷生物合成更为相关。竹节参<sup>[12]</sup>、桑黄<sup>[13]</sup>、野三

七<sup>[14]</sup>中鲨烯环氧酶也均被克隆并进行表达分析。Jung等<sup>[15]</sup>研究人参中鲨烯环氧酶后得出PgSE1和PgSE2的表达受到不同方式的调节,PgSE1调节人参皂苷的生物合成,但不会调节人参中植物甾醇的生物合成。玉米SEs基因编码蛋白长度为320~534 aa,大部分定位在细胞质和细胞核,启动子区域含有光响应、茉莉酸响应元件。系统进化树根据蛋白序列的相似性以及保守结构域可以揭示物种之间的亲缘关系,进而推测基因的功能。Motif是蛋白质分子内具有特定空间构象、特定蛋白质功能的结构成分是结构域的亚单位,与特定功能联系在一起<sup>[16]</sup>。本研究结果

显示,玉米 *SEs* 基因家庭包含有 10 个保守基序, Motif1~Motif10 由 32~35 个不等的磷酸位点组成。

## 4 结论

本研究从玉米基因组的筛选鉴定到 10 个 *SEs* 基因,系统进化树分析将玉米中 *SEs* 基因分为三支。基因结构和保守基序分析发现玉米中 *SEs* 基因存在一定的差异性,氨基酸多序列比对发现 *SEs* 基因存在一定的相似性。本文为在分子水平上对玉米药用生物功能物质挖掘提供了一定的理论支持。

## 参考文献:

- [1] LANDL K M, KLÖSCH B, TURNOWSKY F. *ERG1*, encoding squalene epoxidase, is located on the right arm of chromosome VII of *Saccharomyces cerevisiae* [J]. Yeast Mapping Reports, 1996, 12: 609-613.
- [2] SCHAFER U, REED D, HUNTER D G, et al. An example of intron junctional sliding in the gene families encoding squalene monooxygenase homologues in *Arabidopsis thaliana* and *Brassica napus* [J]. Plant Molecular Biology, 1999, 39: 721-728.
- [3] 邢朝斌, 王一曼, 陈正恒, 等. 三萜皂苷的生物合成[J]. 生命的化学, 2005, 25(5): 420-422.
- [4] XING Z, CAO L, CHEN L, et al. Cloning and sequence analysis on cDNA of squalene epoxidase gene in *Eleutherococcus senticosus* [J]. China Journal of Chinese Materia Medica, 2012, 37(2): 172.
- [5] MOROZOVA S, SUCROYER I, AUWERX J. Cholesterol metabolism modulators and the future of atherosclerosis therapy[J]. Medecine Sciences; M/S, 2004, 20(6-7): 685.
- [6] HAN J Y, IN J G, KWON Y S, et al. Regulation of ginsen-

oside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in *Panax ginseng* [J]. Phytochemistry, 2010, 71(1): 36.

- [7] 张朵朵, 林丽梅, 国红玉, 等. 人参与三七鲨烯环氧酶家族的适应性进化[J]. 华北理工大学学报(自然科学版), 2022, 44(1): 98-106.
- [8] SUZUKI H, ACHNINE L, XU R, et al. A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula* [J]. Plant Journal, 2002, 32(6): 1033-1048.
- [9] JONES P, MACDOUGALL D, NTANIOS F, et al. Dietary phytosterols as cholesterol-lowering agents in humans[J]. Canadian Journal of Physiology and Pharmacology, 1997, 75: 217-227.
- [10] 刘艳. 雷公藤三萜合成途径关键酶基因克隆与表达分析[D]. 西安: 西北农林科技大学, 2018.
- [11] 龙月红, 李非非, 杨果, 等. 刺五加鲨烯环氧酶基因家族成员表达特性的分析[J]. 中国中药杂志, 2015, 40(1): 59-62.
- [12] 梁娥, 齐敏杰, 张来. 竹节参鲨烯环氧酶基因的克隆与生物信息学分析[J]. 基因组学与应用生物学, 2021, 40(1): 788-794.
- [13] 孙婷婷, 邹莉, 张林芳, 等. 桑黄鲨烯环氧酶基因克隆与序列分析[J]. 中草药, 2015, 46(18): 2768-2773.
- [14] 王宝婕, 朱灵英, 周青青, 等. 野三七鲨烯环氧酶基因的克隆及原核表达[J]. 中国实验方剂学杂志, 2019, 25(22): 147-153.
- [15] JUNG Y, JUN G, YONGN S, et al. Regulation of ginsenoside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in *Panax ginseng* [J]. Phytochemistry, 2010, 71: 36-46.
- [16] 赵晶, 李旭彤, 梁学忠, 等. 陆地棉漆酶基因家族鉴定及在黄萎病菌胁迫下的表达分析[J]. 作物学报, 2019, 45(12): 1784-1795.

# Bioinformatics Analysis of *SEs* Gene Family in Maize

WU Li-ren<sup>1</sup>, ZHANG Dong-xue<sup>2</sup>, TANG Gui<sup>2</sup>, SUI Dong-hua<sup>2</sup>, WU Xin-juan<sup>2</sup>, GAO Jia-yuan<sup>2</sup>, WANG Teng<sup>2</sup>, WANG Lu-yao<sup>2</sup>

(1. Institute of Economic Crops, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China; 2. Institute of Rural Revitalization Science and Technology, Heilongjiang Academy of Agricultural Sciences, Harbin 150023, China)

**Abstract:** In order to learn the function and phylogenetic relationship of *SEs* gene family in maize, bioinformatics methods were used to identify the maize *SEs* gene family, and the physicochemical properties, protein secondary structure, evolutionary relationship, gene structure, conserved structure, amino acid sequence, chromosomal localization and promoter. The results showed that a total of 10 members of the maize *SEs* gene family were identified in maize. The protein-encoding amino acid sequences ranged in length from 320 to 534, the molecular weight ranged from 33.85 to 56.94 kDa, and the isoelectric points varied in the range of 6.57 to 9.15. The results of subcellular localization prediction showed that most of the maize *SEs* proteins were located in the cytoplasm and protoplasm, the phylogenetic tree divided the maize *SEs* gene family into three branches, the conserved motif analysis of the gene structure showed that there were certain differences in the maize *SEs* gene family. The members of the maize *SEs* gene family were mainly distributed on Chr1, Chr5 and Chr9, amino acid multiple sequence alignment analysis showed that 10 maize *SEs* gene family members had certain similarity, there were 10 Motifs in protein sequence, promoter *cis*-acting element analysis showed that maize *SEs* genes are mainly regulated by light response and jasmonic acid signal response.

**Keywords:** maize; squalene epoxidase; bioinformatics