

大豆 *DMP* 基因的生物信息学及基因多样性分析

周雅^{1,2}, 张祥^{1,2}, 赵权^{2,3}, 孙建强^{2,4}, 王晓波¹, 邱丽娟²

(¹安徽农业大学农学院, 合肥 230036; ²中国农业科学院作物科学研究所/农作物基因资源与基因改良国家重大科学工程/农业农村部北京大豆生物学重点实验室, 北京 100081; ³东北林业大学生命科学学院, 黑龙江哈尔滨 150040; ⁴东北农业大学农学院, 黑龙江哈尔滨 150030)

摘要: 双单倍体技术被广泛应用于加速植物育种, 近年来, 利用含有功能未知结构域 DUF679 膜蛋白 (*DMP*) 突变的玉米 (*Zea mays* L.) 株系作为单倍体诱导株系。本研究搜索与玉米 *DMP* 基因氨基酸序列同源率 60% 以上的大豆 *DMP* 基因, 对其进行生物信息学分析; 并结合 2214 份重测序数据库探究 *GmDMP* 基因在大豆中的分子机理及生物多样性。分析结果表明: *GmDMP1* (*Glyma.18G097400*) 与 *GmDMP2* (*Glyma.18G098300*), 与玉米 *DMP* 基因的进化关系亲缘度极高, 基因全长均为 645 bp, 氨基酸序列同源率高达 95% 以上, 编码氨基酸数目与等电点完全一致, 磷酸化分布仅有 1 个位点差异。 *GmDMP1* 与 *GmDMP2* 基因均有 1 个相同的结构域 DUF679, 定位于内质网的可能性较大, 均为跨膜的不分泌亲水蛋白。 *GmDMP1* 与 *GmDMP2* 基因在 2214 份种质资源中分别发生了 3 个与 1 个非同义突变, 各组成 3 种和 2 种单倍型。 *GmDMP1*^{H3} 在驯化中受到了强烈的选择, *GmDMP1*^{H1} 与 *GmDMP1*^{H2} 单倍型的突变位点在结构域 DUF679 上。将大豆的两个 *DMP* 基因突变后可能获得单倍体诱导系, 从而缩短大豆育种年限。

关键词: 大豆; *GmDMP1*; *GmDMP2*; 单倍型分析; 单倍体诱导系

Bioinformatics and Gene Diversity Analysis of the *DMP* Genes in Soybean

ZHOU Ya^{1,2}, ZHANG Xiang^{1,2}, ZHAO Quan^{2,3}, SUN Jianqiang^{2,4}, WANG Xiaobo¹, QIU Lijuan²

(¹School of Agronomy, Anhui Agricultural University, Hefei 230036; ²Institute of Crop Sciences,

Chinese Academy of Agricultural Sciences/National Key Facility for Crop Gene Resources and

Genetic Improvement / Key Laboratory of Soybean Biology in Beijing, Ministry of Agriculture and Rural Affairs,

Beijing 100081; ³College of Life Sciences, Northeast Forestry University, Harbin 150040,

Heilongjiang; ⁴College of Agriculture, Northeast Agricultural University, Harbin 150030, Heilongjiang)

Abstract: Double haploid techniques have been widely used to accelerate plant breeding, and in recent years, maize (*Zea mays* L.) lines containing mutations in the membrane protein (*DMP*) of the functionally unknown domain DUF679 have been used as haploidy-induced lines. In this study, soybean *DMP* genes with more than 60% homology with maize *DMP* gene were searched and bioinformatic analysis was performed. The molecular mechanism and biodiversity of *GmDMP* gene in soybean were investigated by using 2214 resequencing databases. The analysis results show that: *GmDMP1* (*Glyma.18G097400*) and *GmDMP2* (*Glyma.18G098300*) have high evolutionary affinity with maize *DMP* gene. The full length of genes is 645 bp, the homology rate of amino acid sequence is over 95%, and the number of encoded amino acids is completely consistent with the isoelectric point. The phosphorylation distribution differs at only one site. Both *GmDMP1* and *GmDMP2* genes have the same domain DUF679, which is more likely to locate in the ER, and both are transmembrane non-secreting hydrophilic proteins. There were 3 and 1 non-synonymous mutations of *GmDMP1*

收稿日期: 2023-04-19 修回日期: 2023-05-13 网络出版日期: 2023-06-14

URL: <https://doi.org/10.13430/j.cnki.jpgr.20230419003>

第一作者研究方向为大豆优异基因挖掘, E-mail: 985874731@qq.com

通信作者: 邱丽娟, 研究方向为大豆基因资源挖掘与利用, E-mail: qiulijuan@caas.cn

基金项目: 中国农业科学院创新工程基金项目

Foundation project: The Innovation Project of Chinese Academy of Agricultural Sciences

and *GmDMP2* genes in 2214 germplasm resources, respectively, consisting of 3 and 2 haplotypes. *GmDMP1^{H3}* has been strongly selected during domestication, and the mutation site of the *GmDMP1^{H1}* and *GmDMP1^{H2}* haplotypes is located in the DUF679 domain. It is possible to obtain a haploid inducible line by mutating two *DMP* genes in soybean, thus shortening the breeding life of soybean.

Key words: soybean; *GmDMP1*; *GmDMP2*; haplotype analysis; haploid induction line

大豆 [*Glycine max* (L.) Merr.] 的原产国是中国, 属一年生草本植物, 作为中国的四大粮食作物之一, 大豆富含油脂、蛋白质、碳水化合物和其他有益营养物质, 作为人类食物和动物饲料的主要来源^[1-2], 是重要的经济作物和油料作物^[3]。随着中国人口持续不断地增多, 造成国内大豆市场需求量的持续增高^[4]。但由于国内种植的大豆面积较少, 单产水平不高, 从而造成了严重的供给不足^[5]。

传统的大豆育种方法需经过7~8代的选育过程, 已不能满足人们的需求, 纯系育种是作物育种十分重要的组成部分^[6]。如今, 单倍体技术的发展为育种提供了新思路, 具有重要的技术优势与实用育种价值。单倍体是指仅由原生物体染色体组一半的染色体组数所构成的个体, 和运用多代自交或回交得到纯合株系的传统育种模式相比较, 单倍体育种技术利用杂交诱导的育种方式诱导产生单倍体植株, 再利用化学药剂进行染色体加倍, 只需要2代约1~2年的时间得到纯合株系, 为实现快速提升作物成为纯系提供了高效的育种途径^[7]。自然发生的突变与人工诱导突变两种育种途径均可产生单倍体。自然突变促使作物产生单倍体的概率极低, 当作物受到侵害或者发育出现异常状况时才会产生。人工诱导突变则是对作物进行人为技术处理, 除了原始的体外花药离体培养等方法之外^[8], 单倍体体内诱导系是目前诱导作物产生单倍体最有效的方法。最早运用体内诱导系诱导出单倍体的作物是玉米^[9]。1959年美国Northrup King种子公司的Coe^[10]从玉米中发掘出诱导系Stock6, 作为父本诱导系可以诱导母本产生单倍体但该诱导系存在自身不足, 诱导率仅有1%~2%。后经过玉米研究者的优化, 将诱导率提升至7%~15%, 但Stock6诱导系具体的分子机能尚未得到解释。研究者利用图位克隆的方法在Stock6诱导系中定位到两个控制玉米产生单倍体的数量性状位点: *qhir1* 与 *qhir8*。经过对第一个重要数量性状基因 *qhir1* 的探索研究, 克隆得到的基因命名为 *ZmMTL/ZmPLA1/ZmNLD*^[11-12], 该基因编码一种花粉特异性磷脂酶, 在玉米的精细胞

中表达, 但并未能在双子叶植物中得到运用。另一个重要数量性状基因 *qhir8* 定位于玉米9号染色体的一个789 kb的区域, 该区域编码一个DUF679结构域膜蛋白, 并将此基因命名为 *ZmDMP*^[13]。DMP蛋白在单子叶与双子叶植物中都具有保守功能和相似的表达模式, 而在双子叶作物中的氨基酸相似性高达74%。随着现代生物分子技术的发展运用, 可以对目标基因DNA序列进行调控从而达到改变靶基因功能的目的^[14]。拟南芥 (*Arabidopsis thaliana*)^[15]、番茄 (*Solanum lycopersicum*)^[16]、甘蓝型油菜 (*Brassica napus*)^[17] 和紫花苜蓿 (*Medicago sativa*)^[18] 都利用生物技术改变玉米单倍体诱导基因 *ZmDMP* 同源基因的基因功能, 最终都成功诱导产生单倍体, 说明 *DMP* 基因在诱导植物产生单倍体的道路中起到了关键的作用, 研究该基因的蛋白结构与生物信息多样性分析为大豆育种的研究提供了新方法。

本研究首先利用玉米 *ZmDMP* 基因的蛋白序列经过同源序列比对得到大豆的同源基因 *GmDMP1* 与 *GmDMP2*, 分析大豆两个同源基因的结构差异、进化关系及理化性质。通过2214份重测序数据分析大豆两个同源基因的突变位点及突变类型, 为进一步对 *GmDMP1* 与 *GmDMP2* 基因的单倍型特性分析以及基因的分子机理与多样性的研究分析提供基础依据。

1 材料与方法

1.1 试验材料

从Phytozome (<https://phytozome-next.jgi.doe.gov/>) 和NCBI (<https://www.ncbi.nlm.nih.gov/>) 网站中下载玉米、拟南芥、番茄、甘蓝型油菜、紫花苜蓿、大豆等作物的 *DMP* 基因序列与氨基酸序列。农业农村部北京生物学重点实验室已完成重测序的2214份国内外大豆种质数据, 均保存在NCBI序列读取档案中, 登录号为PRJNA681974^[19]。

1.2 试验方法

1.2.1 DMP基因进化分析 在GSDS (<http://gsds.gao-lab.org>) 分析绘制 *DMP* 基因家族外显子和内含

子结构图。利用MEGAX软件中的Muscle算法对DMP基因家族序列进行比对,获得比对结果后用邻接法(Neighbor-joining, NJ)构建系统演化树,利用泊松模型将Bootstrap值设置为“1000”。分析大豆与其他植物DMP基因的亲缘关系。

1.2.2 DMP同源基因氨基酸序列分析 利用DNAMAN对玉米、拟南芥、甘蓝型油菜、紫花苜蓿,以及大豆*GmDMP1*(*Glyma.18G097400*)和*GmDMP2*(*Glyma.18G098300*)基因的氨基酸序列进行多序列比对,绘制这几个基因的氨基酸序列结构图。

1.2.3 GmDMP蛋白高级结构分析 通过prabi(<https://npsa-prabi.ibcp.fr/>)在线网站预测和分析了大豆*GmDMP1*和*GmDMP2*蛋白质二级结构模型图;利用SWISS-MODEL(<https://swissmodel.expasy.org/>)构建两个蛋白质的三级结构模型图;利用CDD(<https://www.ncbi.nlm.nih.gov/cdd>)网站分析蛋白功能结构域。

1.2.4 GmDMP蛋白质特性分析 运用GenScript(<https://www.genscript.com/tools/psort>)网站中的PSORTII对*GmDMP1*及*GmDMP2*基因氨基酸序列进行亚细胞定位预测。利用在线软件ExPasy中的ProtParam(<https://web.expasy.org/protparam/>)工具预测*GmDMP1*及*GmDMP2*基因所编码蛋白质的理化性质,包括蛋白质分子量、等电点、不稳定性、脂肪量和氨基酸含量。使用TMHMM Server v.2.0网站(<http://www.cbs.dtu.dk/services/TMHMM-2.0/>)预测*GmDMP1*及*GmDMP2*基因是否拥有跨膜结构。利用在线网站SignalP4.1 Server(<http://www.cbs.dtu.dk/services/SignalP-4.1/>)对*GmDMP1*及*GmDMP2*基因氨基酸序列进行信号肽预测。利用网站NetPhos-3.1(<https://services.healthtech.dtu.dk>)预测*GmDMP1*及*GmDMP2*基因氨基酸的磷酸化位点。应用ExPasy中的ProtScale(<https://web.expasy.org/protscale/>)网站预测*GmDMP1*及*GmDMP2*基因氨基酸序列的亲疏水性。

1.2.5 GmDMP在大豆中各组织的表达分析 从Phytozome(<https://phytozome-next.jgi.doe.gov/>)在线网站下载大豆*GmDMP1*与*GmDMP2*基因在大豆各组织器官中的表达量数据进行比较分析获得差异性并运用TBtools软件绘制表达量热图表达模式图。

1.2.6 GmDMP的多态性分析 运用Phytozome网站查询得到*GmDMP1*及*GmDMP2*基因的位置信息,根据实验室所拥有的2214份重测序数据中的

SNP数据对其进行多态性分析,获得其变异位点、变异类型及变异数目。

2 结果与分析

2.1 大豆*GmDMP1*和*GmDMP2*基因序列与蛋白结构分析

2.1.1 GmDMP基因进化树与编码序列分析 利用Phytozome(<https://phytozome-next.jgi.doe.gov/>)网站对玉米(*Zea mays*)、拟南芥、番茄、紫花苜蓿、甘蓝型油菜、大豆的DMP基因进行同源比对,并运用MEGA11软件构建系统进化树,GSDS(<http://gsds.gao-lab.org/>)网站绘制基因结构图如图1所示。大豆*GmDMP1*(*Glyma.18G097400.1*)和*GmDMP2*(*Glyma.18G098300.2*)与玉米、拟南芥、甘蓝型油菜、紫花苜蓿这4个物种的同源性较高、亲缘关系较近,聚为一类,结果表明*GmDMP1*(*Glyma.18G097400.1*)和*GmDMP2*(*Glyma.18G098300.2*)两个基因敲除后可能同样具有诱导出大豆单倍体的能力。

2.1.2 DMP基因编码序列比对 大豆*GmDMP1*(*Glyma.18G097400*)和*GmDMP2*(*Glyma.18G098300*)两个基因结构均位于18号染色体上,基因全长均为645 bp,无内含子。编码区核苷酸序列长度均为645 bp,编码蛋白长度均为215 aa,基因结构相似度极高,仅有少量的氨基酸差异。两个基因的氨基酸序列同源性高达97%。玉米、拟南芥、甘蓝型油菜、紫花苜蓿以及大豆*GmDMP1*和*GmDMP2*基因的多序列比对如图2所示。比对序列结果可以看出各作物存在一定的相似度,关联性较大。

2.1.3 GmDMP1和GmDMP2氨基酸序列二级结构与三级结构分析 蛋白二级结构主要有 α 螺旋、 β 折叠、 β 转角、无规则卷曲和延伸链。大豆*GmDMP1*和*GmDMP2*氨基酸序列预测出的二级结构如图3A、B所示,结果发现,*GmDMP1*和*GmDMP2*氨基酸主要由 α 螺旋、延伸链、 β 转角、无规则卷曲构成。*GmDMP1*的 α 螺旋、延伸链、 β 转角、无规则卷曲构成比例分别为38.32%、14.02%、4.21%、43.46%,*GmDMP2*的 α 螺旋、延伸链、 β 转角、无规则卷曲构成比例分别为42.52%、14.49%、4.21%、38.79%。可以确定*GmDMP1*和*GmDMP2*氨基酸二级结构主要以 α 螺旋和无规则卷曲为主要构成部分。大豆*GmDMP1*和*GmDMP2*氨基酸序列预测出的三级结构如图3C、D所示,两个基因的氨基酸三级结构基本无差异。

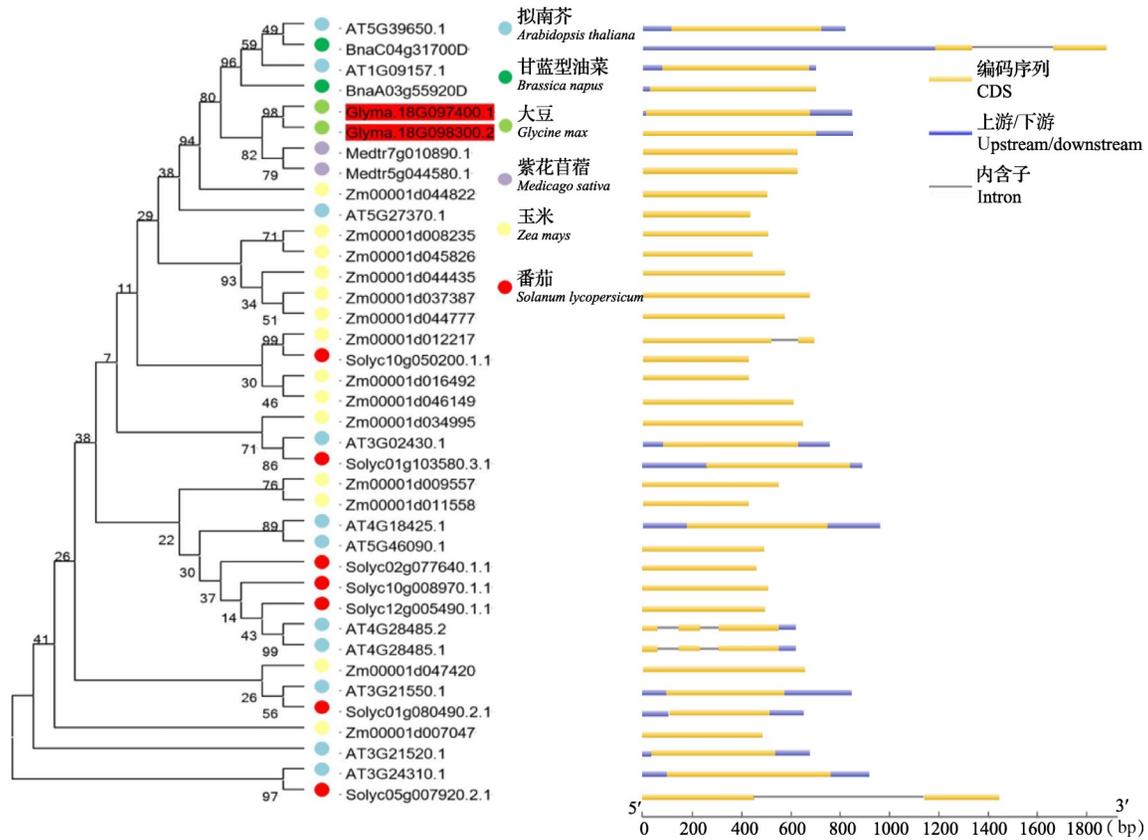
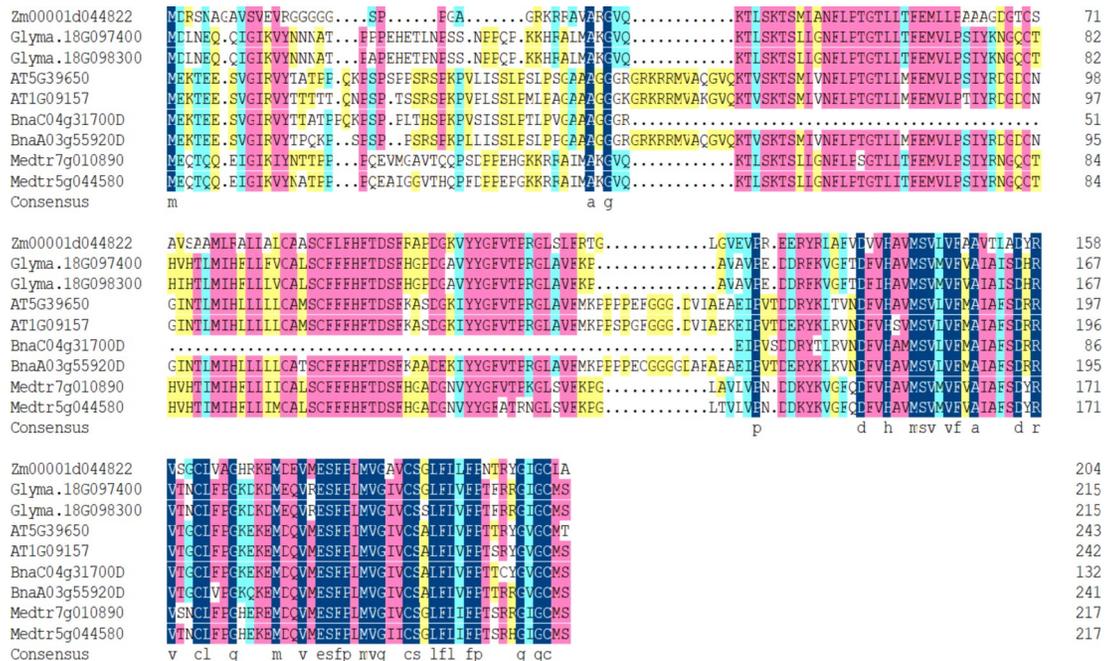


图1 植物DMP基因系统进化树与基因结构

Fig.1 Phylogenetic tree and gene structure of plant DMP genes

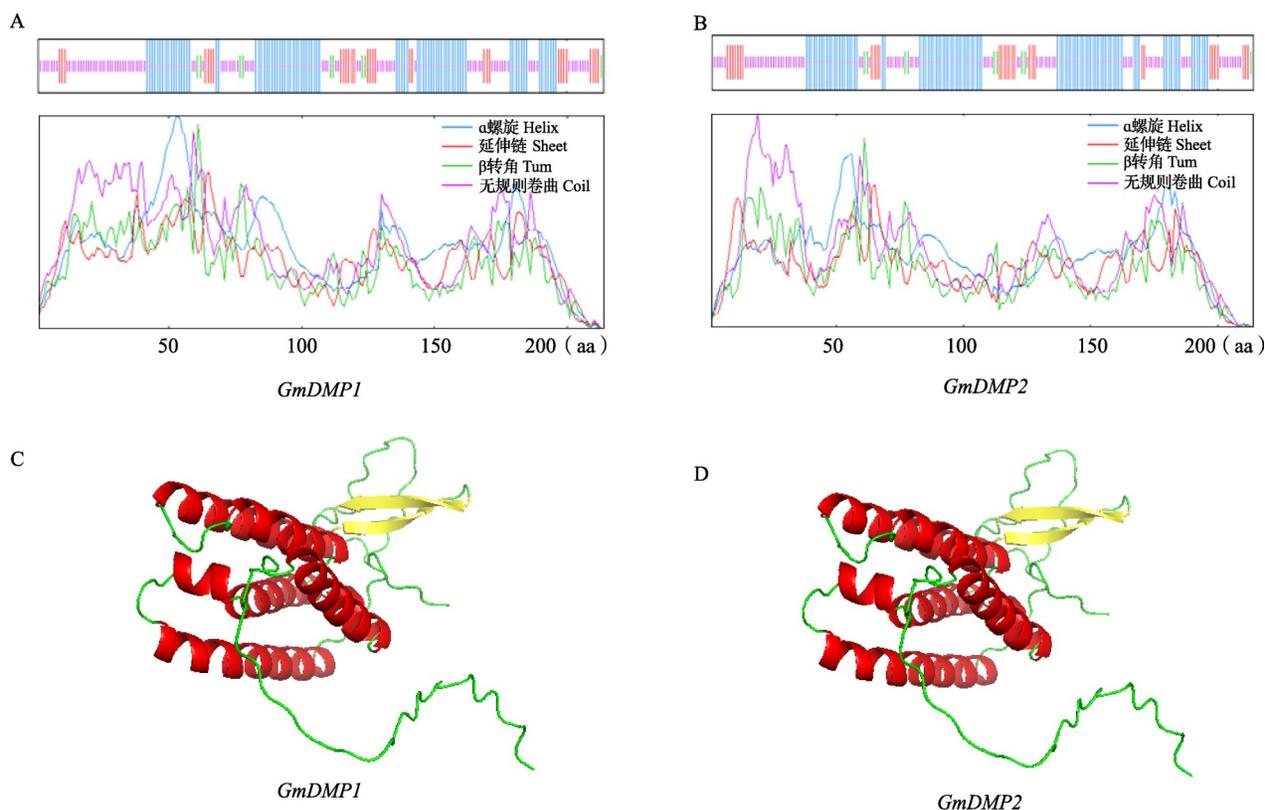


深蓝色部分为比对相似度值为100%,粉色部分为比对相似度值大于等于75%,
淡蓝色部分为比对相似度大于等于50%,黄色部分为比对相似度大于等于33%

The dark blue part indicates that the comparison similarity value is 100%, the pink part indicates that the comparison similarity value is greater than or equal to 75%, the light blue part indicates that the comparison similarity value is greater than or equal to 50%, and the yellow part indicates that the comparison similarity value is greater than or equal to 33%

图2 植物DMP基因多序列比对

Fig.2 Multiple sequence alignment of plant DMP genes



图C、D红色部分为 α 螺旋区,黄色部分为 β 折叠区,绿色部分为环区

In Figure C and D, the red part is the α helix region, the yellow part is the β pleated sheet region, and the green part is the loop region

图3 *GmDMP1* 和 *GmDMP2* 氨基酸序列二级、三级结构预测图

Fig.3 Secondary and tertiary structure prediction of *GmDMP1* and *GmDMP2* amino acid sequences

2.1.4 *GmDMP1* 和 *GmDMP2* 氨基酸序列结构域分析 氨基酸的保守结构域分析可以辅助确定基因的功能研究。运用在线网站对 *GmDMP1* 和 *GmDMP2* 氨基酸序列进行 BLAST 得到的功能结构域分析如

图4所示。*GmDMP1* 和 *GmDMP2* 具有相同的结构域,且所在位置相同,均为 *DMP* 基因家族所特有的 DUF679 保守结构域,说明 *GmDMP1* 和 *GmDMP2* 均属于 *DMP* 基因家族。

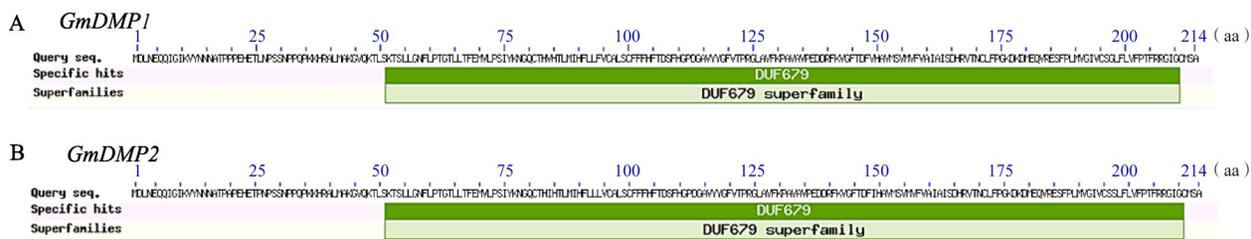


图4 *GmDMP1* 和 *GmDMP2* 氨基酸序列结构域

Fig.4 Amino acid sequence domains of *GmDMP1* and *GmDMP2*

2.2 大豆 *GmDMP1* 和 *GmDMP2* 蛋白特性分析

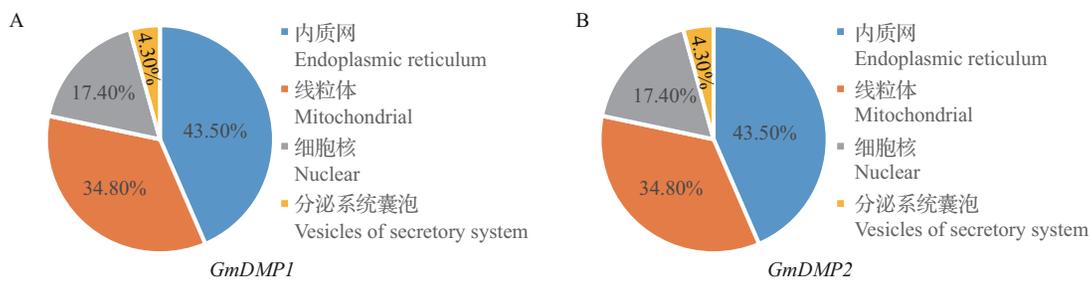
2.2.1 理化性质 大豆 *GmDMP1* 和 *GmDMP2* 基因的蛋白理化性质分析结果如表1所示,两个蛋白的编码氨基酸数目均为214,分子量差值约为18Da, pI 值均为8.22,不稳定指数均小于40,为稳定蛋白,脂肪指数差值为1.4。蛋白序列中含量丰富的氨基酸为缬氨酸、苯丙氨酸、亮氨酸、脯氨酸、甘氨酸、苏氨酸、丙氨酸。两个基因中亮氨酸、脯氨酸和苏氨酸

的含量均一致,剩余氨基酸含量差值都较小,两个基因的理化性质存在较高的相似度。

2.2.2 亚细胞定位 *GmDMP1* 和 *GmDMP2* 基因氨基酸的亚细胞定位预测结果如图5所示,两个基因的亚细胞定位预测结果相同,在内质网、线粒体、细胞核、分泌系统囊泡的百分比分别为43.5%、34.8%、17.4%、4.3%。可得出结论,两个基因均定位于内质网的可能性最大。

表 1 *GmDMP1* 和 *GmDMP2* 蛋白特性分析Table 1 Physical and chemical properties of *GmDMP1* and *GmDMP2* protein

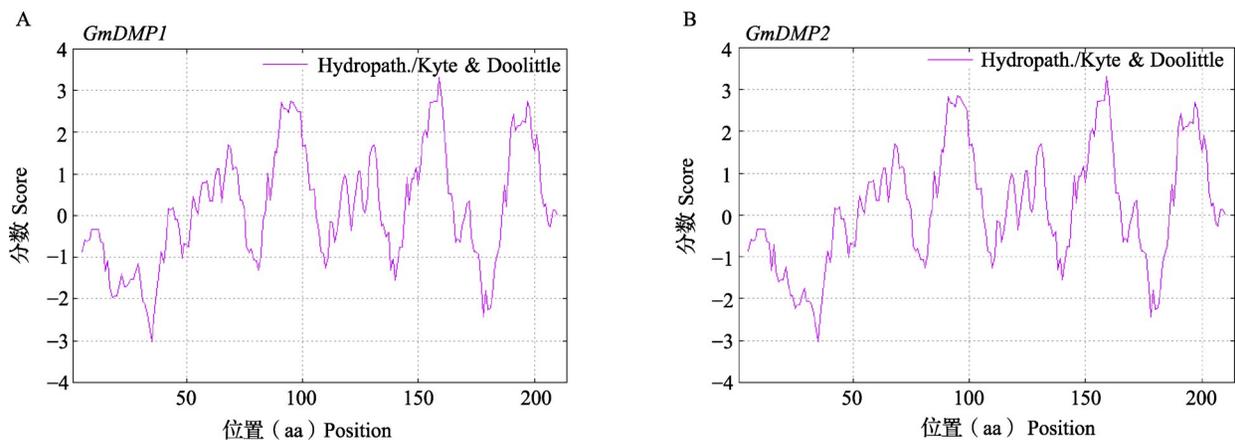
蛋白名称 Protein name	编码氨基酸数目 Number of encoding amino acids	分子量 Molecular weight	等电点 pI	不稳定指数 Instability index	脂肪指数 (%) Aliphatic index	含量最丰富的氨基酸 (%) The most abundant amino acid						
						缬氨酸 Valine	苯丙氨酸 Phenylalanine	亮氨酸 Leucine	脯氨酸 Proline	甘氨酸 Glycine	苏氨酸 Threonine	丙氨酸 Alanine
<i>GmDMP1</i>	214	23746.73	8.22	28.04	83.27	9.8	9.3	8.9	7.5	7	6.5	5.6
<i>GmDMP2</i>	214	23728.71	8.22	32.14	84.67	8.9	8.9	8.9	7.5	6.5	6.5	6.1

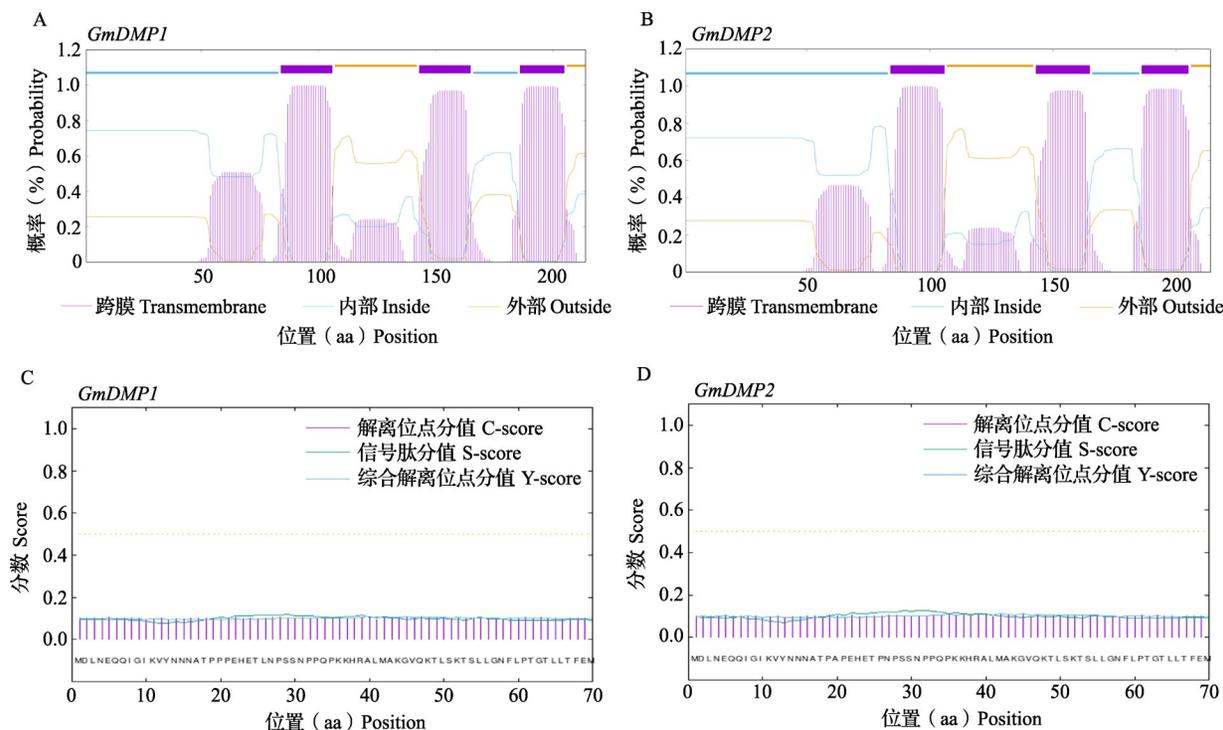
图 5 *GmDMP1* 和 *GmDMP2* 蛋白亚细胞定位Fig.5 Subcellular localization of *GmDMP1* and *GmDMP2* proteins

2.2.3 亲疏水性 蛋白质的亲疏水性是影响和形成蛋白质二级结构稳定性的重要因素之一,从而形成发挥作用的结构域和三级结构。利用工具预测出的大豆 *GmDMP1* 和 *GmDMP2* 基因蛋白的亲疏水性分析如图 6 所示, *GmDMP1* 蛋白的最高值为 3.322, 位于 159 aa 处, 疏水性最强; 最低值为 -3.033, 位于 35 aa 处, 亲水性最强; 且大于 0 的值多于小于 0 的值。 *GmDMP2* 基因氨基酸亲疏水性预测结果与 *GmDMP1* 基因氨基酸亲疏水性预测结果完全一致。理化性质中的脂肪指数也可以作为氨基酸亲疏水性的判定指标, 百分比小于 100% 的为亲水、反之百

分比大于 100% 的为疏水。由表 1 中的脂肪指数以及图 6 中的低峰数可知 *GmDMP1* 和 *GmDMP2* 基因均为亲水蛋白。

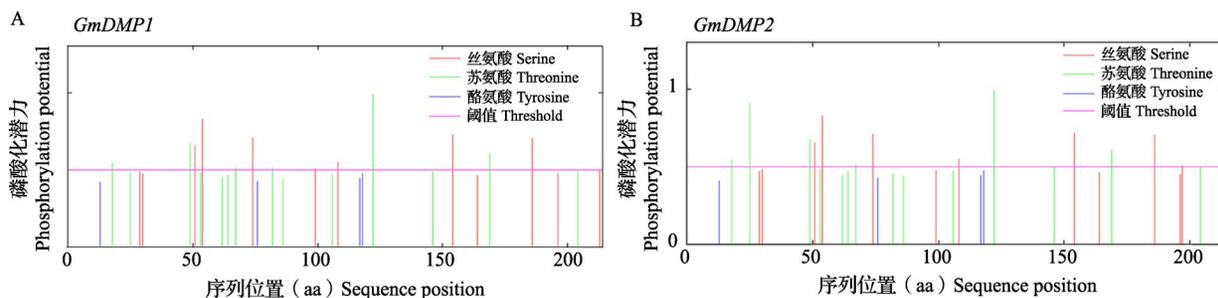
2.2.4 跨膜结构域和信号肽 跨膜结构域是蛋白与膜脂结合的主要区域。通过在线网站对 *GmDMP1* 和 *GmDMP2* 蛋白序列预测的跨膜结构域与信号肽分析结果如图 7 所示。 *GmDMP1* 和 *GmDMP2* 基因蛋白具有跨膜结构域, 属于跨膜蛋白。 *GmDMP1* 和 *GmDMP2* 基因蛋白并不具有信号肽。总结得出 *GmDMP1* 和 *GmDMP2* 基因蛋白均为跨膜的不分泌蛋白。

图 6 *GmDMP1* 和 *GmDMP2* 蛋白亲疏水性Fig.6 *GmDMP1* and *GmDMP2* proteins are hydrophilic and hydrophobic

图7 *GmDMP1*和*GmDMP2*蛋白跨膜结构与信号肽Fig.7 Transmembrane structure and signal peptide of *GmDMP1* and *GmDMP2* proteins

2.2.5 磷酸化位点 蛋白质磷酸化是生物体内重要的共价修饰之一,与信号转导、细胞周期、生长发育及癌症机理有重要关系。*GmDMP1*和*GmDMP2*基因氨基酸序列的磷酸化位点分析结果如图8所示。*GmDMP1*共有30个磷酸化位点,酪氨酸磷酸化位点有4个,分别为Y13、Y76、Y117、Y118;苏氨酸磷酸化位点有14个,分别为T18、T25、T49、T53、T62、T64、T67、T82、T86、T106、T122、T146、T169、T204;丝氨酸磷酸化位点有12个,分别为S29、S30、S51、S54、S74、S99、S108、S154、S164、S186、S196、S213。其中Y118、T122、S54在所属位点中磷酸化

潜力较高,分别为0.477、0.985、0.827。*GmDMP2*共有31个磷酸化位点,酪氨酸磷酸化位点有4个,分别为Y13、Y76、Y117、Y118;苏氨酸磷酸化位点有14个,分别为T18、T25、T49、T53、T62、T64、T67、T82、T86、T106、T122、T146、T169、T204;丝氨酸磷酸化位点有13个,分别为S29、S30、S51、S54、S74、S99、S108、S154、S164、S186、S196、S197、S213。*GmDMP2*中的丝氨酸磷酸化位点比*GmDMP1*的丝氨酸磷酸化位点多一个S197,且磷酸化潜力较高的位点与*GmDMP1*的一致,分别为Y118(0.477)、T122(0.985)、S54(0.827)。

图8 *GmDMP1*和*GmDMP2*蛋白磷酸化位点Fig.8 *GmDMP1* and *GmDMP2* protein phosphorylation sites

2.2.6 大豆 *GmDMP1*和*GmDMP2*位于各组织器官的表达分析 将运用Phytozome的表达数据在Tbtools软件分析得出大豆*GmDMP1*和*GmDMP2*基因不同组织部分的相对表达量(图9)。Phytozome

表达数据显示*GmDMP1*和*GmDMP2*基因均只在花中表达且均在始花期表达量较高,其他组织部位未表达。*GmDMP1*在始花期与盛花期的表达量均高于*GmDMP2*,但数值相差不大。

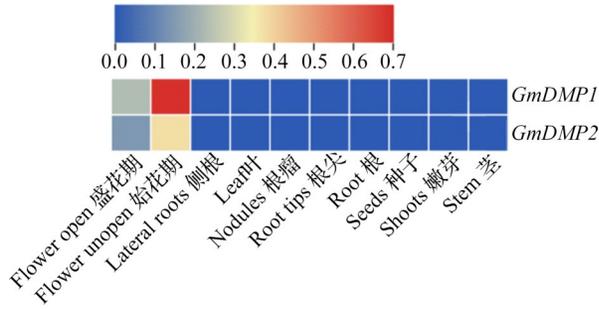


图9 大豆各组织部位 *GmDMP1* 和 *GmDMP2* 基因表达分析
Fig.9 Analysis of *GmDMP1* and *GmDMP2*

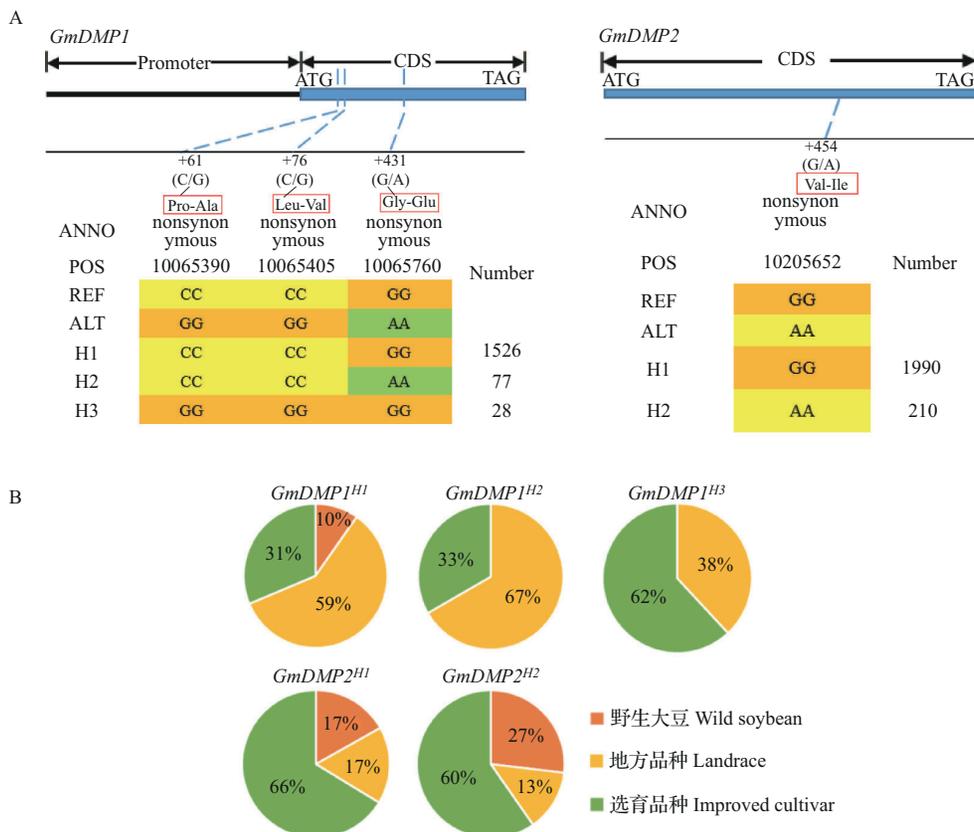
gene expression in soybean tissues

2.2.7 大豆 *GmDMP1* 和 *GmDMP2* 多态性分析 将实验室的 2214 份重测序数据中分析 *GmDMP1* 和 *GmDMP2* 基因组序列的变异情况(表 2)。发现 *GmDMP1* 基因外显子发生了 3 个非同义突变; *GmDMP2* 基因外显子位置发生了 1 个非同义突变。 *GmDMP1* 形成了 3 种单倍型, *GmDMP2* 形成了两种单倍型(图 10)。 *GmDMP1^{H1}* 存在于栽培大豆(包括地方品种与选育品种)和野生大豆中,其中栽培大豆地方品种占比

表 2 *GmDMP1* 和 *GmDMP2* 基因多态性分析

Table 2 Polymorphism analysis of *GmDMP1* and *GmDMP2* genes

基因名称 Gene name	突变类型 Mutation type	碱基变化 Base change			氨基酸序列变化 Amino acid sequence change		
		位点 (bp) Site	原碱基 The original base	突变碱基 Mutant base	位点 (bp) Site	原氨基酸 The original amino acid	突变氨基酸 Mutant amino acid
<i>GmDMP1</i>	非同义突变	+61	C	G	+21	脯氨酸	丙氨酸
	非同义突变	+76	C	G	+26	亮氨酸	缬氨酸
	非同义突变	+431	G	A	+144	甘氨酸	谷氨酸
<i>GmDMP2</i>	非同义突变	+454	G	A	+152	缬氨酸	异亮氨酸



A: *GmDMP1* 与 *GmDMP2* 基因单倍型分析; B: *GmDMP1* 与 *GmDMP2* 基因在野生大豆、栽培大豆地方品种、栽培大豆选育品种间的分布频率
A: The haplotype analysis of *GmDMP1* and *GmDMP2* genes; B: Distribution frequency of *GmDMP1* and *GmDMP2* genes in wild soybean, cultivated soybean local varieties and cultivated soybean selected varieties

图 10 *GmDMP1* 和 *GmDMP2* 单倍型分析

Fig.10 Haplotype analysis of *GmDMP1* and *GmDMP2*

59%,栽培大豆选育品种占比31%,野生大豆占比10%。而 *GmDMP1^{H2}* 和 *GmDMP1^{H3}* 单倍型仅存在于栽培大豆中,其中 *GmDMP1^{H2}* 栽培大豆地方品种占比67%,栽培大豆选育品种占比33%,*GmDMP1^{H3}* 栽培大豆地方品种占比38%,栽培大豆选育品种占比62%。表明该基因在驯化过程中受到了强烈选择。相比之

下,*GmDMP2*在野生大豆与栽培大豆中两种单倍型比例的变化较小。将不同的单倍型进行亲疏水性、结构域、信号肽和磷酸化位点分析,结果表明除了 *GmDMP1^{H3}* 单倍型结构域在54~76 bp位置多了一个跨膜结构域之外,其余的单倍型特性均无变化(表3)。推测该位点的突变可能导致单倍体诱导系功能的产生。

表3 *GmDMP1* 与 *GmDMP2* 基因各单倍型特性的变化情况

Table 3 Changes of haplotype characteristics after *GmDMP1* and *GmDMP2* gene mutations

基因 Gene	单倍型 Haplotype	亲疏水性 Hydrophilic	结构域 Domain	信号肽 Signal peptide	磷酸化位点 Phosphorylation site
<i>GmDMP1</i>	<i>GmDMP1^{H1}</i>	亲水蛋白	与参考基因组一致	无信号肽	与参考基因组一致
	<i>GmDMP1^{H2}</i>	亲水蛋白	与参考基因组一致	无信号肽	与参考基因组一致
	<i>GmDMP1^{H3}</i>	亲水蛋白	多一个跨膜结构域(54~76 bp)	无信号肽	与参考基因组一致
<i>GmDMP2</i>	<i>GmDMP2^{H1}</i>	亲水蛋白	与参考基因组一致	无信号肽	与参考基因组一致
	<i>GmDMP2^{H2}</i>	亲水蛋白	与参考基因组一致	无信号肽	与参考基因组一致

3 讨论

3.1 作物单倍体育种技术创新

原始的作物育种技术是将外源DNA导入到植物基因组中实现基因表达的调节,使得基因沉默或过表达后获得优良性状^[20]。因为大豆的基因组十分复杂,从而造成大豆育种需要消耗很长的时间。双单倍体技术作为大豆遗传改良的技术手段,在大豆品质提高^[21]、生育期调控^[22]、提高优异大豆品种抗逆性能力^[23]等方面提供了重要的技术支持^[24],尤其是加快品系纯化,有助于提高育种效率。通过花粉诱导形成的单倍体诱导系,是创造纯系的主要方法。诱导单倍体产生的方法主要有花药离体培养、小孢子培养、远缘杂交等技术。单倍体植株在杂草曼陀罗中被首次自然发现^[25],曼陀罗也是第一个通过花药培养技术产生单倍体植株的物种^[26]。随后运用小孢子培养技术在小麦^[27]、大麦^[28]、水稻^[29]等作物中诱导出单倍体,而烟草^[30]、小麦^[31]和大麦^[32]则是运用远缘杂交技术获得了母体单倍体株系。

3.2 大豆 *DMP* 单倍体诱导基因

DMP 基因的出现开辟了单倍体育种的新篇章,更多的作物利用该基因创制出单倍体诱导系以达到缩短育种年限的目标,从而改变了传统的育种方法,实现减少时间与成本消耗的高效快速的育种新模式^[33-35]。*DMP* 基因编码 DUF679 结构域,突变后会造成作物产生单倍体植株,是诱导作物产生单倍体的关键基因之一。本研究在大豆中筛选出两个

编码 *DMP* 的同源基因 *GmDMP1* 与 *GmDMP2*。经过生物信息学分析发现两个基因的功能与结构存在较高的相似度,均编码 DUF679 结构域膜蛋白。进化树分析表明这两个基因的蛋白序列与玉米、拟南芥、甘蓝型油菜、紫花苜蓿的 *DMP* 基因序列聚为一类,说明其亲缘关系较近。据预测 *GmDMP1* 与 *GmDMP2* 基因定位于内质网中的可能性最大,而在玉米中 *DMP* 的同源基因蛋白定位于质膜与细胞核中,所以不同物种的 *DMP* 基因可能存在功能差异。在大豆不同组织器官中的表达分析得出该基因仅在花中表达,这与玉米 *DMP* 基因在花器官中表达结论相一致^[15]。因此大豆 *DMP* 基因功能缺失可能影响植物受精从而诱导了母体单倍体的产生。研究表明,大豆的基因遗传多样性会影响油脂与蛋白质^[36]、百粒重^[37]与种子大小与质量等性状^[38]。根据对大豆 *DMP* 基因多态性分析发现突变后 *GmDMP1^{H3}* 单倍型结构域多出一个跨膜结构域,研究发现玉米 *DMP* 基因的多态性变化影响了其单倍体诱导率^[13],该位点的变化可能同样影响大豆单倍体的诱导效率,后续将通过实验对此推测进行进一步研究验证。该基因的发现为尝试打破大豆传统育种方法拓展了新的研究方向。*DMP* 基因在单子叶与双子叶植物中都有诱导单倍体的功能^[15],为利用分子生物育种技术手段解决大豆传统育种周期长的问题提供了新思路。

4 结论

本研究通过对可诱导单倍体株系的 *DMP* 基因

进行大豆同源基因搜索,获得 *GmDMP1* 与 *GmDMP2* 基因,二者的蛋白序列、基因结构、结构域、二三级结构基本一致,相似率达95%以上。经过生物信息学分析表明, *GmDMP1* 与 *GmDMP2* 基因编码的蛋白均属于跨膜的不分泌亲水蛋白,亚细胞定位于内质网的可能性最大。邻接法构建的多物种系统进化树以及编码序列分析表明大豆的 *DMP* 基因与玉米等作物处在同一分支上,亲缘关系较近。利用2214份重测序数据对 *GmDMP1* 与 *GmDMP2* 基因进行多态性分析发现两个基因分别在外显子位置发现了3个和1个非同义突变,形成了3种和2种单倍型,并将突变后单倍型亲疏水性、结构域等特性进行比对分析得到只有 *GmDMP1*^{H3} 单倍型多出了1个54~76 bp的跨膜结构域之外,其余突变后的单倍型特性均无变化。所以本研究将 *GmDMP1* 与 *GmDMP2* 基因定为候选基因,探究其是否拥有构成大豆单倍体诱导系的功能,为后续获得大豆的单倍体纯合诱导系提供了稳定的前期基础。

参考文献

- [1] Chaudhary J, Patil G B, Sonah H, Deshmukh R K, Vuong T D, Valliyodan B, Nguyen H T. Expanding omics resources for improvement of soybean seed composition traits. *Frontiers in Plant Science*, 2015, 6: 1021
- [2] Clemente T E, Cahoon E B. Soybean oil: Genetic approaches for modification of functionality and total content. *Plant Physiology*, 2009, 151(3):1030-1040
- [3] Chen L, Yang H, Fang Y, Guo W, Chen H, Zhang X, Dai W, Chen S, Hao Q, Yuan S, Zhang C, Huang Y, Shan Z, Yang Z, Qiu D, Liu X, Tran L P, Zhou X, Cao D. Overexpression of *GmMYB14* improves high-density yield and drought tolerance of soybean through regulating plant architecture mediated by the brassinosteroid pathway. *Plant Biotechnology Journal*, 2021, 19(4): 702-716
- [4] Liu S, Zhang M, Feng F, Tian Z. Toward a "green revolution" for soybean. *Molecular Plant*, 2020, 13(5): 688-697
- [5] Ainsworth E A, Yendrek C R, Skoneczka J A, Long S P. Accelerating yield potential in soybean: Potential targets for biotechnological improvement. *Plant Cell Environment*, 2012, 35(1): 38-52
- [6] Zhang M, Liu S, Wang Z, Yuan Y, Zhang Z, Liang Q, Yang X, Duan Z, Liu Y, Kong F, Liu B, Ren B, Tian Z. Progress in soybean functional genomics over the past decade. *Plant Biotechnology Journal*, 2022, 20(2): 256-282
- [7] Chaikam V, Molenaar W, Melchinger A E, Boddupalli P M. Doubled haploid technology for line development in maize: Technical advances and prospects. *Theoretical and Applied Genetics*, 2019, 132(12): 3227-3243
- [8] Gajecka M, Marzec M, Chmielewska B, Jelonek J, Zbieszczyk J, Szarejko I. Changes in plastid biogenesis leading to the formation of albino regenerants in barley microspore culture. *BMC Plant Biology*, 2021, 21(1): 22
- [9] Jacquier N M A, Gilles L M, Pyott D E, Martinant J P, Rogowsky P M, Widiez T. Puzzling out plant reproduction by haploid induction for innovations in plant breeding. *Nature Plants*, 2020, 6(6): 610-619
- [10] Coe E H. A line of maize with high haploid frequency. *American Naturalist*, 1959, 93(873): 381-382
- [11] Kelliher T, Starr D, Richbourg L, Chintamanani S, Delzer B, Nuccio M L, Green J, Chen Z, McCuiston J, Wang W, Liebler T, Bullock P, Martin B. MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature*, 2017, 542(7639): 105-109
- [12] Liu C, Li X, Meng D, Zhong Y, Chen C, Dong X, Xu X, Chen B, Li W, Li L, Tian X, Zhao H, Song W, Luo H, Zhang Q, Lai J, Jin W, Yan J, Chen S. A 4 bp insertion at *ZmPLA1* encoding a putative phospholipase a generates haploid induction in maize. *Molecular Plant*, 2017, 10(3): 520-522
- [13] Zhong Y, Liu C, Qi X, Jiao Y, Wang D, Wang Y, Liu Z, Chen C, Chen B, Tian X, Li J, Chen M, Dong X, Xu X, Li L, Li W, Liu W, Jin W, Lai J, Chen S. Mutation of *ZmDMP* enhances haploid induction in maize. *Nature Plants*, 2019, 5(6): 575-580
- [14] Liu G, Lin Q, Jin S, Gao C. The CRISPR-Cas toolbox and gene editing technologies. *Molecular Cell*, 2022, 82(2): 333-347
- [15] Zhong Y, Chen B, Li M, Wang D, Jiao Y, Qi X, Wang M, Liu Z, Chen C, Wang Y, Chen M, Li J, Xiao Z, Cheng D, Liu W, Boutilier K, Liu C, Chen S. A *DMP*-triggered in vivo maternal haploid induction system in the dicotyledonous *Arabidopsis*. *Nature Plants*, 2020, 6(5): 466-472
- [16] Zhong Y, Chen B, Wang D, Zhu X, Li M, Zhang J, Chen M, Wang M, Riksen T, Liu J, Qi X, Wang Y, Cheng D, Liu Z, Li J, Chen C, Jiao Y, Liu W, Huang S, Liu C, Boutilier K, Chen S. *In vivo* maternal haploid induction in tomato. *Plant Biotechnology Journal*, 2022, 20(2): 250-252
- [17] Li Y, Li D, Xiao Q, Wang H, Wen J, Tu J, Shen J, Fu T, Yi B. An in planta haploid induction system in *Brassica napus*. *Journal of Integrative Plant Biology*, 2022, 64(6): 1140-1144
- [18] Wang N, Xia X, Jiang T, Li L, Zhang P, Niu L, Cheng H, Wang K, Lin H. In planta haploid induction by genome editing of *DMP* in the model legume *Medicago truncatula*. *Plant Biotechnology Journal*, 2022, 20(1): 22-24
- [19] Li Y H, Qin C, Wang L, Jiao C Z, Hong H L, Tian Y, Li Y F, Xing G N, Wang J, Gu Y Z, Gao X P, Li D L, Li H Y, Liu Z X, Jing X, Feng B B, Zhao T, Guan R X, Guo Y, Liu J, Yan Z, Zhang L J, Ge T L, Li X K, Wang X B, Qiu H M, Zhang W H, Luan X Y, Han Y P, Han D Z, Chang R Z, Guo Y L, Reif J C, Jackson S A, Liu B, Tian S L, Qiu L J. Genome-wide signatures of the geographic expansion and

- breeding of soybean. *Science China Life Sciences*, 2023, 66 (2): 350-365
- [20] Turner-Hissong S D, Mabry M E, Beissinger T M, Ross-Ibarra J, Pires J C. Evolutionary insights into plant breeding. *Current Opinion in Plant Biology*, 2020, 54: 93-100
- [21] Nagamine A, Ezura H. Genome editing for improving crop nutrition. *Frontiers in Genome Editing*, 2022, 4: 850104
- [22] Lin X, Liu B, Weller J L, Abe J, Kong F. Molecular mechanisms for the photoperiodic regulation of flowering in soybean. *Journal of Integrative Plant Biology*, 2021, 63 (6): 981-994
- [23] Zhang M, Liu S, Wang Z, Yuan Y, Zhang Z, Liang Q, Yang X, Duan Z, Liu Y, Kong F, Liu B, Ren B, Tian Z. Progress in soybean functional genomics over the past decade. *Plant Biotechnology Journal*, 2022, 20(2): 256-282
- [24] Chaudhary M, Mukherjee T K, Singh R, Gupta M, Goyal S, Singhal P, Kumar R, Bhusal N, Sharma P. CRISPR/Cas technology for improving nutritional values in the agricultural sector: An update. *Molecular Biology Reports*, 2022, 49(7): 7101-7110
- [25] Weyen J. Applications of doubled haploids in plant breeding and applied research. *Methods in Molecular Biology*, 2021, 2287: 23-39
- [26] Kelliher T, Starr D, Su X, Tang G, Chen Z, Carter J, Wittich P E, Dong S, Green J, Burch E, McCuiston J, Gu W, Sun Y, Strebe T, Roberts J, Bate N J, Que Q. One-step genome editing of elite crop germplasm during haploid induction. *Nature Biotechnology*, 2019, 37(3): 287-292
- [27] Ferrie A M R, Bhowmik P, Rajagopalan N, Kagale S. CRISPR/Cas9-mediated targeted mutagenesis in wheat doubled haploids. *Methods in Molecular Biology*, 2020, 2072: 183-198
- [28] Cistué L, Echávarri B. Barley isolated microspore culture. *Methods in Molecular Biology*, 2021, 2287: 187-197
- [29] Kuniyoshi D, Masuda I, Kanaoka Y, Shimazaki-Kishi Y, Okamoto Y, Yasui H, Yamamoto T, Nagaki K, Hoshino Y, Koide Y, Takamura I, Kishima Y. Diploid male gametes circumvent hybrid sterility between asian and african rice species. *Frontiers in Plant Science*, 2020, 11: 579305
- [30] Burk L G, Gerstel D U, Wernsman E A. Maternal haploids of *nicotiana tabacum* L from seed. *Science*, 1979, 206 (4418): 585
- [31] Kasha K J, Kao K N. High frequency haploid production in barley (*Hordeum vulgare* L.). *Nature*, 1970, 225 (5235): 874-876
- [32] Satpathy P, de la Fuente S A, Ott V, Müller A, Büchner H, Daghma D E S, Kumlehn J. Generation of doubled haploid barley by interspecific pollination with *hordeum bulbosum*. *Methods in Molecular Biology*, 2021, 2287: 215-226
- [33] Cao P, Zhao Y, Wu F, Xin D, Liu C, Wu X, Lv J, Chen Q, Qi Z. Multi-omics techniques for soybean molecular breeding. *International Journal of Molecular Sciences*, 2022, 23 (9): 4994
- [34] Thudi M, Palakurthi R, Schnable J C, Chitikineni A, Dreisigacker S, Mace E, Srivastava R K, Satyavathi C T, Odeny D, Tiwari V K, Lam H M, Hong Y B, Singh V K, Li G, Xu Y, Chen X, Kaila S, Nguyen H, Sivasankar S, Jackson S A, Close T J, Shubo W, Varshney R K. Genomic resources in plant breeding for sustainable agriculture. *Journal of Plant Physiology*, 2021, 257: 153351
- [35] Varshney R K, Bohra A, Roorkiwal M, Barmukh R, Cowling W A, Chitikineni A, Lam H M, Hickey L T, Croser J S, Bayer P E, Edwards D, Crossa J, Weckwerth W, Millar H, Kumar A, Bevan M W, Siddique K H M. Fast-forward breeding for a food-secure world. *Trends in Genetics*, 2021, 37 (12): 1124-1136
- [36] Zhang H, Goettel W, Song Q, Jiang H, Hu Z, Wang M L, An Y C. Selection of *GmSWEET39* for oil and protein improvement in soybean. *PLoS Genetics*, 2020, 16 (11): e1009114
- [37] Qi Z, Song J, Zhang K, Liu S, Tian X, Wang Y, Fang Y, Li X, Wang J, Yang C, Jiang S, Sun X, Tian Z, Li W, Ning H. Identification of *QTNs* controlling 100-seed weight in soybean using multilocus genome-wide association studies. *Frontiers in Genetics*, 2020, 11: 689
- [38] Duan Z, Zhang M, Zhang Z, Liang S, Fan L, Yang X, Yuan Y, Pan Y, Zhou G, Liu S, Tian Z. Natural allelic variation of *GmST05* controlling seed size and quality in soybean. *Plant Biotechnology Journal*, 2022, 20(9): 1807-1818