

小麦磷脂酶 D α 的基因克隆及其编码序列的生物信息学分析

王俊斌^{1,2} 李明¹ 丁博¹ 包曙光¹ 王锐¹ 谢晓东¹

(1. 天津农学院 天津-布里斯托环境变化对农作物影响研究中心 天津 300384; 2. 天津农学院 基础科学系 天津 300384)

摘要: 为获得更多小麦抗逆基因资源,利用 RT-PCR 方法克隆了磷脂酶 D 基因(PLD),命名为 *TaPLD α* ,并对基因及其蛋白进行了生物信息学分析。结果表明,*TaPLD α* 开放阅读框为 2 394 bp,编码 812 个氨基酸残基,等电点 5.30,分子量为 92 kDa。序列分析显示,*TaPLD α* 基因编码的氨基酸序列含有 N 端 C2 结构域及 2 个保守的活性中心。预测 *TaPLD α* 蛋白是一个亲水性稳定蛋白且定位于细胞质,其二级结构含 28.82% 的 α -螺旋、20.07% 的延伸链、6.03% 的 β -转角和 45.07% 的不规则卷曲。该蛋白不存在信号肽,无跨膜区。*TaPLD α* 与水稻、玉米和拟南芥的 PLD α 氨基酸序列之间具有高度的保守性,进化分析显示,小麦 PLD α 序列与毒麦、水稻和玉米 PLD 序列亲缘关系密切。

关键词: 小麦; 磷脂酶 D α ; 基因克隆; 生物信息学

中图分类号: Q78 文献标识码: A 文章编号: 1000-7091(2013)01-0117-06

Cloning and Sequence Bioinformatics Analysis of Phospholipase D α Gene from Wheat

WANG Jun-bin^{1,2} LI Ming¹ DING Bo¹ BAO Shu-guang¹ WANG Rui¹ XIE Xiao-dong¹

(1. Tianjin-Bristol Research Center for the Effects of the Environment Change on Crops Agronomy, Tianjin Agricultural University, Tianjin 300384, China; 2. Department of Basic Sciences, Tianjin Agricultural University, Tianjin 300384, China)

Abstract: To obtain more wheat (*Triticum aestivum* L.) stress-response genes, a phospholipase D gene, named *TaPLD α* , was cloned by RT-PCR. And the protein of this gene was also predicted and studied through the bioinformatics analysis method. The results showed that the largest open reading frame of *TaPLD α* gene has 2 394 bp in length and encoded a polypeptide of 812 amino acid residues. The estimated molecular weight and isoelectric point of the putative protein were 92 kDa and 5.30, respectively. Sequence analysis showed that the amino acid sequence encoded by *TaPLD α* gene contained N-terminal C2 domain and two HKD motifs. *TaPLD α* protein was a stable hydrophilic protein and located in the cytoplasm. The predicted secondary structure composition for the protein has about 28.82% alpha helices, 20.07% extended strand, 6.03% beta turn and 45.07% random coil. The *TaPLD α* had no signal peptide and transmembrane helices. Compared with rice, maize and Arabidopsis, amino acid sequence encoded by *TaPLD α* was almost conserved. The phylogenetic tree showed that *TaPLD α* was most similar to PLD α protein from dandelion, rice and maize.

Key words: Wheat; Phospholipase D α ; Gene cloning; Bioinformatics

磷脂是构成细胞膜的基本骨架,也是细胞对外界环境刺激作出应答的物质基础。激活磷脂酶是细胞响应刺激启动信号转导的初始步骤。磷脂酶 D (Phospholipase D, PLD) (EC 3.1.4.4) 能水解磷脂,

产生磷脂酸 (Phosphatidic acid, PA) 和一个极性头部。1947 年, Hanahan 和 Chaikoff^[1] 首次在胡萝卜根和白菜叶中发现了 PLD,但直到 1994 年有活性的植物 PLD 基因被分离克隆成功后^[2],人们才对其进

收稿日期: 2012-11-17

基金项目: 国家“863”计划项目(2011AA100104); 天津市应用基础与前沿技术研究计划项目(11JCYBJC09100); 国家转基因新品种培育重大专项(2009ZX08009-084B)

作者简介: 王俊斌(1979-),男,内蒙古呼和浩特人,副研究员,博士,主要从事植物逆境生理研究。

通讯作者: 谢晓东(1975-),男,内蒙古突泉人,研究员,博士,主要从事植物抗逆机理与分子育种的研究。

行了较为深入细致的研究。

PLD 是植物中最主要的磷脂酶,参与植物种子形成、植株发育和对生物及非生物胁迫响应等多个过程^[3]。迄今为止,已经从蓖麻、拟南芥、草莓、豇豆、甘蓝、烟草、棉花、水稻和玉米等多种植物中克隆到 PLD 基因。PLD 基因属于多基因家族,具有多分子异质性。拟南芥中已经确认了 12 个 PLD 基因,分为 PLD α (1, 2, 3)、PLD β (1, 2)、PLD γ (1, 2, 3)、PLD δ 、PLD ϵ 及 PLD ζ (1, 2) 6 类^[4-6]。PLDs 在底物选择特异性、亚细胞定位和基因时空表达特性等方面均有明显差异,暗示不同 PLD 的生理功能的多样性^[7]。

作为重要的信号分子,PLD 和 PA 能够响应干旱、盐胁迫、低温冷害和脱落酸(Absciscic acid, ABA)而激活,在植物抵抗逆境胁迫中发挥重要作用^[8-10]。PLD α 1 缺失的拟南芥气孔不能响应 ABA 正常关闭,外加 PA 可以恢复气孔关闭^[11-12]。抑制 PLD α 1 导致多个参与干旱条件下转录、信号转导、脂类信号、激素响应和代谢的基因表达发生变化^[13]。这些结果表明,PLD α 1 和 PA 在 ABA 对植物缺水的影响过程中发挥正调控作用。同时也说明 PLD 在农作物抗逆遗传改良方面有很高的潜在应用价值。然而,作为重要的粮食作物,到目前为止,对小麦 PLD 基因的克隆及功能鉴定的研究还未见报道。本研究克隆了小麦 PLD 家族基因 PLD α ,并进行生物信息学分析,以为小麦的基因功能研究和分子育种提供参考。

1 材料和方法

1.1 材料和主要试剂

供试小麦品种扬麦 158 是典型的异源六倍体普通小麦(*Triticum aestivum* L., 染色体组为 AABBDD, $2n=6x=42$),目前是长江中下游地区主栽品种,由本中心保存;感受态细胞 DH5 α 、植物总 RNA 提取试剂盒购于 TIANGEN 有限公司; cDNA 合成试剂盒购于 Invitrogen 公司; Phusion 超保真 DNA 聚合酶购于 NEB 公司; TA 克隆试剂盒、pGEM-T Vector Kit、DNA 回收纯化试剂盒购于 Takara 公司;测序由北京博迈德生物有限公司完成。

1.2 序列获得及引物设计

以模式生物的 PLD 序列作为查询序列,检索小麦 EST 数据库,对检索到的序列进行拼接和再检索,直到没有新的 PLD 基因的 EST 序列可供拼接为止。以拼接得到的序列为基础,设计上游引物: ATGGCTCAGATCTTGCTC,下游引物: CTATGTGGTGAGGATGGG。

1.3 总 RNA 提取与目的基因的克隆

对长到两叶一心期的小麦进行干旱处理,取材后液氮冷冻,储藏在 -80°C 冰箱中备用。总 RNA 提取和 cDNA 合成均按说明书进行。以反转录获得的 cDNA 为模板扩增目的基因,反应体系为 cDNA 模板 1 μL , 5 \times Phusion HF Buffer 10 μL , 10 mmol/L dNTP 1 μL , 10 $\mu\text{mol/L}$ 上下游引物各 1 μL , ddH₂O 35.5 μL 。反应条件为 98°C 预变性 30 s, 98°C 变性 10 s, 58°C 退火 20 s, 72°C 延伸 2 min, 35 个循环; 72°C 延伸 5 min。PCR 产物经 1% 琼脂糖电泳分离后回收纯化,回收片段连接到 pGEM-T Vector,并转化到感受态大肠杆菌 DH5 α 中,挑选阳性克隆进行测序。

1.4 生物信息学分析

利用 CDD 程序(<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)对测定序列进行保守结构域预测;采用 ORF finder 程序(<http://www.ncbi.nlm.nih.gov/gorf>)寻找开放阅读框;利用 Pfam 26.0(<http://pfam.sanger.ac.uk/search>)软件中的 SEARCH 功能对基因进行家族预测。通过蛋白质分析系统 ProtParam tool(<http://www.Expasy.ch/tools/protparam.html>)对蛋白质的各种理化性质进行分析;在线软件 SOPMA(<http://npsa-pbil.ibcp.fr/>)预测蛋白二级结构。利用 CELLO(<http://cello.life.nctu.edu.tw/>)网站对蛋白质进行亚细胞定位的预测;通过 SignalP 4.0(<http://www.cbs.dtu.dk/services/SignalP/>)预测该蛋白的信号肽;利用 ProtScale(<http://web.expasy.org/protscale/>)以默认算法(Hphob./Kyte&Doolittle)对该蛋白进行疏水性预测。利用在线软件 TMHMM Server 2.0(<http://cbs.dtu.dk/services/TMHMM/>)预测蛋白跨膜结构域。利用 ClustalX 2.0 软件进行蛋白质多重序列比对,用 MEGA 5.0 软件,以邻近法构建进化树。

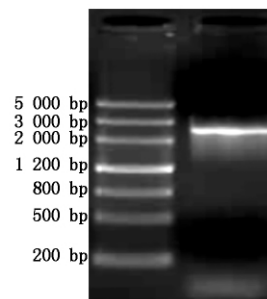


图1 *TaPLD α* 开放阅读框扩增产物

Fig.1 Products of *TaPLD α* ORF

2 结果与分析

2.1 小麦 PLD α 基因的克隆

以干旱处理后小麦 cDNA 为模板进行 RT-PCR

扩增,产物经 1% 琼脂糖凝胶电泳检测,获得约 2 500 bp 单一条带,与预期相符,结果如图 1 所示。将目的片段与 pGEM-T Easy 载体连接,获得重组质粒 pGEM-T Easy-PLD,经测序表明该片段长 2 574 bp。利用 NCBI 中相关软件(包括 blastp 和 blastx)进行搜索,其与拟南芥 PLD α 同源性较高,命名为 TaPLD α 。

2.2 TaPLD α 编码氨基酸序列分析

经 NCBI 的 ORF finder 程序分析, TaPLD α 编码

的氨基酸序列(图 2) 含有一个完整的开放阅读框,长 2 439 bp,编码 812 个氨基酸残基。在起始密码子 ATG 上游 -3 位核苷酸为 A, +4 位为 G,符合典型的 Kozak 序列规则。TaPLD α 序列包含两个 HKD 基序,分别为 HQKIVCVD 和 HTKMMIVD。两基序间隔 320 个氨基酸。采用 CDD 程序分析小麦 PLD α 氨基酸序列的结构域,结果显示,存在 3 个保守结构域,分别为 C2 结构域和 2 个 PLDc 结构域(图 3)。

```

MAQILLHGNLHVTFEASSLSHPGRASGGAPKFKFVEGIEETVGVGKGSSKLY
ATIDLEKARVGRTRMLGNEPVNPRWYESFHIYCAHLAADVIFTMKADNAIGATL
IGRAYLPVAELLEGEEDRWLEICDDKREPVGESKIHVKLQYFGVDKDRNWARG
VRSVKFPFVPYTFFSQRQGCNVRLYQDAHVDPNFIPIPLADGKNYEPARCWED
IFDAISNAQHLYIITGWSVHTEITLIRD TNRPKPGDVTLGELLKRKASEGVRVL
MLVWDDRTSVGLLKRDLGMLATHDEETANYFQGTDVHCVLCPRNPDDSGSIVQ
DLQISTMFTHHQKIVCVDALPSQSGSEQRRLSFVGGIDLCGRYDTQYHSLFRT
LDTVHDDDFHQPNFATASITKGGPREPWHDIHSRLEGPIAWDVLNFEQRWRKQ
GGKDLLVQLRDLSDIIPSPVMFPEDRDTWNVQLFRSIDGGAAGFPDTPPEAA
RAGLVSGKDQIHDRSIQDAYINAIRAKDFIYIENQYFLGSSYCWKPEGIKPEEIG
ALHVIPKELSLKIVSKIEAGERFTVYVVPWMPGEMPEASVQAILDWQRRTME
MMYTDITQALEAKGIEANPKEYLTFFCLGNREVKQDGEYEPQEQPEDTDYVR
AQEARRFMIYVHTKMMIVDDEYIIIGSANINQRSMDGARDSEIAMGAYQPYHLA
NREPARGQIHGFRMALWYEHGLMLDDVFQRPESVECVQKVNRIAEKYWDIYSS
DDLEQDLPGHLLSYPIGVASDGVVTELPGMEFFPDTRARILGAKSDYLPILT
  
```

横线部分为 C2 结构域,方框内部分为 HKD 基序。

The sequence with underline means C2 domain and the conserved HKD motifs are boxed.

图 2 TaPLD α 编码的氨基酸序列

Fig. 2 Amino acid sequences of TaPLD α

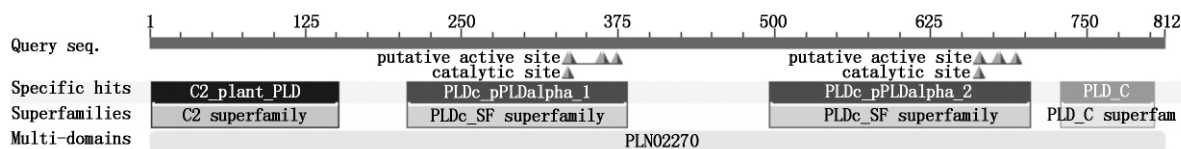


图 3 TaPLD α 编码蛋白的保守结构域分析

Fig. 3 Conserved domain of TaPLD α protein

2.3 TaPLD α 编码氨基酸序列生物信息学分析

通过 ProtParam 推测 TaPLD α 蛋白的基本理化特性,分子量为 92 094.3 Da,等电点为 5.30,为酸性蛋白。负电荷残基(Asp + Glu)总数为 119 个,正电荷残基(Arg + Lys)总数为 88 个。理论推导半衰期是 30 h,不稳定系数是 39.77,为稳定型蛋白质。疏水性分析(图 4)结果表明, TaPLD α 亲水性区域均匀分布在整条肽链中,且分值高,而疏水性区域较少,分值低,因此, TaPLD α 在一级结构上以亲水性为主。

利用 Pfam 26.0 软件中的 SEARCH 功能预测 TaPLD α 基因所属的蛋白家族,结果表明,该基因为 PLD 基因家族,具有 C2 结构域和 PLD 活性位点。通过 SOPMA 对该蛋白的二级结构进行预测(图 5)。结果表明, α -螺旋的比例达到了 28.82%, β -转角占 6.03%,不规则卷曲占 45.07%,另外还有 20.07% 为延伸链。

亚细胞定位工具 CELLO 推测 TaPLD α 蛋白定位于细胞质。TMHMM Server 2.0 预测其没有跨膜

区域。信号肽分析表明, TaPLD α 不存在信号肽酶切位点,可以推断, TaPLD α 在细胞质中合成后,不进行蛋白转运。

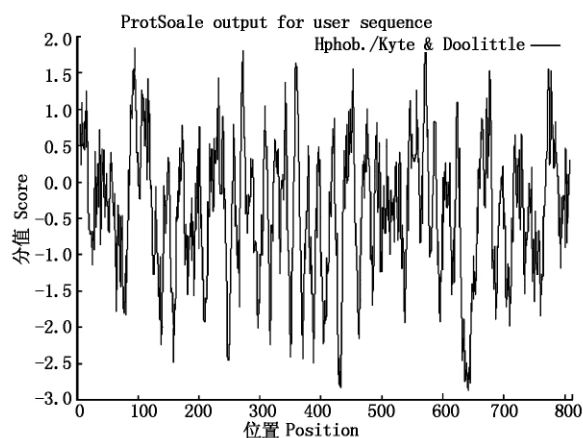


图 4 TaPLD α 蛋白质的疏水性/亲水性分析

Fig. 4 Hydrophobicity/hydrophilicity analysis of TaPLD α protein

利用 ClustalX V2.0 软件将 TaPLD α 与玉米(Zea mays)、水稻(Oryza sativa)和拟南芥(Arabidop-

蛋白质与单子叶植物毒麦(*Lolium temulentum*)、水稻和玉米亲缘关系较近(图7)。

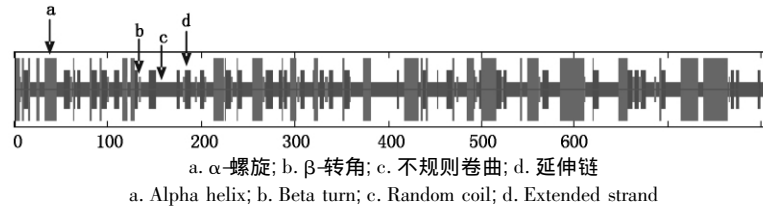


图5 SOPMA 预测的 TaPLD α 蛋白二级结构

Fig. 5 Predicted secondary structure of TaPLD α protein by SOPMA

[illegible]

“*”、“.”和“.”分别代表各序列之间完全一致、保守和半保守的氨基酸残基。

The symbols “*”, “.” and “.” represented completely identical, conservative and semi-conservative amino acid residues respectively.

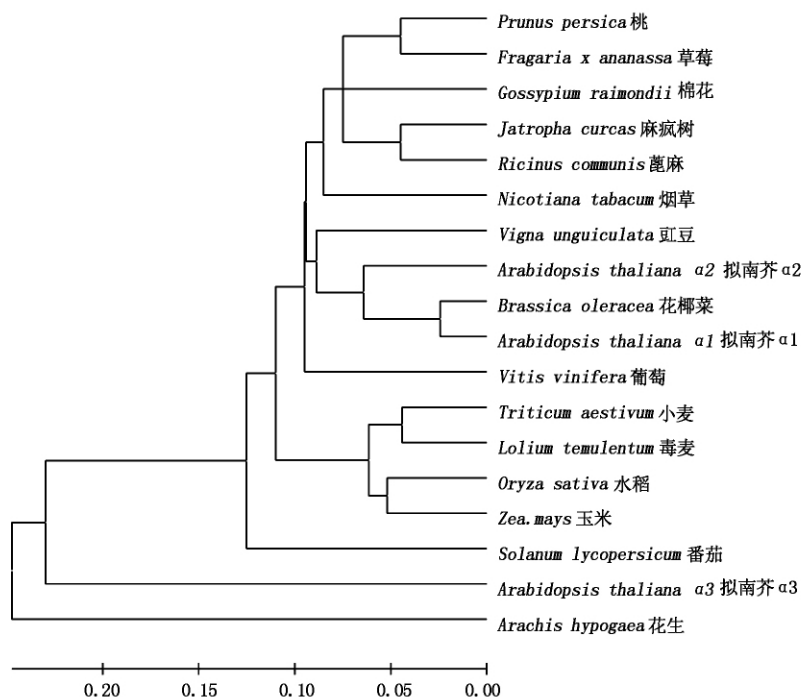
图6 TaPLD α 蛋白序列与其他植物同源蛋白质序列的比较

Fig. 6 The Multi-alignment of TaPLD α amino acid sequences with other plant phospholipase D α s

3 讨论

PLD 是植物体内磷脂代谢途径的关键酶,在植物细胞信号转导、膜运输及降解、脂降解、胞分化 and 生殖、细胞防御反应中均发挥重要作用^[14]。*PLD α* 是植物中分布最广泛的基因型,在幼嫩叶片中较丰富,在细胞内主要分布于质膜、内膜和胞质部

分^[15]。尽管它在烟草、拟南芥等模式植物中的研究已相对比较成熟,展现出了十分诱人的前景,并取得了许多预期的理想结果,但它真正在农业中应用的还比较少,特别是在重要作物小麦中尚未见成功的报道。在本研究中,我们克隆到了小麦 *PLD α* 全长序列并进行生物信息学分析。序列分析表明,小麦 *PLD α* 蛋白由 812 个氨基酸组成,与 NCBI 中已经登

图7 小麦与其他植物 PLD α 的系统进化树Fig. 7 Phylogenetic tree of wheat and other homologous PLD α sequences

录的其他植物的 PLD 核苷酸相似性高于 80% ,氨基酸序列相似性高于 90% 。该基因编码的氨基酸序列具有植物 PLD 的典型结构域 ,如 N 端的 C2 结构域及 2 个 HKD 活性中心。C2 结构域是植物 PLD 特有的结构 ,是一个 Ca^{2+} 与磷脂结合的折叠区 ,该结构域中二分环 (Bipartite loop) 上的 4 ~ 5 个氨基酸的性质与 Ca^{2+} 的结合有关^[16]。在拟南芥中 , Ca^{2+} 结合到 PLD α 1 的 C2 域 ,引起酶的构象变化 ,从而能够增强 C2 结构域对磷脂酰胆碱 (PC) 的亲合力^[17]。植物 PLD 具有高度保守性 ,所有 PLD 氨基酸序列都包括 2 个 HxKxxxxD 基序 ,即 HKD1 ,HKD2 ,两基序间隔约 320 个氨基酸。H、K、D 分别为 His(组氨酸)、Lys(赖氨酸)、Asp(天冬氨酸)。HKD 基序是 PLD 的标志序列 ,也是催化水解的活性部位^[18-19]。H、K、D 残基在 PLD 活性中的作用已在酵母 PLD 特异位点诱变试验中得到证明 ,改变其中任何 1 个残基都会导致 PLD 活性的丧失^[20]。

由于 PLD 基因在植物中属于一个多基因家族 ,且在不同种类中保守性较高 ,因此 ,将 TaPLD α 的氨基酸序列与水稻、玉米和拟南芥的 PLD α 氨基酸序列进行了比对 ,结果表明同源性很高。进化分析表明 ,小麦 PLD α 氨基酸序列与毒麦、玉米和水稻亲缘关系近。推测这是由于小麦与毒麦、水稻、玉米同属单子叶禾本科植物 ,因此 PLD α 基因的氨基酸序列同源性相对较高。

小麦是我国和世界范围内主要的粮食作物之

一^[21] ,对 TaPLD α 基因进行克隆和生物信息学分析 ,为研究该基因在小麦生长发育和逆境响应过程中的作用奠定了基础。利用生物信息学软件对基因进行克隆、结构和功能预测已成为生物学研究领域的热点和方向之一 ,但由于生物大分子结构和功能的复杂性 ,预测结果仍然需要进行验证^[22]。在模式物种中 PLD α 基因属于多基因家族 ,普通小麦是一个由 A、B、D 这 3 个基因组构成的异源六倍体物种 ,本研究克隆的 TaPLD α 基因只是小麦 PLD 基因家族之一 ,对其他基因家族成员的克隆和基因组定位有待进一步研究。

参考文献:

- [1] Hanahan D J ,Chaikoff I L. A new phosphorus-containing lipids of the carrot [J]. J Biol Chem ,1947 ,168: 233 - 240.
- [2] Wang X ,Xu L ,Zheng L. Cloning and expression of phosphatidylcholine-hydrolyzing phospholipase D from *Ricinus communis* L [J]. J Biol Chem ,1994 ,269: 20312 - 20317.
- [3] Wang X. The role of phospholipase D in signal transduction cascade [J]. Plant Physiol ,1999 ,121: 456 - 462.
- [4] Li M ,Hong Y ,Wang X. Phospholipase D- and phosphatidic acid-mediated signaling in plants [J]. Biochim Biophys Acta 2009 ,1791: 927 - 935.
- [5] Qin C ,Wang X. The *Arabidopsis* phospholipase D family: Characterization of a calcium independent and phosphatidylcholine-selective PLD γ 1 with distinct regulatory domains [J]. Plant Physiol 2002 ,128: 1057 - 1068.

- [6] Wang C ,Wang X. A novel phospholipase D of *Arabidopsis* that is activated by oleic acid and associated with the plasma membrane [J]. *Plant Physiol* ,2001 ,127: 1102 – 1112.
- [7] Wang X. Regulatory functions of phospholipase D and phosphatidic acid in plant growth , development , and stress responses [J]. *Plant Physiol* ,2005 ,139: 566 – 573.
- [8] Hong Y ,Zhang W ,Wang X. Phospholipase D and phosphatidic acid signaling in plant response to drought and salinity [J]. *Plant Cell Environ* 2010 ,33: 627 – 635.
- [9] Bargmann B O ,Laxalt A M ,ter Riet B ,*et al.* Multiple PLDs required for high salinity and water deficit tolerance in plants [J]. *Plant Cell Physiol* 2009 ,50: 78 – 89.
- [10] Hong Y ,Pan X ,Welti R *et al.* Phospholipase D α 3 is involved in the hyperosmotic response in *Arabidopsis* [J]. *Plant Cell* 2008 ,20: 803 – 816.
- [11] Sang Y ,Cui D ,Wang X. Phospholipase D and phosphatidic acid-mediated generation of superoxide in *Arabidopsis* [J]. *Plant Physiol* 2001 ,126: 1449 – 1458.
- [12] Mishra G ,Zhang W ,Deng F *et al.* A bifurcating pathway directs abscisic acid effects on stomatal closure and opening in *Arabidopsis* [J]. *Science* ,2006 ,312: 264 – 266.
- [13] Mane S P ,Vasquez-Robinet C ,Sioson A A *et al.* Early PLD α 1-mediated events in response to progressive drought stress in *Arabidopsis*: A transcriptome analysis [J]. *J Exp Bot* 2007 ,58: 241 – 252.
- [14] 万嗣宝 ,张 斌 ,战吉成 ,等. 桃果实磷脂酶 D α 基因的电子克隆及序列分析 [J]. *中国农业科学* 2009 ,42 (5) : 1705 – 1712.
- [15] Zhang W ,Wan X ,Hong Y *et al.* Plant phospholipase D [M]// Munnik D T . *Lipid signaling in plants* (plant cell monographs) . Berlin: Springer publishing company , 2010: 39 – 62.
- [16] Ponting C P ,Parker P J. Extending the C2 domain family: C2s in PKCs δ , ϵ , η , θ ,phospholipases ,GAPs ,and perforin [J]. *Protein Sci* ,1996(5) : 162 – 166.
- [17] Zheng L ,Krishnamoorthi R ,Zolkiewski M *et al.* Distinct Ca²⁺-binding properties of novel C₂ domains of plant phospholipase D α and β [J]. *J Biol Chem* ,2000 ,275: 19700 – 19706.
- [18] Munnik T ,Arisz S A ,de Vrije T *et al.* G protein activation stimulates phospholipase D signaling in plants [J]. *The Plant Cell* ,1995(7) : 2197 – 2210.
- [19] Sung T C ,Roper R L ,Zhang Y *et al.* Mutagenesis of phospholipase D defines a defines a superfamily including a trans-golgi viral protein required for poxvirus pathogenicity [J]. *The EMBO Journal* ,1997 ,16(15) : 4519 – 4530.
- [20] Pointing C P ,Kerr I D. A novel family of phospholipase D homologues that include phospholipid synthases and putative endonucleases: Identification of duplicated repeats and potential active site residues [J]. *Protein Sci* , 1996(5) : 914 – 922.
- [21] 时晓伟. CIMMYT 小麦品质育种策略 [J]. *天津农业科学* 2011 ,17(3) : 108 – 111.
- [22] 李 萌 ,张爱华 ,白雪飞 ,等. 甜瓜玉米黄质环氧化酶基因的电子克隆及生物信息学分析 [J]. *华北农学报* 2011 ,26(S) : 23 – 28.