

羊草 *CDPK1* 基因的克隆及序列分析

郭园艺, 慕小倩*, 麻鹏达*

(西北农林科技大学生命科学学院, 陕西 杨凌 712100)

摘要: 羊草是一种优质的天然牧草, 对于盐碱胁迫有着很强的耐受性。本研究通过羊草 (*Leymus chinensis*) EST (Expressed Sequence Tags, ESTs) 序列和 RACE (Rapid-amplification of cDNA Ends, RACE) 技术, 获得了 *LcCDPK1* 基因序列, 通过生物信息学分析, 得到 *LcCDPK1* 的全长基因, 含 1 599 bp 的开放读码框 (Open Reading Frame, ORF), 编码含有 532 个氨基酸的不稳定弱酸性亲水蛋白质, 同时对得到的 *LcCDPK1* 基因进行了进化分析。本研究有助于了解 *LcCDPK* 基因家族的功能, 为进一步应用 *LcCDPK* 基因进行植物的耐盐碱性遗传改良奠定了基础。

关键词: 羊草; CDPK 蛋白; RACE; 生物信息学

中图分类号: Q943.2

文献标识码: A

文章编号: 2096-5877(2022)01-0044-07

Cloning and Sequence Analysis of *CDPK1* Gene from *Leymus chinensis*

GUO Yuanyi, MU Xiaoqian*, MA Pengda*

(College of Life Science, Northwest A&F University, Yangling 712100, China)

Abstract: *Leymus chinensis* is an important native forage grass with strong adaptability to saline-alkali. In this study, the full-length gene of *LcCDPK1* was cloned by EST and RACE, with an ORF of 1 599 bp, which encodes an unstable, weakly acidic hydrophilic protein containing 532 amino acids. And a phylogenetic tree is constructed with the *LcCDPK1* sequence. This study would help to further understand the function of the CDPK protein family of *Leymus chinensis* and provide a basis for genetic improvement of plants' saline-alkali tolerance with *LcCDPK1*.

Key words: *Leymus chinensis*; CDPK protein; RACE; Bioinformatics analysis

羊草 (*Leymus chinensis*) 又名碱草, 禾本科赖草属, 是一种优质牧草, 在我国多分布于东北三省及西北地区^[1]。羊草是多年生根茎型禾草, 适应盐碱条件下的生长环境^[2-3], 说明其具有耐盐碱的生理机制, 在面对非生物胁迫时能够高效表达帮助其适应不同逆境的物质。调控这些物质表达的相关信号通路、各级转录因子和酶类基因是其抗胁迫的关键环节^[4]。目前, 草地盐碱化、沙漠化面积日益扩大, 而关于羊草抗逆相关基因的报道较少, 羊草对盐碱的耐受性分子机制还不明确^[5], 因此, 挖掘羊草抗旱耐盐碱相关基因, 对培育抗旱耐盐碱的转基因作物, 增进我国科技原创, 提升我国农业科技生产力水平具有重要意义。

类钙调素结构域的蛋白激酶 (Calcium-like

Protein Kinase, CDPK) 属于 Ser/Thr 型蛋白激酶, 在植物的各种组织器官均有表达, 这种特异性表达受到干旱、低温、盐碱等多种非生物胁迫的影响^[6]。CDPK 在激酶家族中相对研究较多、了解较为清楚, 其同系物众多, 结构上相对保守^[7]。典型的 CDPK 蛋白含有 N 端可变区、ATP 结合的激酶区、连接区和调控区 4 个功能域^[5-6]。CDPK 蛋白独特的结构使它同时具备 Ca^{2+} 感应器和效应器的功能, 并通过解除自抑制使之具有活性^[7]。外界胁迫能够引起细胞质 Ca^{2+} 浓度发生变化, 产生钙信号, 在不需要钙调素的情况下通过钙信号激活 CDPK, 并完成对底物蛋白的磷酸化, 随后通过特定的通路调控下游基因表达, 从而影响植物的多种生理过程^[7-9]。目前, 拟南芥、小麦、玉米、水稻等作物中已分离得到多个 *CDPK* 基因^[7-10]。

有关羊草的 *CDPK* 基因研究报道较少。2015 年, 崔喜艳等报道羊草中克隆得到了一个 *CDPK* 基因, 通过模拟干旱、盐碱条件分析该 *CDPK* 基因的表达量, 发现其表达量对干旱、盐碱胁迫非常敏感^[11-12]。本研究根据 EST 的参考序列, 通过 RT-

收稿日期: 2019-12-16

基金项目: 中央高校基本科研业务费专项基金 (2452018150)

作者简介: 郭园艺 (1996-), 女, 在读硕士, 主要从事植物次生代谢研究工作。

通讯作者: 慕小倩, 女, 硕士, 教授, E-mail: muxiaoqian@126.com

麻鹏达, 男, 博士, 副教授, E-mail: mapengda@163.com

PCR技术和RACE技术从羊草中克隆到另一个钙依赖蛋白激酶,命名为LcCDPK1,并对其进行生物信息学分析。本研究不仅可以丰富禾本科植物CDPK基因家族的信息,同时为进一步研究羊草的抗逆机制奠定基础^[4]。

1 材料与方法

1.1 植物材料

挑选大小均一饱满的羊草种子置于盛满蛭石的花盆中,人工气候培养箱萌发,25 °C恒温培养^[3],光强度为11 000~13 000 lx,光照周期为16 h/d,出苗后每天浇1次0.5倍Hoagland营养液至长出4叶期。

1.2 羊草RNA的提取及反转录

羊草总RNA的提取按照天根生化科技(北京)有限公司的总RNA提取试剂盒(RNAprep pure Plant Kit)操作,用Nanodrop 2000检测所提RNA的浓度,用1%的琼脂糖凝胶电泳检测所提RNA的纯度和完整性。总RNA的反转录按照普洛麦格(北京)生物技术有限公司的反转录试剂盒(Reverse Transcription System)进行操作。

1.3 RACE克隆

通过Genbank检索,得到LcCDPK1蛋白的EST序列(登录号:CD808794)。经过NCBI BLASTX比对发现所得基因序列为LcCDPK1基因的3'端。基于LcCDPK1基因的EST序列,通过Primer Primer 5.0设计得到LcCDPK1基因的5'RACE特异性引物。根据Clontech公司的RACE试剂盒(SMARTTM RACE cDNA Amplification Kit)说明书进行PCR反应。通过DNAMAN拼接序列得到LcCDPK1基因的全长cDNA。

1.4 序列拼接及基因阅读框分析

通过NCBI(<http://www.ncbi.nlm.nih.gov/>)对LcCDPK1基因5'RACE的结果和原有EST序列进行拼接。通过NCBI中的ORF finder在线程序(<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>)查找序列的开放阅读框。

1.5 蛋白理化性质分析

通过ProtParam(<https://web.expasy.org/prot-param/>)进行LcCDPK1基因核酸及氨基酸序列的理化性质分析,通过ProtScale(<https://web.expasy.org/protscale/>)在线软件对其进行亲疏水性分析,通过NetPhos 3.1(<http://www.cbs.dtu.dk/services/Net-Phos/>)在线软件对LcCDPK1蛋白序列进行激酶磷酸化修饰位点预测,通过NetOGlyc 3.1(<http://www.cbs.dtu.dk/services/NetOGlyc-3.1/>)在线软件进行糖

基化修饰位点预测,通过SignalP 4.1(<http://www.cbs.dtu.dk/services/SignalP/>)在线软件预测其所含的信号肽,通过TMHMM(<http://www.cbs.dtu.dk/services/TMHMM/>)在线软件预测其所含的跨膜结构域,通过Psort Prediction(<http://psort1.hgc.jp/form.html>)在线软件预测蛋白的亚细胞定位,通过COILS(http://embnet.vital-it.ch/software/COILS_form.html)在线软件对序列进行卷曲螺旋预测,通过ScanProsite(<https://prosite.expasy.org/scanprosite/>)在线软件预测其所包含的保守结构域,通过PSIPRED(<http://bioinf.cs.ucl.ac.uk/psipred/>)在线软件和SOPMA(https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)在线软件对LcCDPK1蛋白的二级结构进行分析,并通过Phyre2(<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>)在线服务器和PyMol软件对其三维结构进行预测分析及处理^[4]。

1.6 系统发育树构建

通过NCBI中的BLAST,对LcCDPK1蛋白进行序列比对,分析序列的同源性。通过DNAMAN 6.0进行多序列比对,MEGA 5.0采用邻位相连法构建系统进化树(Neighbor-Joining,NJ)(图1),设置重复1 000次,其他均为默认设置。

2 结果与分析

2.1 LcCDPK1基因全长cDNA的获得

实验通过5'RACE,得到长656 bp的5'末端TA克隆产物。将5'RACE扩增序列与LcCDPK1基因EST序列进行拼接,根据拼接序列设计引物进行全长cDNA扩增,产物经测序长为2 136 bp。通过“ORF finder”在线程序分析,结果显示克隆得到的LcCDPK1基因序列包括一个1 599 bp的开放阅读框,294 bp的5'UTR和243 bp的3'UTR,该LcCDPK1基因编码具有532个氨基酸的蛋白质(图1)。

2.2 LcCDPK1蛋白的生物信息学分析

2.2.1 LcCDPK1基因编码蛋白质的理化性质

通过ProtParam在线分析工具对LcCDPK1基因编码的蛋白质氨基酸序列进行分析,结果(表1)显示LcCDPK1蛋白由碳、氢、氧、氮、硫5种元素组成,其分子式为C₂₆₀₆H₄₁₃₆N₇₂₂O₈₀₀S₂₉,相对分子质量为59 311.61,理论等电点pI=5.79<7,是一个弱酸性蛋白质。从氨基酸组成特点来看,CDPK1蛋白包含20种常见氨基酸,其中含量最高的为Ala(7.89%),其次为Glu(7.71%)和Gly(7.52%),含量

1 CTAATACGACTACTATAGGGCAAGCAGTGGTATCAACGAGAGTACGGGGAATCGCGGCCAACCTCCAA
 76 TCACCTCTCGGGTCCCGTCCGAGCTGTTCTTGGCTCTCTCACTCAAGATTGTTCTTTAGAGCC
 151 TGCAGCTTCGGTCCGAGGAGCGCGGAGACTCGGGACTTGGTTCAAGCTTAGGATCTTTCGATTAGGTT
 226 TTGATTGGTTTCGCTCGCTCGCTGACTGGTGTAAATTTGGAATCGAAGGAAGATTTCAGCAGATGGGC
 M G

301 CAGTGTTCAGCAGAGTACGTCGCCGATCTTCCCAAGGAGGCAATGGCTATGCTATCTCACACGCA
 Q C C S R A T S P D S A Q G G A N G Y G Y S H Q P

376 AAACAAGCACAACACCTCTAGTTCACAACATGCTCAGCCACCAAGCTGAGGTGAGGTACACACCGCC
 K Q A Q T P P S V Y N N A Q P P P Q A E V R Y T F P P

451 GCGATGAACCTCTCTAGTCCCACTGTGGTTCGCCCTCAAAGCCATGCCTGACACGATCTCGGCAAGCAG
 A M N P P V V P P V V A P S K P M P D T I L G K Q

526 TACGAGATGTCGCTCTGTCTACTCCCTTGGCAAGAACTGGTAGGGGAGTTCGGGTGACGTACCTTTC
 Y E D V R S V Y S L G K E L G R G Q F G V T Y L C

601 ACTGAGATTAGCACTGGCAGGCTACGCTTCAAGTCCATATCCAAGCCAGCTCATGAGTAAGGCGCAAG
 T E I S T G R Q Y A C K S I S K R K L M S K A D K

676 GAGGATATCCCGAGGAGTCCAGACTGACGACCTATCTGGACAGCAACATAGAGAGTTCGTCGAGCA
 E D I R R E I Q I M Q H L S G Q P N I V E F R G A

751 TATGAGGACAAGGCGCTGATCTGCTGATGAGCTCTGTGCAAGCGGAGTGTGGTATCGGATTATTCGC
 Y E D K S V H V V M E L C A G G E L F D R I I A

826 AAGGGCACTATTCAGAGCGCTGACACTGCTACAATGTCAGAGGAGTGTGAATGTTCATGTTTCCGATTTC
 K G H Y S E R A A A T I C R G V V N V N V C H F

901 ATGGAGTATGCAAGCTGATCGAAGCCGGAACCTTCTGCTCGCAACAGGATGAGATGCAAGTCAAG
 M G V M H R D L K P E N F L L A T K D E N A V L K

976 GCCACTGATTTGGGCTTCGCTTTCATGGAAGGAAAATGTACAGGACATCGTGGAGTGCTATTATT
 A T D F G L S V F I E E G K M Y R D I V G S A Y Y

1051 GTTCTCCGAGTCTTAGGAGAATATGTTAAAGAGATAGATGTCGGAGTGCAGGTTATTTTGTATAT
 V A P E V L R R N Y G K E I D V R S A G V I L Y I

1126 CTTCTCAGCGGTTCCCTCCCTTCGGGTGAACTGAGAGGGGATTTGATGCTATTTCGCAAGGGAGATT
 L L S G V P P F W G E T E K G I F D A I L Q K E I

1201 GACTTGTAGAGTCAAGCATGCCAATTTCTGAGAGCGCTAAAGACTTGTGAGGAAGTGTGGCAGGAT
 D F E S Q P W P S I S E S A K D L V R K M L A Q D

1276 CCAAAGAAAAGATTAGTTCAGCCCAAGTCTTCAACATCCATGGCTCAGGGAAGGGAAGCATCAGATAAOCCT
 P K K R I S S A Q V L Q H P W L R E G E A S D K P

1351 ATTGACAGTCTGCTTCTTCTAGGATGAAGCAATCAGAGCAATGAATAACTGAAAAGATGGCTCTAAAGGTG
 I D S A V L S R M K Q F R A M N K L K K M A L K V

1426 ATAGCTTCAATCTTAATGAGGAAGAGATCAAGGGCTGAGCAATGTTTATGACATGGAACAGCAGACAGT
 I A S N L N E E E I K G L K Q M F M N M D T D N S

1501 GGGCAATCACAATGAGAACTCAAAGCAGGACTAGCCAACTTGGATCGAAGCTGTGAGAACTGAAATGAA
 G T I T Y E E L K A G L A K L G S K L S E A E V K

1576 CAGTTGATGATCGCGTATGGTGGCAATGGATCAATGACTACCTTGGATTCATAACAGCAGCGATGAT
 Q L M D A A D V D G N G S I D Y V E F I T A A M H

1651 AGACACAAGCTGAAAGAGAGCAGCAATTTGTTCAAGGGTCCAGTACTTGTGACAAAGACAAGCTGGCTTATC
 R H K L E R D E L F K A S Q Y F D K D N S G F I

1726 ACAAGAGATGAAGTCTGCTTGTGATGAGCATGAAATGGGCGATCGGATACCATAAGGACATCATATCA
 T R D E L E T A L I E H E M G D A D T I K D I I S

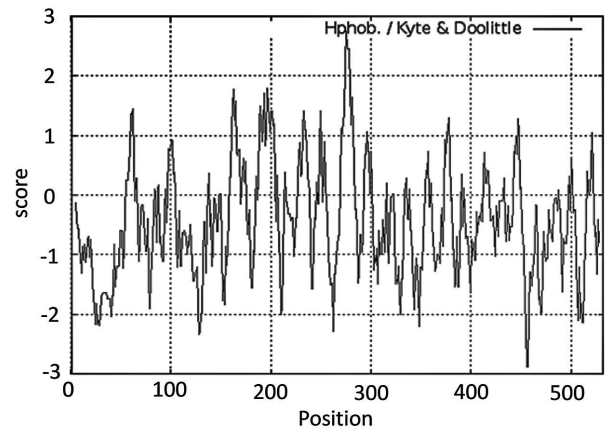
1801 GAAGTCGACACGGATAATGATGGAGGATTAACACGAGGAATTTGCGCTTGTATGAGAGAGGATGACAGCAG
 E V D T D N D G R I N Y E E F C A L M R G G M Q Q

1876 CCAATAAGGCTCAAGTACTTCTTCTTGTAGAATGGTGTACCAATAGCAACTGTCTTGGTCTTTTCTTCT
 P I R L K *

1951 TCCATGGTCCGAGTGCACGATGAGAAGTCTCTTGGCGTGTGAGATGATGTGTATCTTGGGCTT
 2026 TTGCTCGTCAAGTGTGCTCTAGTGTTCGCACTACTTTTCTGTAAGTGGTTTTTTCGAAGTACATCTAA
 2101 TAAAAGAGTATTTTTCAGCTCAATGTCATATTT

图1 *LcCDPK1*基因的全长cDNA序列以及编码的氨基酸序列

最低的为 Trp (0.56%), 其负电荷残基数 (Asp+Glu) 为 75, 正电荷残基数 (Arg+Lys) 为 66, 脂溶性系数 (AI) 为 78.48, 其不稳定系数 (Instability index) 为 42.72, 属于不稳定型蛋白。LcCDPK1 蛋白总平均亲水性 (Grand average of hydropathicity, GRAVY) 为 -0.437, 初步推测为亲水性蛋白。根据 ProtScale 分析得到的亲/疏水信号图 (图 2) 可知, CDPK1 蛋白 N 端与 C 端含有亲水头部, 中间氨基酸部分表现有较多的亲水性, 其中以 A456\A457 残基亲水性最强 (-2.889), 以 A275 残基疏水性最强 (2.744)。整体看来, LcCDPK1 蛋白的亲/疏水



注:横轴为氨基酸位置,纵轴为疏水性得分 (疏水性和亲水性分别用正值和负值表示)

图2 *LcCDPK1*蛋白的疏水性/亲水性预测

表1 *LcCDPK1*基因编码蛋白质的理化性质分析

名称	原子总数	分子量/KDa	理论等电点	正负电荷氨基酸残基数		不稳定系数	脂肪系数	总平均亲水性
				+	-			
<i>LcCDPK1</i>	8 293	59.3	5.79	66	75	42.72 (不稳定)	78.48	- 0.437

信号图峰值分布在-0.5以下的信号明显多于0.5以上的信号,再次证明LcCDPK1蛋白为亲水性蛋白。

2.2.2 *LcCDPK1*基因编码蛋白质的修饰位点预测

蛋白激酶的活性常受到如磷酸化等蛋白翻译后修饰的影响。利用 NetPhos 3.1 对 LcCDPK1 蛋白进行磷酸化位点分析,在系统预设的阈值内 LcCDPK1 蛋白预测到 27 个丝氨酸 (Serine) 磷酸化位点, 15 个苏氨酸 (Threonine) 磷酸化位点, 6 个酪氨酸 (Tyrosine) 磷酸化位点。通过 NetOGlyc 3.1 进行蛋白质糖基化位点分析,结果表明 LcCDPK1 蛋白存在 9 个潜在的糖基化位点。

2.2.3 *LcCDPK1*基因编码蛋白质的信号肽、跨膜结构域及亚细胞定位预测

利用 SignalP 4.1 分析了 LcCDPK1 蛋白的信号

肽序列,结果表明 LcCDPK1 蛋白没有信号肽,不存在剪切位点。同时利用 TMHMM 2.0 预测了 LcCDPK1 蛋白的跨膜结构域,预测结果表明, LcCDPK1 蛋白不存在跨膜结构域和膜内的氨基酸序列,非膜蛋白。利用 Psort Prediction、Softberry 和 Cello 三个在线软件预测 LcCDPK1 蛋白的亚细胞定位。根据 Psort Prediction 预测结果, LcCDPK1 蛋白主要定位于叶绿体和微体中。Softberry 的预测结果显示, LcCDPK1 蛋白定位于质膜上。而 Cello 的预测结果显示,该蛋白定位于细胞质。由于 CDPK 蛋白不同亚型可能分别调控细胞的不同功能,即其亚型间功能存在差异,在细胞中各有特定的亚细胞定位^[5],因此对其预测可能存在多种结果。

2.2.4 *LcCDPK1* 基因编码蛋白质的功能位点及结构域预测

通过NCBI在线分析工具CD Search和在线软件InterPro中检索的EBI数据库分析*LcCDPK1*蛋白的功能位点和保守结构域,预测结果显示*LcCDPK1*蛋白含有蛋白激酶结构域和EF手型结构域(图3)。其中蛋白激酶结构域部分含有

CDPK家族特征性的ATP结合激酶区和活性结合位点(图4),该激酶结构域属于类PK超级蛋白激酶家族特征性的结构域STKc_CAMK。C-端类钙调素蛋白结合结构域有明确的4个EF手型结构负责与钙离子的结合,其结构与功能类似于CaM的氨基酸序列,这是CDPK对Ca²⁺高度亲和而不依赖于CaM的原因^[16]。

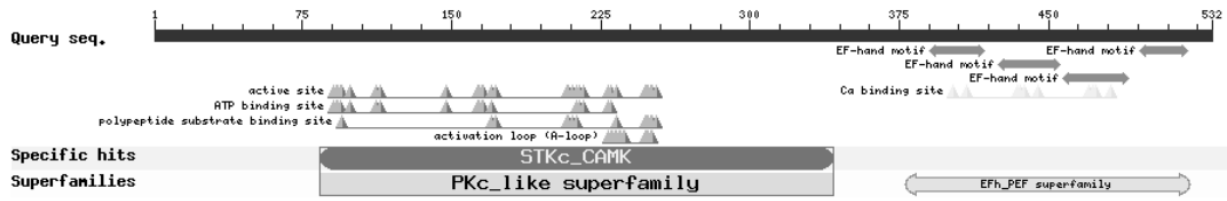


图3 *LcCDPK1* 蛋白保守结构域预测

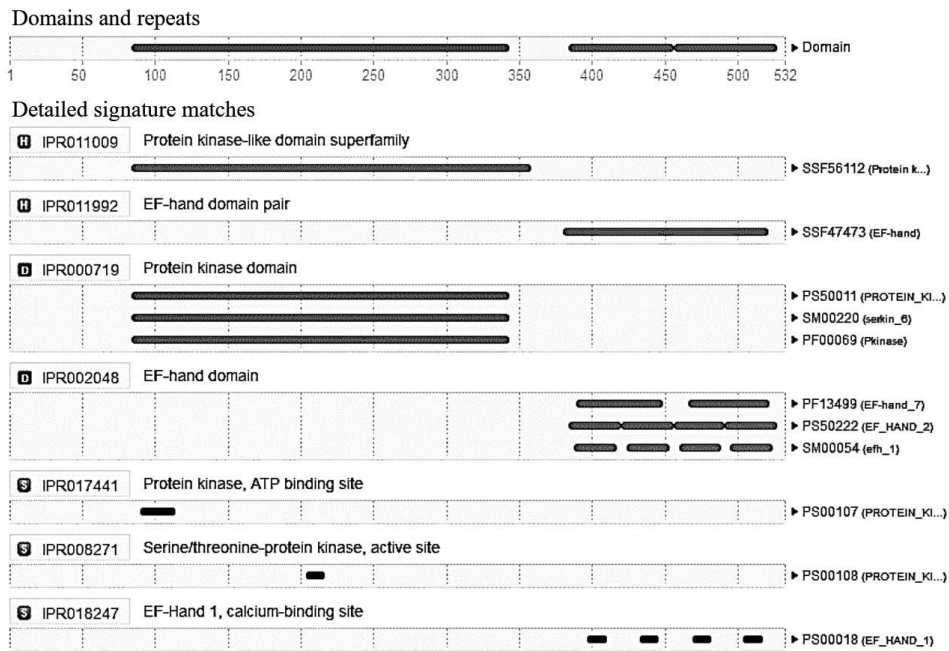


图4 *LcCDPK1* 蛋白功能位点预测

2.2.5 *LcCDPK1* 基因编码蛋白质的二级结构预测

利用PSIPRED在线软件和SOPMA在线软件对*LcCDPK1*蛋白的二级结构进行分析,分析结果(图5)显示*LcCDPK1*蛋白中氨基酸残基组成以 α -螺旋为主(42.29%),同时存在部分的 β -折叠(11.09%)。

2.2.6 *LcCDPK1* 基因编码蛋白质的三级结构预测

蛋白的高级结构对研究蛋白的功能有重要的意义,通过在线软件Phyre 2进行同源建模软件模拟*LcCDPK1*蛋白的三维结构。使用PyMOL三级结构显示软件分析*LcCDPK1*蛋白的同源建模结果,以二级结构标记颜色,同时分别以黄色和橙色标注蛋白的N端和C端(图6)。

2.2.7 *LcCDPK1* 基因编码蛋白质的进化分析预测

以*LcCDPK1*基因序列为探针,在NCBI中进行Blast比对,下载与探针蛋白相似性较高的小麦(*Triticum aestivum*: ABD98803)、短芒大麦草(*Hordeum brevisubulatum*: ABB54456)、羊草(*L.chinensis*: CD808794)、二穗短柄草(*Brachypodium distachyon*: XP_024312381)、日本稻(*Oryza sativa Japonica Group*: XP_015646655.1)、高粱(*Sorghum bicolor*: XP_021309517.1)、玉米(*Zea mays*: NP_001105542.2)、小米(*Setaria italica*: AGG54988.1)、菠萝(*Ananas comosus*: OAY67230.1)、中华猕猴桃(*Actinidia chinensis var. chinensis*: PSR98421.1)、木薯(*Manihot esculenta*: XP_021607153.1)、枣

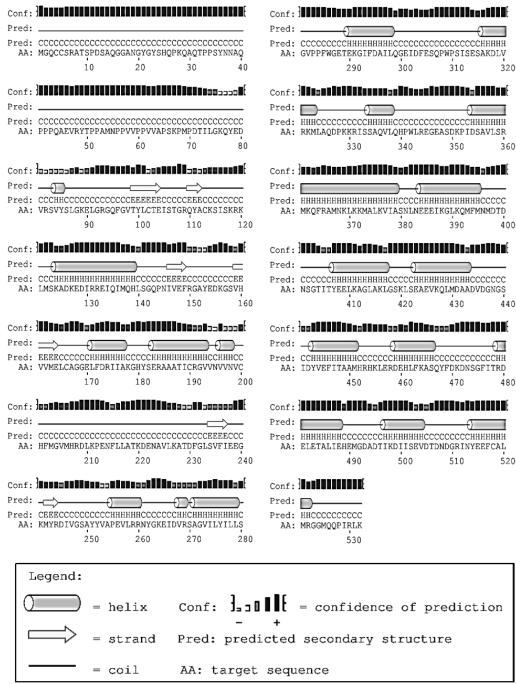


图5 LcCDPK1蛋白二级结构域预测

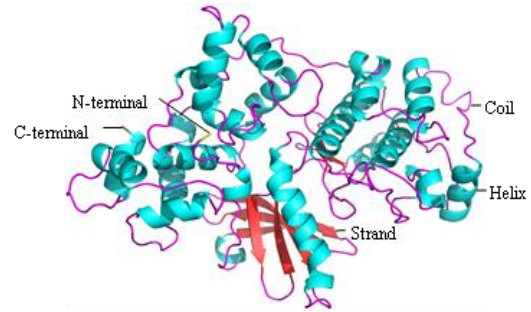
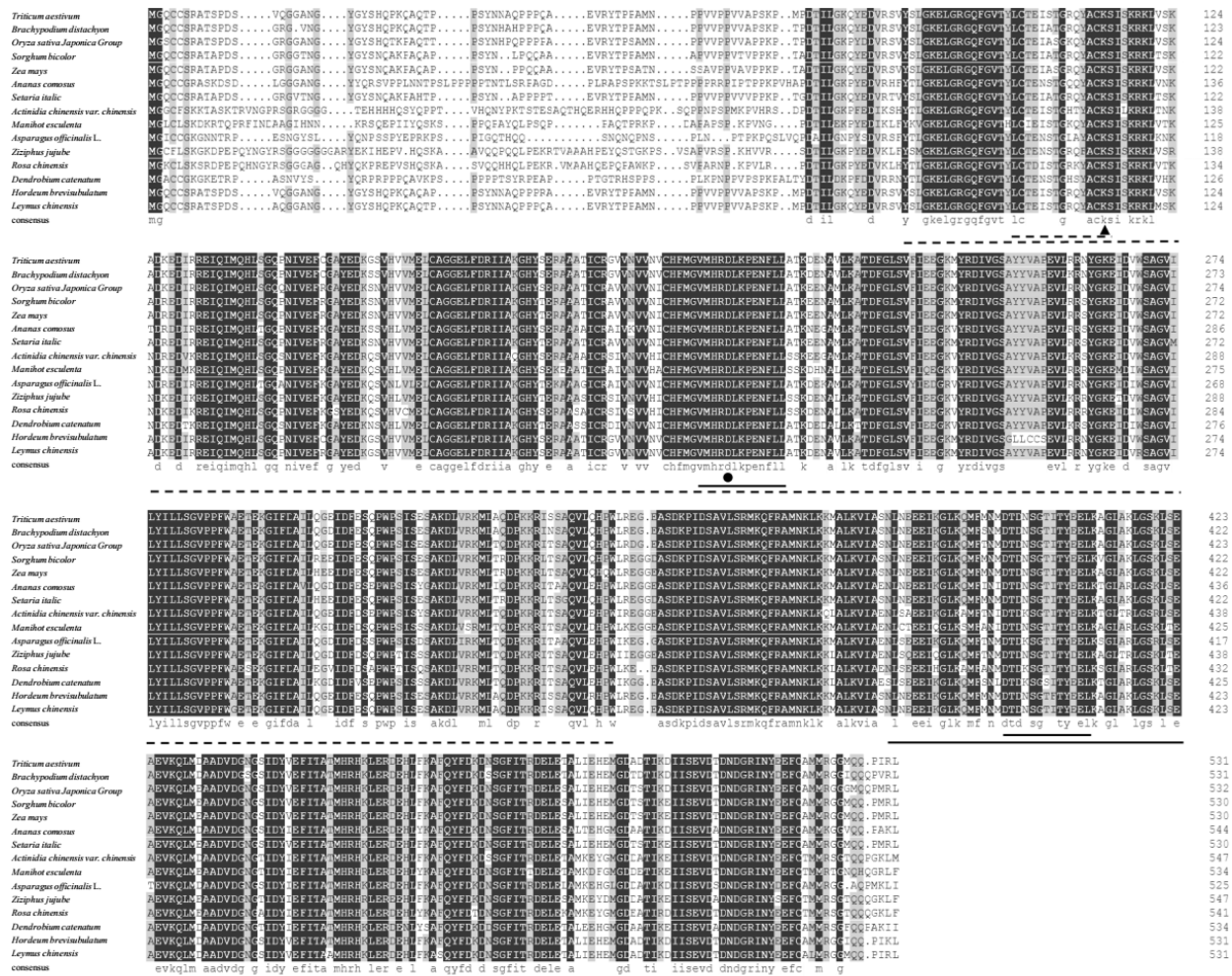


图6 LcCDPK1蛋白的同源建模空间结构

(*Ziziphus jujuba*: XP_015899753.1)、月季(*Rosa chinensis*: XP_024156502.1)、石刁柏(*Asparagus officinalis*: XP_020273550.1)、铁皮石斛(*Dendrobium catenatum*: XP_020698035.1)进行氨基酸序列同源比对。

通过DNAMAN进行序列比对(图7),发现比对选取的物种其CDPK蛋白的N端可变区长短不一且同源性极低,但N端起始氨基酸序列均具有豆蔻酰化和棕榈酰化作用所必需的保守序列MGXXC(S/Q)XXT^[17],本研究得到的LcCDPK1蛋白



注:虚线区域,蛋白酶酶结构域;实线区域,EF-hand结构域;双虚线区域,蛋白激酶ATP结合区;虚实双线区域,丝氨酸/苏氨酸蛋白激酶活性位点;双实线区域,EF-hand基序的核心序列

图7 羊草LcCDPK1与其他植物多序列比对

多肽链1~9位氨基酸序列为“MGQCCSRAT”,符合这一序列特征。蛋白激酶区由300个左右的氨基酸组成,该区域含有催化关键位点Ser/Thr蛋白激酶序列,序列保守性较高。LcCDPK1蛋白的85~343位氨基酸为其蛋白激酶区,91~114位为蛋白激酶ATP结合区,其中第114位赖氨酸(K)为ATP的结合位点,205~217位为丝氨酸/苏氨酸蛋白激酶活性位点,其中209位的天冬氨酸(D)是其催化活性中心的核心氨基酸,这些关键区域的氨基酸序列与其他物种之间有着极高的保守性。连接区负责连接激酶区与调控区,由30个左右氨基酸组成,在没有Ca²⁺时,CDPK的蛋白激酶区可能与连接区结合,使其激酶活性被抑制。C端调控区是CDPK区别其他

类型激酶的特有区域,通常含有4个EF手型结构,少数含有2个或6个等数量不等的手型结构。LcCDPK1蛋白的381~521位为其调控区,其4个EF手型结构分别在398~410、434~446、470~482和505~517位氨基酸处。根据序列比对结果,CDPK基因的序列特征非常明显,不同种属的催化区同源性达80%以上,活性部位同源性达到100%。

采用邻位相连(Neighbor-Joining, NJ)算法,利用软件MEGA 5.0构建LcCDPK1蛋白与其他几种植物CDPK蛋白的NJ进化树(图8),结果表明,LcCDPK1蛋白与小麦和短芒大麦草在同一进化枝上,表明其亲缘关系较近,次之为二穗短柄草。说明CDPK蛋白在长期的生物进化过程中有着较强的保守性。

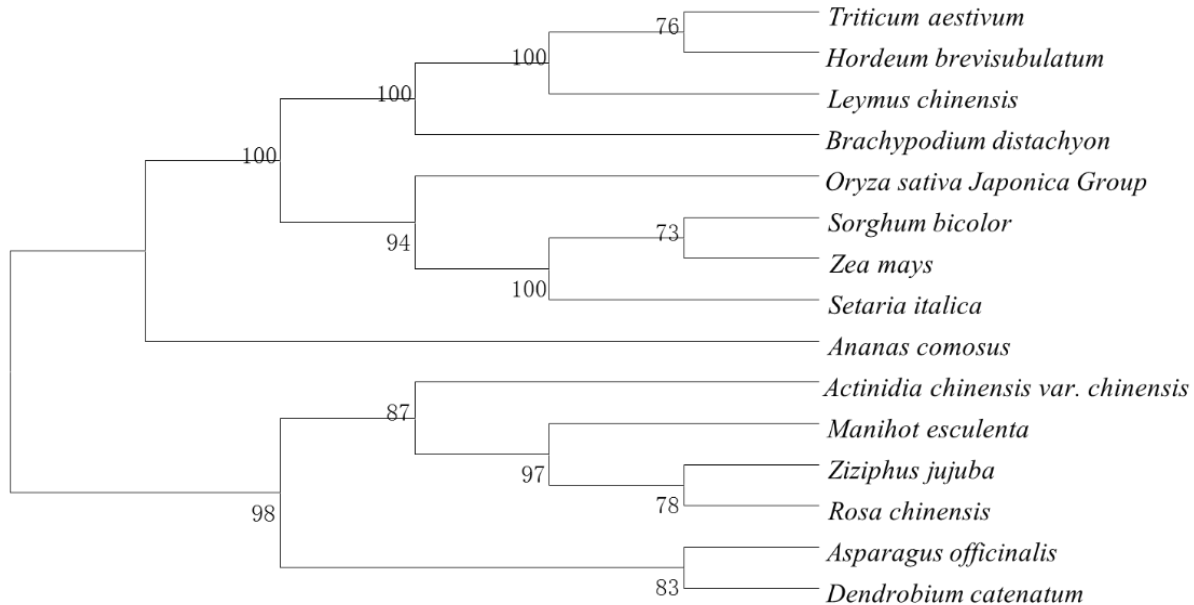


图8 LcCDPK1蛋白的系统进化分析

3 讨论与结论

CDPK是具有多种功能的重要酶类,其参与植物的多种逆境调控,对其基因的克隆及序列分析是帮助研究羊草对逆境尤其是盐碱和干旱胁迫响应机制的基础^[6]。本研究以耐盐碱的羊草作为试验材料,通过已知的EST序列采用RT-PCR和RACE技术等分子生物学手段对基因的同源区和全序列进行克隆,成功克隆得到了一个CDPK基因,将其命名为LcCDPK1,全长基因为2136 bp,具有1599 bp的开放读码框(Open Reading Frame, ORF),编码532个氨基酸。

崔喜艳等2015年报道了一个LcCDPK基因,其氨基酸第81~339位为STKc_CAMK结构域,368~500位包含4个EF手型结构^[12]。将本研究克

隆得到的LcCDPK1基因和LcCDPK进行对比,发现两个基因编码的蛋白理论等电点和总平均亲水性较接近,均没有信号肽,不存在剪切位点和跨膜结构域。LcCDPK蛋白和LcCDPK1蛋白的保守结构域预测结果也很一致,均含有CDPK蛋白特征性的STKc_CAMK结构域和4个EF手型结构。两者的区别在于,本研究得到的LcCDPK1蛋白略小,且LcCDPK蛋白不稳定系数为32.57,是稳定蛋白,而LcCDPK1蛋白不稳定系数为42.72,是不稳定蛋白。LcCDPK蛋白亚细胞定位预测结果在质膜上,而LcCDPK1蛋白在不同的亚细胞定位分析软件中呈现出了不同结果,包括质膜、细胞质、叶绿体和微体。2003年报道的拟南芥AtCPK3和AtCPK4蛋白,其预测结果定位不在细胞核,但实验中发现两蛋白激酶存在靶向细胞核的现象,这

种现象可能是由于CDPK蛋白家族不同亚型的作用底物不同,调控机制不同造成的^[15]。

进一步分析LcCDPK1蛋白的特征序列,该蛋白多肽链N端起始氨基酸序列为“MGQCCSRAT”,是豆蔻酰化和棕榈酰化作用所必需的保守序列“MGXXC(S/Q)XXT”^[17]。其205~217位的Ser/Thr蛋白激酶活性位点含有交感序列“DLKPEN”^[17]。对比LcCDPK蛋白序列,发现也具有这两个保守性极高的关键序列。目前对CDPK蛋白的研究发现,其N端的豆蔻酰化和棕榈酰化与蛋白间的互作及蛋白的膜定位有关,关键酰基化位点为甘氨酸(Gly)或半胱氨酸(Cys)残基^[18]。但酰基化只是膜定位的特征之一,目前已知定位质膜的CDPK蛋白并没有其他明显的保守序列。因此,有可能许多定位质膜的亚型实际上被整合到不同的特异性调控复合物中发挥作用^[15]。对LcCDPK1与LcCDPK进行序列比对,发现其相似性仅为46.11%,其中可变区的相似性是最低的,调控区和连接区相似性次之,蛋白激酶区是保守性较高的部分。由于植物CDPK基因可能来自蛋白激酶和CaM基因的融合,因此会出现不同亚型在结构与序列组成上有较大差异的情况^[16,19]。

拟南芥、烟草、水稻、玉米和梨等植物的CDPK蛋白家族进化分析结果均聚类为4个亚族,推测LcCDPK和LcCDPK1蛋白序列相似性低的原因是两个蛋白分属不同亚族^[15]。后续可能会发现更多羊草的CDPK基因,这些基因分属哪一亚族,各亚族序列的差异与其翻译后修饰、亚细胞定位等之间的关系,这种差异对其功能的影响,不同亚型的CDPK如何响应干旱盐碱胁迫这些方面有待进一步研究。

参考文献:

- [1] 李永宏. 放牧影响下羊草草原和大针茅草原植物多样性的变化[J]. *Journal of Integrative Plant Biology*, 1993, 35(11): 877-884.
- [2] 王 萍, 殷立娟, 李建东. 中性盐和碱性盐对羊草幼苗胁迫的研究[J]. *草业学报*, 1994, 3(2): 37-43.
- [3] 王志锋, 张世忠. 吉林省草业经济发展优势及其途径分析[J]. *吉林农业科学*, 2003, 28(5): 38-43.
- [4] 刘滨硕, 康春莉, 王 鑫, 等. 羊草对盐碱胁迫的生理生化响应特征[J]. *农业工程学报*, 2014, 30(23): 166-173.
- [5] 郭 坚, 王 涛, 薛 娟, 等. 松嫩沙地荒漠化现状和原因[J]. *干旱区资源与环境*, 2007, 21(5): 99-103.
- [6] 张 丽, 张 磊, 康 乐, 等. 绒毛状烟草钙依赖蛋白激酶基因家族分析[J]. *中国烟草科学*, 2012, 33(3): 56-62.
- [7] Cheng S H, Willmann M R, Chen H C, et al. Calcium signaling through protein kinases: The Arabidopsis calcium-dependent protein kinase gene family[J]. *Plant Physiology*, 2002, 129(2): 469-485.
- [8] Harmon A C, Gribskov M, Gubrium E, et al. The CDPK superfamily of protein kinases[J]. *New Phytologist*, 2001, 151(1): 175-183.
- [9] Yusuke S, Shingo H, Junko K, et al. Over-expression of a single Ca²⁺-dependent protein kinase confers both cold and salt/drought tolerance on rice plants[J]. *The Plant Journal*, 2000, 23(3): 319-327.
- [10] Urao T, Katagiri T, Mizoguchi T, et al. Two genes that encode Ca²⁺-dependent protein kinases are induced by drought and high-salt stresses in *Arabidopsis Thaliana*[J]. *Molecular and General Genetics*, 1994, 244(4): 331-340.
- [11] Li A L, Zhu Y F, Tan X M, et al. Evolutionary and functional study of the CDPK gene family in wheat (*Triticum aestivum* L.)[J]. *Plant Molecular Biology*, 2007, 66(4): 429-443.
- [12] 崔喜艳, 张继伟, 秦思余, 等. 羊草钙依赖型蛋白激酶基因生物信息学及表达特性分析[J]. *中国草地学报*, 2015, 37(3): 11-18.
- [13] 赵 英, 刘晶莹, 杨梦丹, 等. 羊草14-3-3基因的克隆及序列分析[J]. *西北农业学报*, 2018, 27(2): 269-274.
- [14] 云巾宴, 任春宇, 车 达, 等. 气肿疽梭菌细胞毒素CctA基因的克隆及生物信息学分析[J]. *吉林农业科学*, 2015, 40(1): 82-85.
- [15] Christian D, Audrey I, Bimei H, et al. Subcellular targeting of nine calcium-dependent protein kinase isoforms from *Arabidopsis*[J]. *Plant Physiology*, 2003, 132(4): 1840-1848.
- [16] 麻 浩, 王 爽, 周亚丽. 植物中钙依赖蛋白激酶的研究进展[J]. *南京农业大学学报*, 2017, 40(4): 565-572.
- [17] 杨洪强, 梁小娥. 高等植物中的CDPK及其生理生化功能[J]. *山东农业大学学报(自然科学版)*, 2003, 34(2): 186-190.
- [18] Lu S X, Hrabak E M. An *Arabidopsis* calcium-dependent protein kinase is associated with the endoplasmic reticulum[J]. *Plant Physiology*, 2002, 128(3): 1008-1021.
- [19] 孟义淳, 曲柏宏, 王甜元, 等. 苹果梨MYB108基因的克隆与序列分析[J]. *东北农业科学*, 2020, 45(2): 60-63.
- (责任编辑: 刘洪霞)
- ~~~~~
- (上接第30页)
- [20] 付 艳, 高树仁, 王振华. 玉米种质苗期耐盐性的评价[J]. *玉米科学*, 2009, 17(1): 36-39, 50.
- [21] 单长建. 抗逆转基因玉米苗期耐盐碱评价方法的初步建立及功能验证[D]. 哈尔滨: 东北农业大学, 2012.
- [22] 杨书华, 张春宵, 朴明鑫, 等. 大田苗期耐盐碱鉴定指标筛选及玉米自交系盐碱耐性分析[J]. *种子*, 2012, 31(1): 9-13.
- [23] 汤 华, 柳晓磊, 罗秋芸. 玉米耐盐早期筛选体系的初步研究[J]. *海南大学学报(自然科学版)*, 2007(2): 169-172, 176.
- [24] 于永涛, 刘 成, 吕 玲. 玉米品种耐旱性评价及相关鉴定指标的研究[J]. *作物杂志*, 2008, 25(4): 55-58.
- [25] 吴文荣. 玉米不同品种芽苗期抗旱性及指标的研究[D]. 北京: 中国农业科学院, 2008.
- (责任编辑: 王 昱)