

中间锦鸡儿 *CiUPF0114* 克隆与表达及 豆科 *UPF0114* 基因家族成员分析

杨 杞^{1,2,3}, 牛肖翠^{1,2,3}, 王瑞刚^{1,2,3}, 李国婧^{1,2,3}

¹ 内蒙古农业大学生命科学学院, 呼和浩特 010018; ² 内蒙古自治区植物逆境生理与分子生物学重点实验室, 呼和浩特 010018;

³ 内蒙古自治区抗逆植物遗传资源利用与分子改良科技创新团队, 呼和浩特 010018)

摘要: 本研究克隆了中间锦鸡儿一个 *CiUPF0114* 基因, 该基因编码框长 876 bp, 编码 291 个氨基酸的蛋白质, 预测等电点 9.83, 分子量 32.13 kD。利用荧光定量 PCR 检测发现, 在干旱、脱水和冷处理后, *CiUPF0114* 基因转录水平的表达量均明显上调。通过全基因组筛选, 获得了 29 个来自于 8 种豆科植物以及模式植物拟南芥和水稻中的 *UPF0114* 基因家族成员, 并对其进行了系统进化分析、基因结构和保守基序分析。系统进化分析将 *UPF0114* 基因分成 A 和 B 两个亚家族, A 亚家族又分成 A1 和 A2 两个组, 本研究中克隆的中间锦鸡儿 *CiUPF0114* 属于 B 亚家族成员。通过对蛋白长度、结构域数量、分子量、等电点、基因结构和保守基序的分析发现, 29 个 *UPF0114* 基因差异不大, 说明该基因家族成员在进化上相对保守。该研究为进一步验证 *CiUPF0114* 基因的功能奠定了基础。

关键词: *CiUPF0114*; 基因克隆; 中间锦鸡儿; 基因家族; 进化分析

Cloning and Expression of *CiUPF0114* in *Caragana intermedia* and Bioinformatics Analysis of *UPF0114* Genes from *Leguminosae*

YANG Qi^{1,2,3}, NIU Xiao-cui^{1,2,3}, WANG Rui-gang^{1,2,3}, LI Guo-jing^{1,2,3}

¹ College of Life Sciences, Inner Mongolia Agricultural University, Hohhot 010018; ² Inner Mongolia Key Laboratory of

Plant Stress Physiology and Molecular Biology, Hohhot 010018; ³ Inner Mongolia Scientific Innovaton Team of

Genetic Resource Utilization and Molecular Improvement of Stress Resistant Plants, Hohhot 010018)

Abstract: This study reported the isolation of *CiUPF0114* gene from *Caragana intermedia* Kuang & H.C.Fu, which contained an 876 bp open reading frame, encoding for a protein with 291 amino acids. The predicted isoelectric point and molecular weight of *CiUPF0114* were 9.83 and 32.13 kD, respectively. The expression level of *CiUPF0114* gene was detected by Real-time PCR, the result shown that *CiUPF0114* gene was significantly up-regulated under drought, dehydration or cold treatments. Furthermore by genome-wide screening, a total of 29 *UPF0114* gene family members were identified from ten plant species, including eight legumes (*Medicago truncatula* Gaertn., *Glycine max* (L.) Merr., *Cicer arietinum* L., *Phaseolus vulgaris* L., *Vigna radiata* (L.) R.Wilczek var. *radiata*, *Trifolium pratense* L., *Cajanus cajan* (L.) Huth, *Arachis ipaensis* Krapov. & W.C.Greg.) and two model plant (*Arabidopsis* and rice). The 29 *UPF0114* genes were divided into subfamily A and subfamily B according to phylogenetic analysis. The A subfamily was further divided into two groups, A1 and A2. *CiUPF0114*, a gene from *Caragana intermedia* Kuang & H.C. Fu cloned in this study, belongs to the B subfamily. By analyzing the length of the protein, the number of domains, the molecular weight, the isoelectric point, the gene structure and the conserved motif, the 29 *UPF0114* genes were not significantly different,

收稿日期: 2018-08-29 修回日期: 2018-10-25 网络出版日期: 2018-12-05

URL: <http://kns.cnki.net/kcms/detail/11.4996.s.20181204.1454.001.html>

第一作者研究方向为植物抗逆生理与分子生物学, E-mail: atp_yangqi@imau.edu.cn

通信作者: 李国婧, 研究方向为生物化学与分子生物学, E-mail: liguojing@imau.edu.cn

基金项目: 国家自然科学基金项目 (31560199)

Foundation project: National Natural Science Foundation of China (31560199)

indicating that the members of *UPF0114* gene family are relatively conservative in evolution. These results might lay the foundation for further study on the functions of *CiUPF0114*.

Key words: *CiUPF0114*; gene clone; *Caragana intermedia* Kuang & H.C.Fu; gene family; phylogenetic analysis

中间锦鸡儿 (*Caragana intermedia* Kuang & H.C. Fu) 属豆科 (Leguminosae) 蝶形花亚科 (Papilionoideae) 亚科, 是分布于中国西北干旱荒漠地区重要的饲用灌木, 广泛用于植物造林和生态保护^[1-2]。中间锦鸡儿抗逆性强、耐贫瘠, 通过对中间锦鸡儿抗逆基因的研究, 可以使我们深入了解豆科植物抗逆的分子机制, 从而为大豆 (*Glycine max* (L.) Merr.)、苜蓿 (*Medicago* L.) 等重要经济作物的抗逆基因工程提供帮助。

随着分子生物学的飞速发展, 和其他各类植物一样, 对豆科植物的研究也深入到了基因组、转录组和蛋白组的多组学阶段。目前, 已经有多种豆科植物的基因组被测序, 包括大豆^[3]、蒺藜苜蓿 (*Medicago truncatula* Gaertn.)^[4]、三叶草 (*Trifolium pratense* L.)^[5]、鹰嘴豆 (*Cicer arietinum* L.)^[6] 和木豆 (*Cajanus cajan* (L.) Huth)^[7] 等, 这为利用生物信息学, 从数据库中挖掘新基因、研究基因功能奠定了基础。同时, 利用转录组研究豆科植物抵抗逆境的分子机制, 挖掘抗逆基因也成为了研究的热点, 如沙冬青 (*Ammopiptanthus mongolicus* (Maxim. ex Kom.) S.H.Cheng) 冷处理下基因表达的转录组研究^[8]、中间锦鸡儿逆境胁迫下 microRNA^[9] 和 *WRKY* 基因^[10] 的鉴定和功能分析, 鹰嘴豆中大规模转录组测序的研究^[11]、绿豆 bHLH 转录因子家族的鉴定与分析^[12] 等。

近年来, 有关中间锦鸡儿抗逆分子机制和相关基因功能研究的报道逐渐增多, 但是大多集中在一些常见的抗逆基因上, 对于在植物抗逆中发挥作用的未知功能基因却鲜有报道。我们在实验室前期构建的中间锦鸡儿转录组数据库 (NCBI 收录号: SRP121096) 中发现 1 条脱水处理后转录水平表达上调的未知功能基因, 同源比对发现该基因编码蛋白属于 *UPF0114* 蛋白 (Uncharacterized protein family 0114)。*UPF0114* 是一个功能未知的蛋白质家族, 广泛存在于各类生物中^[13]。在 Pfam 数据库中该蛋白结构域的编号为 PF03350, 截止到 2018 年 6 月, Pfam 数据库中共收录了 846 个物种中 1074 条 *UPF0114* 蛋白序列, 其结构域的平均长度是 119.90 个氨基酸残基。从植物比较基因组数据库 PLAZA4.0 中检索后发现, 从藻类到被子植物中具有 *UPF0114* 结构域的基因数量几乎没有变化, 每个

物种中有 2~4 个 *UPF0114* 基因, 并没有随着物种的演化而发生明显的基因家族扩张。目前为止, 该家族基因的功能还没有任何相关的报道。

本研究对一条来源于中间锦鸡儿的 *UPF0114* 基因进行了克隆和非生物胁迫下的表达分析。为了进一步了解该基因家族成员在植物中的进化与分布, 对 8 种豆科植物及模式植物拟南芥和水稻的 *UPF0114* 基因进行了全基因组的筛选和生物信息学分析, 这为今后研究 *UPF0114* 基因的生物功能, 特别是在植物抵抗逆境胁迫中的功能奠定了基础。

1 材料与方法

1.1 植物材料

中间锦鸡儿种子采自内蒙古自治区四子王旗坡底村。选取饱满无虫蛀种子种植于装有营养土和蛭石的培养钵中, 置于植物培养室 (温度 22 °C, 光照/黑暗为 16 h/8 h, 相对湿度 30%~40%) 中生长 30 d, 挑选生长均一的幼苗用于 RNA 提取和胁迫处理。

1.2 中间锦鸡儿 *CiUPF0114* 基因 cDNA 序列的克隆

根据在转录组数据库中找到的中间锦鸡儿 *UPF0114* 基因序列, 用 Primer Premier 5.0 设计特异性引物, 扩增基因完整的编码框。引物序列如下: F-*CiUPF0114*: ATGGCAACCACTAGATTGTTG R-*CiUPF0114*: AGCTCTGAATCTGGAATATTAATAATACGT

利用 RNA 提取试剂盒 (天根, code DP419) 提取中间锦鸡儿总 RNA, 并用反转录试剂盒 (全式金, code AT311-03) 反转录获得 cDNA。以中间锦鸡儿 cDNA 为模板, F/R-*CiUPF0114* 为引物, 利用高保真酶 Primer STAR (TaKaRa) 进行 PCR 的扩增: 98 °C 预变性 1 min, 98 °C 变性 10 s, 50 °C 退火 10 s, 72 °C 延伸 1 min, 30 个循环, 72 °C 补充延伸 10 min。将扩增得到的目的片段连接到克隆载体 pEASY-Blunt simple cloning vector (全式金, code CB111-02), 转化到大肠杆菌 Trans1-T1 (全式金, code M47023) 感受态细胞中, 菌液涂含有卡那 (Kan) 抗性的 LB 平板上, 37 °C 过夜培养, 菌落 PCR 和酶切挑选鉴定阳性克隆, 送上海生工测序。

对克隆得到的中间锦鸡儿 *CiUPF0114* 基因进行序列分析, 利用软件 DNAMAN 6.0 绘制序列图

并进行亲水性分析,利用 ExPASy (<https://www.expasy.org/>) 中的 Compute pI/MW 进行等电点和分子量计算,GOR4 分析二级结构组成^[14]。

1.3 植物 *UPF0114* 基因家族成员分析

以 UPF0114 为关键词在植物基因组数据库 PLAZA4.0^[15] (<https://bioinformatics.psb.ugent.be/plaza/>) 和 Phytozome^[16] (<https://phytozome.jgi.doe.gov/pz/portal.html>) 中搜索植物 *UPF0114* 家族候选基因,下载基因的核酸序列和蛋白序列。本研究中,8 种豆科植物藜苜蓿、大豆、鹰嘴豆、木豆、三叶草、菜豆、豇豆 (*Vigna radiata* (L.) R.Wilczek var. *radiata*) 和落花生 (*Arachis ipaensis* Krapov.& W.C.Greg.), 以及模式植物拟南芥 (*Arabidopsis thaliana* (L.) Heynh.) 和水稻 (*Oryza sativa* L.) 的 *UPF0114* 基因被用做分析。利用 Pfam 数据库 (<http://pfam.xfam.org/>) 和 NCBI (<https://www.ncbi.nlm.nih.gov/>) 中的 Batch CD-Search 对候选基因进行结构域的确证^[13,17],去除结构域不完整的基因。利用 ExPASy (<https://www.expasy.org/>) 中的 Compute pI/MW 进行等电点和分子量计算^[14]。

1.4 基因结构、保守基序和系统进化分析

根据鉴定出的 8 种豆科植物及拟南芥、水稻和中间锦鸡儿 UPF0114 蛋白全长序列,以大肠杆菌 UPF0114 蛋白 (YQHA_ECOLI UPF0114 protein) 作为外类群,利用软件 MEGA7 构建系统发育树^[18]。具体条件设置如下:邻接法 (Neighbor-joining),模型采用 p-distance,空位采用 Complete deletion, Bootstrap method 验算 1000 次。

根据下载的 *UPF0114* 基因的基因组序列、CDS 序列,利用 GSDS 2.0 (<http://gsds.cbi.pku.edu.cn/>) 绘制基因结构图^[19]。根据下载的 UPF0114 蛋白序列,利用 MEME (<http://meme-suite.org/index.html>) 分析 UPF0114 的共有保守基序, motif 数量为 6,其余各参数采用默认设置^[20]。

1.5 植物胁迫处理和基因表达检测

土壤干旱处理:将植物充分浇水后不再浇水,进行干旱处理,取处理后的 0 d、6 d、9 d、12 d 的植物地上部分各 5 株,液氮速冻,保存于 -80 °C 冰箱。

脱水处理:从培养钵中小心取出植物,避免损伤,自来水清洗干净根部蛭石后放于滤纸上,22 °C,相对湿度 30%~40% 环境下自然脱水干燥。

低温胁迫处理:将中间锦鸡儿连培养钵放入 4 °C 光照培养箱中处理 (光照 / 黑暗: 16 h/8 h)。

脱水和低温处理取样时间点为处理后 0 h、1 h、3 h 和 12 h,每个时间点都随机取 5 株植物混合,液

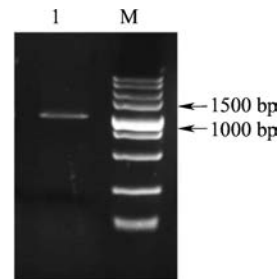
氮速冻,放置于 -80 °C 冰箱保存。

荧光定量方法及内参基因参照实验室以前发表文献^[21]。提取总 RNA,并反转录获得 cDNA。将获得的 cDNA 稀释 16 倍作为模板,以 RT-CiUPF0114-F、RT-CiUPF0114-R 为目的基因引物;以 CiEF1 α -RT-F、CiEF1 α -RT-R 为内参引物,利用 Roche480 仪器进行实时定量荧光 PCR。反应体系为:模板 5 μ L, SYBR[®]Premix Ex Taq[™] 10 μ L, 10 μ mol/L 引物各 0.8 μ L, DEPC 水补至 20 μ L。反应程序如下:95 °C 预变性 1 min, 95 °C 变性 5 s, 60 °C 退火 30 s, 72 °C 延伸 10 s, 40 个循环,循环结束后做熔解曲线。试验结果采用 $2^{-\Delta\Delta CT}$ 法计算,每个反应 3 次技术重复。每个处理 3 次重复。荧光定量所用引物如下。CiEF1 α -RT-F: AGCAATCGTTCTTCCTAATGATCTAA, CiEF1 α -RT-R: CAAAAAGTCCCCTCGTTGTCTC, RT-CiUPF0114-F: AGATATCCCTTGGTCACAGGTGTTG, RT-CiUPF0114-R: CGATGAATCTGGAGAGGACAAAATAG。

2 结果与分析

2.1 中间锦鸡儿 *CiUPF0114* 基因的克隆

从中间锦鸡儿脱水诱导转录组数据库中得到一条注释为 *UPF0114* 的基因,在 NCBI 数据库中比对后确定该基因包含完整的编码框。以中间锦鸡儿 cDNA 为模板,设计引物对该基因进行克隆 (图 1)。将 PCR 产物连接克隆载体,筛选鉴定阳性克隆送测序。



M: DNA ladder, DL5000 (TaKaRa)

图 1 中间锦鸡儿 *CiUPF0114* 基因 cDNA 扩增电泳结果
Fig.1 PCR amplification of the cDNA fragment of *CiUPF0114*

2.2 中间锦鸡儿 *CiUPF0114* 基因序列分析

测序后得到 1 条包含有完整编码框的 899 bp 的 cDNA 序列,经过 NCBI 数据库中 Blast 比对及 CDD-search 确定该基因具有完整的结构域,将其命名为 *CiUPF0114*。*CiUPF0114* 基因编码框长 876 bp,起始密码子 ATG,终止密码子 TGA,编码 291 个氨基酸 (图 2)。CDD-search 显示编码蛋白的 119~257 个氨基酸是 UPF0114 结构域位置,不存在其他的蛋白结构域。该基因编码蛋白质预测等

电点 9.83, 分子量 32.13 kD, 二级结构中 α 螺旋占 46.74%, β 折叠占 14.78%, 不规则卷曲占 38.49%。

```

1  ATGGCAACCAGTAAITGTTGGCCACCTCTAGGCCAATTCAACTCCCTCACTTCTOCTAT
1  M A I T T R L L R I S R P F Q L P S S S Y
61  TCITCCATTTCTTCTTCGTTTCATTGAGAGTGTACGGTATCAAAGCAGGGCCAATTTGAGT
21  S S I S S S F I E S V R Y Q S R A N L S
121  GGAGATACTATCAAAGCTATGGAAAAGGAGAAAGGAAACCGGTAGTATCCGTGAAAAGCG
41  G D T I K A I G K G E R K P V V S V K A
181  TCGATGGCCACCACCAATACATAACCACATCACAACCCAGTGGTTCTGTAAGGATGTCA
61  S M A I T T N H I T T S Q P V V R E G L S
241  AATTTAGCTGCGATTATAGCAAGTATTCGCAATGCTATCTTGTCTTTTGGGACTGCT
81  N L A A I I A S I R N A I L V L L R T A
301  GTCAGAGGAGCAATGAGGAATTTACAAACCCAGATGCTCAITGAAAGGGCAATCATC
101  V K R K Q L R N L Q P Q M L I E R A I I
361  GAITCCAGATTTTTTATATTGTTGGCAGTAGCAGGATCTTTACTTGGCTCCGTAITGTGT
121  D C R F F I L F A V A G S L L G S V L C
421  TTTTTGAGGGCTGCTCTTCTTATTTAGTCAATAGCGCATTATTCCACAGTGTCTCT
141  F F E G C L L V I E S Y A H Y F H T L S
481  CAAAGGTTGGATCAAGGACATCTTGGCATCTACTCATCGAAGCCATAGATATGTTTCATG
161  Q R L D Q G H L V H L L I E A I D M F M
541  GTAGAAACAGCCCTGCTCATTITTTGGGGTGGGTTTGACGTCATGTTTGGGATCAGCG
181  V G T A L L I F G V G L Y V M F V G S R
601  ACTACTACTAGCAAGAAAGAAACCATGGCTTTTGGGTCAAAACCTATTITGGTCTTTTTC
201  I T I S K E R E P W L C G S N L F G L F
661  TGATGAAAGTCACCGCTCGGTGGGTTGGGATGCAATCAATGAGCAGGCAAGTGAAG
221  C M K S P P R W V G M Q S I E Q A K S K
721  ATTGGCAGTGGGATGATGATCTTTCAGGTGGGACTAATAGAGAAGTTCAAGATATT
241  I G H A V M M I L Q V G L I E K F K D I
841  CCCTTGGTCAAGGTTGATCTTGGCTTGGCTTGGCTGACGCTCTACTACATCTCAGCT
261  P L V T G V D L A C F A A A L L T S S A
841  ACTAATTTTGTCTCCCAAGATTCATGATATGACGCTATTATTAATATTCAGATTC
281  T I F V L S K I H R Y *

```

图 2 *CiUPF0114* 基因 cDNA 序列和推导的氨基酸序列
Fig.2 cDNA and deduced amino acids sequences of *CiUPF0114*

2.3 豆科植物 *UPF0114* 基因家族成员的鉴定

通过对全基因组的搜索,共得到 8 种豆科植物和拟南芥、水稻中具有完整结构的 29 个 *UPF0114* 家族成员,如表 1 所示。其中水稻和大豆中的 *UPF0114* 基因最多,分别有 4 个,而其他植物中都存在 2~3 个。结构域鉴定后发现,除了三叶草 TPR.G28382 蛋白有 2 个 *UPF0114* 结构域以外,其他 *UPF0114* 蛋白都仅具有 1 个完整的结构域。绝大多数 *UPF0114* 蛋白中不存在其他的蛋白结构域,仅豇豆 *Vradi05g17050* 和三叶草 TPR.G28382 具有其他的结构域,如 pfam05089、pfam12972 和 pfam12971 等。因此,这 2 个基因的基因长度和编码蛋白也远大于其他基因。仅具有 1 个 *UPF0114* 结构域的 27 个 *UPF0114* 基因长度有较大变化,但编码蛋白长度、预测的等电点和分子量都比较接近,只有落花生 *Araip.9Y1W2* 蛋白长度明显较短,相应的等电点和分子量也与其他基因差异较大(表 1)。中间锦鸡儿 *CiDUF0114* 基因编码 291 氨基酸,与其他植物 *UPF0114* 蛋白长度相当。

表 1 8 种豆科植物、拟南芥和水稻中的 *UPF0114* 基因家族成员

Table 1 *UPF0114* genes identified from eight species of *Leguminosae*, *Arabidopsis* and rice

物种 Species	基因号 Gene ID	基因长度 (bp) Gene length	蛋白长度 (bp) Protein length	等电点 Isoelectric point	分子量 (kD) Molecular mass
拟南芥 <i>Arabidopsis thaliana</i> (L.) Heynh.	<i>AT5G13720</i>	1351	262	28.91	9.03
水稻 <i>Oryza sativa</i> L.	<i>Os07g46330</i>	2739	286	30.29	9.25
	<i>Os06g49640</i>	1780	257	27.64	9.12
	<i>Os03g52910</i>	1761	344	35.66	10.35
	<i>Os07g46320</i>	1644	266	28.43	8.54
蒺藜苜蓿 <i>Medicago truncatula</i> Gaertn.	<i>Medtr4g068770</i>	2605	262	28.82	9.18
	<i>Medtr4g082590</i>	1741	292	32.02	9.33
	<i>Medtr5g025750</i>	4432	267	29.21	8.63
大豆 <i>Glycine max</i> (L.) Merr.	<i>Glyma.20G027900</i>	4803	292	31.91	9.68
	<i>Glyma.15G276700</i>	15050	266	29.7	9.05
	<i>Glyma.09G217900</i>	3065	264	28.72	9.19
	<i>Glyma.01G152700</i>	3914	264	28.73	9.18
鹰嘴豆 <i>Cicer arietinum</i> L.	<i>Ca_22871.g</i>	1664	249	27.71	9.46
	<i>Ca_10659.g</i>	2854	265	29.30	8.81
菜豆 <i>Phaseolus vulgaris</i> L.	<i>Phvul.002G103300</i>	1592	288	31.74	9.28
	<i>Phvul.002G122800</i>	2760	265	29.05	9.14
	<i>Phvul.005G023200</i>	3820	268	29.2	8.89
豇豆 <i>Vigna radiata</i> var. <i>radiata</i> (L.) R.	<i>Vradi07g06650</i>	1310	240	26.42	8.99
	<i>Vradi11g07670</i>	2974	265	29.05	9.25
Wilczek	<i>Vradi05g17050</i>	25216	1089	122.90	7.07

表 1(续)

物种 Species	基因号 Gene ID	基因长度 (bp) Gene length	蛋白长度 (bp) Protein length	等电点 Isoelectric point	分子量 (kD) Molecular mass
三叶草 <i>Trifolium pratense</i> L.	<i>TPR.G7858</i>	2533	269	29.63	8.80
	<i>TPR.G7060</i>	3442	293	32.25	9.55
	<i>TPR.G28382</i>	21549	2124	240.84	7.63
木豆 <i>Cajanus cajan</i> (L.) Huth	<i>C.cajan_43907.g</i>	2598	259	28.30	9.11
	<i>C.cajan_10332.g</i>	7349	262	28.51	9.03
	<i>C.cajan_04896.g</i>	2083	276	30.2	9.43
落花生 <i>Arachis ipaensis</i> Krapov.& W.C.Greg.	<i>Araip.9Y1W2</i>	938	158	17.24	7.81
	<i>Araip.55PE5</i>	2311	253	28.38	8.99

2.4 系统进化分析

利用得到的豆科植物、拟南芥、水稻的 UPF0114 蛋白和中间锦鸡儿 CiUPF0114 共 30 条序列构建系统发育树,大肠杆菌 UPF0114 蛋白作为外类群(图 3)。可以看出,UPF0114 家族基因明显可以分成 A 和 B 两个亚家族,其中 A 亚家族又可以分成 A1 和 A2 两个分支。其中 B 亚家族成员较少,共有 10 个,包括本研究克隆获得的中间锦鸡儿 CiUPF0114。其余 20 个成员则被分在 A 亚家族, A1 和 A2 各有 10 个。拟南芥中的两个 UPF0114 蛋白都在 A 亚家族,而其他物种的 UPF0114 则 A 和 B 两个亚家族中都存在。来自于豆科植物的 UPF0114 大多聚在一起,同为双子叶植物的拟南芥 UPF0114 与豆科的关系较近,而作为单子叶植物水稻的 UPF0114 距离则大多相对较远,这与植物的分类关系保持一致。中间锦鸡儿 CiUPF0114 与蒺藜苜蓿、鹰嘴豆和三叶草的 UPF0114 聚在 1 个小枝上。

2.5 基因结构和保守基序分析

用基因组全长序列和 CDS 序列绘制出 29 个 UPF0114 基因的内含子外显子分布情况(图 4)。结果显示,三叶草 *TPR.G28382* 和豇豆 *Vradi05g17050* 因为具有其他结构域,内含子、外显子的数量远多于其他基因。落花生 *Araip.9Y1W2* 和拟南芥 *At5g13720* 分别有 3 个和 4 个外显子,而其他 26 个 UPF0114 基因虽然内含子长度各异,但是都具有 5 个外显子,2 个亚家族之间的基因并没有区别。

对 8 种豆科植物、拟南芥、水稻和中间锦鸡儿的 UPF0114 共 30 条蛋白序列进行保守基序(motif)分析,获得共有的 6 个保守基序及其在每条蛋白序列中的分布情况(图 5)。分析发现, A 亚家族的 UPF0114 中,除了水稻 *Os07g46320* 和 *Os06g49640*、

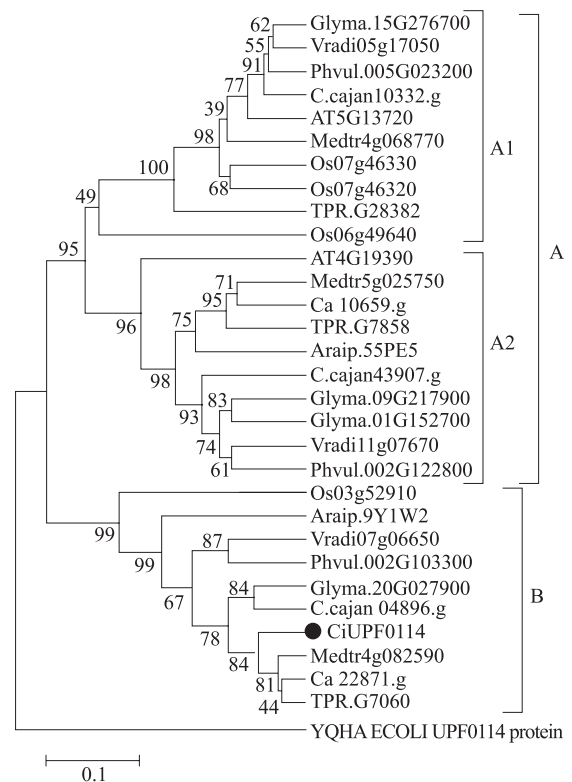
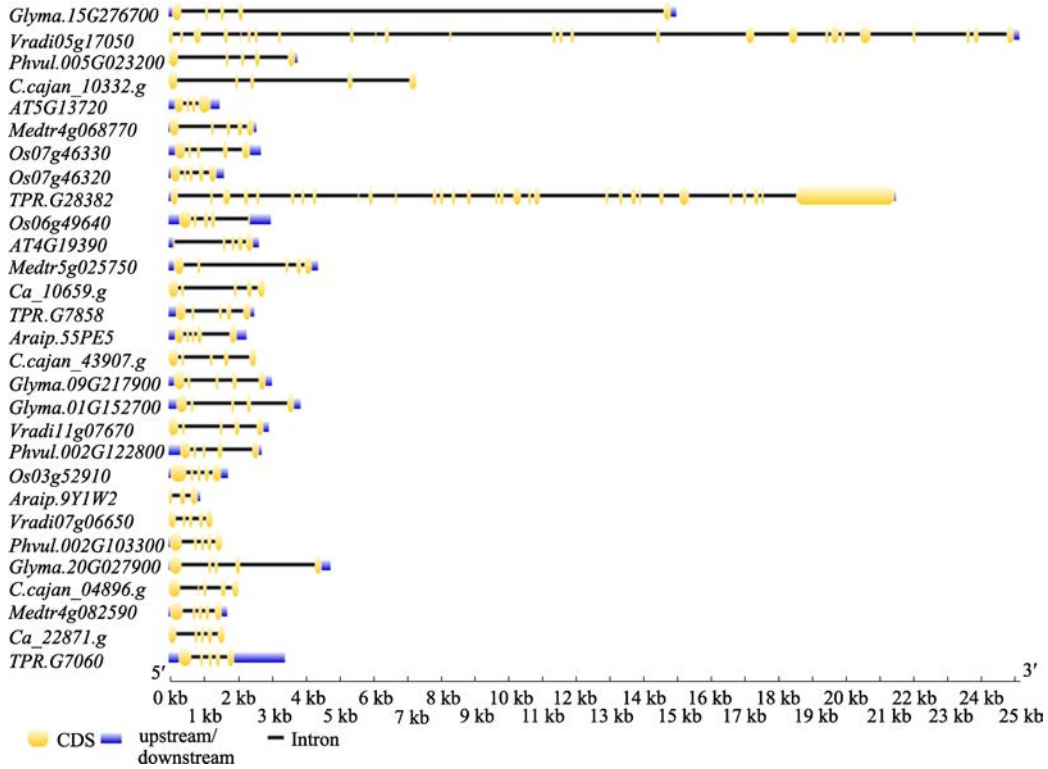


图 3 8 种豆科植物、拟南芥、水稻和中间锦鸡儿 UPF0114 系统进化分析

Fig.3 Phylogenetic analysis of UPF0114 proteins among eight legumes, Arabidopsis and rice

三叶草 *TPR.G28382*, 其他的 17 条蛋白序列都具有全部 6 个保守基序。而在 B 亚家族中,保守基序数量变化差异较大,落花生 *Araip.9Y1W2* 和鹰嘴豆 *Ca_22871.g* 分别仅有 3 个和 4 个保守基序,其他 8 条蛋白序列则有 5 个或 6 个保守基序。其中 motif 6 主要存在于 A 亚家族的全部成员中,而 B 亚家族中只有大豆 *Glyma.20G027900*、木豆 *C.cajan_04896.g* 和中间锦鸡儿 CiUPF0114 中有 motif 6。中间锦鸡儿 CiUPF0114 具备从 motif 1~6 全部的保守基序。

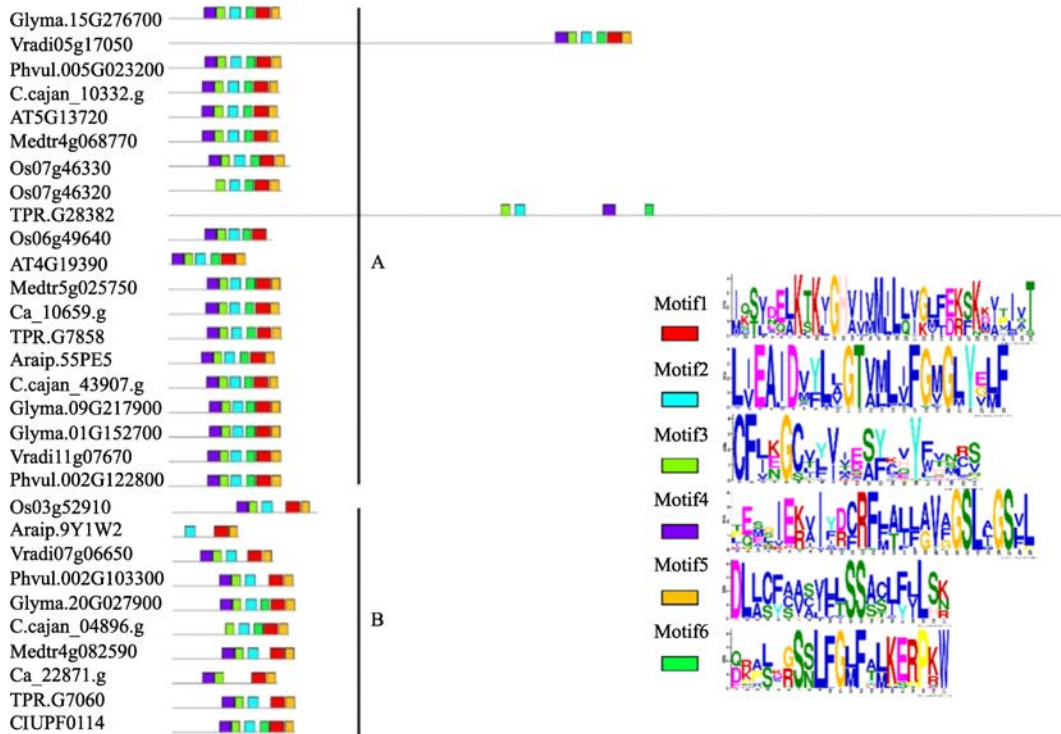


图中黑色的线表示内含子, 黄色方块代表外显子, 蓝色方块代表非编码区

For all genes, black lines represent introns, yellow boxes represent exons and blue boxes represent UTRs

图 4 8 种豆科植物、拟南芥和水稻 *UPF0114* 基因内含子、外显子结构

Fig.4 The exon-intron structure of *UPF0114* genes from eight legumes, Arabidopsis and rice



灰色线代表没有保守基序的序列, 每个彩色的方块代表一个保守基序

Grey lines represent the non-conserved sequences, and each motif is indicated by a colored box numbered on the right side of the figure. Sequence logos for the conserved motifs of *UPF0114* proteins

图 5 8 种豆科植物、拟南芥和水稻 *UPF0114* 蛋白保守基序

Fig.5 Conserved motifs of *UPF* proteins from eight legumes, Arabidopsis and rice

2.6 *CiUPF0114* 基因在非生物胁迫下的表达分析

转录组数据显示 *CiUPF0114* 受到脱水处理的诱导,推测该基因与植物的非生物胁迫响应有关。利用荧光定量 PCR 检测 *CiUPF0114* 在冷、干旱和脱水处理下的表达量,结果如图 6 所示,3 种胁迫处理后 *CiUPF0114* 基因的转录水平表达量比未处理

时都明显上调。其中, *CiUPF0114* 基因在 4 °C 冷处理和脱水处理后,转录水平的表达量逐渐升高,在处理 12 h 后都达到最高,约为未处理的 20 倍左右。在土壤干旱试验中,处理后第 6 天基因表达量开始上升,在处理 9 d 和 12 d 后则达到干旱处理前的 10 倍左右。

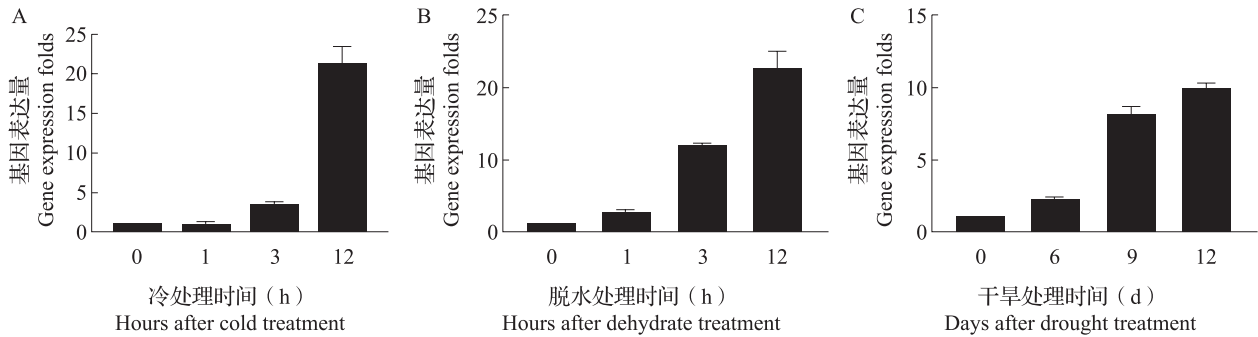


图 6 中间锦鸡儿 *CiUPF0114* 基因表达分析

Fig.6 The expression of *CiUPF0114* under drought, dehydration and cold treatment

3 讨论

非生物胁迫是造成世界粮食减产的最主要因素^[22]。对于在农牧业生产中有重要地位的豆科作物,如何增强其抗逆能力,培育耐逆新种质,是豆科作物育种研究中的一个重大课题。很多野生豆科植物具有较强的耐逆、耐贫瘠能力,环境适应性强,例如沙冬青^[9,23]、鹰嘴豆^[7,24]、骆驼刺 (*Alhagi sparsifolia* Shap.ex Keller & Shap.)^[25]、中间锦鸡儿和柠条锦鸡儿 (*Caragana korshinskii* Kom.)^[26] 等,从这些植物中克隆耐逆基因,研究其抗逆的分子机制和生理功能,可以为培育耐逆品种提供基因源。

本研究从抗逆豆科饲用灌木中间锦鸡儿中克隆得到 1 条功能未知基因,其编码蛋白具有 1 个 UPF0114 结构域,命名为 *CiUPF0114*。为了进一步了解豆科植物中 UPF0114 家族成员的数量、基因结构和进化关系,从植物基因组数据库中获得了 8 种豆科植物和模式植物拟南芥、水稻中的 29 个 UPF0114 基因(表 1)。系统进化分析显示这些 UPF0114 基因可以分成 A 和 B 两个大的亚家族, *CiUPF0114* 属于 B 亚家族的成员,单子叶的水稻及双子叶的拟南芥中都存在 A 和 B 两个亚家族的成员,这说明 A 和 B 亚家族在单子叶双子叶植物分离前便已经出现(图 3)。通过对基因编码蛋白长度、结构域数量、分子量、等电点、基因结构和保守基序

分析发现,除个别基因外,29 个 UPF0114 基因差异不大,说明该基因家族成员在进化上相对保守(表 1,图 4 和图 5)。

近些年来,植物中一些功能未知的基因逐步被研究者关注,很多报道证明了它们在植物抵抗逆境中有着重要的作用,例如水稻的 *DUF221* 和 *DUF1644* 家族基因参与植物对盐胁迫的响应^[27-28],紫花苜蓿 (*Medicago sativa*) 的 *MsDUF* 可以增强转基因番茄耐渗透胁迫的能力^[29],拟南芥中 *AtPep3* 在盐胁迫耐受和免疫应答中扮演短肽激素的角色^[30]。随着测序技术的发展,研究者发现越来越多的功能未知基因可能与植物抵抗逆境胁迫有关, UPF0114 基因家族就是其中的一类。具有 UPF0114 结构域的基因还没有任何功能方面的研究报道,仅在一些模式植物的表达谱数据中显示受到非生物胁迫的诱导,例如拟南芥 *AT4G19390* 基因在冷、盐和渗透胁迫下表达量明显上调 (AtGenExpress Visualization Tool, <http://jsp.weigelworld.org/expviz/expviz.jsp>)。对中间锦鸡儿转录组测序数据的分析发现, *CiUPF0114* 在脱水处理后表达量上调,处理 0 h、1 h、3 h 和 12 h 的表达量分别是 2.85、28.8、56.19 和 180.2 (RPKM 值),这与拟南芥中同源基因有类似的表达模式,预示着该基因可能与植物抗逆相关。进一步通过荧光定量 PCR 技术验证该结果,发现中间锦鸡儿 *CiUPF0114* 基因确实受到干旱、脱水和低温诱导(图 6),并且表达量随着处理时间延长逐渐升高,脱

水、低温处理 12 h 及干旱处理 12 d 时达到未处理时的 10~20 倍, 说明了 *CiUPF0114* 可能参与了中间锦鸡儿对非生物胁迫的响应。这些结果为进一步研究 *CiUPF0114* 基因在植物抗逆中的功能提供了依据。

参考文献

- [1] Zhu J F, Zhang L F, Li W F, Han S Y, Yang W H, Qi L W. Reference gene selection for quantitative real-time PCR normalization in *Caragana intermedia* under different abiotic stress conditions. *PLoS One*, 2013, 8 (1): e53196
- [2] 安超平, 王兴, 宋乃平, 周娟, 随金明. 荒漠草原中间锦鸡儿林土壤养分效应对群落组分和多样性的影响. *西北植物学报*, 2016, 36 (9): 1872-1881
An C P, Wang X, Song N P, Zhou J, Sui J M. Effect of soil nutrient on community composition and diversity of *Caragana intermedia* in desert steppe. *Acta Botanica Boreali-Occidentalia Sinica*, 2016, 36 (9): 1872-1881
- [3] Schlueter J. Genome sequence of the palaeopolyploid soybean. *Nature*, 2010, 463 (7278): 178-183
- [4] Young N D, DeBellé F, Oldroyd G E, Geurts R, Cannon S B, Udvardi M K, Benedito V A, Mayer K F, Gouzy J, Schoof H. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature*, 2011, 480 (7378): 520-524
- [5] Vega J J D, Ayling S, Hegarty M, Kudrna D, Goicoechea J L, Ergon Å, Rognli O A, Jones C, Swain M, Geurts R. Red clover (*Trifolium pratense* L.) draft genome provides a platform for trait improvement. *Scientific Reports*, 2015, 5: 17394
- [6] Varshney R K, Song C, Saxena R K, Azam S, Yu S, Sharpe A G, Cannon S, Baek J, Rosen B D, Tar'An B. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology*, 2013, 31 (3): 240-246
- [7] Varshney R K, Chen W, Li Y, Bharti A K, Saxena R K, Schlueter J A, Donoghue M T A, Azam S, Fan G, Whaley A M. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nature Biotechnology*, 2012, 30 (1): 83
- [8] Pang T, Ye C Y, Xia X, Yin W. De novo sequencing and transcriptome analysis of the desert shrub, *Ammopiptanthus mongolicus*, during cold acclimation using Illumina/Solexa. *Bmc Genomics*, 2013, 14 (1): 488
- [9] Zhu J F, Li W F, Yang W H, Qi L W, Han S M. Identification of microRNAs in *Caragana intermedia* by high-throughput sequencing and expression analysis of 12 microRNAs and their targets under salt stress. *Plant Cell Reports*, 2013, 32 (9): 1339-1349
- [10] Wan Y Q, Mao M Z, Wan D L, Yang Q, Yang F Y, Mandlaa, Li G J, Wang R G. Identification of the *WRKY* gene family and functional analysis of two genes in *Caragana intermedia*. *Bmc Plant Biology*, 2018, 18 (1): 31
- [11] Hiremath P J, Andrew F, Cannon S B, Jimmy W, Himabindu K, Reetu T, Ashish K, Amindala B P, Benjamin M, Neha G. Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnology Journal*, 2011, 9 (8): 922-931
- [12] 陈红霖, 胡亮亮, 王丽侠, 王素华, 程须珍. 绿豆 bHLH 转录因子家族的鉴定与生物信息学分析. *植物遗传资源学报*, 2017, 18 (6): 1159-1167
Chen H L, Hu L L, Wang L X, Wang S H, Cheng X Z. Genome-wide identification and bioinformatics analysis of bHLH transcription factor family in Mung Bean (*Vigna radiata* L.). *Journal of Plant Genetic Resources*, 2017, 18 (6): 1159-1167
- [13] Finn R D, Coggill P, Eberhardt R Y, Eddy S R, Mistry J, Mitchell A L, Potter S C, Punta M, Qureshi M, Sangrador-Vegas A. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, 2016, 44 (D1): D279-D285
- [14] Gasteiger E, Hoogland C, Gattiker A, Duvaud SE, Wilkins M R, Appel R D, Bairoch A. Protein identification and analysis tools on the ExPASy Server. *Humana Press*, 2005: 571-607
- [15] Goodstein D M, Shu S, Howson R, Neupane R, Hayes R D, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar D S. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, 2012, 40 (D1): D1178-D1186
- [16] Proost S, Bel M V, Sterck L, Billiau K, Parys T V, Peer Y V D, Vandepoele K. PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *The Plant Cell*, 2009, 21 (12): 3718-3731
- [17] Marchlerbauer A, Bo Y, Han L, He J, Lanczycki C J, Lu S, Chitsaz F, Derbyshire M K, Geer R C, Gonzales N R. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, 2017, 45 (Database issue): D200-D203
- [18] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology & Evolution*, 2016, 33 (7): 1870-1874
- [19] Hu B, Jin J, Guo A Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, 2015, 31 (8): 1296-1297
- [20] Bailey T L, Boden M, Buske F A, Frith M, Grant C E, Clementi L, Ren J, Li W W, Noble W S. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research*, 2009, 37 (suppl_2): W202-W208
- [21] Han X M, Feng Z Q, Xing D D, Yang Q, Wang R G, Qi L W, Li G J. Two NAC transcription factors from *Caragana intermedia* altered salt tolerance of the transgenic *Arabidopsis*. *BMC Plant Biology*, 2015, 15: 208
- [22] Zhu J K. Abiotic stress signaling and responses in plants. *Cell*, 2016, 167 (2): 313-324
- [23] Wei Q, Guo Y J, Cao H M, Kuai B K. Cloning and characterization of an /H antiporter gene from *Ammopiptanthus mongolicus* (Leguminosae) and its ectopic expression enhanced drought and salt tolerance in *Arabidopsis thaliana*. *Plant Cell Tissue & Organ Culture*, 2011, 105 (3): 309-316
- [24] Zeng F J, Song C, Guo H F, Liu B, Luo W C, Gui D W, Arndt S,

- Guo D L. Responses of root growth of *Alhagi sparsifolia* Shap. (Fabaceae) to different simulated groundwater depths in the southern fringe of the Taklimakan Desert, China. *Journal of Arid Land*, 2013, 5(2): 220-232
- [25] 马丽娟, 张巨松, 李利民, 陈晓露, 曾繁明, 杨忠芳, 羌松. 鹰嘴豆种质资源对 *Ascochyta rabiei* 的抗性评价及遗传多态性分析. *植物遗传资源学报*, 2013, 14(6): 1179-1184
- Ma L J, Zhang J S, Li L M, Chen X L, Zeng F M, Yang Z F, Qiang S. Resistance evaluation and genetic diversity analysis of Chickpea (*Cicer arietinum* L.) to *Ascochyta rabiei*. *Journal of Plant Genetic Resources*, 2013, 14(6): 1179-1184
- [26] Peng L, Lang S, Wang Y, Pritchard H W, Wang X. Modulating role of ROS in re-establishing desiccation tolerance in germinating seeds of *Caragana korshinskii* Kom. *Journal of Experimental Botany*, 2017, 68(13): 3585-3601
- [27] Ganie S A, Pani D R, Mondal T K. Genome-wide analysis of DUF221 domain-containing gene family in *Oryza* species and identification of its salinity stress-responsive members in rice. *PLoS One*, 2017, 12(8): e0182469
- [28] Li M, Guo L, Guo L J, Wang L J, Chen L. Over-expression of a DUF1644 protein gene, *SIDP361*, enhances tolerance to salt stress in transgenic rice. *Journal of Plant Biology*, 2016, 59(1): 62-73
- [29] Wang Y F, Zhang Z Q, Liu H M, An Y R, Han B, Wu Y J, Chang L Q, Hu T M, Yang P Z. Overexpression of an alfalfa (*Medicago sativa*) gene, *MsDUF*, negatively impacted seed germination and response to osmotic stress in transgenic tobacco. *Plant Cell Tissue & Organ Culture*, 2018, 132(3): 525-534
- [30] Nakaminami K, Okamoto M, Higuchi-Takeuchi M, Yoshizumi T, Yamaguchi Y, Fukao Y, Shimizu M, Ohashi C, Tanaka M, Matsui M, Shinozaki K, Seki M, Hanada K. AtPep3 is a hormone-like peptide that plays a role in the salinity stress tolerance of plants. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(22): 5810-5815