

doi:10.11937/bfy.20171745

## 马蔺 *DFR* 基因的克隆及生物信息学特征分析

虎娟<sup>1,2</sup>, 安韶雅<sup>1,2</sup>, 林哲<sup>1,2</sup>, 张虹<sup>1,2</sup>, 包晗<sup>1,2</sup>, 陈任<sup>1,2</sup>

(1. 宁夏大学 西部特色生物资源保护与利用教育部重点实验室, 宁夏 银川 750021;

2. 宁夏大学 生命科学学院, 宁夏 银川 750021)

**摘要:**以马蔺花瓣为试材,通过转录组测序,应用PCR技术克隆马蔺花青素生物合成关键酶二羟黄酮醇-4-还原酶(dihydroflavonol-4-reductase, DFR)基因并获得其生物信息学特征。结果表明:马蔺 *DFR* 基因全长 1 427 bp,共编码 357 个氨基酸(该基因命名为 *IIDFR*, GenBank 登录号为 KY907171)。该基因在氨基酸水平上与多种植物的 DFR 具有较高的同源性,在系统进化上与同属的荷兰鸢尾(*Iris hollandica*)的 DFR 亲缘关系较近;预测其蛋白质相对分子质量为 39.99 kDa,等电点为 5.89,主要由  $\alpha$ -螺旋和  $\beta$ -折叠构成,定位于细胞质,属于酸性亲水不含信号肽的不稳定类蛋白质。马蔺 *IIDFR* 有一段氨基酸序列‘VTGASGYVGSWLVMKLLRDGY’与大部分物种的 DFR 特有相对保守的 NADPH 结合域非常相似,只有 4 个氨基酸残基有所差异。

**关键词:**马蔺;转录组测序;花青素;二羟黄酮醇-4-还原酶;生物信息学分析

**中图分类号:**S 682.1<sup>+</sup>9 **文献标识码:**A **文章编号:**1001-0009(2017)24-0109-07

马蔺(*Iris lactea* Pall. var. *chinensis* (Fisch.) Koidz, 以下简称 *I. lactea* var. *chinensis*)属鸢尾

**第一作者简介:**虎娟(1991-),女,硕士研究生,研究方向为植物分子生理与生物技术。E-mail:380128339@qq.com.

**责任作者:**陈任(1963-),男,博士,教授,现主要从事植物分子生物学与基因工程等研究工作。E-mail:chenren511@163.com.

**基金项目:**2013 年度宁夏回族自治区科技支撑计划资助项目。

**收稿日期:**2017-07-13

科鸢尾属多年生草本宿根植物,是白色马蔺的变种又名马兰花、马莲花等,广泛分布于我国东北、华北、西北等地<sup>[1]</sup>。由于马蔺具有独特的生态生物学特性和利用价值,近年来逐渐被用作水土保持、园林地被、山体护坡等用途的优良材料<sup>[2]</sup>。其花为浅蓝色、蓝色或蓝紫色,花青素是构成其花瓣颜色的主要色素之一,在植物的生理进程中扮演重要的作用<sup>[3]</sup>。目前有关马蔺的植物学特性、生物学特性、形态特征、栽培及应用研究较多<sup>[4]</sup>,而关于马蔺花青素生物合成途径的研究文献较少,

**Abstract:** Two different light adaptation of Taiwan *Zoysia matrella* with pre-treatment under strong radiation and extreme shade were employed to study the effects of 0.5 mol · L<sup>-1</sup> NaCl for 24 days on photochemical efficiency by using chlorophyll fluorescence technology. The results showed that the Fv/Fm, PhiPS II, Fv'/Fm', qP and ETR in leaves of two light adaptation types gradually reduced under 0.5 mol · L<sup>-1</sup> NaCl, NPQ gradually increased, photochemical efficiency gradually reduced and photoinhibition appeared, and photoinhibition increased with the extension of time. Adaptability of the sun-type to salt stress was stronger than the shade-type.

**Keywords:** Taiwan *Zoysia matrella*; light-adaptation; salt stress; fluorescence parameters; photochemical efficiency

尚鲜见马蔺花青素生物合成途径关键酶的文献报道。

二羟黄酮醇-4-还原酶(dihydroflavonol-4-reductase, DFR)是花青素生物合成途径中的一个关键酶,其选择性地催化二氢杨梅黄酮(dihydromyricetin, DHM)二氢山奈黄酮醇(dihydrokaempferol, DHK)和二氢槲皮黄酮(dihydroquercetin, DHQ)分别生成无色飞燕草色素、无色天竺葵色素和无色矢车菊色素。这3种物质随后在花青素合酶(ANS)和各种糖苷转移酶(UDP-glycosyltransferases, UDPGs)的作用下转化为3种类型的花青素,即飞燕草色素(蓝紫色)、天竺葵色素(红色)和矢车菊色素(品红色)。该酶是花青素从无色转为有色的一个调控点,在花青素合成途径中是重要的节点基因<sup>[5]</sup>,能使花青素在植物体内积累。O'REILLY等<sup>[6]</sup>1985年采用转座子标签技术从玉米和金鱼草中分离出了DFR基因,现已从矮牵牛、非洲菊、番茄、葡萄、苜蓿、紫苏、罗莎、牵牛花、白杨、海小麦、百脉根、大白菜、茶叶和莲花等植物中克隆到该基因<sup>[7-8]</sup>。

该试验通过马蔺花瓣转录组测序,根据测序结果利用PCR技术从马蔺花瓣中克隆获得DFR基因,并通过生物信息学方法对该基因进行特征分析,为进一步研究马蔺花青素生物合成相关基因、遗传改良和分子育种奠定理论基础。

## 1 材料与方法

### 1.1 试验材料

马蔺花瓣采自宁夏大学周边,新鲜花瓣装入离心管中放入超低温冰箱(-80℃)保存备用。大肠杆菌(*Escherichia coli*)菌株DH5 $\alpha$ 和pKAFCR0载体由宁夏大学西部特色生物资源保护与利用教育部重点实验室提供。

### 1.2 试验方法

#### 1.2.1 马蔺花瓣总RNA的提取

从超低温冰箱中取出先前保存的马蔺花瓣,放入液氮中迅速冷却后放入细胞破碎仪Tissue Lyser 85200(德国Qiagen)中进行破碎,采用RNA prep Pure Plant Kit试剂盒(北京天根生化科技,DP441)提取总RNA。分别采用Nanodrop

2000(美国Thermo Fisher Scientific)、Qubit 2.0(美国Life Technologies)和Agilent 2100(美国Agilent Technologies)检测RNA样品的纯度(OD<sub>260</sub>/OD<sub>280</sub>比值)、浓度和完整性(rRNA 28S/18S比值)等。

#### 1.2.2 马蔺转录组测序

转录组测序由北京百迈客生物科技有限公司完成。测序的试验流程:用带有Oligo(dT)的磁珠富集真核生物mRNA;随后将mRNA进行随机打断,利用随机引物合成第一、二条cDNA链,然后加A尾连接测序接头,通过PCR富集得到cDNA文库。

cDNA文库检验合格后,采用HiSeq 2500(美国Illumina)进行高通量测序,测序读长为PE125。利用Trinity软件(英国Broad)进行序列拼接组装,获得该物种的Unigene库。利用BLAST(美国NCBI)将Unigene序列与NR、Swiss-Prot、KEGG、COG、KOG、GO和Pfam数据库比对,获得Unigene的注释信息。利用Trans Decoder软件(英国Broad和澳大利亚CSIRO)预测以上Unigene的编码区序列(Coding Sequence, CDS)。

#### 1.2.3 马蔺cDNA的合成

利用从马蔺花瓣提取的总RNA,参照PrimerScript<sup>TM</sup> II 1st Strand cDNA Synthesis Kit(日本Takara, 6210A)说明书进行cDNA的合成。

#### 1.2.4 马蔺IIDFR基因的PCR扩增

基于转录组测序数据,利用Primer Premier 5.0软件(美国Premier Bio)设计一对特异性PCR扩增引物(上游引物ILDFR-S: 5'-GCCGCACTATACTTATACTACACTTTTCT-3';下游引物ILDFR-A: 5'-CTATCTTCATATTTCACTTGACCTTTTG-3')。以合成的cDNA(50 ng)为模板,采用Phusion High-Fidelity DNA聚合酶(美国New England Bio Labs, M0530S)进行马蔺IIDFR基因的PCR扩增。反应条件为98℃预变性30 s;98℃变性10 s,60℃退火30 s,72℃延伸45 s,共30个循环;最终72℃延伸7 min。

#### 1.2.5 马蔺IIDFR基因的克隆及测序

PCR产物以及经Sma I酶切,CIAP(calf intestine alkaline phosphatase,日本Takara, 2250A)脱磷处理的pKAFCR0载体<sup>[9]</sup>,经1%琼

脂糖凝胶电泳后,利用 Wizard® SV Gel and PCR Clean-Up System 试剂盒(美国 Promega, A9281)回收,之后利用 Ligation High 试剂盒(日本 Toyobo, LGK101)连接,连接产物转化至大肠杆菌 DH5 $\alpha$  感受态细胞;随后进行菌落 PCR 鉴定,检测其阳性重组子,采用 IIDFR-S、IIDFR-A 引物检测目的基因片段是否插入;采用 ILDFR-A、CR0 MCS insert 35S-S (5'-CACAATCCCAC-TATCCTTCGCA-3')及 ILDFR-S、CR0 MCS insert NOS-A (5'-TCCACTCTAATCATA-AAAACCCATCTC-3')分别检测插入基因片段的正反向;采用 Plasmid Mini Kit I(美国 Omega Bio-Tek, D6943-01)试剂盒提取其质粒后送至昆泰锐生物技术有限责任公司进行测序验证。

#### 1.2.6 马蔺 IIDFR 基因的序列分析

利用 NCBI 在线 BLAST 搜索 IIDFR 基因的同源序列,从而判断该基因氨基酸序列的完整性和与其它物种的相似性;使用 Lasergene Protean(美国 DNA Star)软件包中的 MegAlign 软件和 TreeView 软件进行同源性分析及构建分子系统进化树<sup>[10]</sup>。

#### 1.2.7 马蔺 IIDFR 蛋白的物理特性预测

使用 Lasergene Protean 软件包中的 Editseq 软件、Protean 软件和 ProtParam(<http://web.expasy.org/protparam/>)预测其蛋白质的二级结构与理化性质,包括相对分子质量、等电点等<sup>[11]</sup>;使用在线亚细胞定位工具 PSORT II Prediction(<http://psort.hgc.jp/form.html>)预测蛋白质的亚细胞定位<sup>[12]</sup>。利用 SignalP 软件(<http://www.cbs.dtu.dk/services/SignalP/>)分析信号肽<sup>[13]</sup>。

#### 1.2.8 马蔺 IIDFR 蛋白的三级结构预测

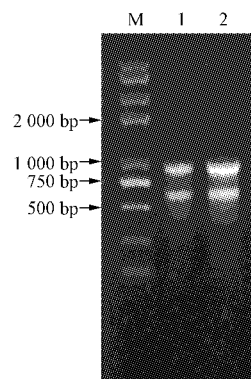
通过 Swiss-Model Workspace(<http://swissmodel.expasy.org>)在线分析工具对其蛋白质进行同源建模,预测其三级结构<sup>[14-16]</sup>。

## 2 结果与分析

### 2.1 马蔺花瓣总 RNA 的提取

从马蔺花瓣提取的总 RNA 的 rRNA 28S 和 18S 条带清晰,28S/18S 为 1.7, RIN 值为 7.3, OD<sub>260</sub>/OD<sub>280</sub> 为 2.09,表明提取的总 RNA 完整性

较好(图 1),可以用于下一步的 cDNA 合成和转录组测序。



注:M. Marker 2 000, 下同;1~2. 总 RNA。

Note: M. Marker 2 000, the same below; 1~2. Total RNA.

图 1 马蔺花瓣总 RNA 琼脂糖凝胶电泳

Fig. 1 Agrose gel electrophoresis of total RNA extracted from the petals of *I. lactea* var. *chinensis*

### 2.2 马蔺 IIDFR 基因的 PCR 扩增

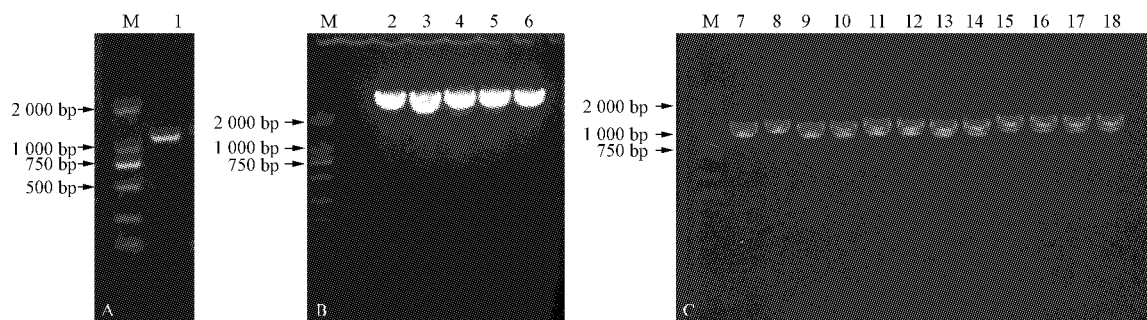
用反转录得到的 cDNA 为模板,根据转录组测序数据设计 DFR 基因引物,通过 PCR 技术扩增出大小约为 1 400 bp 的 PCR 条带(图 2A),所得大小符合预期。

### 2.3 马蔺 IIDFR 基因的克隆及测序

PCR 产物与经酶切后的 pKAFCR0 载体(图 2B)连接、大肠杆菌转化、菌落 PCR 检测插入片段以及正反向(图 2C)后,提取重组质粒进行测序,测序结果表明马蔺 DFR 基因全长 1 427 bp,起始密码子 ATG 位于 175 bp 处,终止密码子位于 1 248 bp 处,CDS 区全长 1 074 bp,共编码 357 个氨基酸(图 3)。该基因命名为 IIDFR,基因信息在 GenBank 的登录号为 KY907171。IIDFR 有一段氨基酸序列‘VTGASGYVGSWLV MKLL-RDGY’与大部分物种的 DFR 特有、相对保守的 NADPH 结合域‘VTGASGFVGSWLV MRLL-HGY’非常相似,只存在 4 个氨基酸残基(下划线标注)有所差异。

### 2.4 马蔺 IIDFR 基因序列分析

根据 BLAST 分析结果表明,马蔺 IIDFR 基因所编码的氨基酸与多种植物的 DFR 蛋白具有较高的同源性,其中与荷兰鸢尾(*Iris hollandica*,



注:1. PCR 产物;2~6. pKAFCR0 载体经 *Sma* I 酶切;7~18. 马蔺 *IIDFR* 基因连接至 pKAFCR0 载体后菌落 PCR 鉴定(7~10 采用 *IIDFR*-S, *IIDFR*-A 引物;11~14 采用 CR0 MCS insert 35S-S, *IIDFR*-A 引物;15~18 采用 *IIDFR*-S, CR0 MCS insert NOS-A 引物)。

Note:1. PCR products;2~6. pKAFCR0 vector digested by *Sma* I. 7~18 Colony PCR identification of *IIDFR* gene constructed into pKAFCR0 vector (7~10 used *IIDFR*-S and *IIDFR*-A primer pair;11~14 used CR0 MCS insert 35S-S and *IIDFR*-A primer pair;15~18 used *IIDFR*-S and CR0 MCS insert NOS-A primer pair).

图2 马蔺 *IIDFR* 基因重组质粒构建过程

Fig. 2 Construction progress of *IIDFR* vector

```

0001 AGGAAGAGGAGCAGTTGTGCTTGCTCTTTTCGAAATAGGGAACCTGTCACACAGATGTTGGTCTCTCTTCGTATAAAAAGGACGTAC
0091 ACCACCATAAGCTAGCTAATAACTAAGCTTTACCACACTATACTTATACACTTTTCTTTCTTTCTTTATATATCGAGCTACATGATC
保守序列 conserved region
0181 ATGAGCCCGGTTCTGTGACCGGAGCGGCTACGTCGGTTCATGGCTTGTATGAAGCTTCTCGCGACGGCTACGTCGGTTCGAGCC
M S P V L V T G A S G Y V G S W L V M K L L R D G Y V V R A
0271 ACTGTCAGAGATCCAACTAATGTAAGTAGAGAGGCTCTGTGGACCTCCGAGAGCTGACGAGCTGCTCAGCATCTGGAAGGCGGAC
T V R D F T N V K K T K P L L D L P R A D E L L T I W K A D
0361 CTGCGCCAGGACGGAAGCTTCGACGAGGCGGTCGACGGTGCACCGCGGCTCTTCACGTCGCGCACGCCATGGAATTCGAGTCCAAGGAC
I G Q D G S F D E A V A G C T A V F H V A T P M D F E S K D
0451 CCGMAAAGCAAGTGATCAAGCGGACCATAAATGGCGTTTAAAGCATCTGAGGTCCTGTAAGAAGGCCAGGACGGTCAMCGCGTCGTC
P E N E V I K P T I A G V I S I L R S C K K A R T V K R V V
0541 TTCACTTCGTCGCTCGGACGGTGGAGCTGAAGGAACATCAGCAGATGGAGTACGAGCAGAGCTCGTGGAGCGATGTCGAGTTCTGAGA
F T S S A G T V D V K E H Q Q M E Y D E S S W S D V E F C R
0631 CGCGTCAAGATGACTGGCTGGATGTATTTTGTGTCAGAGCTCTGGCTGAGAAAGCAGCGCTGGGATTTTGAAGAGAGAATGGCATAGAC
R V K M T G W M Y F V S K T L A E K A A W D F A R E N G J D
0721 TTGATAAGCATCATACCAACGATGGTCTGTCGTCCTTCATCACCAGCACCATGCCACGAGCATGGTCACTGCGTTATCATTCATGACA
L I S I I P T M V V G P I T T I T V P P S M V T A I S F M T
0811 GGAACAAAGCACTATCATAATCAAGCAAGCGAGCTCATCCACCTCGAGCAGCTTCGCGCGCTCAGATCTACTCTCTGAATCGC
G A K A H Y H I I K Q A Q I I H I D D I C A A H I Y I I A R
0901 CCGAGGCGAAGCGAGGTACATATGCTCCTCGACGAGGCCACCATCTACGATCTAGCGAGGATGGTCAAGAGAGGACCCCTGCTGT
P E A N G R Y I C S S H E A T I Y D L A R M V R E R H P W C
0991 TCCATACCCAAAAGTTTGAAGGATCGAGGAGGCGATCGAAGTCTGCACTTCTCTTGAAGAGGCTATTGGACCTCGGATTCGAGTTC
S I P K K F D G I E E G I E V V H F S S K R L L D L G F E P
1 081 AAGCACACGGTGAAGAGATGTTGACGAGGCGATCCAGTCGTGATCGAAAGAAGCTCATACCCCTCCCTGGAATGGCAAAGTGGTG
K H T V E E M F D E A J Q S C I E K K L I P L P E N G K V V
1 171 GATGCTGGGGAGGAAGTTGCGGCTAAAGCATAGTTTCATGGAGAGAGGAACGTGCCGCAATTGCTATAGGCTAAATCAAAAGTCA
D A G E E E V G A K D I V H G E E E R A R I A I G *
1 261 AGTGAATATGAAGATAGAGCATTTATGCATATGGGATATTCCTAATGTCAGATGTTCAAGGTTTTTCAGATTATATGTTAAAGTTGAGTT
1 351 ATGTTGACACTCAGAAGTTTCACTGAAATAATGAATGAAAAGGATTTATCTAATCTTAGAATAAACTATAGTTTGC

```

图3 马蔺 *IIDFR* 基因全长序列

Fig. 3 Complete sequence of *IIDFR*

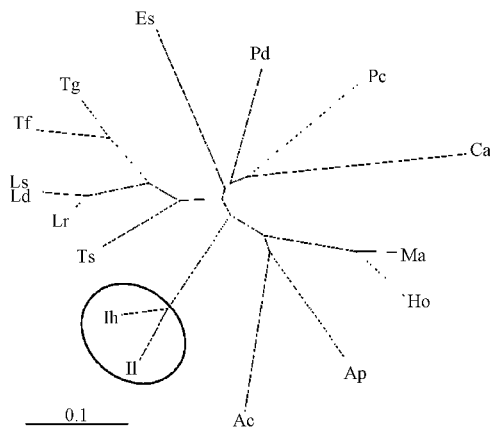
BAF93856.1)、白子莲 (*Agapanthus praecox*, BAE78769.1)、岷江百合 (*Lilium regale*, BAU29941.1)和风信子 (*Hyacinthus orientalis*,

AFP58815.1)的同源性分别达到 88%、69%、76%和 69%;通过构建分子系统进化树表明该基因编码的蛋白与同属的荷兰鸢尾 DFR 蛋白亲缘

关系最为较近(图 4)。

## 2.5 马蔺 IIDFR 蛋白物理特性预测

根据 Lasergene Protean 和 ProtParam 软件预测,马蔺 DFR 基因编码的蛋白相对分子质量为 39.99 kDa,等电点为 5.89,为酸性蛋白质,分子式为  $C_{1784}H_{2821}N_{477}O_{525}S_{20}$ ;氨基酸中 Val (9.0%)、Glu(8.1%)及 Ala(7.6%)的频率较高,极性氨基酸占 19.37%,疏水性氨基酸占 35.11%,带电荷氨基酸占 39.26%(表 1);平均亲水性为-0.124,说明该蛋白属亲水性蛋白<sup>[17]</sup>,其不稳定系数为 40.37,属不稳定类蛋白质<sup>[18]</sup>;预测其亚细胞定位表明该蛋白主要在细胞质(56.5%)



注.: Ac. 洋葱 (AAO63026.1); Ap. 白子莲 (BAE78769.1); Ca. 姜荷花 (ADK62520.1); Es. 三枝九叶草 (AFU90826.1); Ho. 风信子 (AFP58815.1); Ih. 荷兰鸢尾 (BAF93856.1); Ls. 药百合 (BAE79202.1); Ld. 鹤望兰 (BAL15045.1); Lr. 岷江百合 (BAU29941.1); Ma. 亚美尼亚葡萄风信子 (AIC33028.1); Pc. 同色兜兰 (AFS60083.1); Pd. 海枣 (XP\_008797532.1); Ts. 毛油点草 (BAN62761.1); Tg. 郁金香 (BAH98155.1); Tf. 福斯特郁金香 (AGJ50590.1)。

Note: Ac. *Allium cepa* (AAO63026.1); Ap. *Agapanthus praecox* (BAE78769.1); Ca. *Curcuma alismatifolia* (ADK62520.1); Es. *Epimedium sagittatum* (AFU90826.1); Ho. *Hyacinthus orientalis* (AFP58815.1); Ih. *Iris hollandica* (BAF93856.1); Ls. *Lilium speciosum* (BAE79202.1); Ld. *Lilium hybrid division* (BAL15045.1); Lr. *Lilium regale* (BAU29941.1); Ma. *Muscari armeniacum* (AIC33028.1); Pc. *Paphiopedilum concolor* (AFS60083.1); Pd. *Phoenix dactylifera* (XP\_008797532.1); Ts. *Tricyrtis sp. Shinonome* (BAN62761.1); Tg. *Tulipa gesneriana* (BAH98155.1); Tf. *Tulipa fosteriana* (AGJ50590.1)。

图 4 马蔺 IIDFR 基因的系统进化树

Fig. 4 Phylogenetic tree of IIDFR

中起作用,其次在线粒体和细胞核中分布比例均为 13%,因此定位于细胞质中;SignalP 软件分析表明 DFR 蛋白不存在信号肽。具有 17 个  $\alpha$ -螺旋,21 个  $\beta$ -折叠,12 个  $\beta$ 转角,8 个无规则卷曲。

表 1 马蔺 IIDFR 的氨基酸组成

Table 1 Composition of amino acid in IIDFR

氨基酸	数量	频率	氨基酸	数量	频率
Amino acid	Number	Frequencies/%	Amino acid	Number	Frequencies/%
Ala (A)	27	7.6	Lys (K)	24	6.7
Arg (R)	18	5.0	Met (M)	13	3.6
Asn (N)	8	2.2	Phe (F)	14	3.9
Asp (D)	22	6.2	Pro (P)	16	4.5
Cys (C)	7	2.0	Ser (S)	22	6.2
Gln (Q)	6	1.7	Thr (T)	21	5.9
Glu (E)	29	8.1	Trp (W)	6	1.7
Gly (G)	22	6.2	Tyr (Y)	8	2.2
His (H)	11	3.1	Val (V)	32	9.0
Ile (I)	25	7.0	Pyl (O)	0	0
Leu (L)	26	7.3	Sec (U)	0	0

## 2.6 马蔺 IIDFR 蛋白三级结构预测

采用 SWISS-MODEL 软件对马蔺 DFR 建模获得三级结构模型(图 5),说明马蔺 DFR 主要由  $\alpha$ -螺旋和  $\beta$ -折叠构成,这与二级结构预测基本一致。

## 3 讨论与结论

该试验根据马蔺花瓣转录组测序数据,利用 PCR 技术克隆了马蔺 IIDFR 基因全长并对其进行生物信息学特征分析。马蔺 IIDFR 基因全长 1 427 bp,CDS 区全长 1 074 bp,共编码 357 个氨基酸(GenBank 登录号为 KY907171)。其编码的蛋白在氨基酸水平上与多种植物的 DFR 蛋白具有较高的同源性,在分子系统进化上与同属的荷兰鸢尾 DFR 亲缘关系较近。预测马蔺 IIDFR 定位于细胞质,属于酸性亲水不含信号肽的不稳定类蛋白质。

DFR 属于 NADPH 依赖性短链还原酶家族或者是 DFR 亚家族,在不同的物种中均存在相对保守的 NADPH 结合域,大部分为‘VTGASG-FVGSWLVMRLLEHGY’,该区域决定底物的特异性<sup>[19]</sup>。不同植物的 DFR 在花青素生物合成中对 DHM、DHK 和 DHQ 3 种底物的选择有差异性,如非洲菊的 DFR 能以 DHM、DHK 和 DHQ

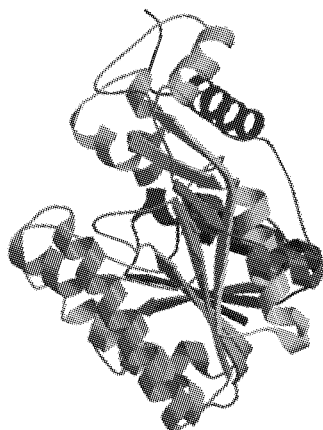


图5 马蔺 IIDFR 蛋白的三级结构预测

Fig. 5 Protein 3D structure prediction of IIDFR

为底物,而矮牵牛 DFR 则主要催化 DHM,而对 DHK 没有转化作用,因而矮牵牛缺乏橙色的天竺葵色素<sup>[20]</sup>,另外,不同物种的 DFR 基因在不同生长时期、不同部位的表达量也不同,如在矮牵牛中分离了 3 种 DFR 基因 (DFRA、DFRB 和 DFRC),但是只有 DFRA 基因在花中转录表达,在胚珠和茎中却只有微量表达<sup>[21]</sup>;对石斛兰的 1 个 DFR 基因的研究发现,DFR 酶活性与石斛兰花瓣的颜色变化密切相关<sup>[22]</sup>。马蔺 IIDFR 的氨基酸序列也同样存在相似、相对保守的 NADPH 结合域‘VTGASGYVGSWLVMKLLRDGY’,其中 4 个氨基酸残基(下划线标注)与大部分物种有所差异。马蔺的花瓣基本为蓝色,是白花马蔺 (*I. lactea*) 的变种,这一结合域的差异是否与底物特异性、花色有关,还有待进一步进行酶活性和底物特异性鉴定。

目前,DFR 基因在很多植物中都已成功克隆,而其在改造花色方面的利用也取得了进展。由于 DFR 具有底物的特异性,利用该特点可以定向改变花的颜色。1987 年,MEYER 等<sup>[23]</sup>首次将玉米的 DFR 基因转入到矮牵牛中,改变了矮牵牛的花色。TANAKA 等<sup>[24]</sup>从玫瑰花瓣中克隆出了 DFR,将其转入淡粉红色矮牵牛中,从而产生了橙红色的矮牵牛。JAMES 等<sup>[25]</sup>通过探针标记技术分离克隆了蔓越橘 DFR 基因,并转到烟草中使其表达。该研究结果为利用转基因技术将马蔺 IIDFR 基因导入到其它花卉中,改变花卉原有的花青

素生物合成途径,创造出自然界不存在、传统育种方法无法培育的新花色品种奠定了基础。

#### 参考文献

- [1] 中国科学院中国植物志编辑委员会. 中国植物志: 鸢尾科 (16 卷第一分册) [M]. 北京: 科学出版社, 1985: 34.
- [2] 孟林, 张国芳, 赵茂林. 水保护坡观赏优良地被植物: 马蔺 [J]. 农业新技术, 2003(3): 38-39.
- [3] 郭凤丹, 王效忠, 刘学英, 等. 植物花青素生物代谢调控 [J]. 生命科学, 2011(10): 938-944.
- [4] 徐玉凤, 史国旭, 金罡, 等. 鸢尾属植物马蔺 (*Iris lactea* Pall. var. *chinensis*) 的研究进展 [J]. 种子, 2011(4): 67-70.
- [5] 刘娟, 冯群芳, 张杰. 二氢黄酮醇-4-还原酶基因 (DFR) 与花色的修饰 [J]. 植物生理学讯, 2005, 41(6): 715-719.
- [6] O'REILLY C, SHEPHERD N S, PEREIRA A, et al. Molecular cloning of the *al* locus of *Zea mays* using the transposable elements *En* and *Mu1* [J]. The EMBO Journal, 1985(4): 877-882.
- [7] 赵启明, 李范, 李萍. 花青素生物合成关键酶的研究进展 [J]. 生物技术通报, 2012(12): 25-32.
- [8] 高燕会, 黄春红, 朱玉球, 等. 植物花青素苷生物合成及调控的研究进展 [J]. 中国生物工程杂志, 2012, 32(8): 94-99.
- [9] 张芮. 甜菊醇糖苷生物合成途径关键基因的功能评价 [D]. 银川: 宁夏大学, 2016.
- [10] 季舒涵, 咎林森, 王洪宝, 等. 秦川牛 *A-FABP* 基因的生物信息学分析 [J]. 西北农林科技大学学报 (自然科学版), 2010, 38(6): 77-81.
- [11] 石宁宁, 杜晓华, 罗玉柱, 等. 甘南牦牛 *NGB* 基因克隆及序列分析 [J]. 西北农林科技大学学报 (自然科学版), 2013, 41(4): 1-7.
- [12] 孙伟, 李达, 苏锐, 等. 绵羊 *YAP1* 基因全长 cDNA 克隆及生物信息学分析 [J]. 中国农业科学, 2013, 46(8): 1725-1735.
- [13] PETERSEN T N, BRUNAK S, von HEIJNE G, et al. SignalP

- 4.0; Discriminating signal peptides from transmembrane regions [J]. *Nature Methods*, 2011(8):785-786.
- [14] 刘虎岐, 刘应保, 宋云鹏, 等. 淡色库蚊氯菊酯抗性相关基因 PR-OP 全长 cDNA 的克隆及生物信息学分析[J]. 西北农林科技大学学报(自然科学版), 2010, 38(9):109-116.
- [15] 陶士珩. 生物信息学[M]. 北京: 科学出版社, 2007: 177-179.
- [16] BLASINI M, BIENERT S, WATERHOUSE A, et al. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information [J]. *Nucleic Acids Research*, 2014, 42: W252-W258.
- [17] KYTE J, DOOLITTLE R F. A simple method for displaying the hydropathic character of a protein[J]. *Journal of Molecular Biology*, 1982, 157:105-132.
- [18] GURUPRASAD K, REDDY B V B, PANDIT M W. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence[J]. *Protein Engineering*, 1990(4):155-161.
- [19] 张波. 杜果 *DFR* 和 *UFGT* 基因的克隆及其对果实着色的影响[D]. 海口: 海南大学, 2015.
- [20] 吴少华, 张大生. 花青素生成相关基因 *DFR* 的研究进展[J]. 福建林学院学报, 2002, 22(2):189-192.
- [21] BELD M, MARTIN C, HUIITS H, et al. Flavanoid synthesis in *Petunia hybrida*: Partial characterization of dihydroflavonoreductase genes[J]. *Plant Molecular Biology*, 1989(13):491-502.
- [22] RICHARD E L. The mango: Botany, production and uses [M]. UK: CABI Head Office, 2005: 3-4.
- [23] MEYER P, HEIDMANN I, FORKMANN G, et al. A new petunia flower colour gene rated by transformation of a mutant with a maize gene[J]. *Nature*, 1987, 330:677-678.
- [24] TANAKA T, FUKUI Y, FUKUCHI M, et al. Molecular cloning and characterization of rose hybrid dihydroflavonoreductase gene[J]. *Plant & Cell Physiology*, 1995, 36(6):1023-1031.
- [25] JAMES J, POLASHOK A, ROBERT J. Cloning of a cDNA encoding the cranberry dihydroflavonol-4-reductase (*DFR*) and expression in transgenic tobacco[J]. *Plant Science*, 2002, 163(2):33-43.

## Cloning and Bioinformatic Analysis of *DFR* Gene From *Iris lactea* Pall. var. *chinensis* (Fisch.) Koidz

HU Juan<sup>1,2</sup>, AN Shaoya<sup>1,2</sup>, LIN Zhe<sup>1,2</sup>, ZHANG Hong<sup>1,2</sup>, BAO Han<sup>1,2</sup>, CHEN Ren<sup>1,2</sup>

(1. Key Lab of Ministry of Education for Protection and Utilization of Special Biological Resources in Western China, Ningxia University, Yinchuan, Ningxia 750021; 2. School of Life Science, Ningxia University, Yinchuan, Ningxia 750021)

**Abstract:** *Iris lactea* Pall. var. *chinensis* (Fisch.) Koidz petals were used as material for transcriptome sequencing. According to the annotation data, a key gene encodes dihydroflavonol-4-reductase (*DFR*) in its anthocyanin biosynthesis was cloned by using PCR amplification and was characterized using bioinformatic analysis. The full-length of *DFR* gene of *I. lactea* var. *chinensis* consisted of 1 427 bp encoding a protein with 357 amino acid residues (designated as *IIDFR*, accession No. KY907171). Based on the comparison of the deduced amino acid sequence, *IIDFR* shared high identity with other plant *DFRs*, and had high homology with congeneric *Iris hollandica* *DFR*. *IIDFR* had a predicted molecular mass of 39.99 kDa and an isoelectric point of 5.89. It mainly composed of  $\alpha$ -helices and  $\beta$ -sheets, and was considered as an acidic, hydrophilic, no signal peptide and labile protein that located in the cytoplasm. *IIDFR* had a 'VTGASGYVGSWLVMKLLRDGY' amino acid sequence that was similar to the NADPH binding domain which was a typical conserved sequence in most of other plant *DFRs*, but differed in 4 amino acid residues.

**Keywords:** *Iris lactea* Pall. var. *chinensis*; transcriptome sequencing; anthocyanin biosynthesis; dihydroflavonol-4-reductase; bioinformatic analysis