

· 中药资源研究 ·

高良姜查尔酮合成酶基因的生物信息学和表达分析

黄琼林 (广东医科大学, 广东 湛江 524023)

摘要: **目的** 获取高良姜查尔酮合成酶(chalcone synthase, CHS)编码基因 cDNA 序列, 明确其序列和功能特征以及在高良姜中的表达模式。**方法** 在前期高良姜转录组测序数据中挖掘 CHS 的 cDNA 全长及其编码区, 利用生物信息学方法对该基因蛋白的特征和功能进行分析, 运用 ClustalX1.83 和 MEGA4.0 软件进行同源性比对和系统进化树构建, 采用实时荧光定量 PCR(qPCR)分析 CHS 在高良姜不同组织中的表达差异。**结果** 高良姜 CHS 编码区为 1 173 bp, 编码 391 个氨基酸。其编码蛋白是分子量为 94.5 kDa 的亲水性蛋白, 不含有信号肽、转运肽、靶肽和跨膜结构域, 具有 CHS 保守功能域并归属于 CHS 超级家族。CHS 在高良姜不同组织中的表达模式为: 根茎 > 茎 > 叶。**结论** 成功获得高良姜 CHS 序列, 并分析其特征和表达模式, 可为高良姜黄酮类有效成分的生物合成研究提供参考。

关键词: 高良姜; 查尔酮合成酶; 生物信息学; 表达模式

中图分类号: R282.1 **文献标志码:** A **文章编号:** 1003-9783(2021)11-1726-05

doi: 10.19378/j.issn.1003-9783.2021.11.019

Bioinformatics and Expression Analysis of Gene Encoding Chalcone Synthase from *Alpinia officinarum*

HUANG Qionglin (Guangdong Medical University, Zhanjiang 524023 Guangdong, China)

Abstract: Objective To obtain cDNA sequence of the gene encoding chalcone synthase (CHS), and clarify its sequence, function characteristics, and expression pattern in *Alpinia officinarum*. **Methods** The sequence of CHS gene was identified in the established transcriptome databases, and then the feature and function of coding protein was analyzed using bioinformatics approaches. The sequence alignment and phylogenetic tree construction were performed by ClustalX1.83 and MEGA4.0 software. The expression difference of CHS gene in various tissues of *A. officinarum* was accomplished using qPCR. **Results** The length of CHS gene was 1 173 bp, which encoded 391 amino acids. The CHS protein with a molecular weight of 94.5 kDa was hydrophilic, possessed none signal peptide, transit peptide, target peptide and transmembrane domain, and contained the functional domains defined as chalcone synthase, and was categorized into chalcone synthase superfamily. The CHS gene expressed the highest in rhizome, followed by stems, and the least in leaves of *A. officinarum*. **Conclusion** A CHS gene was successfully cloned from *A. officinarum*, and its sequence feature and expression pattern were also discussed, which provided a reference for the biosynthesis of the medicinal flavones in *A. officinarum*.

Keywords: *Alpinia officinarum*; chalcone synthase; bioinformatics; expression pattern

高良姜为传统道地南药和大宗药材, 以姜科植物高良姜 *Alpinia officinarum* Hance 的干燥块茎入药, 能散寒止痛、温中止呕, 临床上常用于治疗脘腹冷

痛, 胃寒呕吐, 噎气吞酸等疾病^[1]。资源调查^[2-3]发现, 高良姜野生资源已濒于灭绝, 人工栽培高良姜由于种植周期长、经济收益偏低、加工技术粗放等

收稿日期: 2021-05-06

作者简介: 黄琼林, 男, 博士, 副教授, 研究方向: 医学生物化学研究。Email: perfectql@163.com。

基金项目: 广东省中医药局中医药科研项目(20211224)。

原因,高良姜的栽培面积和产量逐年下降,部分地区甚至已出现供不应求的状况。高良姜含有高良姜素、高良姜素-3-甲醚等黄酮类活性物质,这些化合物是高良姜发挥抗炎镇痛、抗氧化、抗肿瘤等药理作用的物质基础^[4-5]。但是,目前高良姜类黄酮化合物的生物合成及其关键酶基因的研究还未见有相关报道。查尔酮合成酶(chalcone synthase, CHS)是植物黄酮类化合物生物合成途径的第一步关键限速酶,它催化该途径的起始底物——3 分子的丙二酰辅酶 A 和 1 分子的 4-香豆酰辅酶 A 结合形成查尔酮,为各类黄酮类化合物的生物合成提供了最基本的碳架结构^[6]。CHS 还对植物的育性、花色、调节生长素运输和抗逆等生理过程具有调控作用^[7]。目前,CHS 基因已从青天葵^[8]、人参^[9]和白芨^[10]等中药材中得到克隆和鉴定。本研究在高良姜转录组分析的基础上,挖掘和鉴定高良姜 CHS 基因,通过生物信息学分析 CHS 基因及其编码蛋白的序列特征和功能,并分析 CHS 基因在高良姜不同组织中的表达差异,以期为利用基于 CHS 基因的基因工程途径调控高良姜黄酮类活性成分的生物合成与含量积累研究奠定基础。

1 材料与方法

1.1 基因序列来源 在前期研究中,作者^[11]以道地产区高良姜叶片、茎和根茎为材料开展了转录组高通量测序,并通过序列组装获得 147 652 条非重复序列基因(unigenes)。本研究在此基础上,以“chalcone synthase”为主题词,错误值(E value)小于 10 为筛选条件,检索编码高良姜 CHS 的 unigene,并通过同源比对鉴定出含有高良姜 CHS 基因完整编码区的 unigene。

1.2 试剂与仪器 RNA prep Pure Plant Kit 试剂盒,天根生化科技(北京)有限公司;GoScript Reverse Transcription System 和 GoTaq qPCR Master Mix 试剂盒,美国普洛麦格生物技术有限公司;水为灭菌超纯水,其他试剂均为化学纯。Stepone Plus 荧光定量 PCR 仪(美国 Applied Biosystem 公司);TGL-16MS 型离心机(上海卢湘仪离心机仪器公司);JS-power 300 型核酸电泳仪、JS-8600B 型凝胶成像分析仪(上海培清公司)。

1.3 基因序列分析 采用 Omiga 2.0 软件对高良姜 CHS 基因序列进行编码区及开放阅读框(open reading frame, ORF)检索,并将其翻译成蛋白质序列。利用 ProtParam 和 ProtScale 软件分析基因序列的组成和亲/疏水性等理化特征。使用 Conserved

Domains 数据库分析编码蛋白质的保守功能域,利用 Target 1.1 和 TMHMM 软件完成编码蛋白质的转运肽、信号肽和跨膜结构分析。采用 SOPMA 软件分析编码蛋白质的二级结构。运用 ClustalX 和 MEGA 软件完成编码蛋白质的多重序列比对和系统发育分析。

1.4 基因表达分析 采用实时荧光定量 PCR(qPCR)检测 CHS 基因在高良姜不同组织样品中的表达差异。分别取约 100 mg 高良姜根茎、茎和叶片,使用 RNA prep Pure Plant Kit 进行总 RNA 提取,再用 GoScript Reverse Transcription System 反转录合成 cDNA。

根据检索所得的高良姜 CHS 基因序列,并以 β -tubulin 为内参基因,设计用于 qPCR 的特异性引物(见表 1)。参照 GoTaq qPCR Master Mix 试剂盒的说明书,qPCR 反应体系总体积为 20 μ L,其中 2 $\times$ qPCR Master Mix 10 μ L,10 μ mol $\cdot$ L⁻¹上下游引物各 0.3 μ L,cDNA 模板适量,并用灭菌蒸馏水补足体积。qPCR 反应条件为:95 $^{\circ}$ C 预变性 10 min;95 $^{\circ}$ C 变性 15 s,60 $^{\circ}$ C 退火 20 s,72 $^{\circ}$ C 延伸 20 s,40 个循环。采用 2^{- $\Delta\Delta$ Ct} 计算高良姜 CHS 基因的相对表达量。

表 1 高良姜 CHS 基因 qPCR 分析所用的引物

Table 1 Primers used for qPCR analysis of AoCHS

基因	上游引物序列(5'-3')	下游引物序列(5'-3')
CHS	GGAGGAGATGCTGAGGGAGA	CCATTCTTGATAGCCTTCG
β -tubulin	AATGTGCAGAACAAAGAACTC	GAACATCTCTTGGATGGA

2 结果与分析

2.1 基因检索 通过在高良姜转录组数据中检索注释,获得高良姜 CHS 基因的 cDNA 序列,并将其命名为 AoCHS。AoCHS 序列长度为 1 556 bp,其中含有 1 173 bp 编码区、57 bp 5'端非编码区和 323 bp 3'端非编码区。如图 1 所示,AoCHS 编码区包含 1 173 个碱基,GC 含量(即鸟嘌呤和胞嘧啶占全部碱基数量之比)为 63.1%,编码 391 个氨基酸。

2.2 编码蛋白同源比对和功能域检索 将 AoCHS 翻译所得的氨基酸序列提交至 Genbank 数据库进行同源比对,同源性最高的序列是来源于姜科凹唇姜 Boesenbergia rotunda 的 CHS 序列(AD085332),序列匹配率为 100%,相似度高达 90.78%。基于 10 种植物 CHS 蛋白构建的系统发育树见图 2。结果发现,AoCHS 与同为单子叶植物的凹唇姜、姜荷花和浦仔竹聚在一起,与其他双子叶植物来源的 CHS 明显分开,说明 AoCHS 与单子叶植物 CHS 有更高的同源性,与其来源植物的亲缘关系相符。

```

1      ATGGCCACAG TCCAGGAGAT CCGACGGGGT CAGCGGGCGG AGGGTACGGC GACAGTGTCTG
      M A T V Q E I R R G Q R A E G T A T V L
61     GCCATCGGCA CTGCCACGCC GGCCAACGGC GTCTACCAGG CTGATTACGC TGACTACTAC
      A I G T A T P A N G V Y Q A D Y A D Y Y
121    TTCCGCATGA CCAAGAGCGA GCACCTCACT GAGCTCAAGG AGAAGTTTAA AAGAATGTGC
      F R M T K S E H L T E L K E K F K R M C
181    GAAAAGTCGA TGATACGGAA GCGATACATG CACCTGACGG AGGAGATGCT GAGGGAGAAC
      E K S M I R K R Y M H L T E E M L R E N
241    CCGAATATTT GCGCGTACAT GGAGCCGTCG CTGGACGAGC GGCAGGATAT CGTGGTGGTG
      P N I C A Y M E P S L D E R Q D I V V V
301    GAGGTGCGGA AGCTGGGCAA GGAGGCGGCG GCGAAGGCTA TCAAGGAATG GGGGCAGCCC
      E V P K L G K E A A A K A I K E W G Q P
361    AAATCCAAGA TCACCCACCT CATCTTCTGC ACCACCTCTG GCGTAGACAT GCCGGGTGCC
      K S K I T H L I F C T T S G V D M P G A
421    GACTACCAGA TCACCAAGCT CCTCGGTCTC CGCCCTCGG TAAACCGCTT CATGATGTAC
      D Y Q I T K L L G L R P S V N R F M M Y
481    CAGCAGGGCT GCTTCGCCGG CGGCACCGTG CTCCGCATGG CCAAGGACCT GGCGGAGAAC
      Q Q G C F A G G T V L R M A K D L G E N
541    AACC GCGGCG CTGCGGTGCT GGTGGTGTGC TCCGAGATCA CGGCGGTGAC GTTCCGCGGG
      N R G A R V L V V C S E I T A V T F R G
601    CCGTCGGAGT CGCACCTGGA CAGCCTGGTA GGACAGGCCC TGTTCCGGGA CGGAGCAGGG
      P S E S H L D S L V G Q A L F G D G A G
661    GCGGTGATCG TGGGGGCGGA CCCGAACCCG GAAACAGAAC GACCTCTCTT CGAGCTGGTG
      A V I V G A D P N P E T E R P L F E L V
721    TCGGCGAGCC AGACGATCCT GCGGATTCG GAGGGGGCCA TCGACGGGCA CCTGCGGGAA
      S A S Q T I L P D S E G A I D G H L R E
781    GTGGGACTCA CCTTCCACCT GCTCAAGGAC GTGCCGCGCG TCATCTCCAA GAACATCGAA
      V G L T F H L L K D V P A L I S K N I E
841    AAGAGCTTGG TGGAGGCGTT CGAGCCGCTG GGGATCGACG ACTGGAATC CATCTTCTGG
      K S L V E A F E P L G I D D W N S I F W
901    ATTGCCACC CAGGAGGCCC CGCCATCTTG GACCAGGTGG AGGCCAAGCT GGCGCTGCAC
      I A H P G G P A L L D Q V E A K L L Q
961    AAGGAGAAGA TGGTGGCGAC GCGGCACGTG CTCAGCGAGT ACGGCAACAT GTCCAGCGCC
      K E K M V A T R H V L S E Y G N M S S A
1021   TCGGTTTCTGT TCATTCTGGA CGAGATGCGG CGCAGGTCGG CGGAGGAGGG GAAGGCCACC
      C V L F I L D E M R R R S A E E G K A T
1081   ACCGCGGACG GCTTCGAATG GGGCGTCTCT TTCGGCTTCG GCCCTGGACT CACCGTCGAG
      T G D G F E W G V L F G F G P G L T V E
1141   ACCGTGGTCT TGCATAGCAA GCCAATCAAG TACTAA
      T V V L H S K P I K Y
  
```

注：序列始端 ATG 和末端 TAA 分别为编码区启动密码子和终止密码子

图 1 *AoCHS* 编码区核苷酸序列和氨基酸序列

Figure 1 Nucleotide and encoded amino acid sequence of *AoCHS* coding region

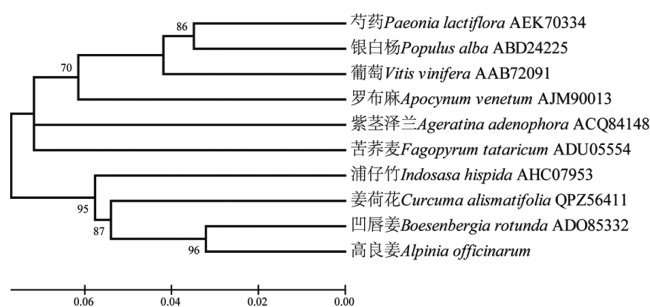


图 2 10 种植物查尔酮合成酶蛋白的系统进化树

Figure 2 Phylogenetic tree of CHS proteins from 10 plants

为了进一步鉴定 *AoCHS*，通过 CDD 在线工具分析 *AoCHS* 编码蛋白所含的保守功能域。如图 3 所示，*AoCHS* 蛋白包含“CHS-Like”和“Chalcone and stilbene synthase”两个特异性保守功能域以及“Chalcone synthase”非特异性保守功能域，归属为查尔酮合成酶超级家族。结果表明，*AoCHS* 蛋白为查尔酮合成酶家族成员，具有查尔酮合成酶的功能。

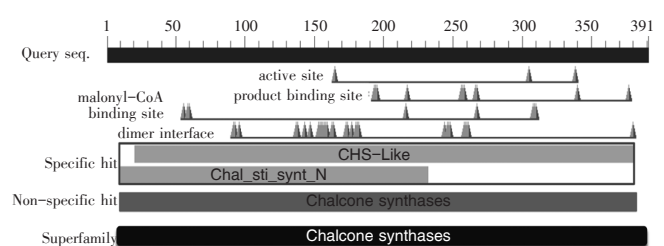


图 3 *AoCHS* 蛋白保守结构域预测

Figure 3 Prediction for conserved domain of *AoCHS* protein

因此，*AoCHS* 即为高良姜查尔酮合成酶基因。

2.3 编码蛋白理化特征分析 ProtParam 软件分析表明，*AoCHS* 编码蛋白的分子量为 94.5 kDa，等电点为 4.98，组成以甘氨酸、丙氨酸、半胱氨酸和苏氨酸为主。蛋白质不稳定性指数为 54.31，大于 40，属于不稳定蛋白质。ProtScale 软件分析发现，*AoCHS* 蛋白的多肽链整体表现为亲水性，并且没有明显的疏水区域，说明其为水溶性蛋白质。

Target1.1 软件分析结果表明, AoCHS 蛋白无信号肽、转运肽和靶肽等结构, 属于非分泌性蛋白质。TMHMM 软件分析发现, AoCHS 蛋白所有氨基酸位点位于细胞膜外的概率均高于其位于细胞膜内或跨膜的概率, 整条多肽链均处于细胞膜外, 因此

AoCHS 蛋白不含有跨膜结构域。

2.4 编码蛋白的高级结构分析 蛋白质二级结构分析如图 4 所示, AoCHS 蛋白二级结构含有 177 个 α -螺旋、61 个延伸链、26 个 β -转角和 127 个不规则盘绕, 说明其是多种元件共同组成的混合结构型蛋白质。

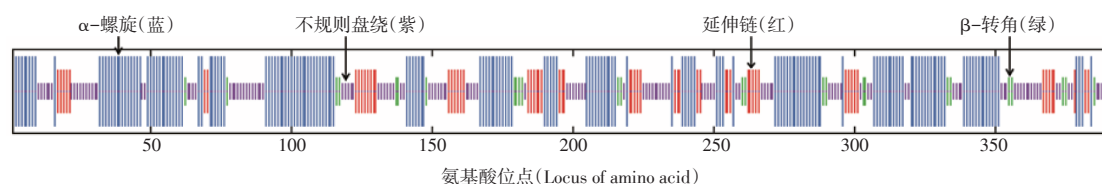


图 4 AoCHS 蛋白二级结构预测

Figure 4 Prediction for secondary structure of AoCHS protein

通过基于 Automatic model 模式为算法的 Swiss-model 同源建模预测 AoCHS 蛋白的三级结构, 以进一步了解其空间排列。结果如图 5 所示, AoCHS 蛋白获得以水稻 CHS1(4yjj.1.A)作为模板构建同源模型, 相似度为 84.75%。以全球模型质量估计(Global Model Quality Evaluation, GMQE)进行建模可信度评估, 该值范围为 0~1, 数值越大表明质量越好。本次建模的 GMQE 值为 0.94, 说明 AoCHS 蛋白三级结构的建模质量良好。AoCHS 蛋白三级结构以 α -螺旋和不规则盘绕居多, 与二级结构预测结果基本相符。

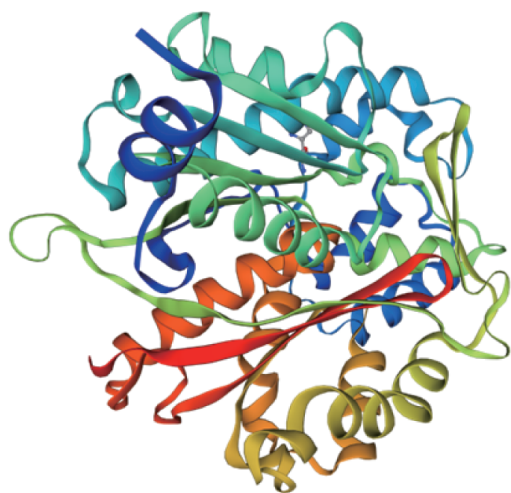


图 5 AoCHS 蛋白三级结构预测

Figure 5 Prediction for tertiary structure of AoCHS protein

2.5 基因表达分析 为了更好地了解 AoCHS 在高良姜中的表达模式, 选用生长旺盛的高良姜根茎、茎和叶片为材料进行 qPCR 分析。结果如图 6 所示, AoCHS 在根茎中的表达量最高, 其次为茎, 最低的是叶片, 提示根茎中黄酮类化合物的生物合成可能比其他两个组织更为活跃。

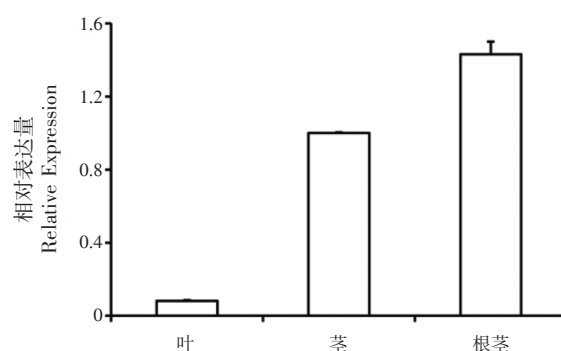


图 6 AoCHS 在高良姜不同组织中表达分析(n=3)

Figure 6 Expression analysis of AoCHS in different tissues of *Alpinia officinarum*(n=3)

3 讨论

查尔酮合成酶(CHS)是植物黄酮类生物合成的关键靶点, 已有大量研究表明其编码基因的表达与黄酮类化合物的积累密切相关。人参发根、根茎、茎和叶片中 CHS 基因表达水平与黄酮含量呈高度一致的正相关性^[9]。刘秀明等^[12]将红花 CHS 基因转入拟南芥中表达, 结果发现, 红花 CHS 基因过表达的 T2 代拟南芥的总黄酮含量比野生植株显著升高, 最高可达后者的 3 倍。Dare 等^[13]发现, CHS 基因被敲除后, 苹果中的黄酮含量明显下降, 甚至没有检测到花青素的存在, 而且出现节间变短、叶片缩小和生长速度减慢等发育异常。因此, 开展 CHS 基因功能和调控研究对促进高良姜黄酮类活性成分合成和品种选育具有重要的科学意义。

本研究基于转录组分析鉴定出 AoCHS 基因, 获得了其 1 173 bp 的编码区序列及编码 391 个氨基酸的蛋白质序列。AoCHS 蛋白与凹唇姜等单子叶物种来源的 CHS 相似度可达 90%以上, 并且含有查尔酮合成酶的典型功能域和活性位点, 推测其具有催化

植物黄酮生物合成途径第一步反应——查尔酮生成的功能。此外,本研究还发现 AoCHS 蛋白为亲水性蛋白质,不含信号肽、转运肽和跨膜结构域,并研究了其高级结构的组成和空间排列。这些结论可为后续高良姜 CHS 蛋白的功能鉴定和表达调控研究提供理论指导。

qPCR 检测发现, AoCHS 在高良姜的不同组织中均有表达,而且在根茎中表达最高,推测黄酮类化合物在高良姜的根茎中有更多的积累,这也与高良姜以根茎入药的事实相符。本研究从基因水平为高良姜根茎入药提供了一定的科学依据。

本研究从转录组测序数据中挖掘和鉴定出高良姜 CHS 基因,分析了其编码蛋白的序列特征和功能,并明确了该基因在高良姜中的表达模式,可为后续利用 CHS 基因工程提高高良姜黄酮类有效成分的积累以及促进高良姜资源可持续利用提供科学基础。

参考文献:

- [1] 国家药典委员会. 中华人民共和国药典: 一部[S]. 北京: 中国医药科技出版社, 2020: 300.
- [2] 谭业华, 陈珍. 南药高良姜生产调查、问题分析及发展对策[J]. 海南师范大学学报(自然科学版), 2014, 27(3): 293-296.
- [3] 杨全, 严寒静, 庞玉新, 等. 南药高良姜药用植物资源调查研究[J]. 广东药学院学报, 2012, 28(4): 382-386.
- [4] 高振虎, 陈艳芬, 杨全, 等. 南药高良姜的研究进展[J]. 广东药学院学报, 2016, 35(6): 817-821.
- [5] JIAO H, XU S, FAN C, et al. Chromatographic fingerprint analysis and quantitative evaluate the rhizomes of *Alpinia officinarum* Hance (lesser galangal) [J]. Journal of Chinese Pharmaceutical Sciences, 2019, 28(10): 728-738.
- [6] PETRUSSA E, BRAIDOT E, ZANCANI M, et al. Plant flavonoids, biosynthesis, transport and involvement in stress responses[J]. International Journal of Molecular Science, 2013, 14(7): 14950-14973.
- [7] FALCONE FERREYRA M L, RIUSS P, CASATI P. Flavonoids: biosynthesis, biological functions and biotechnological application[J]. Frontier Plant Science, 2012, 3: 222.
- [8] 黄琼林, 曾湘达, 卓一南, 等. 青天葵查尔酮合成酶基因的克隆及序列分析[J]. 中药新药与临床药理, 2018, 29(2): 205-210.
- [9] 张变玲, 黄雪梅, 刘心怡, 等. 人参查尔酮合成酶基因PgCHS1的克隆与表达分析[J]. 中国细胞生物学学报, 2018, 40(12): 2010-2017.
- [10] 田艾梅, 许忠民, 张恩慧. 白芨查尔酮合成酶基因的克隆与分析[J]. 西北农业学报, 2014, 23(7): 91-95.
- [11] 黄琼林. 高良姜转录组的特征分析[J]. 中药新药与临床药理, 2020, 31(6): 729-734.
- [12] 刘秀明, 陆皖行, 李佳, 等. 红花CHS基因的过表达提高了拟南芥黄酮含量[J]. 中国细胞生物学学报, 2017, 39(2): 182-190.
- [13] DARE A P, TOMES S, JONES M, et al. Phenotypic changes associated with RNA interference silencing of chalcone synthase in apple (*Malus x domestica*) [J]. Plant Journal, 2013, 74(3): 398-410.

(编辑: 梁进权)