

郝爱平. 高粱查尔酮合成酶的生物信息学分析[J]. 江苏农业学报, 2016, 32(6): 1232-1236.  
doi:10.3969/j.issn.1000-4440.2016.06.006

## 高粱查尔酮合成酶的生物信息学分析

郝爱平

(牡丹江师范学院生命科学与技术学院, 黑龙江 牡丹江 157012)

**摘要:** 为研究高粱查尔酮合成酶(*CHS*)基因的功能,利用生物信息学工具和方法,对 *CHS* 蛋白质理化性质、疏水性/亲水性、跨膜区域、结构功能域、二级结构、三级结构和同源性进行了预测与分析。结果显示,高粱的 *CHS* 蛋白质共有 8 种(*CHS1* ~ *CHS8*),其氨基酸数量都在 400 左右,分子量差别不大;其中 *CHS1* ~ *CHS7* 为较稳定的亲水性蛋白质,*CHS8* 为不稳定的疏水性蛋白质;8 种蛋白质都为非跨膜蛋白。*CHS* 蛋白有 *Chal\_sti\_synt\_N* 和 *ACP\_syn\_III\_C* 2 个结构域,主要构成为  $\alpha$ -螺旋和不规则卷曲。同源性分析结果表明,*CHS1* ~ *CHS7* 蛋白质氨基酸序列间的同源性较高。

**关键词:** 高粱; 查尔酮合成酶; 生物信息学

中图分类号: S514.01

文献标识码: A

文章编号: 1000-4440(2016)06-1232-05

## Bioinformatics analysis of chalcone synthase in *Sorghum bicolor*

HAO Ai-ping

(College of Life Sciences and Technology, Mudanjiang Normal College, Mudanjiang 157012, China)

**Abstract:** To obtain more information about chalcone synthase (*CHS*) gene from *Sorghum bicolor*, the bioinformatics methods and tools were used to analysis the physical and chemical properties, hydrophobic/hydrophilic, transmembrane region, structure and function domain, secondary structure, tertiary structure and homology of *CHS* protein in *Sorghum bicolor*. The results showed that there were eight *CHS* proteins in *Sorghum bicolor* (*CHS1*-*CHS8*). The number of amino acids was about 400 in *CHS* proteins. All *CHS* proteins were stable and hydrophilic protein except for *CHS8*. The analysis also demonstrated that the eight *CHS* proteins were non-transmembrane protein. Its main component was  $\alpha$ -helix and random coil. The structure domains of *CHS* proteins consisted of *Chal\_sti\_synt\_N* and *ACP\_syn\_III\_C*. The sequence homology was higher in *CHS1*-*CHS7* proteins.

**Key words:** *Sorghum bicolor*; chalcone synthase; bioinformatics

高粱(*Sorghum bicolor*)别称蜀黍、木稷、芦粟、荻粱、乌禾等,系禾本科高粱属一年生草本植物<sup>[1]</sup>。中国是世界上高粱的主产国之一,资源十分丰富,广泛种植于黑龙江绥化、吉林松原、辽宁鞍山、内蒙古赤峰、山西太原等地区,仅杂交高粱种植面积就约为  $1.333 \times 10^6 \text{ hm}^2$ ,年产高粱壳  $2 \times 10^6 \text{ t}$ <sup>[2]</sup>。高粱的抗

逆性、适应性强,用途广泛,所产生的经济效益较高,可进行食用、饲用和进一步深加工等应用。高粱壳是高粱的主要副产物,可作为植物色素提取的原材料,既经济又环保<sup>[3]</sup>。

查尔酮合成酶(*CHS*)是一种植物中类黄酮物质生物合成过程中的关键酶,也是植物次生代谢途径中的关键酶,它催化羟基苯丙烯酸辅酶 A 和丙二酰辅酶 A 生成黄酮类物质四羟基查尔酮。查尔酮在花色苷、黄酮、异黄酮等物质生物合成过程中起主要作用,而通过黄酮类物质合成途径合成的色素、花青素、鞣红则是植物色素沉积中的主要化合物。因此查尔酮合成酶的表达对植物色素合成有较为重要

收稿日期:2016-03-16

基金项目:牡丹江师范学院教学名师培养资助项目  
(2014QNGG1805);牡丹江市科学技术计划项目  
(G2015d1974)

作者简介:郝爱平(1979-),女,山东莘县人,硕士,副教授,研究方向为分子生物学。(E-mail)swxhap@126.com

的影响,在类黄酮物质的合成中也起到十分关键的作用<sup>[4-5]</sup>。在植物中查尔酮合成酶基因是一个比较大的多基因家族,其编码序列相对保守,约 1 200 bp,编码 400 个左右的氨基酸,种属间氨基酸序列的同源性为 74% ~ 98%,核苷酸的保守性在 61% 以上<sup>[6]</sup>。对查尔酮合成酶各基因成员关系的研究,有助于进一步了解该酶的功能。查尔酮合成酶的功能多种多样,涉及植物生长发育的诸多过程,例如植物的育性、植物的花色、植物在逆境中的一些反应等等<sup>[7]</sup>。但是查尔酮合成酶有时会受到不同的发育调控和组织特异性调控,所以对外界刺激的敏感程度会有所不同,所表达出来的现象也会有所不同<sup>[8]</sup>。由于查尔酮合成酶广泛存在于植物中,并且在很多生理生化的活动中起十分关键的作用,所以查尔酮合成酶是最近几年植物生理生化和分子生物学研究的主要热点。目前已报道的含有查尔酮合成酶基因的植物大约有 4 000 多种,包括玉米、兰花、拟南芥、水稻、高粱等<sup>[9]</sup>。

本研究利用生物信息学方法对高粱中查尔酮合成酶(*CHS*)的理化性质、跨膜结构、疏水性/亲水性、蛋白结构功能域、二级结构、三级结构以及同源性等方面进行预测和分析,为今后进一步深入开展该酶的功能研究提供理论依据。

## 1 材料与方法

### 1.1 材料

高粱查尔酮合成酶的氨基酸序列 CHS1 (AAD41873.1)、CHS2 (AAD41874.1)、CHS3 (AAD41875.1)、CHS4 (AAD41876.1)、CHS5 (AAD41877.1)、CHS6 (AAD41878.1)、CHS7 (AAD41879.1)和 CHS 8 (AAL49965.1)通过在美国国立生物技术信息中心(National Center for Biotechnology Information, 简称 NCBI)中检索得到。其中 CHS1 氨基酸序列如下:

MAGATVTVEEVRKAQRATGPATVLAIGTATPAN  
CVHQADYDPDYFRITKSEHMTLKEKFKRMCDKSQI  
RKRYMHLTEEYLAENPNMCAYMAPSLDARQDIVVV  
EVPKLGKAAAQKAIKEWGQPKSKITHLVFCTTSGVD  
MPGADYQLTKMLGLRPSVNRLLMMYQQGCFAGGTVL  
RVAKDLAENNRGARVLVVCSEITAVTFRGPSESHLD  
SMVGQALFGDGAAAVIVGADPDERVERPLFQLVSAS  
QRILPDSEGAIDGHLREVGLTFHLLKDVPGLISKNIER

ALEEAFKPLGITDYNISIFWVAHPGGPAILDQVEAKVG  
LEKERMTRATHVLSEYGNMSSACVLFILDEMRKRSA  
EDGQTTTGEFGFDWGVLFQFGPGLTVETVVLHSPITT  
GAAITA

### 1.2 方法

CHS 蛋白质的理化性质用 Protparam 软件(<http://web.expasy.org/protparam>)分析,亲水性/疏水性用 ProtScale 软件(<http://web.expasy.org/protscale>)分析,跨膜结构用 TMHMM 软件(<http://www.cbs.dtu.dk/services/TMHMM-2.0>)分析,蛋白质结构功能域预测用 SMART 软件(<http://smart.embl-heidelberg.de>)分析,二级结构用 SPOMA 软件(<http://npsa-pbil.ibcp.fr/cgi-bin/npsa>)分析,三级结构用 SWISS-model 软件(<http://www.expasy.ch/swissmod/SWISS-MODEL.html>)分析,同源性用 MEGA5.1 分析。

## 2 结果与分析

### 2.1 高粱 CHS 蛋白质的理化性质

用 Protparam 软件分析 CHS 蛋白质的氨基酸序列。结果(表 1)显示,CHS1 ~ CHS7 氨基酸数量均为 400 左右,分子量在 43 500.0 至 47 500.0 之间,等电点在 6.10 至 6.60 之间,不稳定指数都小于 40,属于稳定的蛋白质。在氨基酸组成中丙氨酸(Ala)、缬氨酸(Val)、亮氨酸(Leu)这 3 种氨基酸含量比较丰富。但 CHS8 比较特殊,它含有 398 个氨基酸,分子量为 43 573.1,等电点为 5.92,不稳定指数为 41.09,属于不稳定蛋白质。在氨基酸组成中亮氨酸(Leu)、丙氨酸(Ala)、甘氨酸(Gly)这 3 种氨基酸含量比较丰富。

### 2.2 高粱 CHS 蛋白质的疏水性/亲水性预测

在自然界中常见的 20 种氨基酸都具有疏水性或亲水性,它们的排列顺序和侧链基团之间的相互作用决定了蛋白质的三维空间结构,通过对蛋白质疏水性/亲水性进行预测和分析可以为蛋白质的高级结构预测提供一些参考。用 ProtScale 工具对高粱的 CHS 蛋白质进行疏水性/亲水性预测。预测结果[图 1(CHS2 ~ CHS7 参照图 1)、图 2]表明:CHS1 ~ CHS7 氨基酸的最大分值都为 2.422,最小分值在 -2.533 到 -2.644 之间;CHS8 氨基酸的最大分值为 2.689,最小分值为 -2.622。根据氨基酸分值越低亲水性越强的规律,推测 CHS1 ~ CHS7 都为亲水性蛋白质,CHS8 为疏水性蛋白质。

表 1 高粱 CHS 蛋白质的理化性质

Table 1 The physical and chemical properties of CHS protein in *Sorghum bicolor*

蛋白质	氨基酸数量	分子式	分子量	等电点	不稳定指数	含量丰富的氨基酸
CHS1	401	C <sub>1 931</sub> H <sub>3 085</sub> N <sub>537</sub> O <sub>578</sub> S <sub>21</sub>	43 745.2	6.23	36.57	Ala、Val、Leu
CHS2	401	C <sub>1 930</sub> H <sub>3 084</sub> N <sub>536</sub> O <sub>578</sub> S <sub>21</sub>	43 718.2	6.24	36.98	Ala、Val、Leu
CHS3	401	C <sub>1 929</sub> H <sub>3 080</sub> N <sub>534</sub> O <sub>579</sub> S <sub>20</sub>	43 658.1	6.10	38.06	Ala、Val、Leu
CHS4	401	C <sub>1 929</sub> H <sub>3 093</sub> N <sub>535</sub> O <sub>576</sub> S <sub>21</sub>	43 669.2	6.57	38.10	Ala、Val、Leu
CHS5	401	C <sub>1 930</sub> H <sub>3 079</sub> N <sub>535</sub> O <sub>578</sub> S <sub>21</sub>	43 699.1	6.15	38.25	Ala、Val、Leu
CHS6	401	C <sub>1 930</sub> H <sub>3 085</sub> N <sub>535</sub> O <sub>578</sub> S <sub>21</sub>	43 705.2	6.39	37.45	Ala、Val、Leu
CHS7	400	C <sub>1 927</sub> H <sub>3 082</sub> N <sub>532</sub> O <sub>573</sub> S <sub>22</sub>	43 576.2	6.57	39.50	Ala、Val、Leu
CHS8	398	C <sub>1 932</sub> H <sub>3 069</sub> N <sub>533</sub> O <sub>569</sub> S <sub>22</sub>	43 573.1	5.92	41.09	Leu、Ala、Gly

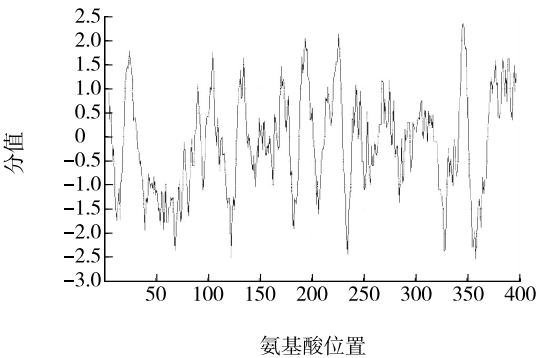


图 1 高粱 CHS1 蛋白质的疏水性/亲水性预测  
Fig. 1 Hydrophobic/hydrophilic prediction of CHS1 protein in *Sorghum bicolor*

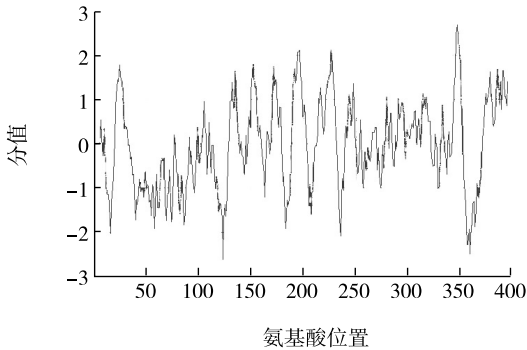


图 2 高粱 CHS8 蛋白质的疏水性/亲水性预测  
Fig. 2 Hydrophobic/hydrophilic prediction of CHS8 protein in *Sorghum bicolor*

2.3 高粱 CHS 蛋白质的跨膜结构预测

使用在线软件 TMHMM 预测 CHS 蛋白质的跨膜结构。以 CHS1 为例,预测结果显示该蛋白质的 401 个氨基酸全部都在膜外,因此该蛋白质为非跨膜蛋白质。CHS2 ~ CHS8 也都为非跨膜蛋白质,但

膜外氨基酸数量有所不同。

2.4 高粱 CHS 蛋白质结构功能域预测

结构功能域是蛋白质亚基结构,是介于二级与三级结构之间的一种有独立结构的功能单位。常见结构域的氨基酸残基数大约在 100 至 400 之间,结构域之间区域结构松散,但是结构域自身是相对独立的单元,而且是具有一定功能的紧密结构单元<sup>[10]</sup>。使用在线软件 SMART 预测 CHS 蛋白质的结构功能域。预测结果显示:该蛋白质有 2 个结构域,Chal\_sti\_synt\_N 在 6 ~ 232 区域,ACP\_syn\_III\_C 在 297 ~ 390 区域。其中 Chal\_sti\_synt\_N 结构域在植物生长发育过程中起关键作用,主要调控色素的合成,而 ACP\_syn\_III\_C 结构域的功能是催化生成抗逆性较高的次生代谢产物。

2.5 高粱 CHS 蛋白质二级结构的预测

用在线软件 SPOMA 预测高粱 CHS 蛋白质的二级结构。结果(表 2)表明:高粱 CHS 蛋白质的二级结构由  $\alpha$ -螺旋、延伸链、 $\beta$ -折叠、不规则卷曲构成,其中  $\alpha$ -螺旋和不规则卷曲为主要构件。

表 2 高粱 CHS 蛋白质的二级结构

Table 2 Secondary structure of CHS protein in *Sorghum bicolor*

蛋白质	序列长度	结构成分百分率(%)			
		$\alpha$ -螺旋	延伸链	$\beta$ -折叠	不规则卷曲
CHS1	401	41.15	18.45	10.47	29.93
CHS2	401	41.15	19.20	10.72	28.93
CHS3	401	41.65	18.95	10.97	28.43
CHS4	401	41.65	18.70	11.22	28.43
CHS5	401	42.14	18.95	10.47	28.43
CHS6	401	40.40	18.95	11.47	29.18
CHS7	400	39.00	19.75	10.50	30.75
CHS8	398	41.21	20.10	11.56	27.14

2.6 高粱 CHS 蛋白质三级结构预测

CHS 蛋白质的三级结构预测结果(图 3)显示, CHS 蛋白质都以 4wum. 1. A 为模板, 其中 CHS1 ~ CHS7 与模板序列的相似性较高, 大约在 83.00% 以

上;CHS8 与模板序列的相似性为 75.65%。通过分析这 8 个模型可知:CHS 蛋白质的三级结构都是由  $\alpha$ -螺旋和不规则卷曲组成, 由于与模板相似性较高, 因此结构较为可靠。

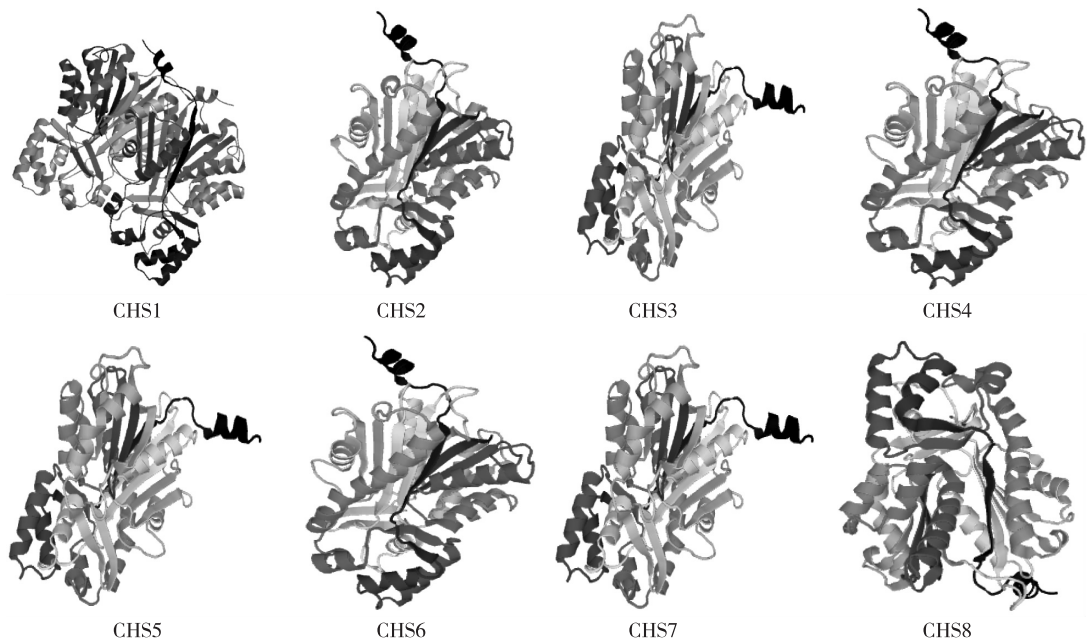


图 3 高粱 CHS1 ~ CHS8 蛋白质的三级结构  
Fig.3 Tertiary structure of CHS1-CHS8 protein in *Sorghum bicolor*

2.7 高粱 CHS 蛋白质氨基酸序列同源性分析

利用 MEGA5.1 软件对高粱 CHS 氨基酸序列进行同源比对, 构建进化树(图 4)。从图 4 中可知 CHS1 ~ CHS7 蛋白质氨基酸序列间的同源性较高, 但它们与 CHS8 蛋白质氨基酸序列间的同源性较低;在 CHS1 ~ CHS7 中 CHS6 蛋白质氨基酸序列与 CHS7 蛋白质氨基酸序列同源性较高, CHS3 蛋白质氨基酸序列与 CHS5 蛋白质氨基酸序列同源性较高。因为 CHS8 的氨基酸数量与其他蛋白质不同, 从而使它的理化性质、亲/疏水性、蛋白质的二级结构和三级结构等都与其他蛋白质有差异。

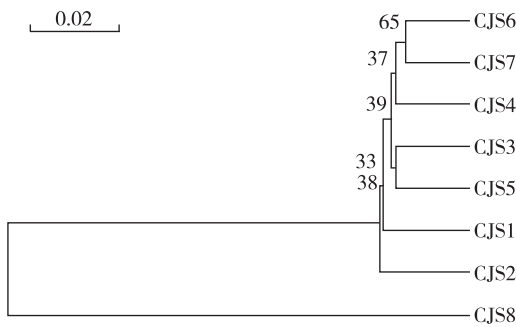


图 4 高粱 CHS 蛋白质同源性分析  
Fig.4 Analysis of homology of CHS protein in *Sorghum bicolor*

3 讨论

查尔酮合成酶在植物苯丙氨酸代谢途径中的作用至关重要, 直接或间接影响着植物代谢产物合成、抗性调节、花色形成等生理生化过程<sup>[11]</sup>。迄今为止, 科学家已经从多种植物中克隆了查尔酮合成酶基因, 包括百合、北美鹅掌楸、大白菜、观赏桃、大豆、

山葡萄等<sup>[12-17]</sup>。随着植物基因组计划的进一步实施, 很多重要基因功能的预测都离不开生物信息技术, 生物信息学是研究基因组信息、基因结构、基因产物功能必不可少的技术手段<sup>[18]</sup>。阮美颖等<sup>[19]</sup>运用生物信息学方法对番茄查尔酮合成酶基因内含子和外显子的结构特征、系统发育关系、序列结构的



保守性进行了分析和预测,结果表明番茄查尔酮合成酶基因(*SlCHS*)是含有 8 个成员的多家族基因,这些基因均含有较少的内含子,这些基因具有较高的保守性。徐婉莉等采用生物信息学方法对辣椒 *CHS* 基因家族进行鉴定及分析,结果表明该基因家族含有 7 个成员,编码的氨基酸序列长度介于 213 aa 至 402 aa 之间,基因间序列相似性变化幅度较大,具有高度遗传多样性<sup>[20]</sup>。

本研究利用生物信息学工具和方法分析高粱查尔酮合成酶(*CHS*)。结果表明:高粱的 *CHS* 蛋白质有 8 种成员(*CHS1* ~ *CHS8*),其氨基酸数量都在 400 左右,分子量差别不大;其中 *CHS1* ~ *CHS7* 为较稳定的亲水性蛋白质,*CHS8* 为不稳定的疏水性蛋白质,都为非跨膜蛋白质。*CHS1* ~ *CHS7* 蛋白质氨基酸序列同源性较高,而 *CHS8* 与它们的同源性较低。说明高粱 *CHS* 基因间相似性变化幅度较大,具有一定的遗传多样性,或者高粱 *CHS* 的调控基因有所差异,或者其他类黄酮合成途径中涉及的基因有所差异,这与其他植物 *CHS* 基因的研究结果类似<sup>[21]</sup>。以上结果为进一步研究高粱查尔酮合成酶(*CHS*)基因功能提供了一定的理论依据。

#### 参考文献:

- [1] 王金亭.天然高粱红色素研究与应用进展[J].粮食与油脂,2012(11):7-11.
- [2] 张兆俊,肖丽娟.天然色素高粱红开发应用[J].粮食与油脂,2015(5):40-41.
- [3] 方芳芳,候秀良,代雅轩,等.高粱壳色素上染毛织物的动力学和热力学[J].纺织学报,2015,36(3):70-76.
- [4] 王 坤,潘怡辰,孙慧君,等.3 种小麦作物查尔酮合成酶(*CHS*)及其基因的生物信息学分析[J].中国农学通报,2013,29(12):39-43.
- [5] MARY L D, BONNIE M C, MICHAEL T C. Molecular evolution of the chalcone synthase multigene family in the morning glory genome[J]. Plant Molecular Biology,2000,42(1):79-92.
- [6] 陈雪菲.三种蕨类植物查尔酮合成酶(*CHS*)基因的克隆与分析[D].上海:上海师范大学,2013.
- [7] 蒋 明,曹家树.查尔酮合成酶基因[J].细胞生物学杂志,2007(29):525-529.
- [8] HARSHAVARDHAN K P K, GRAEME S G, ELIZABETH I B. Genome-wide analysis of the chalcone synthase superfamily genes of *Physcomitrella patens* [J]. Plant Molecular Biology, 2010, 72(3):247-263.
- [9] 张 毅.不同植物查尔酮合成酶 *CHS* 基因的生物信息学分析[J].江西农业学报,2012,24(6):5-8.
- [10] 薛永常,聂会忠,刘长斌,等.木质素合成酶 *C3H* 基因的生物信息学分析[J].生物信息学,2009,7(1):13-18.
- [11] 李 苗,李国旗.查尔酮合成酶基因及其分子进化研究进展[J].中国农学通报,2015,31(18):116-120.
- [12] 陈 洁,安利清,王 涛,等.百合查尔酮合成酶基因克隆及其转化烟草的花色表达分析[J].西北植物学报,2012,32(8):1511-1517.
- [13] 罗群凤,胥 猛,冯源恒,等.北美鹅掌楸 *LuCHS* 基因的克隆及生物信息学与组织表达特征分析[J].林业科学,2015,51(5):37-45.
- [14] 王小霞,刘志勇,李承彧.大白菜查尔酮合成酶基因 *BrCHS1* 的克隆与特征分析[J].沈阳农业大学学报,2012,43(2):137-142.
- [15] 林 玲,汤浩茹,陈 清,等.观赏桃查尔酮合成酶基因的克隆及其序列分析[J].园艺学报,2012,39(3):581-587.
- [16] 单丽伟,汪 勇,王美玲,等.大豆类黄酮生物合成关键酶 *CHS* 基因的克隆及表达分析[J].西北植物学报,2012,32(11):2164-2168.
- [17] 李 娟,曹芳芳,刘海峰.山葡萄查尔酮合成酶基因的克隆与分析[J].贵州农业科学,2015,43(9):1-6.
- [18] 谷瑞升,杜生明.生物信息学研究现状及发展趋势[J].医学信息学杂志,2012,33(5):1-6.
- [19] 阮美颖,万红建,叶青静,等.番茄查尔酮合成酶基因的鉴定及生物信息学分析辣椒查尔酮合成酶基因家族全基因组鉴定及表达特征分析[J].分子植物育种,2013,11(3):379-384.
- [20] 徐婉莉,裴徐梨,荆赞革,等.辣椒查尔酮合成酶基因家族全基因组鉴定及表达特征分析[J].基因组学与应用生物学,2015,34(8):1747-1752.
- [21] 王 坤,潘怡辰,孙慧君.3 种小麦作物查尔酮合成酶(*CHS*)及其基因的生物信息学分析[J].中国农学通报,2013,29(12):39-43.

(责任编辑:张震林)