

甘草 *CHS* 基因多态性与甘草苷含量的相关性分析

张晓冬¹, 周 珊¹, 尹彦超¹, 胡 婷¹, 高智强¹, 韩沂霖¹, 李文东^{2*}, 刘 颖^{1*}

(1. 北京中医药大学生命科学院, 北京 100029;

2. 北京市药品检验所中药成分分析与生物评价北京市重点实验室, 北京 102206)

摘要: 2015 版《中华人民共和国药典》明确规定, 甘草药材中甘草苷含量不得低于 0.5%, 然而目前市售栽培甘草大量存在甘草苷含量不达标现象。由于功能基因多态性对于次生代谢产物的积累至关重要, 因此本文试图对甘草苷生物合成途径中的关键酶——查尔酮合酶 (chalcone synthase, *CHS*) 进行基因多态性分析, 筛选出甘草苷高/低含量样品对应的特异基因型, 为进一步解析甘草苷生物合成的分子机制提供参考。本文利用 HPLC 法测得 60 株甘草样品中 4 种主要黄酮类化合物 (甘草苷、异甘草苷、甘草素、异甘草素) 的含量, 应用 Spearman 相关性分析及 χ^2 检验进行统计学分析, 发现 4 种黄酮类化合物含量间具有相关关系且不同基原间含量差异显著, 乌拉尔甘草 > 光果甘草 > 胀果甘草。因此筛选出含量最高的 5 株乌拉尔甘草及含量最低的 5 株胀果甘草分为黄酮高、低含量组, 对其 *CHS* 基因进行克隆及序列分析。10 株样品共获得了 336 条长度为 1 175 bp 的 *CHS* cDNA 序列, 存在 249 个变异位点, 其中错义突变位点 141 个, 可分为 137 种单倍型; 编码 389 个氨基酸残基, 存在 130 个变异位点, 可分为 102 种氨基酸序列类型; 主流 *CHS* 氨基酸序列类型为 AA-3, 黄酮高含量组甘草样品特有的主流氨基酸序列类型为 AA-35, 黄酮低含量组甘草样品特有主流氨基酸序列类型为 AA-36, AA-35 与 AA-36 在 193 位与 229 位处存在 I/V、V/T 突变。采用 Discovery Studio 2.5 对蛋白三维结构进行分析, 结果表明 AA-35 的 229 位的缬氨酸为底物丙二酰辅酶 A 的结合位点。MEGA 5.0 同源性分析表明甘草 *CHS* 亲缘关系良好, 与其他物种区分度良好。本文筛选获得了黄酮高/低含量组甘草样品的特异 *CHS* 基因型, 对于指导优质甘草的分子育种具有重要意义。

关键词: 甘草; 甘草苷; 查尔酮合酶; 基因多态性

中图分类号: R931

文献标识码: A

文章编号: 0513-4870 (2018) 04-0646-13

Analysis on the correlation between chalcone synthase gene polymorphism and content of liquiritin in licorice

ZHANG Xiao-dong¹, ZHOU Shan¹, YIN Yan-chao¹, HU Ting¹, GAO Zhi-qiang¹,
HAN Yi-lin¹, LI Wen-dong^{2*}, LIU Ying^{1*}

(1. School of Life Sciences, Beijing University of Chinese Medicine, Beijing 100029, China; 2. Beijing Institute for Drug Control, Beijing Key Laboratory of Analysis and Evaluation on Chinese Medicine, Beijing 102206, China)

Abstract: Chinese pharmacopoeia stipulates that the content of liquiritin in licorice slices should be no less than 0.5%. However, there are lots of unqualified licorice slices in the herbal medicine markets. Due to the important role of functional gene polymorphism in secondary metabolism, this study attempts to analyze the influence of chalcone synthase (*CHS*) gene polymorphism on liquiritin biosynthesis and find out the unique

收稿日期: 2017-12-14; 修回日期: 2018-01-12.

基金项目: 北京中医药大学杰出青年人才项目 (2016JYBXJQ002).

*通讯作者 Tel: 86-10-84738646, E-mail: liuyliwd@sina.com;

Tel: 86-10-52779627, E-mail: swlwd76@163.com

DOI: 10.16438/j.0513-4870.2017-1245

haplotypes in licorice samples with high or low content of liquiritin, and to provide a basis for further analysis of molecular mechanism in flavonoid biosynthetic pathway. The contents of the 4 main flavonoids (liquiritin, isoliquiritin, liquiritigenin, isoliquiritigenin) in 60 licorice samples were assayed by HPLC and the results were analyzed by Spearman and χ^2 tests. The contents of the 4 main flavonoids were related to each other and obviously different in different original plants. They were highest in *Glycyrrhiza uralensis* samples and lowest in *Glycyrrhiza inflata* samples. Five *G. uralensis* samples with the highest liquiritin contents and five *G. inflata* samples with the lowest liquiritin contents were selected to clone the *CHS* cDNA sequences. 336 *CHS* cDNA sequences with a full length of 1 175 bp were obtained, 249 variable sites (141 missense mutation sites) were found, and 137 haplotypes were determined. 130 variable sites were found in the 336 *CHS* amino acid sequences and 102 types were determined. AA-3 is the major type of *CHS* in licorice, AA-35 is the special major type of *CHS* in the group with high flavonoids contents and AA-36 is the special major type of *CHS* in the group with low flavonoids contents. The mutation sites between AA-35 and AA-36 are I/V at 193 and V/T at 229. Discovery Studio 2.5 analysis of the three-dimensional structure of the *CHS* protein shows that the valine at site 229 of AA-35 is combined with malonyl-CoA. Homology analysis indicates that the homology of *CHS* among different species is low. This study is significant for identification of the unique haplotypes in licorices with high or low content of liquiritin and guiding the further molecular breeding of high-quantity licorice.

Key words: licorice; liquiritin; chalcone synthase; gene polymorphism

甘草是我国最常用的大宗药材之一, 始载于《神农本草经》, 被列为上品^[1], 享有“十方九草”之美誉, 具有补脾益气、清热解毒、祛痰止咳、缓急止痛、调和诸药的功效。大量现代药理研究表明, 甘草及其提取物具有抗炎^[2]、抗癌^[3]、抗菌^[4]、抗抑郁^[5]、保肝护肝^[6]等多种功效, 用途极为广泛。然而, 由于过度采挖, 我国野生甘草资源蕴藏量已不足 50 万吨^[7]。自 2000 年起, 国务院已禁止采挖野生甘草 (国发[2000] 13 号), 栽培甘草成为中药材市场上的主流商品, 在宁夏、甘肃、新疆、内蒙古等省市区进行大面积种植。《中华人民共和国药典》(2015 版)^[8]规定甘草药材包括 3 种基原植物: 乌拉尔甘草 *Glycyrrhiza uralensis* Fisch.、胀果甘草 *Glycyrrhiza inflata* Bat. 及光果甘草 *Glycyrrhiza glabra* L., 且其质量标准中甘草酸 (glycyrrhizic acid) 含量不低于 2.0%, 甘草苷 (liquiritin) 含量不低于 0.5%。然而本课题组前期对全国 4 个主要中药材集散中心及主要栽培基地的甘草药材进行了深入的调查研究, 发现市售甘草药材质量良莠不齐, 存在大量甘草酸或甘草苷含量未达标的样品^[9, 10]。这将严重影响甘草药材的质量及药效发挥。

植物在长期进化的过程中, 由于自然突变的作用, 遗传基因会发生不同程度的变异, 形成基因多态性, 其中功能基因的多态性对于调控产物的合成具有显著的影响。就药材而言, 药用成分的含量高低是评价其质量的决定性因素。本课题组前期对甘草中甘草酸生物合成途径上的 3 个功能基因^[11-15]进行研究, 功能基因存在单核苷酸多态性 (single nucleotide

polymorphisms, SNPs)、插入/缺失长度多态性 (insertion/deletion length polymorphism, InDel)、选择性剪接多态性 (alternative splicing, AS) 以及拷贝数多态性 (copy number variations, CNVs), 而这些多态性会明确影响到次生代谢产物甘草酸的含量。因此推测, 甘草苷的生物合成途径也势必受关键酶及其编码基因的影响及调控。

查尔酮合酶 (chalcone synthase, *CHS*, EC2.3.1.74) 是甘草苷生物合成途径中的第 1 个关键限速酶^[16]。它催化 3 分子丙二酰辅酶 A (malonyl-CoA) 和 1 分子 4-香豆酰辅酶 A (4-coumaroyl-CoA) 缩合, 生成第 1 个以 C6-C3-C6 为基本骨架的黄酮类化合物——柚皮素查尔酮, 该产物再进一步衍化生成多种黄酮类化合物。*CHS* 基因在植物中广泛存在, Reimold 等^[17]于 1983 年首次报道了欧芹 *Petroselinum crispum* *CHS* 基因序列, 截止目前, GenBank 数据库中已经注册了 9 000 多条植物 *CHS* 的核苷酸序列。此前, 已报道对三基原甘草中 *CHS* 基因的克隆研究^[18, 19], 但样本量较小。有关 *CHS* 基因多态性与甘草苷含量的相关性研究则尚未见诸报道。

本文收集了来自于新疆、黑龙江及内蒙古 3 个省区 7 个不同居群的栽培及野生甘草种子, 在北京中医药大学建立了甘草种质资源圃。利用课题组前期已建立的 HPLC 方法, 对其中 60 份甘草样品 (20 份/基原) 中的 4 种主要黄酮类化合物进行了含量测定。进而, 选取含量最高的 5 株乌拉尔甘草及含量最低的 5 株胀果甘草分为黄酮高、低含量组, 进行 *CHS* 基因的克

隆, 并对其多态性进行分析, 筛选出黄酮高、低含量组特异 *CHS* 基因型及氨基酸序列类型。本文的研究结果将为阐明 *CHS* 基因多态性影响甘草苷生物合成的分子机制, 指导栽培甘草的分子育种奠定基础。

材料与方法

甘草植物材料 从全国3省份7个居群分别收集甘草种子, 统一栽培于北京中医药大学药草园, 生长两年, 由北京中医药大学刘颖副教授对其进行分子鉴定以确定基原。材料详细信息见表1。

实验仪器 Agilent 1100 高效液相色谱仪 (Agilent); CAPCELL PAK C18 MGII 色谱柱 (4.6 mm×250 mm, 5 μm, 日本资生堂株式会社); 十万分之一电子天平 CPA225D (德国赛多利斯科学仪器有限公司); 超声清洗仪 KQ-250E (中国昆山市超声仪器有限公司); 1-13000 型离心机 (SIGMA); TC-3000 PCR 扩增仪 (Techne); DYY-6C 型电泳仪 (北京六一仪器厂); BG-gdsAUTO510 型凝胶成像系统 (上海培清科技有限公司); 酶标仪 (Bio-Tek) 等。

实验试剂 甘草苷对照品 (纯度 98.5%, 批号 MUST-16032801)、异甘草苷对照品 (纯度 99.99%, 批号 MUST-15031204)、甘草素对照品 (纯度 99.07%, 批号 MUST-15021104), 均购于成都曼思特公司; 异甘草素对照品 (纯度 98%, 批号 16061501), 购于成都普菲德公司。乙腈 (色谱纯)、甲醇 (色谱纯) 为 Merck 公司产品, 磷酸 (分析纯) 为北京化工厂产品, 实验用水为屈臣氏蒸馏水。ESPY spin 植物 RNA 快速提取试剂盒、琼脂糖凝胶纯化回收试剂盒、DL2000 核酸分子质量标准、M-MLV (Rnase H⁻)、pMD19-T 载体、大肠杆菌 DH5α 感受态细胞、氨苄青霉素 (ampicillin, Amp)、琼脂糖等购于北京博迈德生物技术有限公司; LA Taq DNA 聚合酶、Oligo dT 等购于 TaKaRa 公司; 引物合成由生工生物工程 (上海) 股份有限公司完成。

HPLC 法分析甘草药材中 4 种黄酮含量 利用本课题组前期已建立的 HPLC 法^[10]对甘草样品中甘草苷、异甘草苷 (isoliquiritin)、甘草素 (liquiritigenin)、异甘草素 (isoliquiritigenin) 4 种主要黄酮类化合物的含量进行测定。运用统计软件 SAS 8.0 及 SPSS 19.0 对甘草中 4 种黄酮类化合物的含量及其相关性进行统计学分析。

甘草 RNA 提取及逆转录 取冻存于 -80 °C 冰箱中的甘草主根, 液氮中充分研磨后, 依据 ESPY spin 植物快速提取试剂盒说明书提取总 RNA; 利用琼脂

糖凝胶电泳检测 RNA 完整性; 利用酶标仪测定 RNA 浓度; 按 M-MLV (Rnase H⁻) 反转录试剂盒说明书, 以 Oligo dT 为引物, 对所提总 RNA 进行逆转录。

甘草 *CHS* 扩增及测序 根据 GenBank 上已注册的胀果甘草 *CHS* 基因序列 (EU706287), 利用 Primer Premier 5.0 软件设计引物如下: 上游引物 GF: 5'-CAA TGGTGAGCGTAGCTGAAATTC-3', 下游引物 GR: 5'-CAATCATATAGCCACACTGTGCAG-3'。25 μL PCR 扩增体系: 2.5 μL 模板 cDNA、0.25 μL 上游引物 GF (10 μmol·L⁻¹)、0.25 μL 下游引物 GR (10 μmol·L⁻¹)、0.25 μL LA TaqDNA 聚合酶、2.5 μL 10×Buffer、2.0 μL dNTP (2.5 mmol·L⁻¹)、17.25 μL RNase-free Water。PCR 扩增程序: 94 °C 5 min; 94 °C 15 s, 56 °C 30 s, 72 °C 90 s (30 个循环); 72 °C 5 min, 4 °C 保存。1.0% 琼脂糖凝胶电泳检测 PCR 扩增结果, 利用胶回收试剂盒对 PCR 产物进行胶回收纯化。将胶回收产物与 pMD19-T 载体连接后转化至大肠杆菌 DH5α 感受态细胞中, 涂于含有 50 mg·L⁻¹ Amp 的 LB 培养基平板上, 于 37 °C 倒置培养 12 h。随机挑取单克隆, 接种于含有 50 mg·L⁻¹ Amp 的 LB 液体培养基中, 于 37 °C、200 r·min⁻¹ 条件下培养 12 h。对菌液进行 PCR 验证, 将阳性克隆送生工生物工程 (上海) 股份有限公司测序, 测至饱和。

生物信息学分析 应用软件 DNAMAN 对全部 cDNA 序列进行比对分析; 应用软件 Editseq 将全部 cDNA 序列翻译成氨基酸序列; 对甘草黄酮高/低含量组中高频出现的序列进行以下生物信息学分析: 运用 SOPMA 进行二级结构预测; 运用 PROSITE 进行蛋白活性位点预测; 运用 Swiss Model 进行三级结构同源模建; 运用 SAVES 服务器对同源模建蛋白结果进行评价; 运用 Discovery Studio 2.5 对三维结构进行功能预测; 运用 MEGA 5.0 进行同源比对并构建系统进化树。

结果

1 三基原甘草样品中 4 种黄酮类化合物的含量分析

HPLC 色谱图如图 1 所示, 曲线 A 为 4 种黄酮类化合物混合对照品的色谱图, 曲线 B、C、D 分别为乌拉尔甘草、光果甘草、胀果甘草样品的色谱图, 由图可知甘草苷、异甘草苷、甘草素、异甘草素的保留时间分别为 9.052、19.393、24.035、35.236 min。甘草苷的标准曲线为: $Y = 5.6447 \times 10^{-4} X + 2.2045 \times 10^{-4}$ ($R^2 = 0.9999$); 异甘草苷的标准曲线为: $Y = 2.4798 \times 10^{-4} X + 2.2144 \times 10^{-4}$ ($R^2 = 0.9999$); 甘草素的标准曲

Table 1 The contents of four main flavonoids in 60 samples

| Code | Species | Location | Liquiritin | Isoliquiritin | Liquiritigenin | Isoliquiritigenin |
|------|---------------------|-----------------------------------|------------|---------------|----------------|-------------------|
| H67 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.930 5% | 0.212 3% | 0.131 1% | 0.045 9% |
| H71 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.927 6% | 0.130 1% | 0.024 6% | 0.005 7% |
| C59 | <i>G. uralensis</i> | Etuoqeqian, Neimeng | 0.882 9% | 0.129 3% | 0.179 0% | 0.044 2% |
| H31 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.817 4% | 0.125 2% | 0.249 9% | 0.038 3% |
| I69 | <i>G. glabra</i> | Dafang cultivation base, Xinjiang | 0.743 9% | 0.089 8% | 0.049 4% | 0.010 6% |
| H95 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.729 6% | 0.130 8% | 0.079 5% | 0.016 6% |
| A37 | <i>G. uralensis</i> | Zhaozhou, Heilongjiang | 0.659 2% | 0.100 4% | 0.063 1% | 0.065 0% |
| H96 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.591 9% | 0.099 4% | 0.068 0% | 0.012 8% |
| H93 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.558 2% | 0.079 1% | 0.032 9% | 0.008 0% |
| H66 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.553 2% | 0.047 9% | 0.049 9% | 0.005 0% |
| L1 | <i>G. uralensis</i> | Zhaozhou, Heilongjiang | 0.553 2% | 0.078 6% | 0.056 2% | 0.013 2% |
| H94 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.550 9% | 0.108 5% | 0.083 7% | 0.021 2% |
| L3 | <i>G. uralensis</i> | Etuoqeqian, Neimeng | 0.537 8% | 0.074 5% | 0.052 2% | 0.012 3% |
| H83 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.537 5% | 0.072 1% | 0.055 0% | 0.007 7% |
| L2 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.509 4% | 0.069 8% | 0.049 5% | 0.010 2% |
| A21 | <i>G. uralensis</i> | Zhaozhou, Heilongjiang | 0.465 7% | 0.072 9% | 0.126 1% | 0.022 4% |
| H70 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.425 4% | 0.072 1% | 0.039 8% | 0.022 8% |
| I68 | <i>G. glabra</i> | Dafang cultivation base, Xinjiang | 0.378 7% | 0.053 9% | 0.035 7% | 0.008 7% |
| C56 | <i>G. uralensis</i> | Etuoqeqian, Neimeng | 0.241 6% | 0.030 4% | 0.036 6% | 0.007 0% |
| H15 | <i>G. uralensis</i> | Hetian, Xinjiang | 0.205 0% | 0.032 7% | 0.019 3% | 0.005 8% |
| C46 | <i>G. uralensis</i> | Etuoqeqian, Neimeng | 0.174 3% | 0.022 5% | 0.029 3% | 0.006 1% |
| J79 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.174 1% | 0.036 6% | 0.016 9% | 0.004 8% |
| C23 | <i>G. uralensis</i> | Etuoqeqian, Neimeng | 0.168 7% | 0.023 9% | 0.023 4% | 0.005 2% |
| J76 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.153 7% | 0.028 5% | 0.014 3% | 0.006 6% |
| N11 | <i>G. glabra</i> | Buerjin, Xinjian | 0.145 0% | 0.018 7% | 0.017 4% | 0.005 2% |
| N10 | <i>G. glabra</i> | Buerjin, Xinjiang | 0.144 5% | 0.020 4% | 0.015 7% | 0.005 2% |
| I87 | <i>G. glabra</i> | Dafang cultivation base, Xinjiang | 0.143 1% | 0.019 7% | 0.013 8% | 0.001 5% |
| J60 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.134 7% | 0.008 8% | 0.010 1% | 0.005 1% |
| N9 | <i>G. glabra</i> | Buerjin, Xinjiang | 0.130 5% | 0.022 5% | 0.017 8% | 0.004 5% |
| I88 | <i>G. glabra</i> | Dafang cultivation base, Xinjiang | 0.129 5% | 0.021 3% | 0.017 9% | 0.010 7% |
| I89 | <i>G. glabra</i> | Dafang cultivation base, Xinjiang | 0.129 4% | 0.021 1% | 0.011 0% | 0.001 5% |
| N1 | <i>G. glabra</i> | Buerjin, Xinjiang | 0.129 2% | 0.020 4% | 0.015 9% | 0.004 5% |
| N4 | <i>G. glabra</i> | Buerjin, Xinjiang | 0.128 2% | 0.024 5% | 0.016 5% | 0.004 9% |
| N2 | <i>G. glabra</i> | Buerjin, Xinjiang | 0.125 0% | 0.021 6% | 0.016 7% | 0.005 3% |
| N5 | <i>G. glabra</i> | Buerjin, Xinjiang | 0.122 5% | 0.021 2% | 0.0165% | 0.005 2% |
| N8 | <i>G. glabra</i> | Buerjin, Xinjiang | 0.122 5% | 0.019 8% | 0.018 6% | 0.004 8% |
| N3 | <i>G. glabra</i> | Buerjin, Xinjiang | 0.122 0% | 0.019 6% | 0.017 9% | 0.004 4% |
| J77 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.120 5% | 0.025 9% | 0.019 3% | 0.009 9% |
| N7 | <i>G. glabra</i> | Buerjin, Xinjiang | 0.118 7% | 0.019 1% | 0.016 9% | 0.004 5% |
| N12 | <i>G. glabra</i> | Buerjin, Xinjiang | 0.115 5% | 0.021 3% | 0.016 7% | 0.003 9% |
| N6 | <i>G. glabra</i> | Buerjin, Xinjiang | 0.099 5% | 0.020 1% | 0.019 8% | 0.004 5% |
| I86 | <i>G. glabra</i> | Dafang cultivation base, Xinjiang | 0.096 6% | 0.014 4% | 0.011 3% | 0.002 8% |
| M5 | <i>G. inflata</i> | Alaer, Xinjiang | 0.086 7% | 0.011 4% | 0.005 6% | 0.002 8% |
| J75 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.084 3% | 0.007 7% | 0.011 5% | missing |
| J82 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.083 5% | 0.016 1% | 0.024 7% | 0.005 4% |
| J65 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.081 8% | 0.012 9% | 0.005 6% | 0.003 4% |
| I17 | <i>G. glabra</i> | Dafang cultivation base, Xinjiang | 0.077 8% | 0.015 4% | 0.021 9% | 0.006 2% |
| M4 | <i>G. inflata</i> | Alaer, Xinjiang | 0.074 5% | 0.007 7% | 0.007 2% | 0.002 6% |
| M1 | <i>G. inflata</i> | Alaer, Xinjiang | 0.074 0% | 0.009 4% | 0.005 6% | 0.003 0% |
| J74 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.073 9% | 0.013 9% | 0.007 6% | 0.001 6% |
| M2 | <i>G. inflata</i> | Alaer, Xinjiang | 0.071 7% | 0.010 6% | 0.006 4% | 0.003 1% |
| J64 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.070 5% | 0.009 7% | 0.004 6% | 0.002 2% |
| M3 | <i>G. inflata</i> | Alaer, Xinjiang | 0.069 7% | 0.008 8% | 0.005 7% | 0.003 5% |
| J63 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.064 3% | 0.003 9% | 0.005 5% | 0.000 7% |
| J62 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.064 1% | 0.009 4% | 0.004 9% | 0.003 0% |
| I90 | <i>G. glabra</i> | Dafang cultivation base, Xinjiang | 0.042 5% | 0.006 2% | 0.004 0% | 0.001 8% |
| J61 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.039 1% | 0.002 9% | 0.004 0% | missing |
| J81 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.033 3% | 0.007 4% | 0.002 2% | 0.000 9% |
| J80 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.021 4% | 0.005 6% | 0.003 9% | 0.001 1% |
| J20 | <i>G. inflata</i> | Dafang cultivation base, Xinjiang | 0.017 7% | 0.003 0% | 0.003 9% | 0.001 2% |

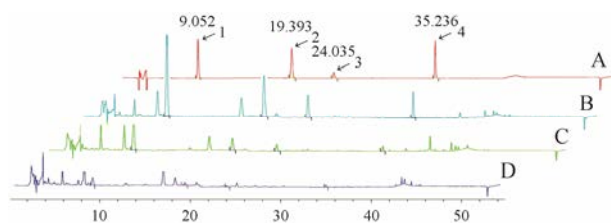


Figure 1 HPLC chromatograms of reference substance and samples. Compound 1–4: liquiritin, isoliquiritin, liquiritigenin, isoliquiritigenin; A: HPLC chromatograms of reference substance; B, C, D: HPLC chromatograms of *G. uralensis*, *G. glabra*, and *G. inflata*

线为: $Y = 3.1796 \times 10^{-4} X - 3.2220 \times 10^{-5}$ ($R^2 = 0.9999$); 异甘草素的标准曲线为: $Y = 1.4238 \times 10^{-4} X - 6.7810 \times 10^{-5}$ ($R^2 = 0.9999$)。60份甘草样品中4种黄酮化合物的百分含量列于表1中,三基原甘草中4种黄酮类化合物含量列于表2中。

2 三基原甘草中黄酮类化合物含量的统计学分析

利用 SPSS 19.0 及 SAS 8.0 对三基原甘草样品中甘草苷、异甘草苷、甘草素及异甘草素含量进行相关性分析,结果列于表3,差异分析结果列于表4。

Spearman 相关性分析表明甘草苷与异甘草苷、甘草素与异甘草素的含量均具有显著相关关系。 χ^2 检验显示甘草苷、异甘草苷、甘草素和异甘草素的含量在三基原甘草之间的差异均具有统计学意义,且通过比较秩均数发现,这4种主要黄酮类成分的含量在三基原甘草中均满足:乌拉尔甘草>光果甘草>胀果甘草。

3 甘草 CHS 基因多态性分析

由于黄酮类化合物的含量在三基原甘草间均存在显著性差异,其中乌拉尔甘草含量最高,而胀果甘草含量最低,因此在上述60份甘草样品中,选取黄酮含量最高的5株乌拉尔甘草(表1中红色标出)及黄酮含量最低的5株胀果甘草(表1中绿色标出)进行多态性分析,按甘草苷含量由高至低的顺序重新编号为H1~H5, L1~L5。10株甘草样品共克隆得到336条长度为1175bp的CHS基因cDNA序列。DNAMAN 比对分析显示336条序列中存在249个变异位点,其中错义突变位点141个,可分为137种单倍型(单倍型1~137),序列一致性为99.98%。变异位点统计于表5中。

Table 2 The contents of four main flavonoids in *G. uralensis* Fisch., *G. inflata* Bat. and *G. glabra* L.

| Species | Quantity | Content/mg·g ⁻¹ | | | |
|---------------------|----------|----------------------------|-------------------|-------------------|-------------------|
| | | Liquiritin | Isoliquiritin | Liquiritigenin | Isoliquiritigenin |
| <i>G. uralensis</i> | 20 | 5.510 0 ± 2.340 6 | 0.856 1 ± 0.461 8 | 0.724 0 ± 0.584 0 | 0.187 2 ± 0.168 4 |
| <i>G. inflata</i> | 20 | 0.796 7 ± 0.404 1 | 0.120 0 ± 0.087 7 | 0.084 9 ± 0.059 5 | 0.033 9 ± 0.023 1 |
| <i>G. glabra</i> | 20 | 1.622 3 ± 1.507 0 | 0.245 5 ± 0.176 1 | 0.185 6 ± 0.092 9 | 0.050 4 ± 0.025 2 |

Table 3 The Spearman correlation test results. **Correlation is significant at the 0.01 level (2-tailed)

| | | Liquiritin | Isoliquiritin | Liquiritigenin | Isoliquiritigenin |
|-------------------|-------------------------|------------|---------------|----------------|-------------------|
| Liquiritin | Correlation coefficient | 1 | 0.946** | 0.887** | 0.834** |
| | Sig. (2-tailed) | | 0.000 | 0.000 | 0.000 |
| | N | 60 | 60 | 60 | 58 |
| Isoliquiritin | Correlation coefficient | 0.946** | 1 | 0.908** | 0.866** |
| | Sig. (2-tailed) | 0.000 | | 0.000 | 0.000 |
| | N | 60 | 60 | 60 | 58 |
| Liquiritigenin | Correlation coefficient | 0.887** | 0.908** | 1 | 0.906** |
| | Sig. (2-tailed) | 0.000 | 0.000 | | 0.000 |
| | N | 60 | 60 | 60 | 58 |
| Isoliquiritigenin | Correlation coefficient | 0.834** | 0.866** | 0.906** | 1 |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | |
| | N | 58 | 58 | 58 | 58 |

Table 4 The χ^2 test results

| Species | Quantity | Liquiritin | | | Isoliquiritin | | | Liquiritigenin | | | Isoliquiritigenin | | |
|---------------------|----------|------------|----------|--------------|---------------|----------|--------------|----------------|----------|--------------|-------------------|----------|--------------|
| | | χ^2 | P | Rank average | χ^2 | P | Rank average | χ^2 | P | Rank average | χ^2 | P | Rank average |
| <i>G. uralensis</i> | 20 | | | 49.450 | | | 49.150 | | | 49.600 | | | 45.725 |
| <i>G. inflata</i> | 20 | 41.431 6 | <0.000 1 | 14.200 | 40.323 2 | <0.000 1 | 14.350 | 42.494 1 | <0.000 1 | 13.850 | 30.811 0 | <0.000 1 | 16.278 |
| <i>G. glabra</i> | 20 | | | 27.850 | | | 28.000 | | | 28.050 | | | 25.175 |

Table 5 Variable sites in the 137 *CHS* haplotypes. Underline: Missense mutation

| Variable site/bp | Variation | Variable site/bp | Variation | Variable site/bp | Variation | Variable site/bp | Variation | Variable site/bp | Variation | Variable site/bp | Variation | Variable site/bp | Variation | Variable site/bp | Variation |
|------------------|-----------|------------------|-----------|------------------|-----------|------------------|-----------|------------------|-----------|------------------|-----------|------------------|-----------|------------------|-----------|
| 25 | G/A | 178 | T/C | 289 | A/G | 466 | A/G | 620 | G/A | 760 | C/T | 891 | T/A | 1057 | A/G |
| 35 | A/G | 180 | T/C | 292 | T/C | 467 | A/G | 623 | T/C | 763 | T/C | 894 | A/G | 1058 | A/G |
| 37 | G/T | | T/A | 296 | G/C | 473 | T/C | 628 | T/C | 770 | T/C | 897 | T/C | 1059 | T/C |
| 44 | A/G | 184 | A/G | 299 | C/T | 474 | A/G | 629 | T/C | 776 | T/C | 912 | C/T | 1061 | A/C |
| 53 | A/G | 185 | T/C | 322 | A/G | 477 | A/G | 640 | C/T | 781 | A/G | 923 | G/A | 1064 | T/C |
| 54 | A/G | 192 | A/G | 325 | A/G | 478 | T/C | 644 | A/G | 782 | A/G | | G/T | 1066 | A/G |
| 58 | T/C | 199 | A/G | 326 | A/G | 480 | T/C | 645 | T/C | 784 | T/C | 928 | T/C | 1067 | A/G |
| 61 | T/C | 204 | C/A | 332 | A/G | 485 | A/G | 651 | G/A | 790 | T/C | 934 | A/G | 1073 | A/G |
| 66 | A/G | 205 | G/A | 337 | A/G | 494 | C/T | 652 | A/G | 792 | A/C | 943 | A/G | 1078 | A/G |
| 68 | T/C | 206 | A/G | 338 | A/G | 497 | T/C | 658 | C/T | 795 | T/C | 948 | A/G | 1080 | A/G |
| 70 | G/A | 208 | A/G | 349 | A/G | 499 | C/T | 659 | A/G | 800 | T/C | 949 | A/G | 1083 | A/G |
| 71 | C/T | 210 | A/G | 359 | A/G | 500 | A/G | 662 | T/C | 814 | T/C | 951 | T/C | 1084 | C/T |
| 74 | T/C | 211 | T/C | 368 | G/A | 505 | G/A | 670 | T/C | 818 | C/T | 954 | G/T | 1085 | T/C |
| 80 | C/T | 213 | T/C | 372 | A/G | 507 | A/G | 685 | C/T | 820 | G/A | 956 | T/C | 1086 | G/A |
| 83 | A/G | 217 | T/C | 374 | T/C | 511 | A/G | 686 | A/G | 824 | T/A | 957 | T/A | 1094 | G/A |
| 89 | C/T | 221 | A/G | 375 | A/G | 526 | A/G | 687 | A/G | 827 | C/A | 958 | T/C | 1096 | T/C |
| 94 | T/C | 223 | A/G | 384 | A/G | 527 | A/G | 688 | T/C | 828 | T/A | 960 | A/G | 1097 | T/C |
| 100 | A/G | 238 | A/G | 386 | C/T | 529 | A/G | 692 | T/C | 831 | A/G | 962 | G/A | 1099 | A/G |
| 101 | G/A | 239 | G/A | 393 | A/G | 536 | G/A | 698 | T/C | 832 | A/G | 969 | A/G | 1107 | G/A |
| 108 | T/G | 241 | A/G | 401 | T/C | 555 | C/T | 702 | A/G | 842 | T/C | 972 | A/G | 1109 | G/A |
| 110 | T/C | 247 | A/G | 404 | T/A | 569 | T/G | 717 | T/G | 845 | A/G | 983 | T/C | 1114 | T/C |
| 121 | A/G | 249 | A/G | | T/G | | T/C | | T/C | 850 | T/C | 991 | T/C | 1115 | C/T |
| 123 | T/C | 254 | T/C | 406 | T/C | 579 | G/A | 722 | T/C | 852 | A/G | 998 | T/C | 1122 | G/A |
| 126 | A/G | 258 | T/C | 409 | A/T | 583 | C/T | 725 | G/T | 854 | T/C | 1005 | G/A | 1124 | A/G |
| 134 | A/G | 259 | A/G | 431 | G/A | 584 | T/C | 726 | A/G | 861 | T/C | 1007 | T/C | 1127 | T/A |
| 137 | T/C | 261 | A/T | 433 | T/C | 594 | T/C | 728 | T/G | 875 | C/T | 1010 | T/C | 1133 | T/A |
| 138 | A/T | 264 | G/A | 439 | A/T | 596 | C/T | | T/C | 876 | A/G | 1019 | T/C | 1136 | C/A |
| 140 | T/C | 266 | G/A | 442 | T/C | 599 | T/A | 733 | A/G | 877 | T/C | 1023 | T/C | 1145 | T/C |
| 148 | A/G | 269 | G/A | 451 | T/C | 608 | T/C | 735 | A/G | | T/A | 1028 | C/T | | T/A |
| 169 | T/C | 277 | A/G | 458 | A/G | | T/A | 740 | A/G | 878 | C/A | 1037 | C/A | 1147 | T/C |
| 170 | T/C | 278 | C/T | 463 | T/A | | T/G | 746 | A/G | 879 | T/C | 1038 | T/C | 1149 | A/G |
| 175 | G/A | 287 | A/G | 465 | A/G | 618 | C/T | 754 | A/G | 881 | T/C | 1043 | T/C | 1152 | C/T |

4 甘草 *CHS* 氨基酸序列多态性分析

上述长度为 1 175 bp 的 *CHS* 序列包含 1 个完整的开放阅读框 (open reading frame, ORF), 编码 1 条包含 389 个氨基酸残基的多肽序列。336 条 *CHS* 氨基酸序列中共存在 130 个变异位点, 可分为 102 种氨基酸序列类型 (AA-1 ~ AA-102), 序列一致性为 99.33%。变异位点统计于表 6 中。其中, 单倍型 3~23 为同义突变序列, 编码氨基酸序列类型 AA-3; 单倍型 25~27 为同义突变序列, 编码 AA-5; 单倍型 28、29 为同义突变序列, 编码 AA-6; 单倍型 32~34 为同义突变序列, 编码 AA-9; 单倍型 60~62 为同义突变序列, 编码 AA-35; 单倍型 63~68 为同义突变序列, 编码 AA-36; 单倍型 92~95 为同义突变序列, 编码 AA-60。

5 黄酮高、低含量组 *CHS* 单倍型特异性分析

将全部氨基酸序列注册号、单倍型来源及不同单倍型、不同氨基酸序列类型出现频次统计于表 7 中。137 种 *CHS* 单倍型中, 黄酮高含量组甘草样品中存在 71 种 *CHS* 单倍型, 特有单倍型 64 种; 黄酮低含量组甘草样品中存在 73 种 *CHS* 单倍型, 特有单倍型 66 种; 高、低含量组共有 *CHS* 单倍型 7 种, 即单倍型 1、6、25、28、35、51 和单倍型 92。黄酮高含量组样品中共克隆获得了 178 条 *CHS* 基因 cDNA 序列, 出现频率最高的单倍型为单倍型 1、10 及 60 (数目: 8 条, 占比: 4.49%), 其中单倍型 10 及 60 为黄酮高含量组甘草样品特有 *CHS* 单倍型。黄酮低含量组样品中共克隆获得 158 条 *CHS* 基因 cDNA 序列, 出现频率最高的单倍型为单倍型 92 (数目: 6 条, 占比:

Table 6 Variable sites in the 102 CHS amino acid sequence types

| Variable site | Variation | Variable site | Variation | Variable site | Variation | Variable site | Variation | Variable site | Variation | Variable site | Variation | Variable site | Variation |
|---------------|-----------|---------------|-----------|---------------|-----------|---------------|-----------|---------------|-----------|---------------|-----------|---------------|-----------|
| 8 | R-H | 64 | M-V | 108 | E-G | 168 | G-D | 234 | K-E | 284 | T-A | 335 | G-S |
| 12 | R-M | 66 | K-R | 112 | K-R | 169 | T-A | 239 | L-V | 287 | F-L | 341 | C-R |
| 18 | N-D | 68 | R-Q | 116 | E-G | 170 | V-A | 241 | W-C | 292 | I-T | 352 | K-R |
| 19 | I-T | 69 | Y-C | 124 | I-V | 175 | K-R | 242 | T-A | | I-K | 353 | S-P |
| 20 | L-S | 70 | M-V | 125 | T-A | 176 | D-G | 244 | Q-R | | I-V | 355 | Q-R |
| 22 | I-V | | M-T | 128 | I-V | 185 | R-H | 245 | T-A | 293 | S-P | 359 | K-R |
| 23 | G-D | 71 | Y-H | 131 | T-A | | R-C | 246 | I-M | 296 | N-T | 360 | T-A |
| 31 | V-A | 72 | L-P | 135 | V-A | 193 | V-I | 251 | E-G | 297 | S-T | 361 | T-A |
| 33 | Q-R | 74 | E-G | 136 | D-V | 194 | T-I | 253 | A-V | 298 | I-V | | T-I |
| 36 | Y-D | 79 | E-G | 144 | L-P | 198 | F-L | 254 | I-T | 299 | F-L | 362 | G-S |
| 40 | Y-C | 80 | N-S | 146 | K-I | 209 | L-P | 260 | E-G | 304 | P-S | 365 | L-P |
| 41 | F-L | 82 | N-S | 147 | L-P | 213 | A-V | 261 | V-A | 309 | I-T | 366 | E-G |
| 42 | K-E | 83 | I-V | 150 | L-P | 215 | F-L | 263 | L-P | 311 | D-G | 369 | V-I |
| 46 | S-C | 86 | Y-C | 154 | V-E | 217 | D-G | 264 | T-P | 314 | E-G | 371 | F-S |
| 49 | K-R | | Y-H | 155 | K-E | | D-N | 265 | F-L | 316 | K-G | 374 | G-R |
| 56 | F-S | 87 | M-L | | K-R | 219 | A-V | 271 | V-A | 318 | A-S | 382 | V-A |
| 58 | R-H | 88 | A-T | 158 | M-V | 223 | I-T | 273 | G-E | 319 | L-S | 383 | V-I |
| 59 | M-T | 92 | D-G | 159 | M-V | 228 | P-L | 276 | S-T | 320 | K-E | | |
| 60 | C-R | 96 | D-G | | M-T | 229 | I-V | 277 | K-R | 323 | K-E | | |
| | C-S | 97 | M-T | 160 | Y-H | | I-A | | K-E | 324 | M-V | | |
| 61 | D-G | 107 | K-R | 166 | A-V | | I-T | 283 | L-P | 330 | V-A | | |

Table 7 Haplotypes-quantity-source-AA sequence types-GenBank accession No. information of 137 haplotypes

| Haplotypes | Quantity | Source | Amino acid sequence type | Quantity | Gen Bank accession No. | Haplotypes | Quantity | Source | Amino acid sequence type | Quantity | Gen Bank accession No. | | |
|------------|----------|----------------|--------------------------|----------|------------------------|------------|----------|------------------------|--------------------------|----------|------------------------|----------|--|
| 1 | 12 | H3, H4, L2, L3 | AA-1 | 12 | KY748116 | 70 | 1 | L4 | AA-38 | 1 | KY810366 | | |
| | | | | | | 71 | 3 | L5 | AA-39 | 2 | KY810367 | | |
| 2 | 1 | H3 | AA-2 | 1 | KY748117 | 72 | 3 | H5 | AA-40 | 3 | KY810368 | | |
| 3 | 4 | H3 | AA-3 | 58 | KY748118 | 73 | 1 | H5 | AA-41 | 1 | KY810369 | | |
| 4 | 3 | H3 | | | KY748119 | 74 | 2 | H5 | AA-42 | 2 | KY810370 | | |
| 5 | 2 | H3 | | | KY748120 | 75 | 2 | H5 | AA-43 | 2 | KY810371 | | |
| 6 | 7 | H3, L2 | | | KY748121 | 76 | 2 | L5 | AA-44 | 1 | KY810372 | | |
| 7 | 1 | L1 | | | KY748122 | 77 | 2 | H4 | AA-45 | 2 | KY810373 | | |
| 8 | 1 | L1 | | | KY748123 | 78 | 3 | H4 | AA-46 | 3 | KY810374 | | |
| 9 | 1 | L1 | | | KY748124 | 79 | 1 | H4 | AA-47 | 2 | KY810375 | | |
| 10 | 8 | H1, H3 | | | KY748125 | 80 | 2 | H4 | AA-48 | 2 | KY810376 | | |
| 11 | 1 | H1 | | | KY748126 | 81 | 1 | H4 | AA-49 | 1 | KY810377 | | |
| 12 | 1 | H1 | | | KY748127 | 82 | 3 | H4 | AA-50 | 3 | KY810378 | | |
| 13 | 4 | H1, H5 | | | KY748128 | 83 | 2 | H4 | AA-51 | 3 | KY810379 | | |
| 14 | 1 | H1 | | | KY748129 | 84 | 1 | H4 | AA-52 | 1 | KY810380 | | |
| 15 | 2 | H1 | | | KY748130 | 85 | 2 | H4 | AA-53 | 2 | KY810381 | | |
| 16 | 3 | L4 | | | KY748131 | 86 | 1 | H4 | AA-54 | 1 | KY810382 | | |
| 17 | 2 | L4 | | | KY748132 | 87 | 3 | L2 | AA-55 | 3 | KY810383 | | |
| 18 | 2 | L4 | | | KY748133 | 88 | 2 | L2 | AA-56 | 2 | KY810384 | | |
| 19 | 3 | L5 | | | KY748134 | 89 | 1 | L2 | AA-57 | 1 | KY810385 | | |
| 20 | 4 | H5 | | | KY748135 | 90 | 3 | L2 | AA-58 | 3 | KY810386 | | |
| 22 | 2 | L3 | | | KY748138 | 91 | 2 | L2 | AA-59 | 2 | KY810387 | | |
| 23 | 3 | H3 | | | KY748139 | 92 | 12 | H2, H4, L1, L2, L3, L5 | AA-60 | 18 | KY810388 | | |
| 24 | 2 | H3 | AA-4 | 2 | KY748140 | | | | | | | | |
| 25 | 6 | H3, H4, L3 | AA-5 | 10 | KY748141 | 93 | 3 | H4 | | | | KY810389 | |

| | | | | | | Continued | | | | | |
|------------|----------|----------------|--------------------------|----------|------------------------|------------|----------|--------|--------------------------|----------|------------------------|
| Haplotypes | Quantity | Source | Amino acid sequence type | Quantity | Gen Bank accession No. | Haplotypes | Quantity | Source | Amino acid sequence type | Quantity | Gen Bank accession No. |
| 26 | 1 | L3 | | | KY748142 | 94 | 1 | H2 | | | KY810390 |
| 27 | 3 | L2 | | | KY748143 | 95 | 2 | H2 | | | KY810391 |
| 28 | 6 | H3, L2, L3, L4 | AA-6 | 6 | KY748145 | 96 | 1 | L2 | AA-61 | 1 | KY810392 |
| 29 | 2 | H2 | | | KY748147 | 97 | 2 | L2 | AA-62 | 2 | KY810393 |
| 30 | 1 | H3 | AA-7 | 2 | KY748148 | 98 | 2 | L2 | AA-63 | 2 | KY810394 |
| 31 | 3 | H3 | AA-8 | 3 | KY748149 | 99 | 3 | H2 | AA-64 | 3 | KY810395 |
| 32 | 6 | H2, H3 | AA-9 | 10 | KY748150 | 100 | 1 | H2 | AA-65 | 1 | KY810396 |
| 33 | 2 | H4 | | | KY748151 | 101 | 2 | H2 | AA-66 | 2 | KY810397 |
| 34 | 2 | L3 | | | KY748152 | 102 | 1 | H2 | AA-67 | 1 | KY810398 |
| 35 | 7 | H3, H4, L2 | AA-10 | 6 | KY748153 | 103 | 2 | H2 | AA-68 | 2 | KY810399 |
| 36 | 2 | H3 | AA-11 | 1 | KY748154 | 104 | 3 | H2 | AA-69 | 3 | KY810400 |
| 37 | 2 | H3 | AA-12 | 3 | KY748155 | 105 | 3 | L3 | AA-70 | 3 | KY810401 |
| 38 | 2 | H3 | AA-13 | 2 | KY748156 | 106 | 3 | L3 | AA-71 | 3 | KY810402 |
| 39 | 3 | L1 | AA-14 | 3 | KY748157 | 107 | 1 | L3 | AA-72 | 1 | KY810403 |
| 40 | 1 | L1 | AA-15 | 1 | KY748158 | 108 | 2 | L3 | AA-73 | 2 | KY810404 |
| 41 | 4 | L1 | AA-16 | 4 | KY748159 | 109 | 3 | L3 | AA-74 | 3 | KY810405 |
| 42 | 1 | L1 | AA-17 | 1 | KY748160 | 110 | 1 | L3 | AA-75 | 1 | KY810406 |
| 43 | 2 | L1 | AA-18 | 2 | KY748161 | 111 | 2 | L3 | AA-76 | 2 | KY810407 |
| 44 | 3 | L1 | AA-19 | 3 | KY748162 | 112 | 1 | L3 | AA-77 | 1 | KY810408 |
| 45 | 3 | L1 | AA-20 | 3 | KY748163 | 113 | 3 | L3 | AA-78 | 3 | KY810409 |
| 46 | 2 | L1 | AA-21 | 2 | KY748164 | 114 | 2 | L3 | AA-79 | 2 | KY810410 |
| 47 | 2 | L1 | AA-22 | 2 | KY748165 | 115 | 1 | L3 | AA-80 | 1 | KY810411 |
| 48 | 3 | L1 | AA-23 | 3 | KY810344 | 116 | 3 | H2 | AA-81 | 3 | KY810412 |
| 49 | 1 | L1 | AA-24 | 1 | KY810345 | 117 | 2 | H4 | AA-82 | 2 | KY810413 |
| 50 | 3 | L1 | AA-25 | 3 | KY810346 | 118 | 1 | H4 | AA-83 | 1 | KY810414 |
| 51 | 3 | H2, L1 | AA-26 | 3 | KY810347 | 119 | 2 | H4 | AA-84 | 2 | KY810415 |
| 52 | 1 | L1 | AA-27 | 1 | KY810348 | 120 | 2 | H4 | AA-85 | 2 | KY810416 |
| 53 | 1 | H1 | AA-28 | 1 | KY810349 | 121 | 2 | H3 | AA-86 | 2 | KY810417 |
| 54 | 3 | H1 | AA-29 | 3 | KY810350 | 122 | 2 | H3 | AA-87 | 2 | KY810418 |
| 55 | 3 | H1 | AA-30 | 3 | KY810351 | 123 | 1 | H3 | AA-88 | 1 | KY810419 |
| 56 | 1 | H1 | AA-31 | 1 | KY810352 | 124 | 3 | H3 | AA-89 | 3 | KY810420 |
| 57 | 2 | H1 | AA-32 | 2 | KY810353 | 125 | 2 | H3 | AA-90 | 2 | KY810421 |
| 58 | 3 | H1 | AA-33 | 3 | KY810354 | 126 | 2 | L2 | AA-91 | 2 | KY810422 |
| 59 | 1 | H1 | AA-34 | 1 | KY810355 | 127 | 2 | L2 | AA-92 | 2 | KY810423 |
| 60 | 8 | H1, H5 | AA-35 | 12 | KY810356 | 128 | 3 | L2 | AA-93 | 3 | KY810424 |
| 61 | 2 | H1 | | | KY810357 | 129 | 2 | L2 | AA-94 | 2 | KY810425 |
| 62 | 2 | H5 | | | KY810358 | 130 | 3 | H2 | AA-95 | 3 | KY810426 |
| 63 | 2 | L4 | AA-36 | 12 | KY810359 | 131 | 3 | L3 | AA-96 | 3 | KY810427 |
| 64 | 3 | L4 | | | KY810360 | 132 | 1 | L3 | AA-97 | 1 | KY810428 |
| 65 | 1 | L4 | | | KY810361 | 133 | 2 | L3 | AA-98 | 2 | KY810429 |
| 66 | 2 | L4 | | | KY810362 | 134 | 1 | L3 | AA-99 | 1 | KY810430 |
| 67 | 2 | L4 | | | KY810363 | 135 | 1 | L3 | AA-100 | 1 | KY810431 |
| 68 | 2 | L2 | | | KY810364 | 136 | 2 | L3 | AA-101 | 2 | KY810432 |
| 69 | 2 | L4 | AA-37 | 2 | KY810365 | 137 | 3 | L3 | AA-102 | 3 | KY810433 |

3.80%), 其次为单倍型 1、6、28、41 (数目: 4 条, 占比: 2.53%), 其中单倍型 41 为黄酮低含量组甘草样品特有单倍型。

6 黄酮高、低含量组 *CHS* 氨基酸序列类型特异性分析

根据表 7 中的统计数据, 102 种氨基酸序列类型中, 黄酮高含量组甘草样品存在 53 种氨基酸序列类

型,特有氨基酸序列类型为 45 种;黄酮低含量组甘草样品中存在 57 种氨基酸序列类型,特有氨基酸序列类型为 49 种;高、低含量组共有的氨基酸序列类型有 8 种,即 AA-1、AA-3、AA-5、AA-6、AA-9、AA-10、AA-26 和 AA-60。黄酮高含量组甘草样品中共计得到 178 条 CHS 氨基酸序列,出现频率最高的为 AA-3(数目:36 条,占比 20.22%),其次为 AA-35、AA-60(数目:12 条,占比 6.74%),其中 AA-35 为黄酮高含量组甘草样品特有 CHS 氨基酸序列类型。黄酮低含量组甘草样品中共计得到 158 条 CHS 氨基酸序列,出现频率最高的也是 AA-3(数目:22 条,占比:13.92%),其次为 AA-36(数目:12 条,占比 7.59%),为黄酮低含量组甘草样品的特异 CHS 氨基酸序列类型。综上,由于 AA-35 是黄酮高含量组甘草样品中占比最高的特有氨基酸序列类型,而 AA-36 是黄酮低含量组甘草样品中占比最高的特有氨基酸序列类型,因此将进一步对这两条序列进行二级结构、三级结构等生物信息学分析,从分子水平上寻找其影响甘草黄酮类化合物生物合成的原因。

7 黄酮高、低含量组特有 CHS 氨基酸序列类型 AA-35 与 AA-36 的二级结构比对分析

对 AA-35 与 AA-36 进行比对分析,发现其在 193 位与 229 位处存在 I/V、V/T 突变。应用 SOPMA 对 AA-35 及 AA-36 进行二级结构预测,结果如图 2 所示:AA-35 序列中含 α -螺旋 41.13%、无规卷曲 30.08%、延长链 17.99% 及 β -折叠 10.80%。AA-36 序列中含 α -螺旋 40.87%、无规卷曲 29.56%、延长链 18.77% 及 β -折叠 10.80%。且 AA-35 序列与 AA-36 序列的二级结构仅在 193 位点处存在无规卷曲/延长链、225 位点处存在无规卷曲/延长链及 237 位点处存在 α -螺旋/延长链的差异,这与氨基酸序列差异较一致。应用 PROSITE 对 AA-35 及 AA-36 进行活性位点

预测,结果如图 3 所示:AA-35 与 AA-36 的活性位点均为 164 位点,活性区域处于 156 位点至 172 位点间。

8 黄酮高、低含量组特有 CHS 氨基酸序列类型 AA-35 与 AA-36 的三级结构比对分析

应用在线软件 Swiss-model 同源建模的 AA-35 与 AA-36 的三级结构如图 4 所示。SAVES 对 AA-35、AA-36 同源建模结果的分析如图 5 所示。AA-35 三维模型的拉氏图的最适残基量为 91.52%,ERRAT 值为 97.900;AA-36 三维模型的拉氏图的最适残基量为 92.54%,ERRAT 值为 97.375,均显示建模结构具有良好立体化学参数和空间结构,可用于后续分析。Discovery Studio 2.5 结合位点预测结果如图 6 所示。结果显示仅 AA-35 的 229 位缬氨酸位于结合位点区域,能够与丙二酰辅酶 A 的结合,而 AA-35 的 193 位异亮氨酸、AA-36 的 193 位缬氨酸、229 位苏氨酸均与底物结合无关,这也可以解释为什么 AA-35 对应样品的黄酮水平较高。

9 CHS 氨基酸序列类型聚类分析

利用 MEGA 5.0 将主流 CHS 氨基酸序列类型 AA-3、黄酮高含量组样品特有氨基酸序列类型 AA-35、黄酮低含量组样品特有氨基酸序列类型 AA-36 与 NCBI 数据库中其他物种的 CHS 氨基酸序列进行聚类分析。18 条氨基酸序列聚类结果如图 7 所示。所有序列分

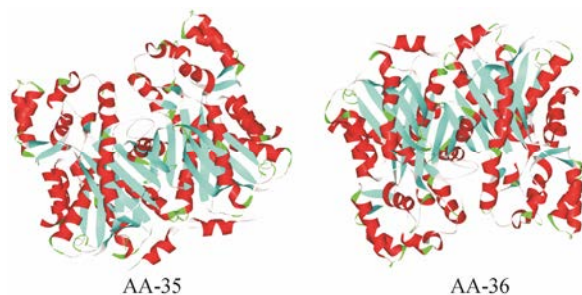


Figure 4 The homology modeling results of AA-35 and AA-36

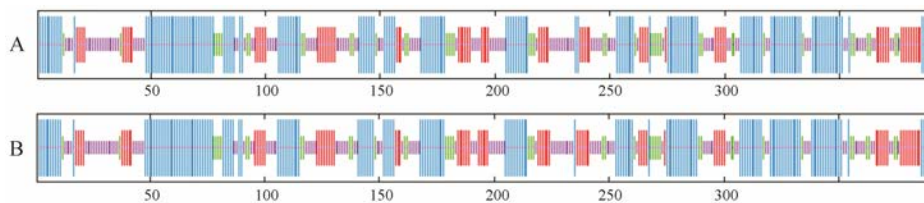


Figure 2 The secondary structure prediction results of AA-35 and AA-36. A: The secondary structure prediction results of AA-35; B: The secondary structure prediction results of AA-36; Blue: Alpha helix; Purple: Random coil; Red: Extended strand; Green: Beta sheet

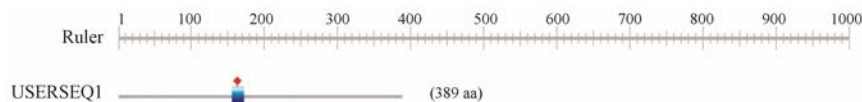


Figure 3 The active site prediction results of AA-35 and AA-36

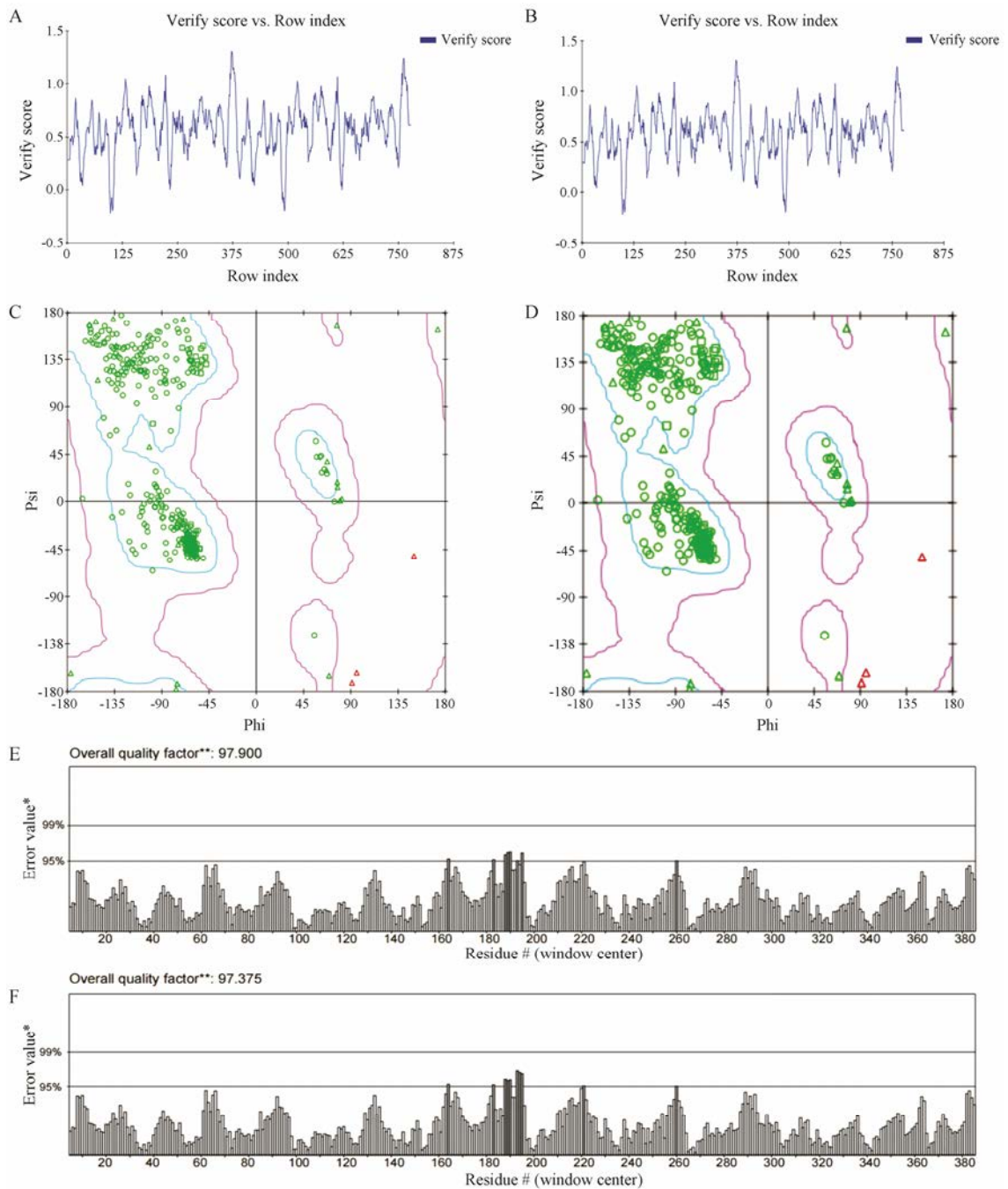


Figure 5 The verify score, Ramachandran plot and ERRAT of AA-35 and AA-36. A: The verify score of AA-35; B: The verify score of AA-36; C: The Ramachandran plot of AA-35; D: The Ramachandran plot of AA-36; E: The ERRAT of AA-35; F: The ERRAT of AA-36

为两大支, 植物聚为一大支, 微生物根瘤菌 *Rhizobium favelukesii* 与芽孢杆菌 *Bacillus sporothermodurans* 单独聚为一支。在植物中, 苔藓类植物小立碗藓 *Physcomitrella patens* 单独聚为一支, 与甘草进化距离最远。豆科植物鹰嘴豆 *Cicer arietinum*、大豆 *Glycine max*、绿豆 *Vigna radiata* 与甘草聚为一支, 且鹰嘴豆与甘草进化关系最近。其余植物依亲缘关系远近, 分别聚为不同小支。甘草的主流 CHS 氨基酸序列类型

AA-3 与黄酮低含量组特有氨基酸序列类型 AA-36 先聚为一支, 再与黄酮高含量组特有氨基酸序列类型 AA-35 聚为一支, 这从进化关系层面上也可推测导致黄酮高含量组样品黄酮大量累积的原因。

讨论

2015 版《中国药典》明确规定栽培甘草的质量标准为甘草酸 $\geq 2.0\%$ 、甘草苷 $\geq 0.5\%$, 即: 三萜类化

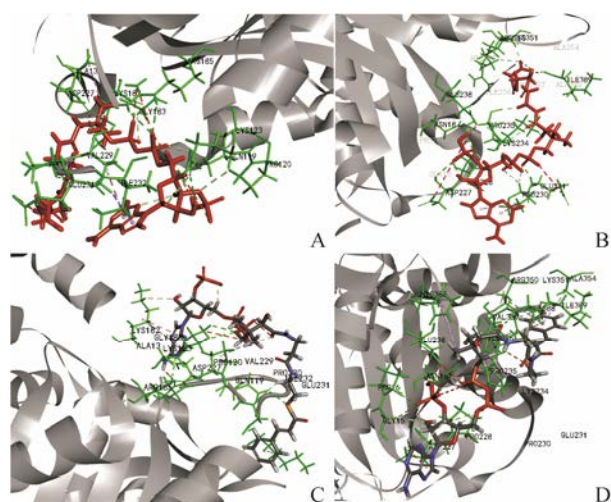


Figure 6 The binding sites prediction results between malonyl-CoA, 4-coumaroyl-CoA and the tertiary structure of AA-35, AA-36. A, B: The binding sites prediction results between malonyl-CoA and the tertiary structure of AA-35, AA-36; C, D: The binding sites prediction results between 4-coumaroyl-CoA and the tertiary structure of AA-35, AA-36

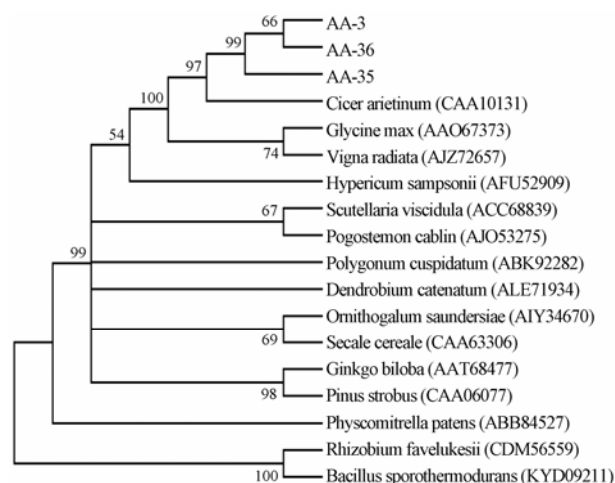


Figure 7 NJ tree based on CHS amino acid sequences from *G. uralensis* Fisch., *G. glabra* L., *G. inflata* Bat. and 15 registered CHS amino acid sequences in Gen Bank. Bootstrap repeated 1 000 times and branch values lower than 50% were hid

合物甘草酸不再作为评价甘草质量的唯一标准，而黄酮类化合物甘草苷作为评价甘草的指标之一，同样具有重要的地位。目前，甘草中发现的黄酮类化合物有 300 多种，结构类型几乎包括了除花青素外的所有黄酮小类^[20]。其中最具代表性的结构是包括甘草素、甘草苷在内的二氢黄酮类及包括异甘草苷、异甘草素在内的查尔酮类。对质量控制成分甘草苷的研究多集中于药理活性^[2, 3]方面，而对其生物合成途径则关注较少。因此，本文从甘草苷生物合成途径中的第一个关键功能基因 *CHS* 入手展开研究。

CHS 为超基因家族成员之一，在不同植物间，

功能及拷贝数均存在差异^[21]。其主要功能与植物花色素积累^[22, 23]、植物的育性^[24]及防御相关^[25, 26]。Courtney-Gutterson 等^[27]利用根癌农杆菌介导法，向花色为粉红色的菊花 *Chrysanthemum morifolium* 中，分别以正义和反义方向转入 *CHS* 基因，获得了开浅红色和白色花的转基因植株。Pollak 等^[28]研究发现，缺乏 *CHS* 的矮牵牛突变植株体内无黄酮类物质积累，且属于雄性不育，可通过补充黄酮醇类物质，使花粉管恢复正常生长。Qi 等^[29]对水稻中与白叶枯病菌互动中的防卫基因 *CHS* 的转录进行研究，发现病菌均不诱导或只是微弱诱导水稻 *CHS* 基因的转录。

在药用植物中，*CHS* 基因的研究多集中于克隆及表达载体构建^[30–32]，红花 *Carthamus tinctorius* L.^[33]、黄芩 *Scutellaria baicalensis* Georgi^[34]、铁皮石斛 *Dendrobium officinale*^[35]等药用植物的 *CHS* 基因已被成功克隆。Liu 等^[33]发现将红花中 *CHS* 基因在拟南芥 *Arabidopsis thaliana* L.中进行过表达，可以导致拟南芥中黄酮含量的提高。Liu 等^[34]成功克隆黄芩 *CHS* 基因，并将其导入根癌农杆菌，获得黄芩 *CHS* 基因高效表达载体，促进了黄芩苷的合成。Meng 等^[35]对 1~3 年生铁皮石斛的根、茎、叶中 *CHS* 基因表达量进行分析，发现 *CHS* 基因在 1 年生铁皮石斛叶片中的表达量最高，且随着植株年龄的增长，叶片中 *CHS* 基因的表达量降低、茎中的表达量升高。因此在药用植物中 *CHS* 基因的功能与药效成分的累积具有明确的相关性。就甘草而言，其 *CHS* 基因的报道仅限于基因克隆及乌拉尔甘草的基因组测序^[36]，这些研究为甘草基因表达谱的建立、基因库的完善、功能基因的挖掘及功能注释奠定了基础，但目前甘草中功能基因与有效成分之间的关联分析尚缺乏系统的研究。

本文利用 HPLC 法对 60 份甘草样品中甘草苷、异甘草苷、甘草素、异甘草素进行含量分析，结果表明乌拉尔甘草黄酮类化合物含量最高，光果甘草次之，胀果甘草最低，这与课题组前期的研究结果一致^[9]。不仅提示在临床用药中，乌拉尔甘草有利于发挥黄酮类化合物独特的药理活性，如：抗氧化^[37]、抗抑郁症^[5]、细胞保护作用^[38]等，更表明乌拉尔甘草中黄酮类化合物生物合成的高效性，而导致此高效性的分子机制则有可能与功能基因的多态性、表达量、拷贝数、转录因子调控、外界环境因子刺激等多种因素相关，本文试图从功能基因的多态性方面寻找原因，而其他方面的调控则亟待进一步研究。

本文的研究结果表明甘草 *CHS* 基因具有丰富的基因多态性，筛选出黄酮高/低含量甘草样品的特异

CHS 基因型及氨基酸序列类型, 并从生物信息学角度对其产生原因进行了分析。发现黄酮高含量组特有主流氨基酸类型 AA-35 的 229 位缬氨酸位于结合位点区域, 能够与丙二酰辅酶 A 结合, 而黄酮低含量组特有的主流氨基酸类型 AA-36 的 229 位苏氨酸则与底物结合无关, 因此推测这一位点的突变有可能与黄酮的高水平积累相关, 但尚需进一步实验对其进行评价。本文可为甘草中其他功能基因的多态性研究提供参考, 为揭示功能基因与次生代谢之间的相关性奠定基础, 并为甘草中优良基因筛选及分子育种提供理论支持。

References

- [1] Sen L. *Shennong's Herbal* (神农本草经) [M]. Shanghai: Shanghai Technology and Health Press, 1959: 35.
- [2] Cui Y, Ao M, Li W, et al. Anti-inflammatory activity of licochalcone A isolated from *Glycyrrhiza inflata* [J]. *Z Naturforsch B*, 2008, 63: 361–365.
- [3] Zhou Y, Ho WS. Combination of liquiritin, isoliquiritin and isoliquiritigenin induce apoptotic cell death through upregulating p53 and p21 in the A549 non-small cell lung cancer cells [J]. *Oncol Rep*, 2014, 31: 298–304.
- [4] Ai-Turki AI, Ei-Ziney MG, Abdel-Salam AM. Chemical and anti-bacterial characterization of aqueous extracts of oregano, marjoram, sage and licorice and their application in milk and labneh [J]. *Int J Food Agric Environ*, 2008, 6: 39–44.
- [5] Zhao Z, Wang W, Guo H, et al. Antidepressant-like effect of liquiritin from *Glycyrrhiza uralensis* in chronic variable stress induced depression model rats [J]. *Behav Brain Res*, 2008, 194: 108–113.
- [6] Abe K, Ikeda T, Wake K, et al. Glycyrrhizin prevents of lipopolysaccharide/*D*-galactosamine-induced liver injury through down-regulation of matrix metalloproteinase in mice [J]. *J Pharm Pharmacol*, 2008, 60: 91–97.
- [7] Huang MJ, Wang WQ, Wei SL, et al. Investigation on medicinal plant resources of *Glycyrrhiza uralensis* in China and chemical assessment of its underground part [J]. *China J Chin Mater Med* (中国中药杂志), 2010, 35: 947–951.
- [8] Chinese Pharmacopoeia Commission. *Pharmacopoeia of the People's Republic of China. Part 1* (中华人民共和国药典一部) [S]. Beijing: China Medical Science Press, 2015: 87.
- [9] Yang R, Li WD, Ma YS, et al. The molecular identification of licorice species and the quality evaluation of licorice slices [J]. *Acta Pharm Sin* (药学学报), 2017, 52: 318–326.
- [10] Yang R, Yuan BC, Ma YS, et al. Simultaneous determination of liquiritin, isoliquiritin, liquiritigenin and liquiritigenin in *Glycyrrhiza uralensis* Fisch., *Glycyrrhiza glabra* L., and *Glycyrrhiza inflata* Bat. by HPLC [J]. *Chin J Pharm Anal* (药物分析杂志), 2016, 36: 1729–1736.
- [11] Liu Y, Zhan XJ, Li WD, et al. Copy number variations of functional genes influence contents of glycyrrhizic acid in *Glycyrrhiza uralensis* [J]. *Acta Physiol Plant*, 2014, 36: 1433–1440.
- [12] Liu Y, Liu DJ, Liu CS, et al. Mechanism of genuineness of liquorice *Glycyrrhiza uralensis* based on CNVs of *HMGR*, *SQS1* and β -*AS* gene [J]. *Acta Pharm Sin* (药学学报), 2012, 47: 250–255.
- [13] Liu Y, Xu QX, Wang XY, et al. Analysis on correlation between 3-hydroxy-3-methylglutaryl-coenzyme A reductase gene polymorphism of *Glycyrrhiza uralensis* and content of glycyrrhizic acid [J]. *China J Chin Mater Med* (中国中药杂志), 2012, 37: 3789–3792.
- [14] Liu Y, Zhang N, Wang XY, et al. Researches on influence of squalene synthase gene polymorphism on catalytic efficiency of its encode enzyme in *Glycyrrhiza uralensis* [J]. *China J Chin Mater Med* (中国中药杂志), 2012, 37: 3777–3783.
- [15] Liu Y, Zhu XQ, Li WD, et al. Enhancing production of ergosterol in *Pichia pastoris* GS115 by over-expression of 3-hydroxy-3-methylglutaryl CoA reductase from *Glycyrrhiza uralensis* [J]. *Acta Pharm Sin B*, 2014, 4: 161–166.
- [16] Koes RE, Francesca Q, Joseph NM. The flavonoid biosynthetic pathway in plants: function and evolution [J]. *Bioessays*, 1994, 16: 123–132.
- [17] Reimold U, Kröger M, Kreuzaler F, et al. Coding and 3' non-coding nucleotide sequence of chalcone synthase mRNA and assignment of amino acid sequence of the enzyme [J]. *Embo J*, 1983, 2: 1801–1805.
- [18] Xu HF. Cell Culture And Cloning of Chalcone Synthase Gene from *Glycyrrhiza inflata* Bat. (胀果甘草细胞培养及查尔酮合成酶基因的 cDNA 克隆) [D]. Baoding: Hebei University, 2011.
- [19] Zhou S, Ma YS, Hu T, et al. Cloning and sequence analysis of chalcone synthase gene from *Glycyrrhiza uralensis* Fisch. [J]. *Lett in Biotech* (生物技术通讯), 2017, 28: 812–817.
- [20] Wang X, Zhang H, Chen L, et al. Licorice, a unique “guide drug” of traditional Chinese medicine: a review of its role in drug interactions [J]. *J Ethnopharmacol*, 2013, 150: 781–790.
- [21] Durbin ML, Mccraig B, Clegg MT. Molecular evolution of the chalcone synthase multigene family in the morning glory genome [J]. *Plant Mol Biol*, 2000, 42: 79–92.
- [22] Kamiishi Y, Otani M, Takagi H, et al. Flower color alteration in the liliaceous ornamental *Tricyrtis* sp. by RNA interference-

- mediated suppression of the chalcone synthase gene [J]. Mol Breeding, 2012, 30:671–680.
- [23] Blokland RV, Geest NVD, Mol JNM, et al. Transgene-mediated suppression of chalcone synthase expression in *Petunia hybrida*, results from an increase in RNA turnover [J]. Plant J, 2010, 6: 861–877.
- [24] Taylor LP, Jorgensen R. Conditional male fertility in chalcone synthase - deficient *Petunia* [J]. J Hered, 1991, 83: 11–17.
- [25] Dao TTH, Linthorst HJM, Verpoorte R. Chalcone synthase and its functions in plant resistance [J]. Phytochem Rev, 2011, 10: 397–412.
- [26] Cushnie TP, Lamb AJ. Antimicrobial activity of flavonoids [J]. Int J Antimicrob Agents, 2005, 26: 243–356.
- [27] Courtney-Gutterson N, Napoli C, Lemieux C, et al. Modification of flower color in florist's chrysanthemum: production of a white-flowering variety through molecular genetics [J]. Biotechnology (NY), 1994, 12: 268–271.
- [28] Pollak PE, Vogt T, Mo Y, et al. Chalcone synthase and flavonol accumulation in stigmas and anthers of *Petunia hybrida* [J]. Plant Physiol, 1993, 102: 925–932.
- [29] Qi FJ, Gao XW, Wang JS, et al. Accumulation of *pal* and *chs* gene transcripts of rice carrying different bacterial blight resistance genes interacted with *Xanthomonas oryzae* pv. *Oryzae* [J]. J Agric Biotechnol (农业生物技术学报), 2000, 8: 337–340.
- [30] Rao C, Lei W, Li P, et al. Cloning of *Scutellaria viscidula* *CHS* gene and construction of its sense and antisense plant expression vectors [J]. J Chin Med Mater (中药材), 2009, 32: 1661–1664.
- [31] Shang J, Chen AL, Pan WZ. Cloning and bioinformation analysis of chalcone synthase gene from *Lycium barbarum* in Ningxia [J]. Chin Tradit Herb Drugs (中草药), 2013, 44: 2898–2903.
- [32] Fu M, Wei L, Yu J, et al. cDNA cloning and protein sequence analysis of chalcone synthase gene in leaves of *Ampelopsis grossdentata* [J]. Chin Tradit Herb Drugs (中草药), 2013, 44: 85–89.
- [33] Liu XM, Lu WH, Li J, et al. Over-expression of chalcone synthase of safflower increases flavonoid content in *Arabidopsis thaliana* [J]. Chin J Cell Biol (中国细胞生物学学报), 2017, 39: 182–190.
- [34] Liu GM, Lei W, Shui XR. Molecular clone, sequence analysis and expression vector construction with high efficiency of *CHS* gene from *Scutellaria baicalensis* Georgi [J]. Genomics Appl Biol (基因组学与应用生物学), 2016, 35: 2814–2818.
- [35] Meng HL, Zhang W, Liu BY, et al. Cloning and expression of chalcone synthase gene from *Dendrobium officinale* [J]. J South Agric (南方农业学报), 2016, 47: 2015–2019.
- [36] Mochida K, Sakurai K, Seki H, et al. Draft genome assembly and annotation of *Glycyrrhiza uralensis*, a medical legume [J]. Plant J, 2017, 89: 181–195.
- [37] Cheel J, Antwerpen PV, Tumboval L, et al. Free radical-scavenging, antioxidant and immunostimulating effects of a licorice infusion (*Glycyrrhiza glabra* L.) [J]. Food Chem, 2010, 122: 508–517.
- [38] Yang Y, Bian GX, Lv QJ. Neuroprotection and neurotrophism effects of liquiritin on primary cultured hippocampal cells [J]. China J Chin Mater Med (中国中药杂志), 2008, 33: 931–935.