

# 基于选择清除分析挖掘连翘异型花柱候选基因

杜晓蓉<sup>1</sup>, 李琦<sup>1</sup>, 李澳旋<sup>1</sup>, 樊阿梅<sup>1</sup>, 何龙娇<sup>1</sup>, 靳亚鑫<sup>1</sup>, 张灵灵<sup>1</sup>, 乔永刚<sup>1,2</sup>, 宋芸<sup>1,2</sup>

(<sup>1</sup>山西农业大学生命科学学院, 太谷 030801; <sup>2</sup>山西农业大学中兽医现代化山西省重点实验室, 太谷 030801)

**摘要:** 连翘(*Forsythia suspensa* (Thunb.) Vahl)为多年生木本植物,花柱二型具有自交不亲和性。挖掘连翘异型花柱发育候选基因,利于揭示连翘异型花柱的进化与发育机理。对40株长花柱型与短花柱型连翘植株进行全基因组重测序,检测样本基因组的SNP变异位点,开展选择清除分析。利用遗传分化系数Fst和核苷酸多态性 $\pi$  ratio结合的方法筛选候选区域(前5%水平),对差异基因进行GO数据库和KEGG数据库功能注释分析,采用qRT-PCR对候选基因在花柱发育不同时期的表达量进行检测。选择清除分析显示受选择基因295个,主要在植物的生长发育以及代谢调控方面发挥重要作用。分析植物激素调节相关的部分通路,确定连翘异型花柱候选基因CYP734A1(EVM0010386)、BRI1(EVM0011829)、CYCD3(EVM0018316),基因表达结果显示CYP734A1基因在连翘短花柱发育的关键时期持续高表达,露冠期BRI1与CYCD3基因表达水平在短花柱型中显著高于长花柱型。本研究挖掘的候选基因在连翘长花柱与短花柱中存在差异表达,油菜素甾醇参与了连翘异型花柱的发育,是产生连翘花柱异型的原因之一。本研究为连翘异型花柱的进化以及遗传发育调控机制解析奠定基础并提供思路。

**关键词:** 连翘;异型花柱;全基因组重测序;选择清除分析;表达分析

## Based on Selective Sweep Analysis Mining Candidate Genes of Heterostyly in *Forsythia suspensa*

DU Xiaorong<sup>1</sup>, LI Qi<sup>1</sup>, LI Aoxuan<sup>1</sup>, FAN Aimei<sup>1</sup>, HE Longjiao<sup>1</sup>, JIN Yaxin<sup>1</sup>,

ZHANG Lingling<sup>1</sup>, QIAO Yonggang<sup>1,2</sup>, SONG Yun<sup>1,2</sup>

(<sup>1</sup>College of Life Sciences, Shanxi Agricultural University, Taigu 030801; <sup>2</sup>Shanxi Key Laboratory of Modernization of Traditional Chinese Veterinary Medicine, Shanxi Agricultural University, Taigu 030801)

**Abstract:** *Forsythia suspensa* (Thunb.) Vahl is a perennial woody plant with dimorphic style and self-incompatibility. Mining the candidate genes for the development of heterostyly in *Forsythia suspensa* is of significance to reveal its evolution and development mechanism. The whole genome resequencing of 40 long style morph and short style morph *Forsythia suspensa* plants were conducted, followed by the SNP variation sites detection in the sample genome and the selection sweep analysis. The candidate regions (top 5% level) were screened by the combination of genetic differentiation coefficient Fst and the nucleotide polymorphism  $\pi$  ratio. Differentially expressed genes (DEGs) were analyzed by GO and KEGG enrichment annotation, and qRT-PCR was used to detect the transcriptional profiles of candidate genes at different periods of style development. Selective sweep analysis showed that 295 genes were under selection, which mainly played an important role in plant growth and development and metabolic regulation. By analyzing the pathways in plant hormone regulation, the candidate genes CYP734A1 (EVM0010386), BRI1 (EVM0011829) and CYCD3 (EVM0018316) of

收稿日期: 2023-09-07 修回日期: 2023-10-14 网络出版日期: 2023-11-16

URL: <https://doi.org/10.13430/j.cnki.jpgr.20230907001>

第一作者研究方向为药用植物遗传学, E-mail: duxiaorong1998@163.com

通信作者: 宋芸, 研究方向为药用植物逆境生理与药用植物遗传学, E-mail: songyun0283@126.com

**基金项目:** 山西省基础研究计划自然科学研究面上项目(20210302123390); 山西省现代农业中药材产业技术体系项目(2023); 山西农业大学科技创新提升工程项目(CXGC2023041); 山西省高等学校大学生创新创业训练计划项目(20220172)

**Foundation projects:** Basic Research Plan of Shanxi Province Natural Science Research Project (20210302123390); Shanxi Province Modern Agricultural Chinese Herbal Medicine Industry Technology System Project (2023); Shanxi Agricultural University Science and Technology Innovation Promotion Project (CXGC2023041); Innovation and Entrepreneurship Training Program for College Students in Shanxi Province (20220172)

heterostyly in *Forsythia suspensa* were analyzed. The gene expression results indicated that the *CYP734A1* gene has sustained high expression during the critical period of development in *F. suspensa*'s S-morph flowers. The *BR11* and *CYCD3* genes in the S-morph were significantly higher than those in the L-morph during the crown exposure period. The candidate genes excavated in this study were differentially expressed in the L-morph and S-morph style of *F. suspensa*. Brassinosteroids was involved in the development of the heterostyly in *F. suspensa*, which was one of the reasons for heterostyly of *F. suspensa*. Collectively, this study laid a foundation for the evolution and genetic development regulation mechanism of heterostyly in *F. suspensa*.

**Key words:** *Forsythia suspensa* (Thunb.) Vahl; heterostyly; whole genome resequencing; selection sweep analysis; expression analysis

连翘(*Forsythia suspensa* (Thunb.) Vahl),又名黄花杆、黄寿丹,为木犀科连翘属植物,具有清热解毒,消肿散结,疏散风热的功效<sup>[1]</sup>,常用于治疗风热感冒、乳痈、瘰疬<sup>[2]</sup>。现代药理研究也表明连翘在抗炎、抗菌、抗病毒和调节免疫功能等方面作用突出<sup>[3]</sup>。连翘主要分布于我国山西、河北、河南等地区,以果实入药,连翘酯苷为主要药用成分<sup>[4]</sup>,以连翘为原料制成的药品多达几十种,如莲花清瘟胶囊<sup>[5]</sup>、双黄连<sup>[6]</sup>、银翘解毒丸<sup>[7]</sup>等,在防治非典、禽流感、新冠病毒<sup>[8]</sup>等方面功效突出。连翘不仅药用价值高,在食用、观赏等方面也具有重要作用。连翘在长期进化中形成异型花柱,具有自交不亲和性,自然界野生资源分布不均匀,影响授粉进而使其结实率和果实产量受到较大影响,关于连翘花柱的进化与发育机制目前研究较少。

异型花柱是被子植物中一种特殊的花多态现象和雌雄异位形式,在避免自交、促进异交传粉方面具有重要的进化意义,可提高亲本适合度<sup>[9]</sup>。异型花柱的相关研究表明花的形态发育与植物激素调控有较大联系,其中对油菜素类固醇(BR, brassinosteroid)的研究最多。油菜素类固醇作为一种甾醇类植物激素,对植物生长发育及生理过程具有调控作用,包括细胞分裂和伸长、光形态建成、生物胁迫等<sup>[10]</sup>。报春花属作为异型花柱研究的模式物种,遗传调控机制与油菜素类固醇密切相关,Huu等<sup>[11]</sup>研究表明编码降解油菜素类固醇的酶CYP734A50仅仅存在于报春花的短花柱中;Matzke等<sup>[12-13]</sup>研究发现只有一个基因*TsBAHD*在时钟花雌蕊中表达,通过在拟南芥中过表达该基因证实引起的矮化表型与油菜素类固醇失活的表型一致。根据之前的研究,连翘异型花柱与油菜素类固醇的关系得到进一步的探索,李政<sup>[14]</sup>对连翘花柱中油菜素类固醇含量进行测定,发现不同花柱类型的栗甾酮(CS, castasterone)含量存在明显差异。栗甾酮是油菜素类固醇合成的前体,对其

发挥功能具有重要影响,这可能是连翘花柱异型的关键因素,目前连翘异型花柱的形成机制与植物激素间的调控关系仍未可知。

新一代测序技术(NGS, next generation sequencing)因其成本较低在作物育种和遗传进化等方面得到广泛应用,其中全基因组重测序技术(WGRS, whole genome resequencing study)是常用的技术手段之一,对已知基因组序列的物种进行不同个体的基因组测序,并在此基础上对个体或群体进行差异性分析,获得单核苷酸多态性(SNP, single nucleotide polymorphisms)位点。SNP作为良好的遗传分子标记,在物种遗传多样性分析、全基因组关联分析方面应用广泛<sup>[15]</sup>。孟歌等<sup>[16]</sup>为了鉴定与桃花型关联的SNP位点并挖掘候选基因,采用重测序的方法,结合全基因组关联分析筛选到关键位点,用于桃品种的分子标记辅助育种。选择清除分析是在全基因组水平分析得到的SNP位点基础上进行<sup>[17]</sup>,基因组水平的某些位点受到选择,在人工选择和对环境的适应中改变了原有的遗传多样性,发生有益突变的基因被保留下来,未被选择的个体经过若干代后逐渐被淘汰<sup>[18-20]</sup>。利用SNP位点进行选择清除分析可以得到被测个体的受选择基因及其功能,为其进化与遗传发育提供帮助。

本研究选择连翘长花柱型与短花柱型植株的叶片混样进行全基因组重测序,采用遗传分化系数Fst<sup>[21]</sup>和核苷酸多态性 $\pi$  ratio<sup>[22]</sup>结合的方法筛选选择清除区域,旨在挖掘连翘在长期进化过程中形成异型花柱的差异基因,对候选基因进行功能注释与表达量水平测定,从而为连翘的异型花柱进化与发育调控提供理论基础。

## 1 材料与方法

### 1.1 试验材料

试验所选材料长花柱型(L)与短花柱型(S)连

翘植株均来自山西农业大学太谷校区植物园, 不同类型的连翘各选 20 株采集新鲜叶片, 每组设置 3 个重复, 酒精擦拭叶片消毒后暂存于  $-80\text{ }^{\circ}\text{C}$  冰箱。材料送往北京百迈客生物科技有限公司进行连翘全基因组重测序。

## 1.2 全基因组重测序

样品基因组 DNA 检测合格后, 用机械打断的方法(超声波)将 DNA 片段化, 然后对片段化的 DNA 进行片段纯化、末端修复、3' 端加 A、连接测序接头, 用琼脂糖凝胶电泳进行片段大小选择, PCR 扩增形成测序文库, 建好的文库先进行文库质检, 质检合格的文库用 Illumina 进行测序。

对测序得到的原始 Reads 进行质量评估并过滤得到 Clean Reads, 用于后续生物信息学的分析。将 Clean Reads 与参考基因组<sup>[23]</sup>序列进行比对, 基于比对结果进行 SNP 变异检测和注释, 并实现 DNA 水平差异基因挖掘和差异基因功能注释等。

## 1.3 SNP 检测与注释

SNP 的检测主要使用 GATK<sup>[24]</sup> 软件工具包实现。根据 Clean Reads 在参考基因组的定位结果, 使用 Picard (<https://sourceforge.net/projects/picard/>) 过滤冗余 Reads (MarkDuplicates), 以保证检测结果的准确性。使用 GATK 的 HaplotypeCaller (局部单体型组装) 算法进行 SNP 的变异检测, 采用 SnpEff<sup>[25]</sup> 软件注释变异 (SNP) 和预测变异影响。根据变异位点在参考基因组上的位置以及参考基因组上的基因位置信息, 可以得到变异位点在基因组的发生区域 (基因间区、基因区或 CDS 区等), 以及变异类型 (同义、非同义突变等)。

表 1 目的基因引物信息

Table 1 Primer information of target genes

基因名称 Gene name	引物名称 Primer name	引物序列 (5'-3') Primer sequence (5'-3')	产物长度 (bp) Product length
<i>HIS</i>	<i>HIS-F</i>	ACAAAGGCGGCGAGAAA	114
	<i>HIS-R</i>	CTCTGTGCTCTTCTGGTACTTC	
<i>CYP734A1</i>	<i>EVM0010386-F</i>	CCACCAGCAGTAGCCATAAT	101
	<i>EVM0010386-R</i>	ACGGCTAGGATCGGTATTAGA	
<i>BRI1</i>	<i>EVM0011829-F</i>	CTCTTGATCTCGCCGGTAATAG	119
	<i>EVM0011829-R</i>	AGCCGGAATTCACCAGTAA	
<i>CYCD3</i>	<i>EVM0018316-F</i>	GGTTCGTCCACCCTGAAT	97
	<i>EVM0018316-R</i>	TGATAGGCGGTTCTTCAATCC	

## 1.4 选择信号检测

使用遗传分化系数  $F_{st}$  和核苷酸多态性  $\pi$  ratio 方法进行选择清除分析。选择显著高的  $F_{st}$  值 (前 5% 的部分, 即  $F_{st} > 0.095$ ) 和具有极端低或高  $\pi$  ratio ( $\pi_L/\pi_S$ ) (5% 左尾和 95% 右尾, 即  $\pi$  ratio  $\leq 0.78$  和  $\pi$  ratio  $\geq 1.29$ ) 作为基因组具有强选择性扫描信号区域的筛选标准。用于参数计算的窗口大小为 100 kb, 步长 10 kb。

## 1.5 GO 注释和 KEGG 富集分析

针对连翘不同花柱类型筛选到的强选择区域, 进行 GO (gene ontology) 注释和相应的 GO 富集分析。同时进行了 KO (KO, KEGG ortholog and pathway) 注释, KO 被映射到相应的 KEGG 通路途径, 从而得到被选择基因对应的显著功能富集信息。

## 1.6 差异基因表达分析

**1.6.1 RNA 提取及 cDNA 合成** 使用华越洋 RNA 提取试剂盒对连翘花柱发育的 3 个关键时期 (鳞片脱落期、现蕾期、露冠期)<sup>[26]</sup> 进行花柱 RNA 的提取, 保存于  $-80\text{ }^{\circ}\text{C}$  冰箱。使用 ToloScript All-in-one RT EasyMix for qPCR 试剂盒反转录得到 cDNA, 反应体系如下:  $5\times$  All-in-one RT Buffer 4  $\mu\text{L}$ , All-in-one Enzyme Mix 1  $\mu\text{L}$ , 总 RNA 1  $\mu\text{g}$ , RNase-free ddH<sub>2</sub>O 补足至 20  $\mu\text{L}$ 。冰上融化各试剂组分后充分混匀, RNA 稀释至统一浓度按体系加入后进行如下反应程序:  $50\text{ }^{\circ}\text{C}$  15 min;  $85\text{ }^{\circ}\text{C}$  5 s。反转录产物于  $-20\text{ }^{\circ}\text{C}$  保存备用。

**1.6.2 引物合成** 筛选得到的目的基因序列经公司提取, 利用 Primer 5.0 设计 qRT-PCR 引物, 引物序列见表 1, 试验所用内参基因为 *HIS*<sup>[27]</sup>, 引物由生物工程 (上海) 股份有限公司进行合成。

**1.6.3 qRT-PCR检测候选基因表达水平** 反转录后的连翘花柱 cDNA 统一稀释 10 倍, 使用 TOLOBIO 荧光定量试剂盒 2×Q3 SYBR qPCR Master Mix 对候选基因表达量进行检测。荧光定量反应体系为 2×Q3 SYBR qPCR Master Mix 5 μL, 正向引物和反向引物 (10 μmol/L) 各 0.2 μL, cDNA 1 μL, ddH<sub>2</sub>O 补足至 10 μL。反应程序为预变性 95 °C 3 min; 循环反应 95 °C 10 s, 60 °C 30 s, 72 °C 30 s 进行 40 次; 熔解曲线使用仪器 (Bio-Rad CFX-96) 默认程序。试验数据根据 2<sup>-ΔΔCt</sup> 方法计算基因相对表达水平, 使用 SPSS Statistics 和 Origin 9 进行显著性分析和柱状图展示。

## 2 结果与分析

### 2.1 连翘全基因组重测序数据

对测序得到的原始序列进行过滤获得 Clean Reads, 用于后续信息分析。连翘不同分组植株的测序数据评估 (表 2) 显示 Q20 含量均在 97.38% 以上,

Q30 含量均在 92.59% 以上, GC 含量处于 33.80% ~ 34.06% 之间。与参考基因组比对后得到 45.82 Gb 的 Clean Data, 样品与参考基因组平均比率为 83.16%, 平均覆盖深度为 34×, 基因组覆盖度为 94.56%。

### 2.2 SNP 变异位点注释结果

SNP 类型的变异分为转换和颠换两种, 同种类型碱基之间突变称为转换 (Ti, transition), 如嘌呤与嘌呤之间、嘧啶与嘧啶之间的变异, 不同类型碱基之间的突变称为颠换 (Tv, transversion), 如嘌呤与嘧啶之间的变异。通常转换比颠换更容易发生, 故转换/颠换 (Ti/Tv) 的比例一般大于 1, 具体数值和所测物种有关。表 3 所示为检测到的 SNP 变异位点统计结果, 转换/颠换的比例均为 1.95。对 SNP 的注释结果显示, 67.45% SNP 位于基因间区域, 4.3% SNP 位于 CDS 区。位于 CDS 区的非同义编码与同义编码比率均大于 1, 说明连翘花型发育受到正向选择。

表 2 连翘测序数据统计

Table 2 Statistics of *Forsythia suspensa* sequencing data

类型 Type	测序数据质量统计 Statistics of the sequencing data quality					与参考基因组比对统计 Alignment statistics with the reference genome					
	过滤后 reads 数 Clean reads	过滤后 碱基数 Clean bases	Q20 含 量 (%) Q20 content	Q30 含 量 (%) Q30 content	GC 含 量 (%) GC content	比对后 reads 数 Total reads	匹配率 (%) Mapped	平均测序 深度 Average sequencing depth	覆盖度比 例 1X (%) Coverage ratio of 1X	覆盖度比 例 5X (%) Coverage ratio of 5X	覆盖度比 例 10X (%) Coverage ratio of 10X
长花柱型 L	78997461	23670292756	97.55	92.97	33.80	157994922	83.36	35	94.59	91.53	88.79
短花柱型 S	73924156	22150916646	97.38	92.59	34.06	147848312	82.96	33	94.54	91.42	88.58

L: Long style morph; S: Short style morph; The same as below

表 3 连翘 SNP 变异结果统计

Table 3 Statistics of SNP variation results in *Forsythia suspensa*

类型 Type	SNP 数目 SNP number	转换 Transition	颠换 Transversion	转换/颠换 Ti/Tv	非同义编码 Nonsynonymous coding	同义编码 Synonymous coding	非同义编码/同义编码 比率 Nonsynonymous coding / synonymous coding ratio
长花柱型 L	3461450	2290570	1170880	1.95	78942	62449	1.26
短花柱型 S	3393630	2245497	1148133	1.95	78417	62125	1.26

全基因组 SNP 突变可分成 6 类, 分别是 C:G>A:T; C:G>G:C; C:G>T:A; T:A>A:T; T:A>C:G; T:A>G:C。以 T:A>C:G 为例, 此种类型 SNP 突变包括 T 突变为 C 和 A 突变为 G, 当 T 突变为 C 类型出现在参考基因组正链上, A 突变为 G 类型即在参考

基因组负链的相同位置。如图 1 所示, 连翘全基因组 SNP 突变类型中 C:G>T:A 以及 T:A>C:G 所占比例显著高于其余 4 类 (C:G>A:T, C:G>G:C, T:A>A:T, T:A>G:C), 长花柱型连翘与短花柱型连翘每种类型突变比例相当, 无明显差异。

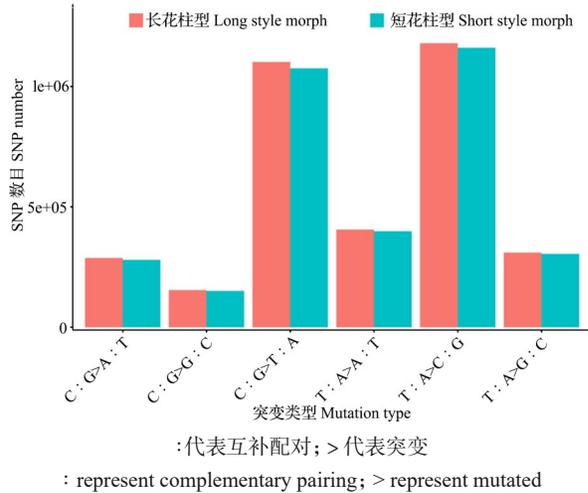


图1 连翘长花柱型与短花柱型SNP变异类型分布图  
Fig.1 Distribution of SNP variation types of long style morph and short style morph in *F. suspensa*

### 2.3 选择清除信号检测

基于重测序数据对SNP变异位点的分析及注释,对不同花柱类型连翘植株进行选择清除分析,遗传分化系数Fst曼哈顿图(图2)显示前5%水平富集显著的20个窗口中包含受选择基因638个,受选择峰值高的在重叠群Contig00086、Contig00315、Contig00521、Contig00612、Contig00646、Contig01032、Contig01113。

核苷酸多态性 $\pi$  ratio主要衡量遗传多样性,通常按照 $\pi_{野生}/\pi_{驯化}$ 的方式计算,受驯化群体的 $\pi$  ratio更低。以 $\pi_L/\pi_S$ 分组计算,图3表明核苷酸多态性 $\pi$  ratio在前5%水平受选择基因为1037个,受选择峰最高的重叠群集中在Contig00313、Contig00315、Contig00517、Contig00522、Contig00612、Contig00824、Contig01032、Contig01113、Contig01160。

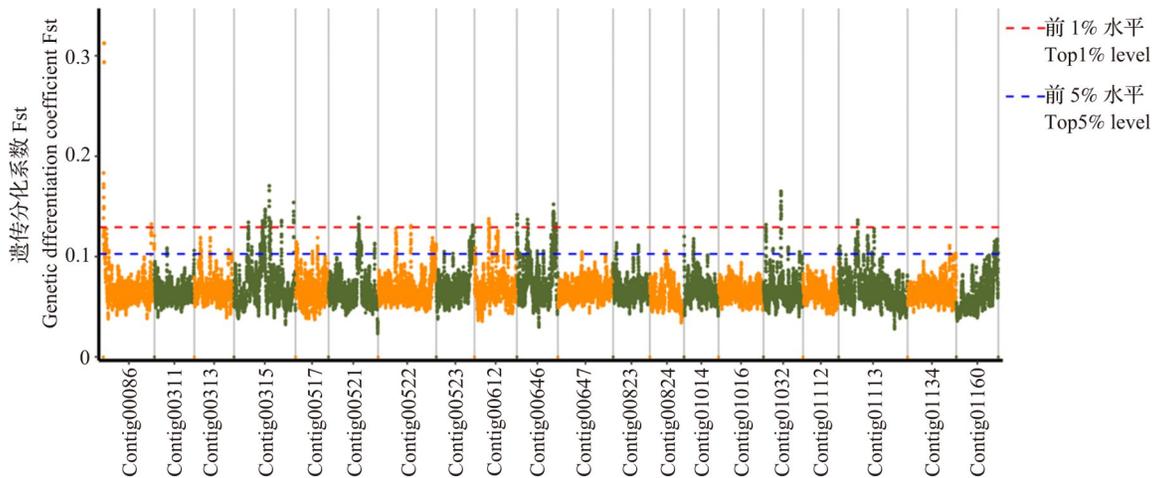


图2 连翘长花柱型与短花柱型遗传分化系数Fst曼哈顿图  
Scatter points represent the values of the corresponding indicators in each window; The blue horizontal line represents the value corresponding to top 5%, and the red horizontal line represents the value corresponding to top 1%; The region above the threshold line is the candidate region; The same below

Fig.2 *F. suspensa* long style morph and short style morph genetic differentiation coefficient Fst Manhattan graph

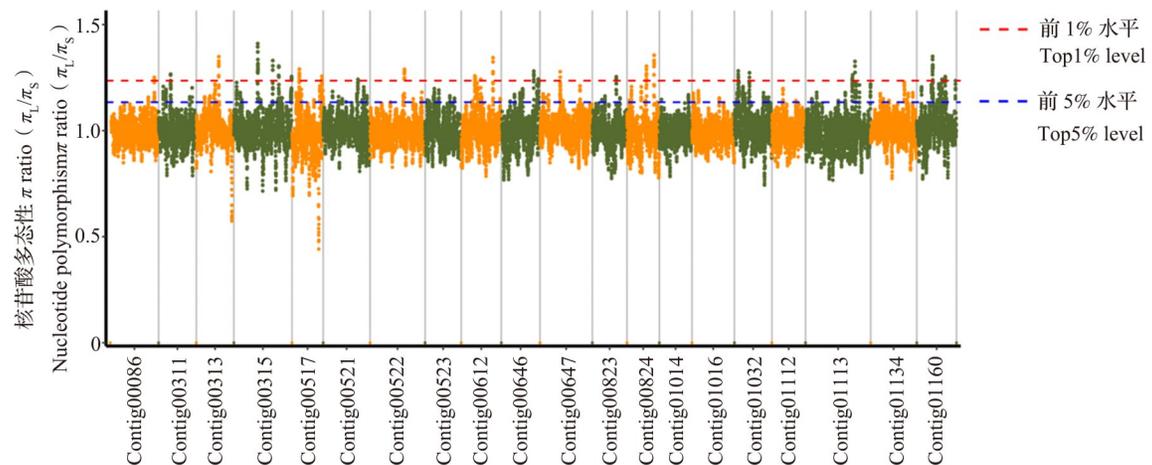


图3 连翘长花柱型与短花柱型核苷酸多态性 $\pi$  ratio( $\pi_L/\pi_S$ )曼哈顿图

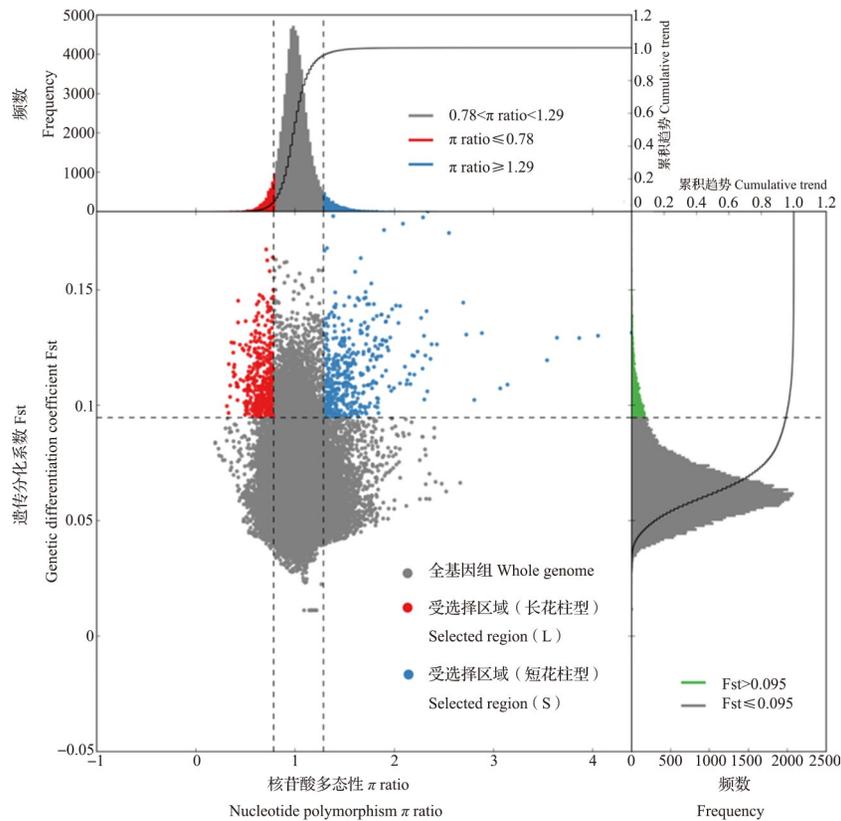
Fig.3 *F. suspensa* long style morph and short style morph nucleotide polymorphism  $\pi$  ratio ( $\pi_L/\pi_S$ ) Manhattan graph

利用遗传分化系数  $F_{st}$  与核苷酸多态性  $\pi$  ratio (前 5% 水平) 结合的方法筛选受选择区域的基因, 图 4 显示绿色部分为遗传分化系数  $F_{st} > 0.095$  受选择区域, 蓝色与红色部分分别表示  $F_{st}$  结合  $\pi$  ratio 在前 5% 水平下短花柱型与长花柱型的受选择区域。在  $F_{st} > 0.095$  以及  $\pi$  ratio  $\geq 1.29$  短花柱型受选择时, 该区域包含 295 个基因, 定位于染色体的 55 个窗口, 其中 284 个基因在 COG、KOG、Swissprot、TrEMBL、Pfam、GO、KEGG 等多个数据库中被注释。

#### 2.4 连翘全基因组受选择信号注释

为进一步分析受选择基因的功能, 对 GO 和 KEGG 数据库的注释结果进行分析。GO 数据库主要分为 3 大类: 生物学过程 (BP, biological process)、

细胞组件 (CC, cellular component)、分子功能 (MF, molecular function)。图 5 显示在生物学过程中富集较显著的条目为光合作用 (Photosynthesis, GO: 0015979)、多生物体过程 (Multi-organism processes, GO: 0051704); 在细胞成分中富集较显著的条目为叶绿体类囊体膜 (Chloroplast thylakoid membrane, GO: 0009535)、光系统 II (Photosystem II, GO: 0009523); 在分子功能中富集较显著的三类条目为 rRNA 结合 (rRNA binding, GO: 0019843)、丝氨酸型内肽酶活性 (Serine-type endopeptidase activity, GO: 0004252)、萜类合成酶活性 (Terpene synthase activity, GO: 0010333)。受选择基因在植物的生长发育以及代谢调控方面发挥了重要作用。

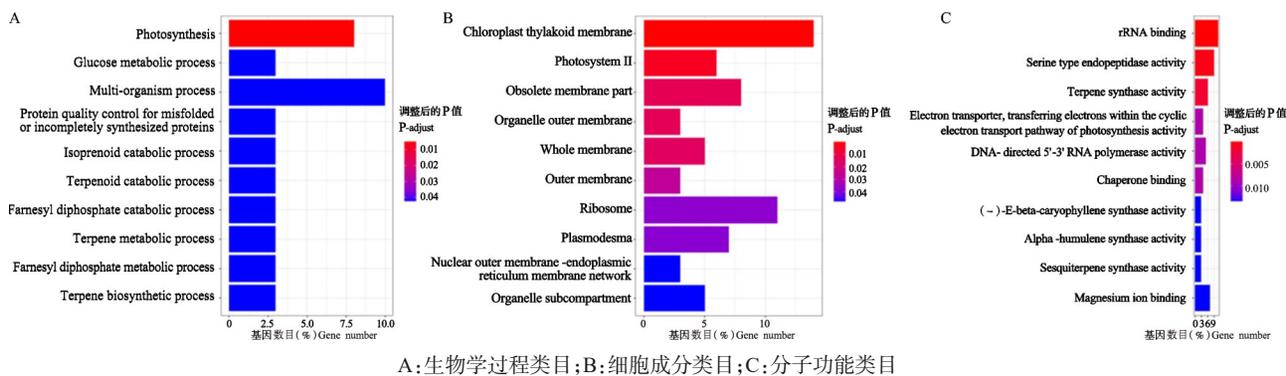


横坐标代表  $\pi$  ratio, 纵坐标为  $F_{st}$  值, 分别对应上方的频率分布图和右侧的频率分布图, 中部的点图则代表不同窗口内的相应的  $F_{st}$  和  $\pi$  ratio。其中最上方红色和蓝色区域为  $\pi$  ratio 选出来的前 5% 区域, 绿色区域为  $F_{st}$  所选择前 5% 区域, 中间蓝色和红色区域为  $F_{st}$  和  $\pi$  ratio 的交集, 即为候选的位点。频数对应图中呈正态分布的部分, 代表对应的 SNP 频数分布。累积趋势对应图中黑色曲线, 代表小于等于该区域数值的 SNP 的比例

The abscissa represents  $\pi$  ratio and the ordinate is the  $F_{st}$  value, corresponding to the upper frequency distribution plot and the right, and the dot plot in the middle represents the corresponding value of  $F_{st}$  and  $\pi$  ratio in different windows. The top red and blue regions are the top 5% regions selected by  $\pi$  ratio, the green regions are the top 5% regions selected by  $F_{st}$ , and the middle blue and red regions are the intersection of  $F_{st}$  and  $\pi$  ratio, which are the candidate loci. The frequency corresponds to the part of the normal distribution in the figure, representing the corresponding SNP frequency distribution. The cumulative trend corresponds to the black curve in the figure, representing the proportion of SNPs less than or equal to the value of the region

图 4 连翘长花柱型与短花柱型的受选择区域

Fig.4 The selection region of long style morph and short style morph of *F. suspensa*



A: Biological process category; B: Cellular component category; C: Molecular function category

图5 受选择信号GO富集分析  
Fig.5 GO enrichment analysis of selected signal

KEGG有助于把基因及表达信息作为一个整体的网络进行研究。连翘不同花柱类型受选择区域注释到50个通路,受选择区域的KEGG通路富集结果显示,细胞进程(Cellular processes)类目富集在胞吞作用(Endocytosis)中;新陈代谢(Metabolism)方面光合作用(Photosynthesis)最显著;环境信息过程(Environmental information processing)只富集到植物激素信号转导通路(Plant hormone signal transduction);在遗传信息过程(Genetic information processing)中,核糖体(Ribosome)功能最显著;在有机系统(Organismal systems)中,植物病原体相互作用(Plant-pathogen interaction)最显著(图6A)。光合作用(Photosynthesis, ko00195)、RNA聚合酶(RNA polymerase, ko03020)、倍半萜和三萜生物合成(Sesquiterpenoid and triterpenoid biosynthesis, ko00909)通路是受选择区域富集通路中最显著的(图6B)。异型花柱相关研究表明,植物激素对其生长发育影响较大,尤其是油菜素甾醇对二型花柱发育的调控已有很大进展,因此重点针对油菜素内酯代谢相关通路的注释信息进行候选基因的筛选。

## 2.5 BR代谢通路候选基因表达水平分析

基于筛选标准对受选择区域所有基因的功能注释与分析进行统计,发现连翘异型花柱发育与植物激素调控、细胞生长发育等有关,选择相关通路油菜素甾醇生物合成通路(ko00905)、植物激素信号转导通路(ko04075)中所有注释到的差异基因,包括

*CYP734A1* (EVM0010386)、*BRI1* (EVM0011829)、*CYCD3* (EVM0018316),对以上3个候选基因进行基因表达量检测。

候选基因表达结果如图7所示,连翘*CYP734A1* (EVM0010386)基因表达量在长花柱型发育不同时期均低于短花柱型,露冠期差异最显著,该时期短花柱中基因表达量达到最高。*BRI1* (EVM0011829)基因在连翘长花柱型与短花柱型发育过程中表达量趋势一致,均显著增加,鳞片脱落期与现蕾期长花柱中基因表达量高于短花柱,露冠期反之。*CYCD3* (EVM0018316)基因表达量在花柱发育的鳞片脱落期与现蕾期中,长花柱型逐渐升高而短花柱型无显著差异,露冠期该基因在短花柱型中表达量显著高于长花柱型。

油菜甾醇(Campesterol)首先转化为6-氧油菜甾醇(6-Oxocampestanol),通过早期C-6氧化途径和晚期C-6氧化途径转化为栗甾酮,进而形成油菜素内酯(BL, brassinolide)<sup>[10]</sup>。结合3个候选基因在连翘不同花期的表达水平与代谢调控途径,其在油菜素甾醇的代谢通路表达水平如图8所示。*CYP734A1*类基因(EVM0010386)可以将栗甾酮和油菜素内酯降解为26-羟基栗甾酮和26-羟基油菜素内酯,短花柱发育过程该基因表达量显著高于长花柱型。同时油菜素内酯信号转导途径中*BRI1*类基因(EVM0011829)通过磷酸化等作用,使得下游*CYCD3*类基因(EVM0018316)表达,促进细胞分裂。



### 3 讨论

植物花柱异型系统的演化及其适应性是植物进化生态学领域关注的主要问题,花柱二型是一类受遗传控制的花多态现象<sup>[28]</sup>。连翘二型花柱在长期进化过程中形成,由于环境适应及人工驯化影响,发生有利突变的基因被保留。全基因组重测序手段已普遍应用,许小涵等<sup>[29]</sup>利用全基因组重测序与转录组测序结合的方法证明忍冬在干旱条件下的甲基化水平上升。本研究对连翘的长花柱型与短花柱型植株进行全基因组重测序,获得45.82 GB的数据,Q30含量达到92.5%以上,染色体覆盖深度分布不均一的地方可能是由于重复序列、PCR偏好性引起的,测序数据可靠。因此利用注释得到的SNP信息进行选择清除分析,筛选连翘的受选择区域。

选择清除分析可以检测物种在进化中受选择的区域,对其系统进化研究具有重要意义。常采用的方法有基于突变频率分布的Tajima's D检验<sup>[30]</sup>,遗传分化系数Fst<sup>[21]</sup>,核苷酸多样性 $\pi$ 法<sup>[22]</sup>,以及基于连锁不平衡的方法<sup>[31]</sup>等。本研究选择遗传分化系数Fst与核苷酸多态性 $\pi$  ratio结合的方法,对连翘异型花柱形成的受选择区域筛选到295个基因,基因功能注释表明受选择基因对植物光合作用、植物激素调控、生长发育以及代谢调控方面作用突出。Li等<sup>[32]</sup>也使用该方法对西藏野猪和中国地域性分布的猪进行全基因组重测序,在被选择的基因组区域发现候选基因大多富集于缺氧、嗅觉、能量代谢和药物反应方面。杨柳<sup>[33]</sup>对枣果实大小进行选择清除分析,同样使用该方法,结果表明大果群组受选择基因聚集于激素和细胞途径,小果群组则聚集于代谢途径。朱红莲等<sup>[34]</sup>对我国莼菜野生品种与栽培品种的核苷酸多样性指数 $\pi$ 和遗传分化系数Fst值计算发现两者之间仅有较低的遗传分化,莼菜栽培品种人工驯化程度较低。选择清除分析对于物种的遗传进化、基因挖掘等研究具有较高的可行性,本研究筛选得到的连翘候选基因对其进化发育具有一定的参考意义。

植物形态建成特别是花芽分化是在基因控制下营养与内源激素综合作用的结果<sup>[35]</sup>。油菜素甾醇被认为是第6类植物内源性激素,其与生长素、赤霉素的特性较相似,对植物的花芽分化和生长发育有一定作用<sup>[36]</sup>,在对荔枝施加外源24-表油菜素内酯后发现,其能够在一定程度上抑制荔枝成花<sup>[37]</sup>。本研究筛选得到的候选基因CYP734A1(EVM0010386)为

油菜素甾醇生物合成通路上的重要基因,CYP734A1类基因对油菜素类固醇的降解机制已有报道<sup>[38]</sup>,该基因在连翘短花柱型花发育的不同时期表达量均显著高于长花柱型,这与连翘花柱发育的形态相一致,短花柱长度显著低于长花柱长度,CYP734A1基因对连翘异型花柱遗传机制的影响至关重要。在植物激素信号转导途径中筛选到的BRI1(EVM0011829)、CYCD3(EVM0018316)分别是单次跨膜的富含亮氨酸重复序列的类受体蛋白激酶<sup>[36]</sup>和信号激活因子细胞周期蛋白<sup>[39]</sup>,两个基因表达水平在露冠期短花柱型显著高于长花柱型;BRI1基因在鳞片脱落期和现蕾期短花柱型中表达量都较低。BRI1是油菜素类固醇的主要受体,在信号转导调控中起关键作用,油菜素类固醇信号在质膜上的感知由BRI1胞外区直接参与<sup>[40]</sup>,该基因在短花柱发育过程表达水平逐渐增加,位于油菜素类固醇调控的下游基因CYCD3可以促使细胞分裂,在连翘短花柱发育晚期也出现高水平表达。本研究挖掘的候选基因对连翘花芽发育的调控机制具有重要作用,但其如何影响连翘异型花柱进化还需深入研究。

### 4 结论

本研究对连翘长花柱型与短花柱型进行选择清除分析,短花柱型受选择时的基因功能注释显示植物生长发育、光合作用等作用显著,连翘异型花柱形成受植物激素的调控,针对油菜素甾醇合成与信号转导途径,筛选到CYP734A1(EVM0010386)、BRI1(EVM0011829)、CYCD3(EVM0018316),CYP734A1基因在短花柱中持续高表达,BRI1、CYCD3基因在露冠期的表达量短花柱型高于长花柱型,促进短花柱型连翘的细胞分裂。候选基因对连翘花柱发育具有重要作用,但其遗传调控机制仍需深入研究。

#### 参考文献

- [1] 国家药典委员会. 中华人民共和国药典: 一部. 2020年版. 北京: 中国医药科技出版社, 2020: 177  
National Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China: Volume I. 2020 Edition. Beijing: China Pharmaceutical Science and Technology Press, 2020: 177
- [2] 阎新佳, 温静, 项峥, 杨波, 吴健, 江培, 崔琳琳, 赵璞. 连翘的化学成分研究. 中草药, 2017, 48(4): 644-647  
Yan X J, Wen J, Xiang Z, Yang B, Wu J, Jiang P, Cui L L, Zhao Y. Chemical constituents from fruits of *Forsythia suspensa*. Chinese Traditional and Herbal Drugs, 2017, 48(4): 644-647

- [3] 夏伟,董诚明,杨朝帆,陈浩.连翘化学成分及其药理学研究进展.中国现代中药,2016,18(12):1670-1674  
Xia W, Dong C M, Yang Z F, Chen H. Research progress on chemical constituents and pharmacology of *Forsythia suspensa*. Modern Chinese Medicine, 2016,18(12):1670-1674
- [4] Yang H X, Liu Q P, Zhou Y X, Chen Y Y, An P, Xing Y Z, Zhang L, Jia M, Zhang H. Forsythiasides: A review of the pharmacological effects. Frontiers in Cardiovascular Medicine, 2022,25(9):971491
- [5] Hu K, Guan W J, Bi Y, Zhang W, Li L, Zhang B, Liu Q, Song Y, Li X, Duan Z, Zheng Q, Yang Z, Liang J, Han M, Ruan L, Wu C, Zhang Y, Jia Z H, Zhong N S. Efficacy and safety of Lianhuaqingwen capsules, a repurposed Chinese herb, in patients with coronavirus disease 2019: A multicenter, prospective, randomized controlled trial. Phytomedicine, 2021,85:153242
- [6] 郭彬,史晗,封海波,刘群.复方双黄连制剂体外抗菌抗病毒活性研究.中国畜牧兽医,2022,49(12):4832-4842  
Guo B, Shi H, Feng H B, Liu Q. Study on *in vitro* antibacterial and antiviral activity of compound Shuanghuanglian preparation. China Animal Husbandry & Veterinary Medicine, 2022,49(12):4832-4842
- [7] 廖强,罗达龙,韦日伟,梁艺坚.在线二维液相色谱法同时测定银翘解毒丸中5种成分.中药材,2017,40(12):2888-2890  
Liao Q, Luo D L, Wei R W, Liang Y J. Simultaneous determination of 5 components in Yinqiao Jiedu Pills by on-line two-dimensional liquid chromatography. Journal of Chinese Medicinal Materials, 2017,40(12):2888-2890
- [8] 陈冉,王婷婷,李开铃,尚锐峰,宋杰,张景勃.免疫调节抗病毒中药的特性与应用.中草药,2020,51(6):1412-1426  
Chen R, Wang T T, Li K L, Shang R F, Song J, Zhang J Q. Characteristics and application of immune-regulating and antiviral Chinese materia medica. Chinese Traditional and Herbal Drugs, 2020,51(6):1412-1426
- [9] 吴凌云.荞麦属植物二型花柱的进化及适应性意义研究.武汉:武汉大学,2017  
Wu L Y. A study of evolution and adaptive significance of distyly in *Fagopyrum* species. Wuhan: Wuhan University, 2017
- [10] 任鸿雁,王莉,马青秀,吴光.油菜素内酯生物合成途径的研究进展.植物学报,2015,50(6):768-778  
Ren H Y, Wang L, Ma Q X, Wu G. Progress in biosynthetic pathways of brassinosteroids. Chinese Bulletin of Botany, 2015,50(6):768-778
- [11] Huu C N, Kappel C, Keller B, Sicard A, Takebayashi Y, Breuninger H, Nowak M D, Bäurle I, Himmelbach A, Burkart M, Ebbing-Lohaus T, Sakakibara H, Altschmied L, Conti E, Lenhard M. Presence versus absence of *CYP734A50* underlies the style-length dimorphism in primroses. eLife, 2016,5:e17956
- [12] Matzke C M, Shore J S, Neff M M, McCubbin A G. The *Turnera* style *S*-locus gene *TsBAHD* possesses brassinosteroid-inactivating activity when expressed in *Arabidopsis thaliana*. Plants (Basel), 2020,9(11):1566
- [13] Matzke C M, Hamam H J, Henning P M, Dougherty K, Shore J S, Neff M M, McCubbin A G. Pistil mating type and morphology are mediated by the brassinosteroid inactivating activity of the *S*-locus gene *BAHD* in heterostylous *Turnera* species. International Journal of Molecular Sciences, 2021,22(19):10603
- [14] 李政.油菜素内酯介导连翘异型花柱发育分子机制的研究.太谷:山西农业大学,2022  
Li Z. Research of the molecular mechanism of heterostyly development mediated by BR of *Forsythia suspensa*. Taigu: Shanxi Agricultural University, 2022
- [15] 范竞升,谢和霞,谢小东,周海宇,程伟东,覃兰秋,江禹奉.基于SNP标记揭示广西糯米地方品种的遗传多样性与群体遗传结构.植物遗传资源学报,2023,24(3):661-670  
Fan J S, Xie H X, Xie X D, Zhou H Y, Cheng W D, Qin L Q, Jiang Y F. Exploring the genetic diversity and population structure of Guangxi waxy maize landraces based on SNP markers. Journal of Plant Genetic Resources, 2023,24(3):661-670
- [16] 孟歌,朱更瑞,方伟超,陈昌文,王新卫,王力荣,曹珂.桃的花型性状相关SNP位点挖掘与候选基因分析.植物遗传资源学报,2022,23(2):505-517  
Meng G, Zhu G R, Fang W C, Chen C W, Wang X W, Wang L R, Cao K. Genome-wide association study identified SNP alleles and candidate genes for flower shape trait in peach (*Prunus persica*). Journal of Plant Genetic Resources, 2022,23(2):505-517
- [17] Varshney R K, Terauchi R, Mccouch S R. Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biology, 2014,12(6):e1001883
- [18] Gross B L, Olsen K M. Genetic perspectives on crop domestication. Trends in Plant Science, 2010,15(9):529-537
- [19] Smith J M, Haigh J. The hitch-hiking effect of a favourable gene. Genetics Research, 2007,89(5-6):391-403
- [20] 张学勇,童依平,游光霞,郝晨阳,盖红梅,王兰芬,李滨,董玉琛,李振声.选择牵连效应分析:发掘重要基因的新思路.中国农业科学,2006,39(8):1526-1535  
Zhang X Y, Tong Y P, You G X, Hao C Y, Gai H M, Wang L F, Li B, Dong Y C, Li Z S. Hitchhiking effect mapping: A new approach for discovering agronomic important genes. Scientia Agricultura Sinica, 2006,39(8):1526-1535
- [21] Weir B S, Cockerham C C. Estimating *F*-statistics for the analysis of population structure. Evolution, 1984,38(6):1358-1370
- [22] Nei M, Li W H. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences, 1979,76(10):5269-5273
- [23] Li L F, Cushman S A, He Y X, Li Y. Genome sequencing and population genomics modeling provide insights into the local

- adaptation of weeping forsythia. Horticulture Researchs, 2020, 7(1):130
- [24] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo M A. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 2010,20(9):1297-1303
- [25] Cingolani P, Platts A, Wang L L, Coon M, Nguyen T, Wang L, Land S J, Lu X, Ruden D M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*, 2012,6(2):80-92
- [26] 乔永刚, 曹亚萍, 贾孟君, 王勇飞, 贺嘉欣, 张鑫瑞, 王文斌, 宋芸. 连翘异型花柱植株花芽生长发育与传粉习性研究. *园艺学报*, 2020,47(4):699-707
- Qiao Y G, Cao Y P, Jia M J, Wang Y F, He J X, Zhang X R, Wang W B, Song Y. Research on flower buds growth development and pollination habits of *Forsythia suspensa* heterostyly. *Acta Horticulturae Sinica*, 2020,47(4):699-707
- [27] 曹亚萍, 王勇飞, 贾孟君, 贺嘉欣, 张鑫瑞, 李政, 乔永刚, 宋芸. 连翘花器官生长阶段内参基因筛选与评估. *山西农业科学*, 2020,48(3):298-303
- Cao Y P, Wang Y F, Jia M J, He J X, Zhang X R, Li Z, Qiao Y G, Song Y. Selecting and evaluation of reference genes during flower organ growth phase in *Forsythia suspensa* (Thunb.) Vahl. *Journal of Shanxi Agricultural Sciences*, 2020, 48(3):298-303
- [28] 孙淑范, 李佳欣, 孙霞玲, 郭杰, 张勃. 连翘花柱二型的繁育系统特征研究. *西北植物学报*, 2021,41(2):317-322
- Sun S F, Li J X, Sun X L, Guo J, Zhang B. Study on characteristics of breeding system in distylous *Forsythia suspensa*. *Acta Botanica Boreali Occidentalia Sinica*, 2021, 41(2):317-322
- [29] 许小涵, 唐志强, 刘谦, 李佳, 刘振华, 张永清, 蒲高斌. 干旱胁迫下忍冬全基因组 DNA 甲基化和转录组分析. *中草药*, 2023,54(16):5339-5349
- Xu X H, Tang Z Q, Liu Q, Li J, Liu Z H, Zhang Y Q, Pu G B. Whole genome DNA methylation and transcriptome analysis of *Lonicera Japonica* in response to drought stress. *Chinese Traditional and Herbal Drugs*, 2023,54(16):5339-5349
- [30] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 1989, 123(3): 585-595
- [31] Flintgarciá S A, Thornsberry J M, Buckler E S, Buckler I V. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology*, 2003, 54(4):357-374
- [32] Li M Z, Tian S L, Jin L, Zhou G Y, Li Y, Zhang Y, Wang T, Yeung C K, Chen L, Ma J D, Zhang J B, Jiang A A, Li J, Zhou C W, Zhang J, Liu Y K, Sun X Q, Zhao H W, Niu Z X, Lou P, Xian L J, Shen X Y, Liu S Q, Zhang S H, Zhang M W, Zhu L, Shuai S R, Bai L, Tang G Q, Liu H F, Jiang Y Z, Mai M M, Xiao J, Wang X, Zhou Q, Wang Z Q, Stothard P, Xue M, Gao X L, Luo Z G, Gu Y R, Zhu H M, Hu X X, Zhao Y, F Plastow G S, Wang J Y, Jiang Z, Li K, Li N, Li X W, Li R Q. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nature Genetics*, 2013,45(12):1431-1438
- [33] 杨柳. 基于选择清除鉴定影响枣果实大小候选基因. 北京: 北京林业大学, 2019
- Yang L. Identification of candidate genes related with jujube fruit size based on selective sweep analysis. Beijing: Beijing Forestry University, 2019
- [34] 朱红莲, 杜娟, 刘正位, 孙亚林, 李明华, 彭静, 周凯, 柯卫东. 我国野生蔬菜考察及遗传多样性研究. *植物遗传资源学报*, 2020,21(6):1586-1595
- Zhu H L, Du J, Liu Z W, Sun Y L, Li M H, Peng J, Zhou K, Ke W D. Investigation and genetic diversity of wild water shield (*Brasenia schreberi*) in China. *Journal of Plant Genetic Resources*, 2020,21(6):1586-1595
- [35] 曾芳, 高娅, 潘鑫, 邹晓勇, 孙雁霞. 调控穗发芽的植物内源激素研究进展. *江苏农业学报*, 2023,39(3):848-858
- Zeng F, Gao Y, Pan X, Wu X Y, Sun Y X. Research progress on plant endogenous hormones regulating pre-harvest sprouting. *Jiangsu Journal of Agricultural Sciences*, 2023, 39(3):848-858
- [36] 孙超, 黎家. 油菜素甾醇类激素的生物合成、代谢及信号传导. *植物生理学报*, 2017,53(3):291-307
- Sun Z, Li J. Biosynthesis, metabolism and signaling of brassinosteroids. *Plant Physiology Journal*, 2017, 53(3): 291-307
- [37] 胡嘉琪. 油菜素内酯对荔枝成花的调控及其与 *LcSVP* 关系研究. 广州: 华南农业大学, 2017
- Hu J Q. Regulation of brassinosteroids on flowering of Litchi and its relationship with *LcSVP*. Guangzhou: South China Agricultural University, 2017
- [38] Turk E M, Fujioka S, Seto H, Shimada Y, Takatsuto S, Yoshida S, Wang H, Torres Q I, Ward J M, Murthy G, Zhang J, Walker J C, Neff M M. *BASI* and *SOB7* act redundantly to modulate Arabidopsis photomorphogenesis via unique brassinosteroid inactivation mechanisms. *The Plant Journal*, 2005,42(1):23-34
- [39] 刘彩霞, 郑唐春, 张鑫鑫, 葛晓兰, 曲冠证. 小黑杨 *CYCD3*; 3 基因启动子序列分析及其在拟南芥中的表达. *东北林业大学学报*, 2016,44(8):29-33
- Liu C X, Zheng T C, Zhang X X, Ge X L, Qu G Z. Promoter sequence analysis of the *CYCD3*; 3 gene and its expression in Arabidopsis. *Journal of Northeast Forestry University*, 2016, 44(8):29-33
- [40] Liu L, Chen G, Li S, Gu Y, Lu L, Qanmber G, Mendu V, Liu Z, Li F, Yang Z. A brassinosteroid transcriptional regulatory network participates in regulating fiber elongation in cotton. *Plant Physiology*, 2023,191(3):1985-2000