

基于全基因组重测序技术的‘红叶’杜仲 SNP 位点开发

杨 赟<sup>1 2</sup> 陈梦娇<sup>3</sup> 杜庆鑫<sup>1</sup> 朱景乐<sup>1</sup> 杜红岩<sup>1</sup> 杨绍彬<sup>1 \*</sup>

(1. 国家林业和草原局泡桐研究开发中心 经济林种质创新与利用国家林业和草原局重点实验室 郑州 450003 ;2. 南京林业大学林学院 , 南京 210000 ;3. 河南农业大学林学院 郑州 450002 )

**摘 要** 为了挖掘与‘红叶’杜仲(*Eucommia ulmoides* ‘Hongye’)红叶性状紧密联系的 SNP 位点,进一步揭示红叶性状的遗传基础和分子机理。以‘红叶’杜仲和普通绿叶杜仲‘小叶’杜仲(*Eucommia ulmoides* ‘Xiaoye’)为研究材料,进行覆盖深度约为 10x 的全基因组重测序。使用 SnpEff 软件预测变异位点对蛋白编码的影响,结合花色苷的代谢通路和关键酶基因,筛选与‘红叶’杜仲叶色形成相关的差异位点。利用 Sanger 测序二代测序筛选的 SNP 位点,分子标记验证群体是‘红叶’杜仲和‘小叶’杜仲。结果表明:‘红叶’杜仲测序产生 Clean data 为 14.16 Gb;‘小叶’杜仲产生 Clean data 为 14.29 Gb。在‘红叶’杜仲中注释到严重影响蛋白质功能的有 1 516 个 SNP,中度影响的 41 328 个 SNP,在‘小叶’杜仲中存在严重影响蛋白质功能的 SNP 为 1 640 个,中度影响功能的 SNP 为 47 192 个。测得 26 722 条基因中有 228 条基因是与花色苷或类黄酮合成相关的酶基因。经过筛选,确定了 12 个特异性的 SNP 位点,均属于外显子区域的错义突变。利用一代测序验证,根据 SNP 位置设计了 7 对引物,SNP 准确率达到 100%。

**关键词** ‘红叶’杜仲;全基因组重测序;花色苷;SNP

中图分类号 S687.9 文献标志码 A doi:10.7525/j.issn.1673-5102.2019.06.018

SNP Sites Developed by Whole Genome Resequencing  
Analysis in *Eucommia ulmoides* ‘Hongye’

YANG Yun<sup>1 2</sup> CHEN Meng-Jiao<sup>3</sup> DU Qing-Xin<sup>1</sup> ZHU Jing-Le<sup>1</sup> DU Hong-Yan<sup>1</sup> YANG Shao-Bin<sup>1 \*</sup>

(1. Paulownia Research and Development Center of State Administration of Forestry and Grassland, Key Laboratory of Non-timber Forest Germplasm Enhancement & Utilization of State Forestry and Grassland Administration Zhengzhou 450003 ;2. Nanjing Forestry University, College of Forestry, Nanjing 210000 ;3. Henan Agricultural University, College of Forestry Zhengzhou 450002 )

**Abstract** We studied the SNP locus and candidate genes closely related to the red leaf traits of *Eucommia ulmoides* ‘Hongye’ for further revealing the genetic basis and molecular mechanism of red leaf traits. We used *E. ulmoides* ‘Hongye’ and the common green leaf *E. ulmoides* to perform whole genome sequencing with a depth of about 10x. With resequencing data and by using SnpEff software we predicted the effect of mutation sites on protein coding, selected the differential sites associated with the formation of *E. ulmoides* ‘Hongye’ leaf color, which were related to anthocyanin metabolic pathways and key enzyme genes. For fine mapping, 14.16 Gb clean data were generated from genome re-sequencing of *E. ulmoides* ‘Hongye’ and 14.29 Gb of clean data was produced by *E. ulmoides* ‘Xiaoye’. There are 1 516 SNPs in the *E. ulmoides* ‘Hongye’ which have serious effects on protein function, and 41 328 moderately affected. There are 1 640 SNPs in the *E. ulmoides* ‘Xiaoye’ that seriously affect protein function, and 47 192 SNPs that moderately affect function. Of the 26 722 genes, 228 were found to be involved in the synthesis of anthocyanin or flavonoids. After screening, 12 specific SNP loci were

基金项目:中国林业科学研究院中央级公益性科研院所基本业务费专项资金(CAFYBB2014QA037)

第一作者简介:杨赟(1995—),女,博士研究生,主要从事林木遗传育种研究。

\* 通信作者 E-mail: ysb1966327@yahoo.com.cn

收稿日期 2019-01-03

Foundation item: Special Fund for Basic Business Expenses of the Central Level Public Welfare Research Institutes of the Chinese Academy of forestry(CAFYBB2014QA037)

First author introduction: YANG Yun (1995—) female, master student, mainly engaged in forest genetic breeding research.

\* Corresponding author E-mail: ysb1966327@yahoo.com.cn

Received date 2019-01-03

identified, all of which belonged to missense mutations in the exon region. Using a generation of sequencing verification, 7 pairs of primers were designed based on the SNP position, and the SNP accuracy rate reached 100%.

**Key words** *Eucommia ulmoides* ‘Hongye’, whole genome resequencing, anthocyanin SNP

杜仲(*Eucommia ulmoides*)隶属于杜仲科(*Eucommiaceae*), 单属单种, 染色体  $2n=2X=34$ , 是我国特有的经济林树种, 在我国北纬  $22^{\circ}\sim 42^{\circ}$ , 东经  $100^{\circ}\sim 121^{\circ}$  的广大范围内均可栽培<sup>[1]</sup>。杜仲因其适应性强, 干型通直, 冠型优美, 主根较深, 侧根发达, 具有较好的水土保持能力, 被广泛用于园林绿化和防止水土流失<sup>[2~3]</sup>。当前, 对于杜仲的研究主要集中在药理功效、成分分析、栽培模式, 干旱胁迫响应等方面<sup>[4~6]</sup>, 对选育观赏型杜仲良种报道较少。‘红叶’杜仲(*Eucommia ulmoides* ‘Hongye’)是中国林业科学研究院选育的观赏型良种, 叶片鲜红至紫红色, 是理想的绿化树种。‘红叶’杜仲叶片呈红色的主要物质是花色苷<sup>[7]</sup>, 其叶片花色苷含量显著高于普通杜仲, 但是‘红叶’杜仲叶片的呈色机理尚不清楚。

随着测序技术的不断发展, 拥有参考基因组的物种越来越多, 全基因组重测序技术在发掘与植物表型相关的单核苷酸多态性(SNP)位点方面发挥着重要作用。研究者利用该技术, 快速得到这一物种某个体的基因组信息和重要功能基因的序列信息, 初步定位和鉴定差异位点, 为从分子水平研究植物基因调控提供了依据<sup>[8~10]</sup>。植物叶片的呈红色的物质是花色苷, 花色苷代谢途径是目前研究的最透彻的通路之一。花色苷由苯丙氨酸经过一系列的酶催化合成, 由上游基因群和下游基因群控制酶的合成, 上游基因群通常编码参与黄酮、黄酮醇和鞣红合成相关的酶, 下游基因群编码合成花青苷和原花青素的酶<sup>[11]</sup>。本研究通过重测序技术检测并分析控制‘红叶’杜仲与普通绿叶杜仲叶色差异的基因位点, 比较与叶色形成相关的酶基因位点差异, 开发杜仲红叶性状 SNP 标记。

## 1 研究方法

### 1.1 研究材料

‘红叶’杜仲(*Eucommia ulmoides* ‘Hongye’)和‘小叶’杜仲(*Eucommia ulmoides* ‘Xiaoye’)均是在河南洛阳地区通过实生选育方法获得的无性系良种。本研究以3年生‘红叶’杜仲和‘小叶’杜仲嫁接苗为对象(砧木均为普通杜仲), 每个无性系选择3个单株, 每个单株选择顶端中上部的3片

叶片, 按无性系将叶片分组后置于液氮中保存。

### 1.2 DNA 提取

按照 Qiagen 公司的 DNeasy Plant Mini Kit 试剂盒的操作步骤提取杜仲 DNA, 使用 Nano Drop 2000 进行 DNA 纯度检测, Picogreen 方法检测 DNA 浓度。1% 的琼脂糖凝胶电泳检测 DNA 完整性。检测前将样品放在冰上融化后, 充分混匀并离心。每个无性系3个单株, 从中取等量的质量较好的 DNA 混合建库, 用于二代测序。

### 1.3 全基因组重测序

将检验合格的基因组 DNA 样品(样品量  $> 2\ \mu\text{g}$ , 样品浓度  $> 50\ \text{ng} \cdot \mu\text{L}^{-1}$ , 无色素污染, 无明显 RNA、蛋白质等杂质污染), 送至上海美吉生物医药科技有限公司进行文库构建。利用超声波将 DNA 序列片段形成随机片段, 经过处理后用磁珠吸附富集基因组长度为 400 bp 左右的片段, 经过 PCR 扩增形成测序文库。质检合格的文库用 Illumina HiSeq™ 平台进行基因组重测序, 测序策略为 Illumina PE150, 总测序读长为 300 bp。

#### 1.3.1 原始测序数据质控与过滤

原始测序数据(Raw Data)经过 Fastp(v0.11.5)进行质控, 包括碱基含量分布和碱基错误率分布。碱基含量分布检查有无 AT、GC 分离现象。测序碱基序列错误率结果由于 Illumina HiSeq™ 测序的技术特点, 测序片段前端几个 Cycles 和末端的错误率会偏高。使用 NGSQCtoolkit(v2.3.3)套件对 raw reads 去除接头等对 reads 进行 trimming, 去除低质量 read 并剪切掉 5' 端及 3' 端 3 bp 碱基并过滤掉长度小于 50 bp 的 reads。最终得到质量较好的 Clean Data。测序数据过滤采用 Fastp 默认参数流程<sup>[12]</sup>。

#### 1.3.2 变异位点检测及注释

对于‘红叶’杜仲和‘小叶’杜仲测序获得的 Clean Data, 利用 BWA(v0.7.16)基于 MEM 算法<sup>[13]</sup>将其比对到杜仲参考基因组序列上(杜仲参考基因组叶片为绿色), 利用 GATK(v3.8.0)对 BAM 文件进行校正<sup>[14]</sup>, 并进行 SNP 检测。利用 GATK 软件的 HaplotypeCaller 算法进行变异位点检测。利用 SnpEff(v4.3)进行 SNP 变异位点功能注释<sup>[15]</sup>, 包括变异位点位置和功能注释, SnpEff 会根

据变异位点对蛋白编码的影响 ,从高到低依次分为 High、Moderate、Low 和 Modifier 4 个等级 ,High 代表严重影响功能的 SNP 个数 ,Moderate 代表中度影响功能突变的 SNP ;Low 代表低影响程度 ;Modifier 表示与功能无关的 SNP。因此 ,只关注 High 和 Moderate 的个数。

1.3.3 变异基因的功能注释和代谢通路分析

利用 Diamond 软件比对到 NR、UniProt 和 Gene Ontology( GO )数据库 ,进行基因功能预测。使用 KOBAS 软件比对到 KEGG 数据库 ,注释基因所涉及的代谢通路。利用 emappe 软件比对 Egg-Nog 数据库 ,注释基因功能。

1.3.4 变异位点的筛选

首先根据变异位点对蛋白编码的影响 ,构建导致蛋白发生 High、Moderate 变化的基因集。根据变异基因的功能注释 ,筛选与类黄酮和花色苷合成途径相关的基因集。为了保证 SNP 位点的准确性 ,过滤掉低质量的 SNP 位点 ,要求 SNP 深度  $\geq 10\times$ 。选取‘红 叶 ’杜仲与参考基因组完全不一致的纯合位点 ;‘小叶 ’杜仲与参考基因组完全一致的纯合位点 ,这些位点被用于 SNP 位点开发。

1.4 基于 Sanger 测序 SNP 验证

参考杜仲基因组序列信息 ,采用 Primer Premier 5 设计引物 ,退火温度在  $60^{\circ}$  左右 ,引物长度 18 ~ 30 bp。PCR 扩增产物长度 :外显检测引物离外显子上下游 150 bp 左右 ;目的基因测序的 PCR 产物条带一般不超过 1 200 bp。PCR 反应体系 :50  $\mu\text{L}$  1 ~ 2  $\mu\text{L}$  的 50  $\text{ng} \cdot \mu\text{L}^{-1}$  cDNA ,上下游引物、dNTP( mix )均 2  $\mu\text{L}$ ( 10  $\text{mmol} \cdot \text{L}^{-1}$  ) , 10  $\times$  Taq Buffer( with  $\text{MgCl}_2$  ) 5  $\mu\text{L}$  ,Taq 酶( 5  $\text{U} \cdot \mu\text{L}^{-1}$  ) 0.5  $\mu\text{L}$  ,加双蒸水至 50  $\mu\text{L}$ 。PCR 反应条件 :95 $^{\circ}\text{C}$  预变性 3 min ;然后 94 $^{\circ}\text{C}$  变性 30 s ,60 $^{\circ}\text{C}$  退火 30 s ,72 $^{\circ}\text{C}$  延伸 90 s ,共 35 个循环 ;最后 72 $^{\circ}\text{C}$  延伸 8 min ,4 $^{\circ}\text{C}$  结束。Sanger 测序由上海生工生物工程公司完成。

2 结果与分析

2.1 原始测序数据统计及质控

本次测序质量较好。‘红 叶 ’杜仲获得 14.16 G 数据量 47 032 423 个 Clean reads ,其中碱基质量  $Q_{30}$  比例达到 91.72% ,GC 含量为 35.54%。‘小叶 ’杜仲获得 14.29 G 数据量 47 476 746 个 Clean reads ,其中碱基质量  $Q_{30}$  比例达到 91.93% ,GC 含量为 35.43%。数据质控分析显示 ,除了 reads 的

前几个碱基存在正常的不平衡现象外 ,碱基质量分布基本无 AT、GC 分离现象 ,测序碱基错误率分布统计低于 0.001%。

2.2 变异位点检测结果

2.2.1 比对参考基因组数据统计

杜仲参考基因组大小为 1.23 Gb ,组装水平是 Scafflod( Scafflod N50 = 1.88 Mb )。比对参考基因组 ;‘红 叶 ’杜仲和‘小叶 ’杜仲测序数据的 Mapped Ratio 均为 99.90%。本研究两样本对参考基因组 ( 排除 N 区 ) 的平均覆盖深度为 10  $\times$  左右 ,1  $\times$  覆盖深度 ( 至少有一个碱基的覆盖度 ) 为 87.77% 以上。本研究覆盖到参考基因组的区域较多 ,可检测到的变异位点较多 ,基因组上碱基的覆盖深度分布较均匀 ,测序的随机性好。

2.2.2 SNP 的检测

在‘小叶 ’杜仲中共检测到 5 711 141 个 SNP ,其中发生转换(  $T_i$  )的 SNP 有 4 139 309 个 ,发生颠换(  $T_v$  )的 SNP 有 1 565 071 个 , $T_i/T_v$  比值为 2.64 ,在‘红 叶 ’杜仲中共检测到 5 407 823 个 SNP ,其中发生转换(  $T_i$  )的 SNP 有 3 915 050 个 ,发生颠换(  $T_v$  )的 SNP 有 1 487 022 个 , $T_i/T_v$  比值为 2.63。

2.2.3 SNP 差异位点对蛋白编码影响的评估

在‘小叶 ’杜仲中存在严重影响蛋白质功能的 SNP 为 1 640 个 ,中度影响功能的 SNP 为 47 192 个。在‘红 叶 ’杜仲中存在严重影响功能的有 1 516 个 SNP ,含有中度影响的 41 328 个 SNP。

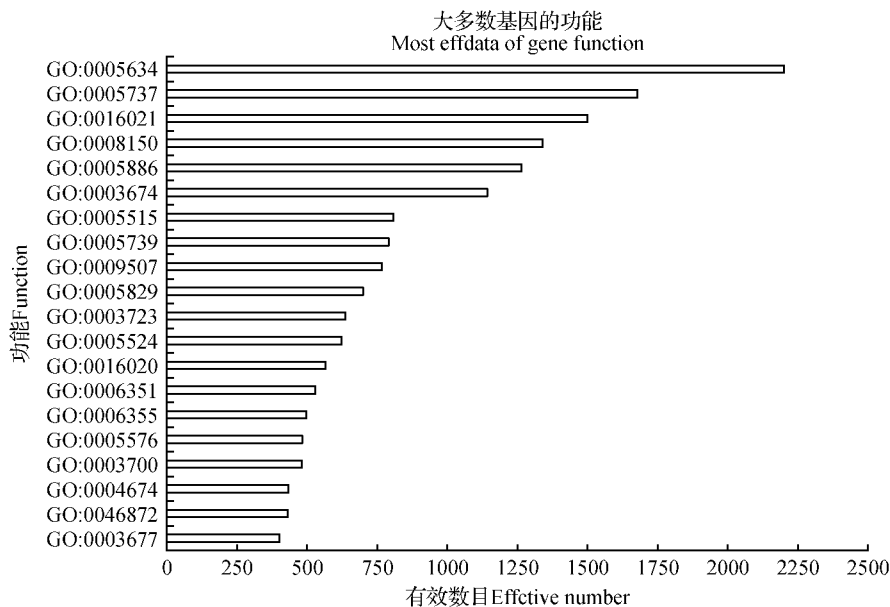
2.2.4 SNP 差异位点功能注释

SnEff 每个变异类型的功能统计如表 1 所示。某一基因在‘红 叶 ’杜仲和‘小叶 ’杜仲间存在碱基差异。‘红 叶 ’杜仲和‘小叶 ’杜仲每个 SNP 类型在数量上差异不大 ,可能它们分布在不同基因上 ,导致了‘红 叶 ’杜仲和‘小叶 ’杜仲在表型上的差异。

2.3 变异基因的功能分类

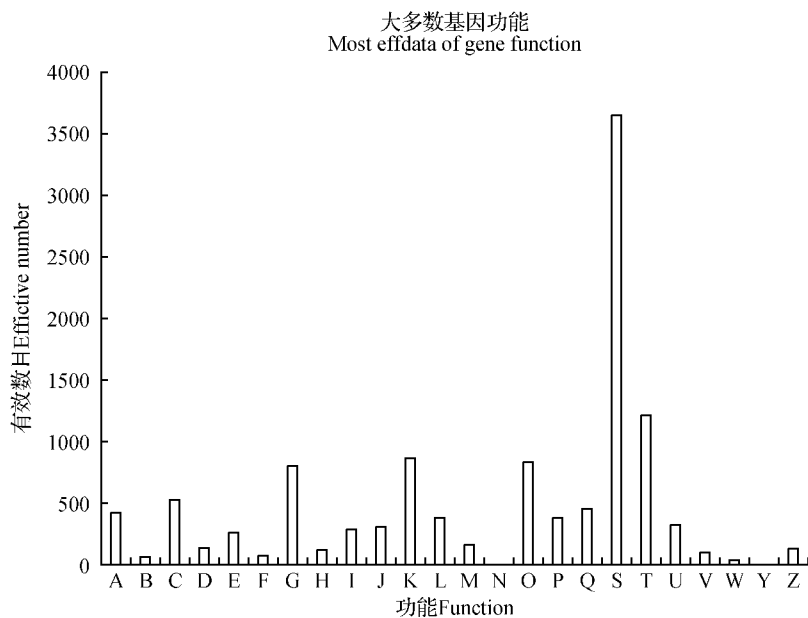
2.3.1 变异基因的 NR、Uniprot、GO、KEGG、EggNOG 注释

含有变异位点的 26 722 基因 ,在 Nr 数据库注释到 13 408 条基因 ,在 Uniport 中注释到 14 418 条基因 ,在 GO 数据库中注释到 21 453 条基因 ,在 EggNOG 数据库中注释 22 660 条基因 ,在 KEGG 数据库中注释到 5 915 条基因。为了进一步研究含有突变位点的基因功能 ,向 GO 数据库的各个 term 进行映射( 图 1 ) ,发现在细胞组分( cellular component )中注释到核、细胞质、膜的有机组成部分 ,在生



**图 1 GO 功能富集分类前 20 位统计图** 横坐标表示位于功能区属于 High 和 Moderate 类型的变异位点的个数 纵坐标表示 GO 项的 ID。

**Fig.1 GO function enrichment classification chart** The abscissa indicates the number of variant sites located in the functional area belonging to the High and Moderate types and the ordinate indicates the ID of the GO Term.



**图 2 COG 功能分类图** A. RNA 加工和修饰 ;B. 染色质结构和动力学 ;C. 能源生产和转换 ;D. 细胞周期控制 细胞分裂 染色体分配 ;E. 氨基酸转运和代谢 ;F. 核苷酸转运和代谢 ;G. 碳水化合物转运和代谢 ;H. 辅酶转运和代谢 ;I. 脂质转运和代谢 ;J. 翻译 核糖体结构和生物发生 ;K. 转录 ;L. 复制 重组和修复 ;M. 细胞壁/膜/包膜生物发生 ;N. 细胞运动性 ;O. 翻译后修饰 蛋白质周转 分子伴侣 ;P. 无机离子转运和代谢 ;Q. 次生代谢产物的生物合成 分解代谢 ;S. 功能未知 ;T. 信号转导机制 ;U. 细胞内运输 分泌和囊泡运输 ;V. 防御机制 ;W. 细胞外结构 ;Y. 核结构 ;Z. 细胞骨架

**Fig.2 COG function classification chart** A. RNA processing and modification ;B. Chromatin structure and dynamics ;C. Energy production and conversion ;D. Cell cycle control cell division chromosome partitioning ;E. Amino acid transport and metabolism ;F. Nucleotide transport and metabolism ;G. Carbohydrate transport and metabolism ;H. Coenzyme transport and metabolism ;I. Lipid transport and metabolism ;J. Translation ribosomal structure and biogenesis ;K. Transcription ;L. Replication recombination and repair ;M. Cell wall/membrane/envelope biogenesis ;N. Cell motility ;O. Posttranslational modification protein turnover chaperones ;P. Inorganic ion transport and metabolism ;Q. Secondary metabolites biosynthesis and catabolism ;S. Function unknown ;T. Signal transduction mechanisms ;U. Intracellular trafficking , secretion and vesicular transport ;V. Defense mechanisms ;W. Extracellular structures ;Y. Nuclear structure ;Z. Cytoskeleton

表 1 SNP 突变位点功能注释统计表

Table 1 SNP mutation site function annotation statistics table

类型 Type	‘红 叶 ’杜仲 <i>E. ulmoides</i> ‘ Hongye ’	‘小 叶 ’杜仲 <i>E. ulmoides</i> ‘ Xiaoye ’
5’ 端非翻译区 5_prime_UTR_variant	5	3
下游 Downstream	707 591	753 545
基因间区 Intergenic region	4 826 842	5 097 746
内含子区 Intron variant	503 340	532 407
外显子区域的错义突变 Missense variant	41 328	43 402
非编码转录本区 Non coding transcript variant	18	10
内含子左侧连接区 Splice acceptor variant	188	197
内含子右侧连接区 Splice donor variant	124	149
距离内含子或外显子 2 bp 的剪切区 Splice region variant	5 527	5 952
突变导致启动子缺失 Start lost	165	181
突变导致终止子获得 Stop gained	753	807
突变导致终止子突变 Stop lost	287	307
突变导致终止子的编码改变 Stop retained variant	82	88
同义突变 Synonymous variant	40 517	42 898
上游 Upstream gene variant	759 898	809 593

物进程( biological process )中注释到等离子体膜 , 在分子功能( molecular function )中注释到蛋白结合、线粒体和叶绿体、ATP 结合及转录过程。根据 KEGG 注释信息 ,如表 2 所示 ,含有差异位点的基因可注释到 286 条代谢途径 ,其中较多的是淀粉和蔗糖代谢 ,苯丙素生物合成途径等 ,与植物叶片生长及颜色密切相关。利用 EggNOG 数据库对比分析 ,如图 2 所示 ,有 12 977 个位点注释到 25 个类别 ,其中信号转导机制( 1 220 ,9. 38% )、转录( 868 ,6. 68% )、蛋白质翻译后修饰与转运、分子伴侣( 838 ,6. 45% )、碳水化合物的运输和代谢( 801 ,6. 16% )等类别注释到的变异位点多 ,其次是能量产生和转换( 527 ,4. 05% )、次生代谢产物的生物合成、转运和分解代谢( 456 ,3. 51% )。

2.3.4 用于分子标记开发的 SNP 位点筛选

根据基因功能注释的 5 大数据库结果 26 722 条含有差异位点的基因中共有 228 条基因注释到与花色苷和类黄酮合成相关。228 条基因内部共有

表 2 杜仲重测序的 KEGG 代谢途径分类图

Table 2 Classification of KEGG metabolic pathways for *Eucommia ulmoides* resequencing

编号 Number	途径 Pathway	Different sites with KEGG annotation ( 12997 )	ko ID
1	淀粉和蔗糖代谢 Starch and sucrose metabolism	183( 1. 41% )	ko00500
2	植物—病原互作 Plant-pathogen interaction	164( 1. 26% )	ko04626
3	苯丙素的生物合成 Phenylpropanoid biosynthesis	128( 0. 98% )	ko00940
4	碳代谢 Carbon metabolism	125( 0. 96% )	ko01200
5	植物激素信号转导 Plant hormone signal transduction	105( 0. 81% )	ko04075
6	氨基酸的生物合成 Biosynthesis of amino acids	103( 0. 79% )	ko01230
7	核糖体 Ribosome	98( 0. 75% )	ko03010
8	内质网蛋白加工 Protein processing in endoplasmic reticulum	98( 0. 75% )	ko04141
9	戊糖和葡萄糖醛酸酯互变 Pentose and glucuronate interconversions	92( 0. 71% )	ko00040
10	氨基糖和核苷酸糖代谢 Amino sugar and nucleotide sugar metabolism	92( 0. 71% )	ko00520
11	RNA 运输 RNA transport	73( 0. 56% )	ko03013
12	糖酵解和糖质新生 Glycolysis/Gluconeogenesis	72( 0. 55% )	ko00010
13	嘌呤代谢 Purine metabolism	67( 0. 52% )	ko00230
14	剪接体 Spliceosome	64( 0. 49% )	ko03040
15	真核生物核糖体生物起源 Ribosome biogenesis in eukaryotes	60( 0. 46% )	ko03008
16	半乳糖代谢 Galactose metabolism	59( 0. 45% )	ko00052
17	内吞作用 Endocytosis	59( 0. 45% )	ko04144
18	氰氨基酸代谢 Cyanoamino acid metabolism	57( 0. 44% )	ko00460
19	细胞周期 Cell cycle	55( 0. 42% )	ko04110
20	抗坏血酸和新陈代谢 Ascorbate and aldarate metabolism	54( 0. 42% )	ko00053

829 个差异位点 ,根据 SnpEff 注释 ,46 个位点严重影响蛋白质结构 ,783 个位点中度影响蛋白质功能。与参考基因组比 ,‘红 叶 ’杜仲中存在的这 228 条基因 ,含有 27 个位点严重影响蛋白质结构 ,其中有 15 个纯合位点 ,有 556 个位点中度影响蛋白质结构 ,269 个纯合位点。经过最终筛选 ,得到 12 个位点在‘红 叶 ’杜仲中完全差异于参考基因组 ,且‘小叶 ’杜仲中的这 12 个位点与参考基因组一致 ,测序深度均 ≥10 ×( 表 3 )。这 12 个位点分布在 6 条基因的内部( 表 4 )。

表 3 12 个与红叶性状相关的 SNP 位点

Table 3 12 SNPs loci associated with red leaf traits

序号 Serial number	序列编号 Scaffold	变异位点 Mutation site	参考基因组 Reference genome	二代测序检测结果		SNP 类型 SNP type
				Second generation sequencing results		
				‘ 红叶 ’杜仲 <i>E. ulmoides</i> ‘ Hongye ’	‘ 小叶 ’杜仲 <i>E. ulmoides</i> ‘ Xiaoye ’	
1	scaffold3882_obj	36515	A	G	A	$T_v$
2	Super-Scaffold_206	1351242	A	T	A	$T_i$
3	Super-Scaffold_433	439367	G	A	G	$T_v$
4	Super-Scaffold_12	15315681	G	C	G	$T_i$
5	Super-Scaffold_12	15316399	T	A	T	$T_i$
6	scaffold187873_obj	382	A	C	A	$T_v$
7	scaffold187873_obj	713	G	T	G	$T_v$
8	scaffold187873_obj	739	G	T	G	$T_v$
9	scaffold187873_obj	750	T	G	T	$T_v$
10	scaffold187873_obj	751	T	C	T	$T_v$
11	scaffold187873_obj	763	A	C	A	$T_v$
12	scaffold155639_obj	43644	C	T	C	$T_v$

表 4 筛选出的 6 条基因的功能注释

Table 4 Functional annotation of 11 selected genes

基因 ID Gene ID	Scaffold 位置 Scaffold position	起始 Start	终止 Stop	基因功能描述 Gene function description
EUC06132	scaffold3882_obj	35 959	37 383	槲皮素 3-O-葡萄糖基转移酶活性 Quercetin 3-O-glucosyltransferase activity
EUC22579	Super-Scaffold_206	1 345 799	1 354 979	在紫外线条件下 积累花色素 Anthocyanin accumulation in tissues in response to UV light
EUC24818	Super-Scaffold_433	438 806	439 402	$\beta$ -1 $\beta$ -葡萄糖基转移酶 Beta-1 $\beta$ -glucosyl transferase
EUC08209	Super-Scaffold_12	15 315 675	15 317 105	黄酮类葡萄糖醛酸化 Flavonoid glucuronidation
EUC10757	scaffold187873_obj	234	1 491	含花青素的化合物生物合成过程 Anthocyanin-containing compound biosynthetic process
EUC23922	scaffold155639_obj	4 3271	44 224	类黄酮生物合成 Flavonoid biosynthesis

表 5 SNP 的验证结果

Table 5 SNP verification results

序号 Serial number	序列编号 Scaffold	变异位点 Mutation site	引物序列( 5′—3′ ) Primer sequence( 5′—3′ )	一代测序验证结果	
				First generation sequencing verification results	
				‘ 红叶 ’杜仲 <i>E. ulmoides</i> ‘ Hongye ’	‘ 小叶 ’杜仲 <i>E. ulmoides</i> ‘ Xiaoye ’
1	scaffold3882_obj	36515	ATCAAGAACAGAATCACCATAAAT	G/G	A/A
2	Super-Scaffold_206	1351242	CTCTTCTAGCTCTGGGATGTCG	T/T	A/A
3	Super-Scaffold_433	439367	CCCGGCTGTTCAGTTCCTAG	A/A	G/G
4	Super-Scaffold_12	15315681	AAACACAACAAACAAGTATACGCTA	C/C	G/G
5	Super-Scaffold_12	15316399	GAAGCACAGAACAATACGGAC	A/A	T/T
6	scaffold187873_obj	382	ATGTCAGGATCAAGGAGAGAAGC	C/C	A/A
7	scaffold187873_obj	713	同 6	T/T	G/G
8	scaffold187873_obj	739	同 6	T/T	G/G
9	scaffold187873_obj	750	同 6	G/G	T/T
10	scaffold187873_obj	751	同 6	C/C	T/T
11	scaffold187873_obj	763	同 6	C/C	A/A
12	scaffold155639_obj	43644	AAATGGCCTAGCTTAGTTAGTGAGA	T/T	C/C

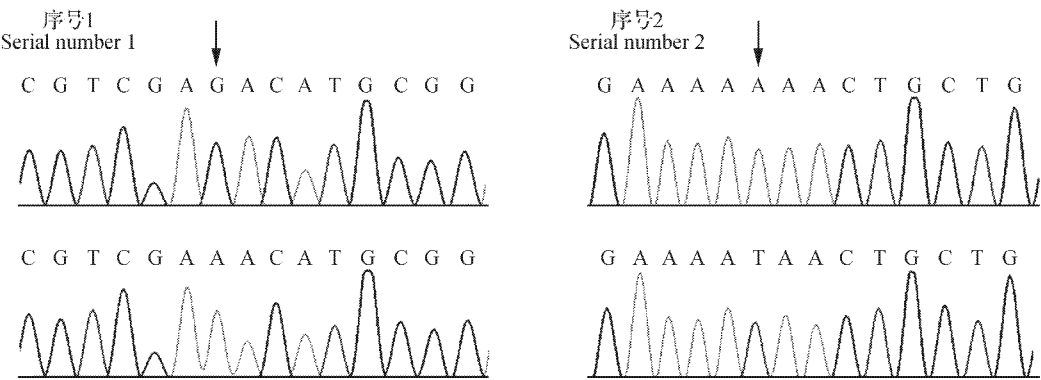


图3 基于 SNP 标记的红叶性状的初步分子鉴定

Fig.3 Preliminary molecular identification of red leaf traits by SNP markers

2.4 Sanger 测序结果

基于重测序数据和 SNP 位置设计了 6 对引物,一代测序的验证结果与重测序筛选结果,匹配率 100%。共筛选出 12 个分子标记用于‘红叶’杜仲红叶性状的早期鉴定(表 5)。图 3 给出了表 5 中序号 1 和 2 的 SNP 验证结果。

3 讨论

杜仲是单科单属单种;‘红叶’杜仲和‘小叶’杜仲均是在河南洛阳地区通过实生选育得到的无性系良种,遗传背景相对较小。本研究在参考杜仲基因组大小(1.18 G)的基础上,前期按 10 × 的测序深度,后期在分析中设定严格的阈值,同时再进行部分 sanger 测序验证,以期尽可能准确地得到 SNP 位点。根据重测序结果,在‘红叶’杜仲中共检测到 5 407 823 个 SNP,583 个位点能够严重或者中度影响蛋白质结构,且与花色苷合成相关。目前也有一些学者采用同样的方法得到类似的结论,如 Ni 等<sup>[16]</sup>对日本杏红绿皮果实 2 个品种的全基因组重测序分析,对 CDS 区域 6 个序列数据集的比对分析发现 181 331 个 SNPs,3 318 个 InDels 和 51 427 个 SVs,筛选出与红色果皮颜色形成相关的 13 个候选基因。Xu 等<sup>[17]</sup>对两种栽培葡萄株系的幼苗叶片进行基因组重测序,筛选出 635 个 sbc 特异性基因和 665 个 sbbm 特异性基因,分析了非同义突变与控制颜色形成的基因的关系,发现 NCED6 表达与 NCED1 有关,并可能影响花青素积累。此次研究可能存在一些未覆盖的位点,得到的数据准确性也可能有一定偏差,今后我们会加大群体和测序规模,继续深入开展相关研究。

‘红叶’杜仲和‘小叶’杜仲除了叶片颜色外

叶形也存在一定差异,我们在分析中锁定的是与花色苷合成相关的代谢通路,以尽量避免其他因素的干扰。目前多种研究方法共同揭示花色苷代谢途径中 SNP 位点差异被广泛使用。Jiao 等<sup>[18]</sup>在 14 种中国野生葡萄品种的样本中克隆了调节花色苷生物合成的 mybA 的两个转录因子 VvmybA1 和 VlmybA2,在测序的片段中检测到总共 121 个 SNP,这些 SNPs 分布在启动子区、内含子区和编码区。Liu 等<sup>[19]</sup>利用 MISA、GATK3 和注释工具,根据转录组数据从毛茛花瓣 151,136 个单基因中,鉴定和定位 SSR 和 SNP 标记。在 8 个花青素合成关键酶基因家族的 116 个单基因中,鉴定出 127 个 SNP。ANS 酶基因中没有 SNP 位点,而 DFR 酶基因具有最大量的 SNP 位点。Oh 等<sup>[20]</sup>利用全基因组重测序检测 SNP 和转录组联合分析黑米花青素的合成调控,发掘到 9 个候选基因中的 3 个基因与花色苷的合成有很好的相关性,尤其是基因 Os06t0192100(UDP 糖基转移酶)与之前报道的 Os04g0557500 基因相似性很高。这样联合分析的方法为我们进一步深入研究‘红叶’杜仲花青素的系统进化和基因通路提供了思路。在本研究结合‘小叶’杜仲的重测序结果与一代测序,发现 12 个纯合差异位点是‘红叶’杜仲特有的,更验证了 12 个与‘红叶’杜仲红叶性状相关的 SNP 位点。下一步将选择叶色有差别的杜仲,通过 PCR 扩增测序的方法对上述 SNP 位点开展进一步验证。

导致‘红叶’杜仲叶片特殊呈色的原因可能会有很多,如基因变异、基因表达量的差异及表观遗传修饰的不同等,本研究仅从基因层面通过 SNP 位点差异初步探讨‘红叶’杜仲呈现红色的可能原因,至于确切的原因,还需进一步的探索。

## 参 考 文 献

- 李芳东, 杜红岩, 杜仲[ M ]. 北京: 中国中医药出版社, 2001.  
Li F D ,Du H Y. *Eucommia ulmoides*[ M ]. Beijing :China Traditional Chinese Medicine Press 2001.
- 黄浩平. 杜仲的开发与利用[ J ]. 水土保持通报, 2001 , 21( 3 ) 31.  
Huang H P. The development and use of *Eucommia*[ J ]. Bulletin of Soil and Water Conservation 2001 21( 3 ) 31.
- 罗海蓉, 姜卫兵, 魏家星, 等. 杜仲的园林特性及开发利用[ J ]. 广东农业科学, 2013 40( 19 ) 25 – 27.  
Luo H R ,Jiang W B ,Wei J X ,et al. Landscape characteristics and developing application of *Eucommia*[ J ]. Guangdong Agricultural Sciences 2013 40( 19 ) 25 – 27.
- 何文. 杜仲开发有前景[ J ]. 资源开发与市场, 2000 ,16 ( 4 ) 221.  
He W. *Eucommia* has a promising development[ J ]. Resource Development & Markets 2000 ,16( 4 ) 221.
- 赵欣, 白伟. 杜仲叶片干旱胁迫响应相关差异蛋白的筛选与鉴定[ J ]. 植物研究, 2018 38( 3 ) 422 – 432.  
Zhao X ,Bai W. Screening and identification of the proteins related to drought response in leaves of *Eucommia ulmoides* [ J ]. Bulletin of Botanical Research 2018 38( 3 ) 422 – 432.
- 杜庆鑫, 刘攀峰, 魏艳秀, 等. 基于主成分与聚类分析的杜仲雄花品质综合评价[ J ]. 植物研究, 2016 36( 6 ) : 846 – 852.  
Du Q X ,Liu P F ,Wei Y X ,et al. Comprehensive evaluation of *Eucommia ulmoides* male flowers quality by principal component and cluster analysis[ J ]. Bulletin of Botanical Research 2016 36( 6 ) 846 – 852.
- 朱景乐, 李芳东, 杜红岩, 等. 三个观赏型杜仲无性系叶片色素含量比较[ J ]. 林业科学研究, 2014 27( 4 ) 562 – 564.  
Zhu J L ,Li F D ,Du H Y ,et al. Comparison of leaf pigment content among three ornamental *Eucommia ulmoides* varieties[ J ]. Forest Research 2014 27( 4 ) 562 – 564.
- 岳桂东, 高强, 罗龙海, 等. 高通量测序技术在动植物研究领域中的应用[ J ]. 中国科学: 生命科学, 2012 42 ( 2 ) 107 – 124.  
Yue G D ,Gao Q ,Luo L H ,et al. The application of high-throughput sequencing technology in plant and animal research[ J ]. Scientia Sinica Vitae 2012 42( 2 ) 107 – 124.
- 陈勇, 柳亦松, 曾建国. 植物基因组测序的研究进展[ J ]. 生命科学研究, 2014 18( 1 ) 66 – 74.  
Chen Y ,Liu Y S ,Zeng J G. Progresses on plant genome sequencing profile[ J ]. Life Science Research 2014 18( 1 ) : 66 – 74.
- 施季森, 王占军, 陈金慧. 木本植物全基因组测序研究进展[ J ]. 遗传, 2012 34( 2 ) 145 – 156.  
Shi J S ,Wang Z J ,Chen J H. Progress on whole genome sequencing in woody plants[ J ]. Hereditas 2012 34( 2 ) : 145 – 156.
- 戴思兰, 洪艳. 基于花青素苷合成和呈色机理的观赏植物花色改良分子育种[ J ]. 中国农业科学, 2016 49( 3 ) : 529 – 542.  
Dai S L ,Hong Y. Molecular breeding for flower colors modification on ornamental plants based on the mechanism of anthocyanins biosynthesis and coloration[ J ]. Scientia Agricultura Sinica 2016 49( 3 ) 529 – 542.
- Chen S F ,Zhou Y Q ,Chen Y R ,et al. Fastp :an ultra-fast all-in-one FASTQ preprocessor[ J ]. Bioinformatics 2018 , 34( 17 ) :i884 – i890.
- Li H ,Durbin R. Fast and accurate short read alignment with burrows-wheeler transform[ J ]. Bioinformatics 2009 , 25( 14 ) :1754 – 1760.
- Uchimura A ,Higuchi M ,Minakuchi Y ,et al. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice[ J ]. Genome Research 2015 25( 8 ) :1125 – 1134.
- Cingolani P ,Platts A ,Wang L L ,et al. 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[ J ]. Fly 2012 6( 2 ) : 80 – 92.
- Ni X P ,Xue S ,Iqbal S ,et al. Candidate genes associated with red colour formation revealed by comparative genomic variant analysis of red- and green-skinned fruits of Japanese apricot (*Prunus mume*) [ J ]. PeerJ 2018 6( 3 ) :e4625.
- Xu Y S ,Gao Z H ,Tao J M ,et al. Genome-wide detection of SNP and SV variations to reveal early ripening-related genes in grape[ J ]. PLoS One 2016 11( 2 ) :e0147749.
- 焦健, 刘崇怀, 樊秀彩, 等. 中国野生种葡萄 mybA 转录因子 SNP 特征分析[ J ]. 植物遗传资源学报, 2013 14 ( 5 ) 885 – 891.  
Jiao J ,Liu C H ,Fan X C ,et al. Characterization of SNP associated with mybA transcription factor in Chinese wild grapes[ J ]. Journal of Plant Genetic Resources 2013 14 ( 5 ) 885 – 891.
- Liu Y F ,Wang J M ,Zhang J H ,et al. Characterization of SSRs and SNPs based on transcriptome of *Ranunculus asiaticus* L. and development of anthocyanin-related SSRs and SNPs[ J ]. Nanoscience and Nanotechnology Letters 2018 , 10( 2 ) 267 – 273.
- Oh J H ,Lee Y J ,Byeon E J ,et al. Whole-genome resequencing and transcriptomic analysis of genes regulating anthocyanin biosynthesis in black rice plants[ J ]. 3 Biotech 2018 8( 2 ) :115.