

DOI: 10.3724/SP.J.1006.2023.24173

茶树新梢中香叶醇樱草糖苷含量的全基因组关联分析

王让剑^{1,2,*} 杨 军^{1,2} 张力岚^{1,2} 高香凤^{1,2}

¹福建省农业科学院茶叶研究所, 福建福州 355012; ²国家茶树改良中心福建分中心, 福建福州 355012

摘 要: 香叶醇是茶树中含有的一种重要的挥发性单萜醇, 在茶树与环境互作中起重要作用, 也是茶叶中关键呈香成分之一, 茶树新梢中的香叶醇主要以香叶醇樱草糖苷形式存在。发掘香叶醇樱草糖苷含量性状显著关联的 SNP 位点与候选基因, 对香叶醇樱草糖苷含量的遗传调控机制研究与茶树遗传改良具有重要意义。以 169 个茶树种质为研究对象, 连续 3 年对其新梢香叶醇樱草糖苷含量进行鉴定, 基于 SLAF 简化基因组测序技术开发 SNP 标记, 然后利用一般线性模型(GLM)对茶树新梢香叶醇樱草糖苷含量性状进行 GWAS 分析, 发掘该性状显著关联的 SNP 位点与候选基因, 最后对极端含量材料间的候选基因编码区碱基差异及其上游顺式作用元件进行分析。结果表明, 参试群体在 3 个环境下香叶醇樱草糖苷含量变异系数范围为 77.6%~81.8%, 广义遗传力为 62.6%。香叶醇樱草糖苷含量性状在基因型、环境间均呈现显著差异, 变异受遗传影响为主。3 个环境下共检测到 340 个 SNP 与香叶醇樱草糖苷含量显著相关, 其中 2 个环境下重复检测到 65 个 SNP。基于参考基因组与连锁不平衡衰减距离, 获得重复检测到的 SNP 两侧各 100 kb 范围内的基因共 88 个, 包含信号蛋白、激酶、磷酸酶、离子转运蛋白、转录因子、热激蛋白、激素相关蛋白、抗性蛋白、萜类代谢酶、糖基转移酶、糖苷酶, 从中初筛出 10 个与香叶醇樱草糖苷含量相关的候选基因。极端含量材料之间的候选基因编码区序列均存在 SNP 非同义突变, 多数候选基因上游 2 kb 区域存在不同数量的与环境胁迫和激素相关的顺式作用元件。本研究为阐明茶树新梢中香叶醇樱草糖苷含量的遗传调控机制提供了新的视角, 为分子标记辅助选择茶树新品种提供了标记与基因资源。

关键词: 茶树; 香叶醇樱草糖苷; 全基因组关联分析; 单核苷酸多态性; 候选基因

Genome-wide association analysis of geraniol primrose glycoside abundance in tender tea shoots

WANG Rang-Jian^{1,2,*}, YANG Jun^{1,2}, ZHANG Li-Lan^{1,2}, and GAO Xiang-Feng^{1,2}

¹Tea Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou 355012, Fujian, China; ²Fujian Branch, National Center for Tea Improvement, Fuzhou 355012, Fujian, China

Abstract: Geraniol is an important volatile monoterpene alcohol contained in tea plant, which plays an important role in the interaction between tea plant and environment and constitutes one of key aroma components in tea. Geraniol from tea plant tender shoots mainly exists in the form of geranyl β -primeveroside. Explore SNP loci and candidate genes significantly associated with geranyl β -primeveroside content is of great significance for the the genetic regulation mechanism of geranyl β -primeveroside content and tea plant genetic improvement. Regarding 169 tea germplasms as the research materials, the geranyl β -primeveroside content of tea plant tender shoots was identified for three consecutive years (three environments), and the SNP markers were developed based on SLAF-seq technology. Then, the geranyl β -primeveroside content of tea plant tender shoots was analyzed by genome-wide association analysis (GWAS) using general linear model (GLM). The candidate genes were further screened based

本研究由国家重点研发计划项目(2019YFD1001601), 福建省与中国农业科学院“5511”协同创新工程项目(XTCXGC2021004)和福建省自然科学基金项目(2020J011366)资助。

This study was supported by the National Key Research and Development Program of China (2019YFD1001601), the “5511” Collaborative Innovation Project between Fujian Province and Chinese Academy of Agricultural Sciences (XTCXGC2021004), and the Natural Science Foundation of Fujian Province (2020J011366).

* 通信作者(Corresponding author): 王让剑, E-mail: wangrj@faas.cn

Received (收稿日期): 2022-08-01; Accepted (接受日期): 2022-11-25; Published online (网络出版日期): 2023-01-17.

URL: <https://kns.cnki.net/kcms/detail/11.1809.S.20230116.1803.001.html>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

on the SNP loci significantly with the geranyl β -primeveroside content. Ultimately, the base differences of candidate gene coding regions and their upstream cis acting elements among the extreme geranyl β -primeveroside content were analyzed. Results showed that the variation coefficient of geranyl β -primeveroside content in three environments ranged from 77.6% to 81.8%, and the broad heritability was 62.6%. The geranyl β -primeveroside content was significant differences among the genotypes and environments, and the variation was mainly affected by heredity. A total of 340 SNP markers significantly correlated with the geranyl β -primeveroside content were detected under three environments, in which 65 SNP markers were detected repeatedly under 2 environments. Based on the tea plant reference genome and linkage disequilibrium decay distance, a total of 88 genes within 100 kb on both sides of the repeated 65 SNP markers were obtained, including signal proteins, kinases, phosphatases, ion transporters, transcription factors, heat shock proteins, hormone related proteins, resistance proteins, terpene metabolic enzymes, glycosyltransferases, and glycosidases. 10 candidate genes were preliminary screened out. SNP nonsynonymous mutations in the coding region sequences of 10 candidate genes between two extreme geranyl β -primeveroside content materials, and different numbers of cis acting elements related to environmental stress and hormones in the upstream 2 kb regions of most candidate genes were discovered. This study provides a new perspective for clarifying the genetic regulation mechanism of geranyl β -primeveroside content in tea plant tender shoots, and also provides the markers and gene resources for molecular marker assisted selection of new tea varieties.

Keywords: tea (*Camellia sinensis* L.); geranyl β -primeveroside; genome-wide association analysis (GWAS); single nucleotide polymorphism (SNP); candidate genes

香叶醇是茶树中含有的一种重要的挥发性单萜醇, 在茶树与环境互动中起重要作用, 如抗菌^[1]、驱虫^[2]、吸引昆虫传粉^[3]、提高相邻茶树对冷胁迫的抗性^[4]等。香叶醇具有温和微甜的玫瑰花香, 也是茶叶中关键呈香成分之一, 其在茶鲜叶中主要以香叶醇樱草糖苷形式存在^[5]。茶鲜叶中的糖苷通常位于液泡中, 糖苷酶存在于细胞壁和细胞间的空腔区域^[6], 正常细胞中二者相互隔离, 在茶树遭受环境胁迫或茶叶付制时, 茶鲜叶细胞遭受破坏, 二者发生接触, 液泡内积累的糖苷会在糖苷酶作用下水解, 释放出香叶醇等挥发性化合物, 从而使茶树提升自身抵抗环境胁迫能力以及使成茶表现出愉悦的花香。

香叶醇等挥发性化合物与糖结合形成糖苷后, 水溶性增大, 稳定性增强^[7]。糖苷形式有利于挥发性物质在植物体内储存运输、信号转导^[8], 还能以低毒或脱毒的形式减少或避免其对植物自身造成毒害^[9]。茶叶中糖苷分离提取与定性定量的报道较多, 热水^[10]、乙醇^[11]、甲醇^[12]浸提等方法均能分离提取茶叶糖苷, 水解^[13]、衍生^[14]、液相色谱质谱联用^[15]等方法均可对茶叶糖苷定性定量。嫩叶和成熟叶中均有萜烯醇糖苷分布, 嫩叶含量高于成熟叶, 春季含量相对较高^[16]。红茶加工过程中, 揉捻后萜烯醇糖苷含量明显降低, 发酵末期樱草糖苷基本消失, 但葡萄糖苷变化不大, 说明樱草糖苷是红茶香气重要来源物质^[17]。研究表明, 茶树体内樱草糖苷可以在糖基转移酶的作用下通过葡萄糖基化和木糖基化顺序合成^[18]。糖苷形成还可能与糖苷酶相关, 糖苷酶除了能水解糖苷键, 还能发生转糖基反应生成键合态糖苷^[19]。

全基因组关联分析 (genome-wide association study, GWAS) 是一种重要的遗传解析方法, 已广泛应用于水稻^[20]、小麦^[21]、玉米^[22]、大豆^[23]等作物重要农艺性状候选基因的鉴定研究之中, 茶树中基于该方法对新梢萌发期^[24]、风味代谢物^[25]等重要性状亦进行了研究。虽然香叶醇樱草糖苷等萜醇类糖苷在茶树中的生物合成、茶鲜叶中的分布以及茶叶付制过程中的变化已见报道, 但鲜见基于群体在全基因组水平上发掘与香叶醇樱草糖苷含量相关的标记位点与候选基因的报道, 鲜叶中香叶醇樱草糖苷含量的遗传调控机制尚不清楚。本研究以 169 个茶树种质构成的自然群体为研究对象, 利用 SLAF-seq 技术 (specific locus amplified fragment sequencing, SLAF-seq) 开发覆盖茶树全基因组的 SNP, 然后在 3 个环境下对茶树新梢香叶醇樱草糖苷含量性状进行 GWAS 分析, 在全基因组水平上发掘与香叶醇樱草糖苷含量性状显著关联的 SNP 位点与候选基因, 以期对香叶醇樱草糖苷含量的遗传调控机制研究及茶树分子标记辅助育种提供参考。

1 材料与方法

1.1 供试材料

茶树种质 FT516 为黄观音(♀)与铁观音(♂)人工杂交创新种质, 该种质制乌龙茶香气馥郁幽雅, 品质优异, 保存于福建省乌龙茶种质资源圃内(27°13'40"N, 119°32'11"E)。供试材料为 FT516 的自然杂交后代, 含 169 个茶树种质。2014 年 11 月份收集 FT516 母树上成熟的自然杂交种子, 温室条件下沙培至发根状态。2015 年 3 月种植到田间, 单株种植, 株距 0.5 m,

行距 1.5 m, 田间管理措施相对一致。2018、2019 和 2020 连续 3 年分别采摘各参试种质春季新梢无损伤无病虫害的单芽, 用于香叶醇樱草糖苷含量测定、SLAF-seq, 其中香叶醇樱草糖苷极端高含量种质 409、极端低含量种质 1605 单芽用于转录组测序、基因组 DNA 重测序。另取相同栽培环境下的已应用品种悦茗香、福云 6 号单芽, 进行转录组测序进一步验证候选基因, 其中悦茗香为香叶醇樱草糖苷高含量品种, 福云 6 号为香叶醇樱草糖苷低含量品种^[26]。

1.2 香叶醇樱草糖苷含量表型性状鉴定

按照文献方法测定参试茶树种质新梢单芽香叶醇樱草糖苷含量^[26]。使用 SAS8.0 进行表型数据统计分析, 利用 PROC VARCOMP 程序进行方差分量评估, 并计算香叶醇樱草糖苷含量性状广义遗传力(h^2), 计算公式: $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$, 其中, σ_g^2 为遗传方差, σ_e^2 为环境方差。

1.3 群体 SNP 分子标记开发

利用 CTAB 法提取茶树基因组 DNA, 选用 *Hae* III 酶切茶树基因组 DNA, 酶切片段长度在 314~364 bp 的序列定义为 SLAF 标签, 然后对检测合格的各样品基因组 DNA 分别进行酶切。对得到的酶切片段进行 3'端加 A 处理、连接 Dual-index 测序接头、PCR 扩增、纯化、混样、切胶选取目的片段, 文库质检合格后基于百迈客 Illumina HiSeq 平台进行测序。利用 BWA^[27]将测序 reads 比对到参考基因组上(参考基因组地址 <http://tpdb.shengxin.ren/>), 使用 GATK^[28]和 samtools^[29]开发 SNP, 以 2 种方法得到的 SNP 交集作为可靠 SNP 数据集。按照完整度大于 0.8, 次要基因型频率大于 0.05 标准, 对 SNP 数据集进行过滤, 使用过滤的 SNP 进行后续遗传分析。

1.4 群体遗传结构与连锁不平衡分析

使用 admixture 软件^[30]进行群体遗传结构分析, 分别假设样品的分群数(K 值)为 1~10, 并对聚类结果进行交叉验证, 根据交叉验证错误率的谷值确定最优分群数。使用 EIGENSOFT 软件^[31]进行 PCA 主成分分析(principal components analysis, PCA)。使用 plink^[32]软件计算两两 SNP 间的连锁不平衡强度(R^2), 将 SNP 间距与 R^2 进行拟合, 绘制 R^2 随距离变化的 LD 衰减图(linkage disequilibrium, LD)。

1.5 全基因组关联分析

基于 Tassel 软件 Glm 模型进行 GWAS 全基因组关联分析, 使用 admixture 软件的 Q 矩阵结果作为固定效应^[33]。显著性阈值 P 设置为: $P < 1.48 \times 10^{-6}$

$[P = (10n)^{-1}, n$ 为使用标记的总数]。根据 GWAS 结果获得显著关联的 SNP 位点, 基于群体连锁不平衡分析 LD 区间和茶树参考基因组注释信息, 获得与香叶醇樱草糖苷含量相关基因。

1.6 极端表型材料 RNA-Seq 与候选基因筛选

分别取极端高香叶醇樱草糖苷含量种质 409 和极端低香叶醇樱草糖苷含量种质 1605 春梢单芽, 每个样品取 3 个生物学重复, 在北京百迈克生物科技有限公司有参全长转录组平台完成转录组测序, 流程包括样品质量检测、文库构建、文库质量检测和文库测序(参考基因组 <http://tpdb.shengxin.ren/>)。样本之间差异表达基因 GO 和 KEGG 富集分析在百迈客云平台完成(<https://international.biocloud.net/>)。另取栽培品种悦茗香、福云 6 号春梢单芽样品同时进行转录组测序, 进一步筛选验证候选基因。筛选差异表达基因条件为: 采用 CPM (counts per million) 作为衡量基因表达水平的指标(CPM=比对到某一转录本上的 reads 数目 $\times 10^6$ /比对到参考转录组的 reads 数目)^[34]。将 GWAS 分析得到的与香叶醇樱草糖苷含量相关的基因在 2 个极端种质以及栽培品种悦茗香和福云 6 号差异表达基因中搜索, 重叠基因视为候选基因。

1.7 候选基因编码区序列差异及上游作用元件分析

在百迈客全基因组重测序分析平台完成 409 和 1605 基因组重测序, 流程包括样品质量检测、文库构建、文库质量检测和文库测序, 然后将重测序获得的无冗余 reads 重新定位到参考基因组上, 进行编码区 SNP 非同义突变分析(参考基因组 <http://tpdb.shengxin.ren/>)。将候选基因起始密码子上游 2 kb 的序列作为启动子区域, 提交到在线工具 Plant CARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>), 默认参数, 进行顺式作用元件的预测。

2 结果与分析

2.1 香叶醇樱草糖苷含量表型变异分析

对参试种质香叶醇樱草糖苷含量性状进行鉴定发现, 各环境下群体香叶醇樱草糖苷含量变异系数在 77.6%~81.8%之间, 说明该表型性状变异较丰富。正态分布检验结果表明(Ryan-Joiner 检验), 各环境下群体香叶醇樱草糖苷含量性状 r 值在 0.855~0.917 之间, 接近正态分布, 说明该性状是由多基因控制的数量性状(表 1)。方差分析结果表明, 香叶醇

樱草糖苷含量性状在基因、环境间均呈现显著差异,依据方差组分估算广义遗传力(h^2)为 62.6%,说明香叶醇樱草糖苷含量性状变异主要受遗传影响(表 2)。

2.2 群体 SNP 开发

基于 SLAF-seq 简化基因组测序技术,共开发群体 SNP 标记 10,802,426 个。根据 SNP 变异位点在基因组发生区域进一步细分类型,大部分检测到的 SNP 变异发生在基因间区、内含子区域、基因上下游区域,基因编码区的同义突变和非同义突变 SNP 数目较少(表 3)。

2.3 群体遗传结构与连锁不平衡分析

经质量控制,筛选出 675,245 个 SNP 用于后续遗传分析,每个标记间的平均距离为 4.47 kb (3.02 Gb/675,245)。利用 admixture 软件分析,当 $K=2$ 时(图 1-A), ΔK 值最大,参试群体被划分为 2 个亚群(图 1-B),利用 EIGENSOFT 软件进行主成分分析(principal components analysis, PCA),参试群体也被划分为 2 个亚群(图 1-C),2 种方法划分结果一致。利用 plink 软件绘制 R^2 随距离变化图发现, R^2 从 0.2 下

降到 0.1 时, LD 衰减距离约为 100 kb (图 1-D)。标记间的平均距离远小于 LD 衰减距离,说明标记密度满足 GWAS 分析要求。

2.4 全基因组关联分析

基于 TASSEL 软件 GLM 模型对香叶醇樱草糖苷含量性状进行 GWAS 分析,3 个环境下共检测到 340 个 SNP 标记与香叶醇樱草糖苷浓度显著相关,其中 2 个环境下均检测到的 SNP 标记共 65 个,各标记贡献率在 17.1%~23.9%之间,分布在 29 个 Scaffold 上,均位于基因间区、基因内含子以及基因下游区域(表 4)。

2.5 香叶醇樱草糖苷含量候选基因预测

基于茶树参考基因组信息(<http://tpdb.shengxin.ren/>)和 LD 衰减距离(100 kb),获得与香叶醇樱草糖苷含量相关 65 个 SNP 标记上、下游 100 kb 范围内的基因共 88 个,涵盖了信号蛋白、激酶、磷酸酶、离子转运蛋白、转录因子、热激蛋白、激素相关蛋白、抗性蛋白、萜类代谢酶、糖基转移酶、糖苷酶类等(表 5)。

表 1 香叶醇樱草糖苷含量表型变异

Table 1 Phenotypic variation of geranyl β -primeveroside content

环境 Environment	均值 Mean value	标准差 Standard deviation	变异系数 Coefficient of variation (%)	Ryan-Joiner 检测 Ryan-Joiner test
2018	0.3585	0.2809	78.4	$r=0.917, P<0.01$
2019	0.2674	0.2187	81.8	$r=0.914, P<0.01$
2020	0.3581	0.2779	77.6	$r=0.855, P<0.01$

参试样本数为 169 个。The number of test samples was 169.

表 2 香叶醇樱草糖苷含量方差分析

Table 2 Variance analysis of geranyl β -primeveroside content

方差来源 Variance	自由度 DF	平方和 SS	均方 MS	F 值 F-value	遗传力 h^2 Heritability h^2 (%)
基因 Genotype	168	34.1299	0.2032	6.12**	62.6
环境 Environment	2	0.9318	0.4659	14.04**	
误差 Error	336	11.3233	0.0337		

参试样本数为 169 个。**表示 $P < 0.01$ 。The number of samples was 169. **: $P < 0.01$.

表 3 群体 SNP 标记统计

Table 3 Statistics of SNP marker types in the population

主要类型 Main types	最大数 Maximum number	最小数 Minimum number	平均数 Average number
基因间区 Intergenic	1,595,407	513,424	1,062,687
内含子 Intron	53,116	21,509	41,091
基因上游(5 kb 内) Upstream in 5 kb	46,686	17,538	37,335
基因下游(5 kb 内) Downstream in 5 kb	42,110	16,067	32,681
同义突变 Synonymous	5004	2161	3537
非同义突变 Nonsynonymous	6703	2931	4911
其他 Other	2899	1161	2030

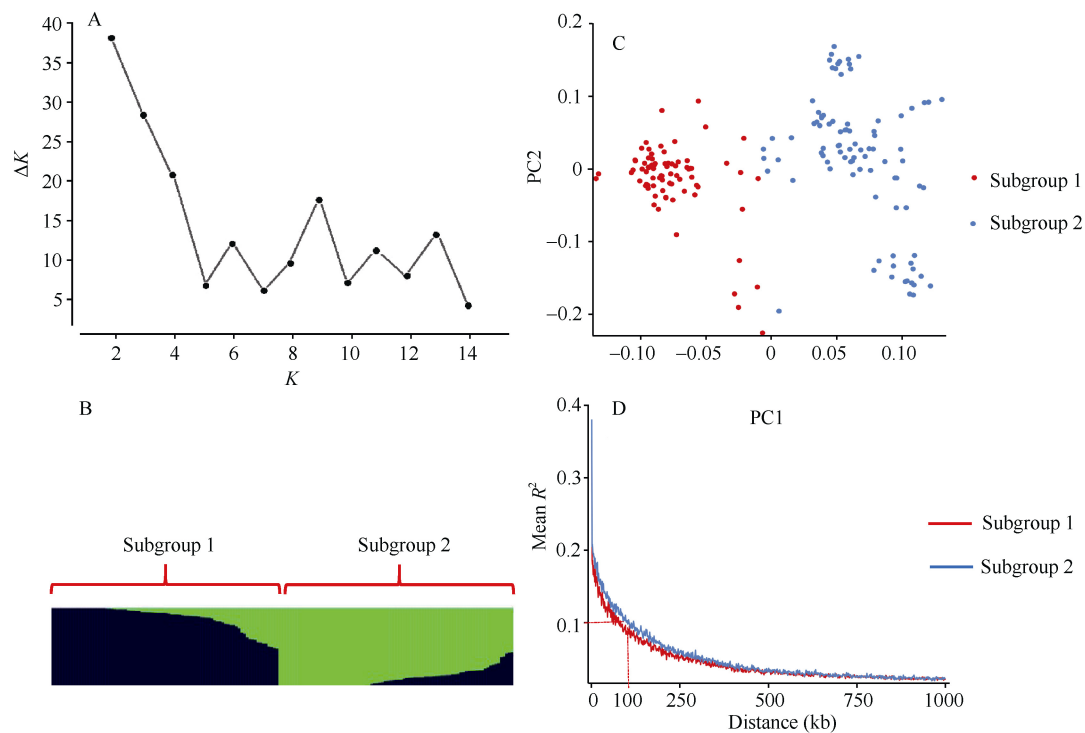


图 1 遗传结构及连锁不平衡分析

Fig. 1 Analysis of genetic structure and linkage disequilibrium

A: 遗传结构分析; B: 主成分分析; C: 连锁不平衡分析。

A: the genetic structure analysis; B: the principal component analysis; C: the linkage disequilibrium analysis.

表 4 香叶醇樱草糖苷含量相关的 SNP 位点

Table 4 SNP sites related to geranyl β -primeveroside content

染色体 Scaffold	SNP 位置 SNP location	SNP 变异 SNP variation	功能区 Functional area	贡献率 Contribution rate (%)	P 值 P-value		
					2018	2019	2020
Scaffold106	940,480	C/T	间区 Intergenic	19.3	—	4.25E-09	7.89E-08
	940,538	A/G	间区 Intergenic	18.1	—	1.01E-08	1.41E-07
Scaffold1262	982,460	G/A	间区 Intergenic	18.2	—	5.56E-09	6.45E-08
	982,581	A/T	间区 Intergenic	18.2	—	5.56E-09	6.45E-08
Scaffold1318	1,188,898	C/A	间区 Intergenic	20.7	2.70E-08	9.11E-10	—
Scaffold1611	610,786	C/T	间区 Intergenic	20.7	1.60E-07	9.52E-09	—
Scaffold1671	583,382	C/T	内含子 Intron	18.4	—	4.92E-09	7.52E-08
	583,566	G/A	内含子 Intron	18.2	—	6.24E-09	8.42E-08
Scaffold209	336,614	G/A	间区 Intergenic	22.6	—	1.88E-09	6.50E-08
Scaffold211	1,175,079	G/A	间区 Intergenic	19.2	1.56E-07	8.21E-09	—
	1,175,109	T/A	间区 Intergenic	19.2	1.56E-07	8.21E-09	—
	1,175,111	C/T	间区 Intergenic	19.2	1.56E-07	8.21E-09	—
	1,175,135	C/T	间区 Intergenic	19.2	1.56E-07	8.21E-09	—
	1,175,177	C/T	间区 Intergenic	19.2	1.55E-07	8.32E-09	—
Scaffold2268	2,777,884	G/A	间区 Intergenic	19.4	1.33E-07	1.43E-08	—
	2,777,990	G/A	间区 Intergenic	20.4	1.26E-07	4.63E-09	—
Scaffold2311	361,209	C/T	间区 Intergenic	18.5	1.00E-07	3.44E-09	—
Scaffold2800	1,171,675	A/T	间区 Intergenic	19.8	8.69E-08	6.93E-09	—
Scaffold2942	113,797	G/A	间区 Intergenic	20.0	1.15E-07	1.00E-08	—
	113,806	C/T	间区 Intergenic	20.0	1.15E-07	1.00E-08	—

(续表 4)

染色体 Scaffold	SNP 位置 SNP location	SNP 变异 SNP variation	功能区 Functional area	贡献率 Contribution rate (%)	P 值 P-value		
					2018	2019	2020
Scaffold30	1,872,412	G/T	间区 Intergenic	20.2	—	6.24E-09	7.68E-08
Scaffold3229	195,605	C/A	间区 Intergenic	17.1	—	1.56E-08	1.53E-07
	1,193,442	C/T	间区 Intergenic	20.7	—	2.28E-09	3.86E-09
Scaffold3317	78,232	C/A	内含子 Intron	23.9	—	2.35E-10	7.93E-10
	78,432	G/A	内含子 Intron	21.8	—	2.12E-09	1.47E-09
Scaffold3395	141,462	C/T	间区 Intergenic	20.4	8.41E-09	3.16E-10	
	141,471	G/A	间区 Intergenic	20.4	8.41E-09	3.16E-10	
	141,479	G/A	间区 Intergenic	20.4	8.41E-09	3.16E-10	
	141,537	G/A	间区 Intergenic	20.4	8.41E-09	3.16E-10	
Scaffold349	3,542,803	G/A	间区 Intergenic	21.5	8.21E-09	4.86E-10	
	3,543,009	G/A	间区 Intergenic	22.0	4.03E-09	3.10E-10	
	3,543,014	G/A	间区 Intergenic	21.0	6.06E-09	9.46E-10	
	3,543,021	G/T	间区 Intergenic	20.8	8.62E-09	1.16E-09	
	3,543,061	C/T	间区 Intergenic	21.0	7.19E-09	9.75E-10	
	3,543,082	G/A	间区 Intergenic	21.0	6.52E-09	9.75E-10	
Scaffold3611	1,622,374	A/C	间区 Intergenic	18.7	—	9.80E-09	8.04E-08
	1,622,376	A/G	间区 Intergenic	18.7	—	9.80E-09	8.04E-08
Scaffold4122	389,378	G/A	间区 Intergenic	19.5	—	2.34E-09	9.22E-10
	444,511	C/T	间区 Intergenic	18.8	—	2.51E-09	7.63E-10
	444,749	C/A	间区 Intergenic	18.8	—	2.51E-09	7.63E-10
	444,760	G/A	间区 Intergenic	19.4	—	2.37E-09	7.33E-10
Scaffold4659	1,061,780	G/A	间区 Intergenic	19.0	3.62E-08	2.72E-09	—
Scaffold4817	276,251	C/T	间区 Intergenic	17.8	6.68E-08	5.92E-09	—
Scaffold522	69,932	T/C	内含子 Intron	20.2	8.67E-08	6.37E-09	—
	69,936	T/C	内含子 Intron	20.2	8.67E-08	6.37E-09	—
Scaffold708	2,098,580	G/A	间区 Intergenic	19.8	—	7.24E-09	1.21E-07
	2,098,607	G/A	间区 Intergenic	20.4	—	4.00E-09	6.13E-08
	2,098,614	C/T	间区 Intergenic	19.9	—	7.01E-09	1.20E-07
	2,098,617	G/A	间区 Intergenic	19.8	—	7.24E-09	1.21E-07
	2,098,661	A/G	间区 Intergenic	19.5	—	8.39E-10	1.67E-08
	2,098,666	G/C	间区 Intergenic	18.0	—	4.86E-09	5.84E-08
	2,098,674	G/A	间区 Intergenic	18.0	—	4.86E-09	5.84E-08
	2,098,695	T/C	间区 Intergenic	18.0	—	4.86E-09	5.84E-08
Scaffold7165	292,911	A/T	下游 Downstream	22.7	9.53E-09	7.52E-10	—
Scaffold8160	383,511	C/T	内含子 Intron	18.6	8.16E-08	5.10E-09	—
Scaffold8614	526,707	A/G	间区 Intergenic	20.8	7.33E-08	3.20E-09	—
	526,802	C/T	间区 Intergenic	19.0	—	1.36E-09	5.85E-08
Scaffold8631	103,446	G/A	间区 Intergenic	21.9	—	1.29E-09	2.71E-08
	103,448	T/C	间区 Intergenic	21.7	—	1.73E-09	3.43E-08
	103,743	G/A	间区 Intergenic	21.9	—	1.29E-09	2.71E-08
Scaffold877	1,740,414	C/T	间区 Intergenic	19.8	—	1.01E-08	1.39E-07
	1,740,416	G/A	间区 Intergenic	22.0	—	1.00E-09	2.50E-08
	1,740,418	G/A	间区 Intergenic	21.2	—	2.37E-09	2.12E-08
Scaffold89	2,029,626	A/G	内含子 Intron	19.1	—	7.64E-09	1.18E-07
Scaffold9179	390,469	A/T	内含子 Intron	19.4	—	9.31E-09	1.22E-07

表 5 候选基因信息

Table 5 Candidate gene information

染色体 Scaffold	SNP 位置 SNP location	距离 Distance (kb)	候选基因 Candidate gene	功能注释 Function annotation
Scaffold106	940,480–940,538	5'-37.4	TEA029079.1	MYB 转录因子 MYB Transcription factor
		5'-65.0	TEA029081.1	丝氨酸/苏氨酸蛋白磷酸酶 Serine/threonine-protein kinase
		5'-76.2	TEA029085.1	脱落酸受体 PYL2 Abscicic acid receptor PYL2
Scaffold1262	982,460–982,581	3'-29.7	TEA016845.1	MYB 转录因子 MYB Transcription factor
Scaffold1318	1,188,898	5'-73.4	TEA023905.1	PTB 调控因子 PTB Regulation factor
		5'-58.4	TEA023908.1	亮氨酸重复受体蛋白激酶 Leucine-rich repeat receptor-like protein kinase
		5'-68.5	TEA023909.1	亮氨酸重复受体蛋白激酶 Leucine-rich repeat receptor-like protein kinase
		5'-11.7	TEA023910.1	莽草酸邻羟基肉桂酰转移酶 Shikimate O-hydroxy cinnamoyl transferase
Scaffold1611	610,716–610,786	5'-17.0	TEA001327.1	丝氨酸/苏氨酸蛋白激酶 Serine/threonine-protein kinase
Scaffold1671	583,382–583,566	5'-47.6	TEA026622.1	抗病蛋白 Disease resistance protein
		3'-87.8	TEA026619.1	乙烯反应转录因子 Ethylene-responsive transcription factor ERF
		5'-53.1	TEA026637.1	抗病蛋白 Disease resistance protein
Scaffold209	336,614	5'-99.2	TEA015065.1	半乳糖醛酸转移酶 Galacturonosyltransferase
		5'-44.9	TEA015072.1	丝氨酸/苏氨酸蛋白激酶 Serine/threonine-protein kinase
Scaffold211	1,175,079–1,175,177	3'-66.4	TEA033378.1	糖基转移酶 Glycosyltransferase
		5'-37.2	TEA033379.1	细胞色素 P450 Cytochrome P450
		3'-12.1	TEA033380.1	未知蛋白 Uncharacterized protein
		3'-21.8	TEA033381.1	MOR1 蛋白 MOR1 Protein
		3'-6.6	TEA033375.1	未知蛋白 Uncharacterized protein
Scaffold2268	2,777,884–2,777,990	5'-53.5	TEA021079.1	抗病蛋白 Disease resistance protein
		5'-59.0	TEA021105.1	亮氨酸重复受体蛋白激酶 Leucine-rich repeat receptor-like protein kinase
		5'-95.5	TEA021068.1	TMV 抗性蛋白 TMV Resistance protein
Scaffold2311	361,209	5'-79.8	TEA030505.1	丝氨酸/苏氨酸蛋白激酶 Serine/threonine-protein kinase
		5'-42.6	TEA030506.1	未知蛋白 Uncharacterized protein
		5'-7.7	TEA030509.1	未知蛋白 Uncharacterized protein
Scaffold2800	1,171,675	3'-23.5	TEA002738.1	钙转运 ATP 酶 Calcium-transporting ATPase
		5'-1.6	TEA002761.1	信号蛋白 Signalling protein
		5'-7.7	TEA002758.1	信号蛋白 Signalling protein
		5'-67.9	TEA002766.1	抗性蛋白 Resistance protein
Scaffold2942	113,797–113,806	5'-19.5	TEA002755.1	丝氨酸羧肽酶 Serine carboxypeptidase
		5'-94.2	TEA018613.1	F-盒子调控蛋白 F-box protein
Scaffold30	1,872,412	5'-32.2	TEA009798.1	淀粉合成酶 Starch synthase
		5'-71.1	TEA009811.1	淀粉合成酶 Starch synthase

(续表 5)

染色体 Scaffold	SNP 位置 SNP location	距离 Distance (kb)	候选基因 Candidate gene	功能注释 Function annotation
		3'-74.6	<i>TEA009793.1</i>	叶绿素结合蛋白 Chlorophyll <i>a-b</i> binding protein
Scaffold3229	1,195,605	3'-15.3	<i>TEA007727.1</i>	微管结合蛋白 Microtubule binding protein
		3'-84.1	<i>TEA007735.1</i>	微管结合蛋白 Microtubule binding protein
		5'-36.8	<i>TEA007726.1</i>	dof 转录因子 dof transcription factor
Scaffold3317	78,232–78,432	5'-18.7	<i>TEA014251.1</i>	赤霉素氧化酶 Gibberellin oxidase
		5'-43.8	<i>TEA014250.1</i>	膜相关激酶调节器 Membrane-associated kinase regulator
Scaffold3395	141,462–141,537	5'-29.2	<i>TEA023356.1</i>	未知蛋白 Uncharacterized protein
		3'-44.7	<i>TEA023352.1</i>	热激蛋白 Heat shock 70 kD protein
Scaffold349	3,542,803–3,543,082	5'-8.6	<i>TEA023826.1</i>	异戊烯基二磷酸合酶/异戊烯基转移酶 Isoprenyl diphosphate synthase/prenyltransferase
		5'-10.9	<i>TEA023811.1</i>	核糖体蛋白 Ribosomal protein
		5'-33.9	<i>TEA023806.1</i>	翻译起始因子 Translation initiation factor
		5'-85.8	<i>TEA023801.1</i>	RNA 聚合酶 II 转录亚基 RNA Polymerase II transcription subunit
		3'-89.3	<i>TEA023809.1</i>	RNA 聚合酶 II 转录亚基 RNA Polymerase II transcription subunit
Scaffold3611	1,622,374–1,622,376	5'-4.5	<i>TEA013393.1</i>	WRKY 转录因子 WRKY Transcription factor
		3'-14.9	<i>TEA013410.1</i>	UDP-糖基转移酶 92A1 UDP-glycosyltransferase 92A1
Scaffold4122	444,511–444,760	3'-56.3	<i>TEA027152.1</i>	F-盒子调控蛋白 F-box protein
		3'-69.0	<i>TEA027153.1</i>	F-盒子调控蛋白 F-box protein
Scaffold4659	1,061,780	3'-30.6	<i>TEA001887.1</i>	细胞壁相关受体激酶 Wall-associated receptor kinase
		5'-87.4	<i>TEA001888.1</i>	Beta-1,3-葡聚糖酶 Beta-1,3-glucanase
Scaffold4817	276,251	3'-71.7	<i>TEA006075.1</i>	富含亮氨酸重复序列的蛋白激酶 Leucine-rich receptor-like protein kinase
		5'-86.7	<i>TEA006076.1</i>	抗病蛋白 RPM1 Disease resistance protein RPM1
		5'-30.9	<i>TEA006078.1</i>	线粒体蛋白 Mitochondria protein
Scaffold522	69,932–69,936	3'-11.3	<i>TEA004895.1</i>	抗病蛋白 Disease resistance protein
		3'-50.6	<i>TEA004897.1</i>	转录因子 ORG2 Transcription factor ORG2
		3'-12.1	<i>TEA004887.1</i>	未知蛋白 Uncharacterized protein
Scaffold708	2,098,580–2,098,695	3'-40.7	<i>TEA027473.1</i>	MLO 蛋白 MLO-like protein
		5'-19.6	<i>TEA027474.1</i>	果胶酯酶/果胶酯酶抑制剂 Pectinesterase/pectinesterase inhibitor
		3'-92.8	<i>TEA027481.1</i>	果胶酯酶抑制剂 Pectinesterase inhibitor
		3'-58.4	<i>TEA027489.1</i>	未知蛋白 Uncharacterized protein
		5'-54.2	<i>TEA027491.1</i>	肽基脯氨酸顺反异构酶 Peptidylprolyl <i>cis/trans</i> isomerase
		5'-65.8	<i>TEA027494.1</i>	uncharacterized protein
		3'-74.6	<i>TEA027468.1</i>	转录激活因子 Transcriptional activator
		5'-66.7	<i>TEA027497.1</i>	未知蛋白 Uncharacterized protein

(续表 5)

染色体 Scaffold	SNP 位置 SNP location	距离 Distance (kb)	候选基因 Candidate gene	功能注释 Function annotation
Scaffold7165	292,911	3'-4.3	TEA009868.1	钙转运 ATP 酶 Calcium-transporting ATPase
		5'-4.5	TEA009869.1	ADP 核糖基化因子 ADP-ribosylation factor
Scaffold8160	383,511	3'-21.1	TEA009860.1	富含脯氨酸受体样蛋白激酶 Proline-rich receptor-like protein kinase
		3'-89.4	TEA009862.1	阳离子逆向转运蛋白 Cation/H(+) antiporter
		5'-26.1	TEA009861.1	抗病蛋白 Disease resistance protein
		3'-15.9	TEA009863.1	未知蛋白 Uncharacterized protein
		3'-12.9	TEA009864.1	核糖体蛋白 Ribosomal protein
		3'-14.1	TEA009865.1	亚硫酸盐输出蛋白 Sulfite exporter protein
		5'-49.9	TEA004103.1	钙调素结合蛋白 Calmodulin-binding protein
		3'-8.8	TEA006901.1	Ras 相关蛋白 Ras-related protein
Scaffold8631	103,446-103,743	5'-8.5	TEA006902.1	Ras 相关蛋白 Ras-related protein
		3'-70.9	TEA006898.1	未知蛋白 Uncharacterized protein
		3'-12.5	TEA030188.1	铜离子转运蛋白 Copper transporter
Scaffold877	1,740,414-1,740,418	3'-46.8	TEA030196.1	阳离子/钙交换剂 Cation/calcium exchanger
		5'-45.1	TEA030182.1	鸟嘌呤核苷酸结合蛋白 Guanine nucleotide-binding protein
		3'-61.6	TEA030186.1	枯草杆菌蛋白酶 Subtilisin-like protease
		3'-47.5	TEA030202.1	阳离子/钙交换剂 Cation/calcium exchanger
		5'-8.7	TEA030204.1	未知蛋白 Uncharacterized protein
		3'-86.2	TEA030207.1	蛋白酶体亚基 Proteasome subunit
		5'-86.0	TEA018925.1	Beta-1,3-葡聚糖酶 Beta-1,3-glucanase
		3'-36.5	TEA027247.1	丝氨酸/苏氨酸蛋白激酶 Serine/threonine-protein kinase
Scaffold9179	390,469	3'-89.2	TEA027245.1	醛酮还原酶 Aldo-keto reductase

2.6 极端表型材料 RNA-Seq 分析

利用 RNA-Seq 获得 2 个香叶醇樱草糖苷极端含量种质差异表达基因共 4393 个, 其中显著上调表达基因 2049 个(图 2-A), 显著下调表达基因 2344 个(图 2-B)。GO 功能分析结果表明, 差异表达基因生物过程最多的 2 类均是细胞过程(cellular process)、代谢过程(metabolic process), 细胞结构最多的 2 类均是膜(membrane)、膜构件(membrane part), 分子功能最多的 2 类均是结合活性(binding)、催化活性(catalytic activity) (图 3-A, B)。KEGG 功能分析结果表明, 上调差异表达基因主要富集在油菜素甾醇、类黄酮、核糖体蛋白、氨基酸等代谢途径(图 4-A), 下调差异表达基因主要富集在单萜类、倍半萜类、花生四烯酸类、玉米素类等代谢途径(图 4-B)。

2.7 香叶醇樱草糖苷含量候选基因筛选

将 GWAS 分析得到的显著关联 SNP 标记上、下游 100 kb 范围内的相关基因在 2 个香叶醇樱草糖苷极端含量种质差异表达基因中搜索, 得到重叠基因 10 个, 均表现为高含量种质中表达量高于低含量种质。利用栽培品种悦茗香、福云 6 号进一步验证 10 个重叠基因的表达水平发现, 10 个重叠基因在香叶醇樱草糖苷高含量品种悦茗香中表达量亦高于香叶醇樱草糖苷低含量品种福云 6 号(表 6), 故将 10 个重叠基因初步视为候选基因。

2.8 候选基因编码区序列差异及上游作用元件分析

对 2 个极端香叶醇樱草糖苷极端含量种质进行基因组 DNA 重测序, 10 个候选基因编码区序列在 2

个极端种质之间均存在 SNP 非同义突变, 造成了不同种质之间编码蛋白氨基酸序列的差异(表 7)。利用 Plant CARE 在线工具对所有 10 个候选基因上游 2 kb 的区域进行了预测(表 8)。在所有类型的元件中, 与激素响应相关的元件及与环境胁迫相关的元件是代表植物核心生理过程的调控元件。候选基因中具有与

激素响应相关的顺式作用元件的基因占 8 个(除 TEA023801.1、TEA009864.1 外), 其中与赤霉素、脱落酸和茉莉酸甲酯有关的顺式作用元件含量较丰富。候选基因中具有与环境胁迫相关的顺式作用元件的基因占 9 个(除 TEA009864.1 外), 其中与防御和应激反应性、干旱诱导、无氧诱导有关的元件含量较丰富。

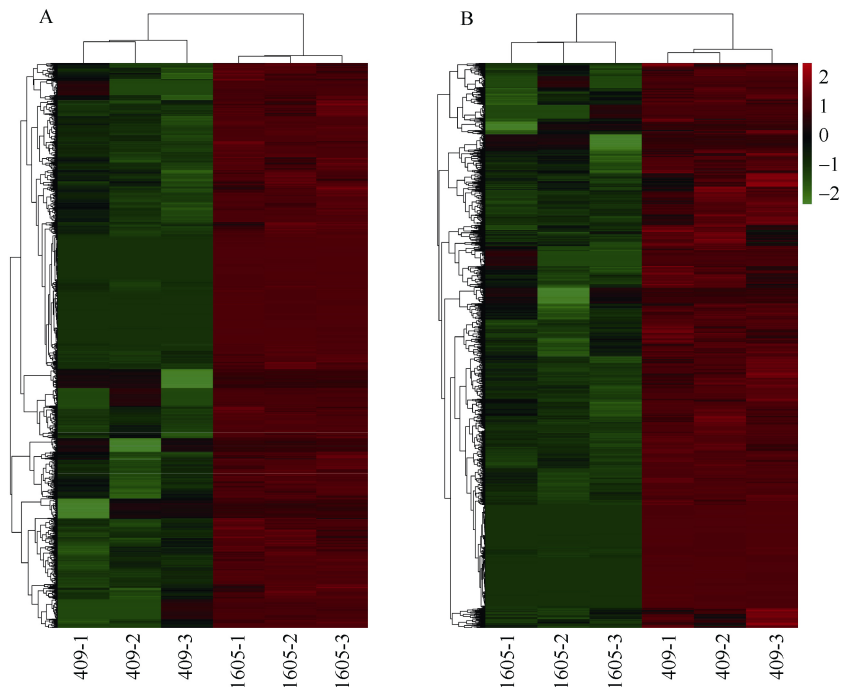
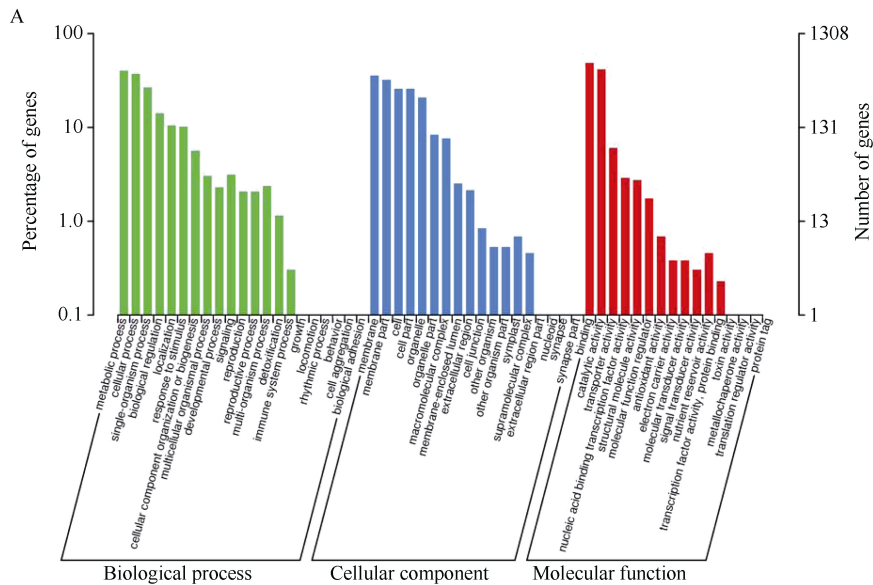


图 2 差异表达基因

Fig. 2 Differentially expressed genes

A: 409 vs 1605 显著上调表达; B: 409 vs 1605 显著下调表达。

A: 409 vs 1605 significantly up regulated; B: 409 vs 1605 significantly down regulated.



(图 3)

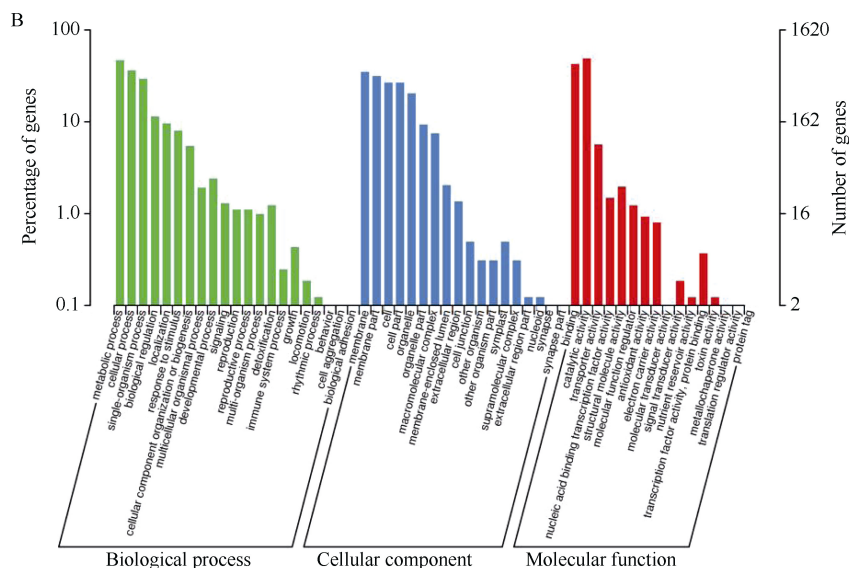


图 3 差异表达基因 GO 功能分析

Fig. 3 GO functional analysis of differentially expressed genes

A: 409 vs 1605 显著上调表达; B: 409 vs 1605 显著下调表达。

A: 409 vs 1605 significantly up regulated; B: 409 vs 1605 significantly down regulated.

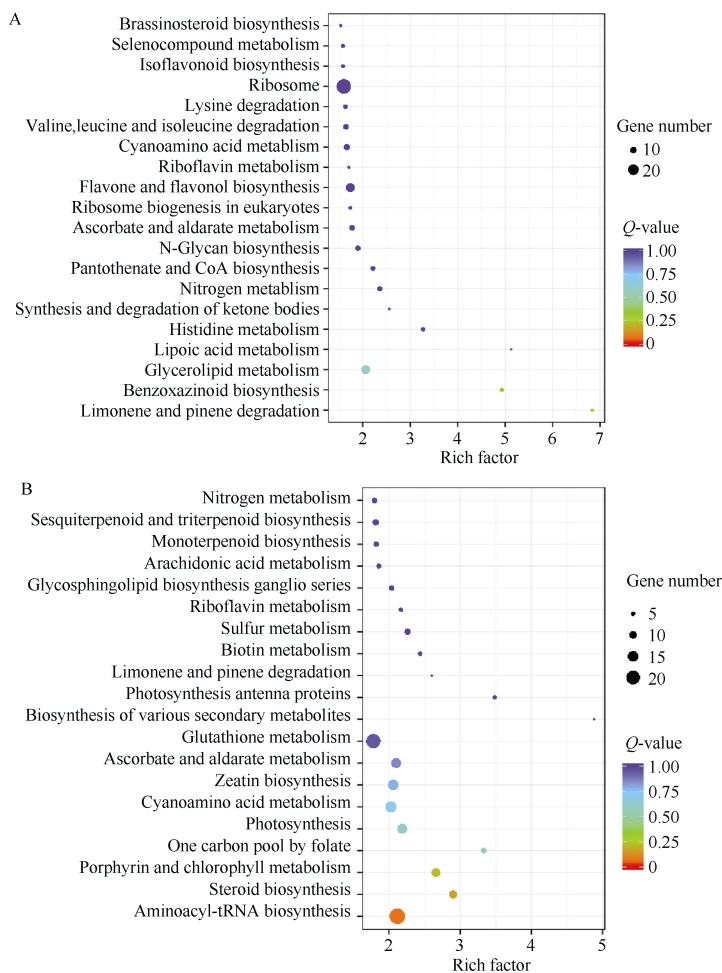


图 4 差异表达基因 KEGG 功能分析

Fig. 4 KEGG functional analysis of differentially expressed genes

A: 409 vs 1605 显著上调表达; B: 409 vs 1605 显著下调表达。

A: 409 vs 1605 significantly up regulated; B: 409 vs 1605 significantly down regulated.

表 6 香叶醇樱草糖苷含量候选基因预测
Table 6 Prediction of candidate genes of geranyl β-primeveroside contents

染色体 Scaffold	候选基因 Candidate gene	基因注释 Gene annotation	每百万比对数目 Counts per million			
			409	1605	悦茗香* Yuemingxiang*	福云 6 号* Fuyun 6*
Scaffold1318	TEA023908.1	富含亮氨酸重复类受体蛋白激酶 Leucine-rich repeat receptor-like protein kinase	7.09	2.31	16.88	1.22
Scaffold4817	TEA006075.1	富含亮氨酸重复类受体蛋白激酶 Leucine-rich repea treceptor-like protein kinase	2.73	0.23	3.89	0.31
Scaffold106	TEA029081.1	丝氨酸/苏氨酸蛋白激酶 Serine/threonine-protein kinase	20.21	8.70	17.13	6.32
Scaffold9179	TEA027247.1	丝氨酸/苏氨酸蛋白激酶 Serine/threonine-protein kinase	11.93	5.26	22.37	4.63
Scaffold349	TEA023811.1	核糖体蛋白 Ribosomal protein	12.40	4.89	21.10	5.69
Scaffold8160	TEA009864.1	核糖体蛋白 Ribosomal protein	65.70	24.41	109.79	33.77
Scaffold349	TEA023801.1	RNA 聚合酶 II 转录亚基 RNA polymerase II transcription subunit	139.33	70.50	147.46	59.41
Scaffold4659	TEA001888.1	Beta-1,3-葡聚糖酶 Beta-1,3-glucanase	7.65	1.09	8.94	1.92
Scaffold89	TEA018925.1	Beta-1,3-葡聚糖酶 Beta-1,3-glucanase	236.17	130.12	190.25	150.34
Scaffold4817	TEA006076.1	抗病蛋白 RPM1 Disease resistance protein RPM1	12.33	0.85	13.07	1.31

悦茗香为适制乌龙茶品种，芽中香叶醇樱草糖苷含量高，福云 6 号为适制绿茶品种，芽中香叶醇樱草糖苷含量低。
Yuemingxiang is a variety suitable for manufacturing oolong tea, the content of geranyl β-primeveroside in buds is high. Fuyun 6 is a variety suitable for manufacturing green tea, the content of geranyl β-primeveroside in buds is low.

表 7 2 个极端种质中差异基因编码区序列变异
Table 7 Sequence variation of differential gene coding region in two extreme germplasms

候选基因 Candidate gene	SNP 位置 SNP location	409 DNA 序列 409 DNA sequence	1605 DNA 序列 1605 DNA sequence	409 氨基酸序列 409 amino acid sequence	1605 氨基酸序列 1605 amino acid sequence
TEA006075.1	348,400	T/A	T	Lys/Ter	Lys
	348,421	C/G	C	Glu/Gln	Glu
	348,436	G/C	G	Leu/Val	Leu
	348,467	G/A	G	Ala/Val	Ala
	348,488	A/G	A	Met/Thr	Met
	348,524	C/T	C	Gly/Glu	Gly
	348,764	C/T	C	Ser/Asn	Ser
	348,797	C/T	C	Cys/Tyr	Cys
	348,830	G/A	G	Ser/Phe	Ser
	349,130	G/A	G	Ser/Phe	Ser
	349,248	A/T	A	Tyr/Ter	Tyr
	349,283	G/A	G	Ser/Phe	Ser
	349,370	T/A	T	His/Leu	His
	349,421	T/C	T	Ter/Trp	Ter
	349,536	G/C	G	Phe/Leu	Phe
	349,550	A/T	A	Ile/Lys	Ile
	349,625	C/T	C	Trp/Ter	Trp

(续表 7)					
候选基因	SNP 位置	409 DNA 序列	1605 DNA 序列	409 氨基酸序列	1605 氨基酸序列
Candidate gene	SNP location	409 DNA sequence	1605 DNA sequence	409 amino acid sequence	1605 amino acid sequence
TEA006075.1	349,633	T/G	T	Asn/His	Asn
	350,501	C/T	C	Trp/Ter	Trp
	350,647	G/A	G	Thr/Ile	Thr
TEA023908.1	1,127,395	A	G	Ile	Thr
	1,127,410	G	G/A	Ala	Ala/Val
	1,127,793	C/G	C	Gln	Glu
	1,127,867	T/A	T	Ser	Arg
	1,127,989	C/T	C	Asn	Ser
	1,128,205	A/G	A	Ala	Val
	1,128,366	G	G/C	Gln	Gln/Glu
	1,128,407	G	G/T	Phe	Phe/Leu
	1,128,445	A	A/G	Ile	Ile/Thr
	1,128,890	T	T/C	Ile	Ile/Met
	1,129,004	G/T	T	His/Gln	Gln
	1,130,137	A	A/G	Ile	Ile/Thr
	1,130,398	A/G	A	Ile/Thr	Ile
	867,063	A/G	A	Leu/Ser	Leu
	867,147	C/T	T	Ser/Asn	Asn
TEA029081.1	867,194	G	C	Pro	Ala
	867,221	C	A	Val	Leu
	867,354	A/G	A	Leu/Ser	Leu
	869,008	C	G	Arg	Thr
	869,038	T/A	T	His/Leu	His
	874,334	T	G	Lys	Gln
	187,415	C/A	C	Leu/Phe	Leu
	187,442	G/T	G	Cys/Ter	Cys
TEA006076.1	189,274	C/G	C	Cys/Ser	Cys
	189,283	T/C	T	Gln/Arg	Gln
	3,449,354	G/C	C	Tyr/Ter	Ter
	3,453,357	A	A/G	Leu	Leu/Pro
TEA023811.1	3,531,058	T	T/A	Lys	Lys/Ter
TEA001888.1	1,154,774	T	T/G	Met	Met/Leu
TEA009864.1	370,433	G/A	A	Ala/Val	Val
TEA027247.1	358,366	A	A/T	Phe	Phe/Ile
TEA018925.1	1,943,138	C	A	Arg	Ser

表 8 茶树香叶醇樱草糖苷候选基因上游与环境相关的顺式作用元件的预测

Table 8 Prediction of environmental related *cis*-elements in the upstream regions of candidate genes for tea geraniol primrose glycoside

候选基因	激素响应元件 Hormone responsive elements					环境胁迫相关元件 Environmental stress-related elements					
	IAA	GA	SA	ABA	MeJA	ASI	WR	LTR	DSR	AI	DI
TEA029081.1			2		2			1		3	2
TEA023908.1	1		2		2				1	2	
TEA023801.1								1		1	
TEA023811.1		2		4						1	

(续表 8)

候选基因 Candidate gene	激素响应元件 Hormone responsive elements					环境胁迫相关元件 Environmental stress-related elements					
	IAA	GA	SA	ABA	MeJA	ASI	WR	LTR	DSR	AI	DI
TEA001888.1	1	1					1		1	4	
TEA006075.1		2	1	5	4			1	1	5	
TEA006076.1	1	1		1	4				1		
TEA009864.1											
TEA018925.1	1	2			2	1			2	1	1
TEA027247.1				3	4			1		3	2

数字表示各种顺式作用元件的总和。IAA: 生长素反应性; GA: 赤霉素反应性; SA: 水杨酸反应性; ABA: 脱落酸反应性; MeJA: 茉莉酸甲酯反应性; ASI: 缺氧特异性诱导; WR: 伤口反应性; LTR: 低温反应性; DSR: 防御和应激反应性; AI: 无氧诱导; DI: 干旱诱导。
The number indicates the sum of various cis-regulatory elements. IAA: auxin responsiveness; GA: gibberellin responsiveness; SA: salicylic acid responsiveness; ABA: abscisic acid responsiveness; MeJA: MeJA (methyl jasmonate) responsiveness; ASI: anoxic specific induction; WR: wound responsive; LTR: low temperature responsiveness; DSR: defense and stress responsiveness; AI: anaerobic induction; DI: drought inducibility.

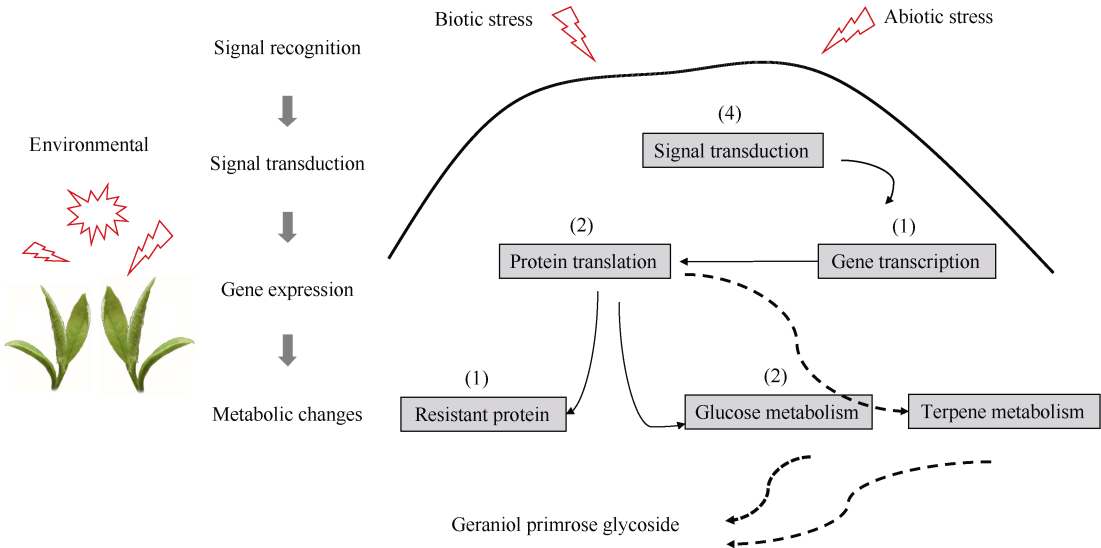


图 5 候选基因示意图
Fig. 5 Schematic diagram of candidate genes
括号中的数字表示基因个数。信号转导: TEA023908.1, TEA006075.1, TEA029081.1, TEA027247.1; 基因转录: TEA023801.1; 蛋白翻译: TEA023811.1, TEA009864.1; 抗病蛋白: TEA006076.1; 糖代谢: TEA001888.1, TEA018925.1。
Numbers in brackets indicate the number of genes. Signal transduction: TEA023908.1, TEA006075.1, TEA029081.1, TEA027247.1; Gene transcription: TEA023801.1; Protein translation: TEA023811.1, TEA009864.1; Resistant protein: TEA006076.1; Glucose metabolism: TEA001888.1, TEA018925.1.

3 讨论

分析了参试茶树种质在 3 个环境下新梢中香叶醇樱草糖苷含量的遗传差异, 其广义遗传力为 62.6%, 说明香叶醇樱草糖苷含量变异主要受遗传控制, 其候选基因编码区序列均存在 SNP 非同义突变, 预示可以通过遗传改良手段定向调控茶树新梢中香叶醇樱草糖苷含量。候选基因中多数基因上游 2 kb 区域存在不同数量的与激素信号转导和环境胁迫相关的顺式作用元件, 预示激素水平和环境变化会影响茶树香叶醇樱草糖苷的积累。候选基因涉及

到信号转导、基因表达、代谢变化等多个过程(图 5), 说明茶树新梢香叶醇樱草糖苷积累的遗传调控机制是一个复杂的体系, 由基因和环境共同决定, 它们之间的相互作用仍有待深入研究。
富含亮氨酸重复类受体蛋白激酶(LRR-RLK)包含 1 个胞外的富含亮氨酸(LRR)结构域、1 个跨膜结构域和 1 个胞内丝氨酸/苏氨酸(Ser/Thr)激酶结构域^[35]。LRR-RLK 定位在细胞质膜上, 能够感知外界环境信号, 然后通过磷酸化将信号传递到胞内相应部位, 可作为信号感知元件参与植物体应答生物

和非生物胁迫过程^[36]。不同的 LRR-RLK 中 LRR 数目和排布有较大的区别, LRR 的多样性使 LRR-RLK 能够特异的识别不同的信号分子, 从而准确传递不同的胞外信号^[37]。水稻接种稻瘟病菌后, *OsBRR1* 在叶片中的表达受强烈诱导, 同时过量表达 *OsBRR1* 转基因植株可以提高稻瘟病抗性^[38]。水稻中的一个 LRR-RLK 蛋白激酶还能正调控与生物胁迫相关的 MAPK 及 WRKY 表达水平, 从而调控植物对昆虫取食的防御机制^[39]。烟草中 *LRR-RLK* 基因参与了盐胁迫反应的调控过程^[40]。辣椒中 LRR 类受体蛋白激酶基因 *CaLRR1* 的表达受病原菌侵染的诱导^[41]。丝氨酸/苏氨酸蛋白激酶在接受到生物和非生物胁迫信号后迅速在其丝氨酸、苏氨酸部位发生磷酸化而被激活, 并通过对下游分子的磷酸化作用, 最终将外界信号传递给细胞核, 调节特异基因的表达^[42]。本研究发现 2 个富含亮氨酸重复类受体蛋白激酶基因(*TEA023908.1*、*TEA006075.1*)和 2 个丝氨酸/苏氨酸蛋白激酶(*TEA029081.1*、*TEA027247.1*)与香叶醇樱草糖苷的含量有关, 推测这些蛋白激酶作为信号转导分子会将环境胁迫信号传递到代谢网络下游, 导致茶树香叶醇樱草糖苷发生代谢变化, 从而引发香叶醇在茶树与环境互作中发挥重要作用。

茶树体内樱草糖苷可在糖基转移酶作用下通过葡萄糖基化和木糖基化顺序合成, 虽然鼠李糖、半乳糖、木糖、葡萄糖醛酸均可作为糖基转移酶的活性糖供体, 但一般认为葡萄糖是最常见的糖基供体^[18]。 β -1,3-葡聚糖酶能够水解 β -1,3-葡聚糖成为葡萄糖和小分子寡糖, 在正常情况下 β -1,3-葡聚糖酶表达水平较低, 当植物遭受生物因子以及非生物因子诱导后, 其表达活性会显著增强^[43]。研究表明, 棉花 β -1,3-葡聚糖酶基因参与了棉花对黄萎病的防卫反应^[44]。烟草 β -1,3-葡聚糖酶基因转化番茄可使番茄发病率降低 36%~58%^[45]。本研究鉴定到 2 个 β -1,3-葡聚糖酶基因(*TEA001888.1*、*TEA018925.1*)与香叶醇樱草糖苷累积相关, 推测 β -1,3-葡聚糖酶通过水解作用产生葡萄糖为香叶醇樱草糖苷生物合成提供糖配基并参与调节香叶醇的累积和释放, 从而抵御生物和非生物胁迫。

植物在生长发育过程中为了应对复杂的环境变化, 进化形成了 2 个不同层次的防御反应, 初级免疫反应(PTI)和次级免疫反应(ETI), 其中 ETI 比 PTI 更为强烈和迅速, 可以通过局部程序性细胞死亡引起植物超敏反应, 能更为有效地遏制病原菌的进一

步扩散^[46]。RPM1 是一种植物抗病蛋白, 可以触发 ETI 使植物获得抗病性^[47]。本研究发现 1 个抗病蛋白 RPM1 基因(*TEA006076.1*)与香叶醇樱草糖苷含量相关, 推断香叶醇樱草糖苷以及香叶醇在与环境互作过程中会协同产生抗病蛋白 RPM1, 从而提高茶树抗病性。

植物萜类化合物的合成途径分为甲羟戊酸(MVA)和 2-C-甲基-D-赤藓糖醇-4-磷酸(MEP)2 种途径, 前者位于细胞质中, 后者位于质体中, 二者虽有所不同, 但共同点都以异戊烯基焦磷酸(IPP)为主要中间产物。IPP 与其异构体二甲丙烯焦磷酸(DMAPP)在异戊烯基二磷酸合酶(IPPS)催化下生成香叶基焦磷酸(GPP), 后者在萜类合成酶(TPS)的作用下生成香叶醇^[48]。本研究在 Scaffold349 上 280 bp 范围内鉴定到 6 个 SNP, 分别位于 3,542,803 bp、3,543,009 bp、3,543,014 bp、3,543,021 bp、3,543,061 bp、3,543,082 bp 位置, 在该区段 10.9 kb 和 85.8 kb 附近区域分别存在 1 个核糖体蛋白基因(*TEA023811.1*)和 1 个 RNA 聚合酶 II 转录亚基基因(*TEA023801.1*), 转录组测序表明这 2 个基因在香叶醇樱草糖苷极端含量种质中均表现显著差异表达, 说明他们可能与香叶醇樱草糖苷含量相关。有趣的是在该区段 8.6 kb 区域附近还存在 1 个异戊烯基二磷酸合酶基因(*TEA023826.1*), 推测核糖体蛋白基因(*TEA023811.1*)和 RNA 聚合酶 II 转录亚基基因(*TEA023801.1*)参与异戊烯基二磷酸合酶基因(*TEA023826.1*)的转录和翻译, 进而导致香叶醇及其樱草糖苷代谢和累积变化。

萜类化合物与单糖(或寡糖)一起作为底物, 可在糖基转移酶的催化下生成不易挥发的糖苷化合物。虽然本研究筛选的候选基因中未包括糖基转移酶基因, 但是在 Scaffold3611 的 1,622,376 bp 位置鉴定到 1 个 SNP, 其附近 14.9 kb 区域存在一个糖基转移酶基因(*TEA013410.1*)和一个 WRKY 转录因子基因(*TEA013393.1*), 他们是否参与香叶醇樱草糖苷的生物合成, 仍有待证实。

4 结论

对茶树新梢中香叶醇樱草糖苷含量进行了 GWAS 分析, 初步鉴定到 10 个与香叶醇樱草糖苷含量相关的候选基因, 其编码区序列在香叶醇樱草糖苷极端含量种质之间均存在非同义 SNP 突变, 并且多数基因上游 2 kb 区域存在不同数量的与环境胁迫

和激素相关的顺式作用元件,说明香叶醇樱草糖苷的含量变异与遗传及环境胁迫相关。该结果对茶树香叶醇樱草糖苷含量的遗传调控机制研究及育种实践具有较好的理论和应用参考价值。

References

- [1] Inouye S, Takizawa T, Yamaguchi H. Antibacterial activity of essential oils and their major constituents against respiratory tract pathogens by gaseous contact. *J Antimicrob Chemother*, 2001, 47: 565–573.
- [2] Wei S, Reuveny H, Bravdo B A, Shoseyov O. Hydrolysis of glycosidically bound volatiles from apple leaves (cv. Anna) by *Aspergillus niger* β -glucosidase affects the behavior of codling moth (*Cydia pomonella* L.). *J Agric Food Chem*, 2004, 52: 6212–6216.
- [3] Magnard J, Roccia A, Caissard J, Vergne P, Sun P, Hecquet R, Dubois A, Oyant L, Jullien F, Nicole F, Raymond O, Huguet S, Baltenweck R, Meyer S, Claudel P, Jeauffre J, Rohmer M, Foucher F, Hugueneyp P, Bendahmane M, Baudino S. Biosynthesis of monoterpene scent compounds in roses. *Science*, 2015, 349: 81–83.
- [4] Zhao M Y, Wang L, Wang J M, Jin J Y, Zhang N, Lei L, Gao T, Jing T T, Zhang S R, Wu Y, Wu B, Hu Y Q, Wan X C, Schwab W, Song C K. Induction of priming by cold stress via inducible volatile cues in neighboring tea plants. *J Integr Plant Biol*, 2020, 62: 1461–1468.
- [5] Wang D M, Yoshimura T, Kubota K, Kobayashi A. Analysis of glycosidically bound aroma precursors in tea leaves: I. Qualitative and quantitative analyses of glycosides with aglycons as aroma compounds. *J Agric Food Chem*, 2000, 48: 5411–5418.
- [6] Mizutani M, Nakanishi H, Ema J, Ma S, Noguchi E, Inohara-ochiiai M, Fukachimizutani F, Nakao M, Sakata K. Cloning of β -primeverosidase from tea leaves, a key enzyme in tea aroma formation. *Plant Physiol*, 2002, 130: 2164–2176.
- [7] Sarry J, Gunata Z. Plant and microbial glycoside hydrolases: volatile release from glycosidic aroma precursors. *Food Chem*, 2004, 87: 509–521.
- [8] Bock K W. The UDP-glycosyltransferase (UGT) superfamily expressed in humans, insects and plants: animal-plant arms-race and co-evolution. *Biochem Pharmacol*, 2015, 99: 11–17.
- [9] Stahlbiskup E, Intert F, Holthuijzen J, Stengele M, Schulz G. Glycosidically bound volatiles—a review 1986–1991. *Flavour Frag J*, 1993, 8: 61–80.
- [10] Guo W, Hosoi R, Sakata K, Watanabe N, Yagi A, Ina K, Luo S. (S)-linalyl, 2-phenylethyl, and benzyl disaccharide glycosides isolated as aroma precursors from oolong tea leaves. *Biosci Biotechnol Biochem*, 1994, 58: 1532–1534.
- [11] Candela L, Formato M, Crescente G, Piccolella S, Pacifico S. Coumaroyl flavonol glycosides and more in marketed green teas: an intrinsic value beyond much-lauded catechins. *Molecules*, 2020, 25: 1765.
- [12] Gu X G, Yao C C, Zhang Z Z, Wan X C, Ning J M, Shao W F. GC-ECD method for determination of glucosidically bound aroma precursors in fresh tea leaves. *Chromatographia*, 2011, 73: 189–193.
- [13] Ogawa K, Moon J H, Guo W F, Yagi A, Watanabe N, Sakata K. A study on tea aroma formation mechanism: alcoholic aroma precursor amounts and glycosidase activity in parts of the tea plant. *Zeitschrift Fur Naturforsch Sect C-J Biosci*, 1995, 50: 493–498.
- [14] Wang D M, Kurasawa E, Yamaguchi Y, Kubota K, Kobayashi A. Analysis of glycosidically bound aroma precursors in tea leaves: II. Changes in glycoside contents and glycosidase activities in tea leaves during the black tea manufacturing process. *J Agric Food Chem*, 2001, 49: 1900–1903.
- [15] Dai W D, Tan J F, Lu M L, Xie D C, Li P L, Lyu H P, Zhu Y, Guo L, Zhang Y, Peng Q H, Lin Z. Nontargeted modification-specific metabolomics investigation of glycosylated secondary metabolites in tea (*Camellia sinensis* L.) based on liquid chromatography-high resolution mass spectrometry. *J Agric Food Chem*, 2016, 64: 6783–6790.
- [16] Rawat R, Gulati A. Seasonal and clonal variations in some major glycosidic bound volatiles in Kangra tea (*Camellia sinensis* (L.) O. Kuntze). *Eur Food Res Technol*, 2008, 226: 1241–1249.
- [17] Cui J L, Katsuno T, Totsuka K, Ohnishi T, Takemoto H, Mase N, Toda M, Narumi T, Sato K, Matsuo T, Mizutani K, Yang Z Y, Watanabe N, Tong H R. Characteristic fluctuations in glycosidically bound volatiles during tea processing and identification of their unstable derivatives. *J Agric Food Chem*, 2016, 64: 1151–1157.
- [18] Ohgami S, Ono E, Horikawa M, Murata J, Totsuka K, Toyonaga H, Ohba Y, Dohra H, Asai T, Matsui K, Mizutani M, Watanabe N, Ohnishi T. Volatile glycosylation in tea plants: sequential glycosylations for the biosynthesis of aroma β -primeverosides are catalyzed by two *Camellia sinensis* glycosyltransferases. *Plant Physiol*, 2015, 168: 464–477.
- [19] Carl S R, Stephen G W. Glycosidase mechanisms. *Curr Opin Plant Biol*, 2000, 4: 573–580.
- [20] 杨飞, 张征锋, 南波, 肖本泽. 水稻产量相关性状的全基因组关联分析及候选基因筛选. 作物学报, 2022, 48: 1813–1821.
Yang F, Zhang Z F, Nan B, Xiao B Z. Genome-wide association analysis and candidate gene selection of yield related traits in rice. *Acta Agron Sin*, 2022, 48: 1813–1821 (in Chinese with English abstract).
- [21] 谢磊, 任毅, 张新忠, 王继庆, 张志辉, 石书兵, 耿洪伟. 小麦穗发芽性状的全基因组关联分析. 作物学报, 2021, 47: 1891–1902.
Xie L, Ren Y, Zhang X Z, Wang J Q, Zhang Z H, Shi S B, Geng H W. Genome-wide association study of pre-harvest sprouting traits in wheat. *Acta Agron Sin*, 2021, 47: 1891–1902 (in Chinese with English abstract).
- [22] 渠建洲, 冯文豪, 张兴华, 徐淑兔, 薛吉全. 基于全基因组关联分析解析玉米籽粒大小的遗传结构. 作物学报, 2022, 48: 304–319.
Qu J Z, Feng W H, Zhang X H, Xu S T, Xue J Q. Dissecting the genetic architecture of maize kernel size based on genome-wide association study. *Acta Agron Sin*, 2022, 48: 304–319 (in Chinese with English abstract).
- [23] Wang L, Yang Y M, Zhang S Y, Che Z J, Yuan W J, Yu D Y. GWAS reveals two novel loci for photosynthesis-related traits in soybean. *Mol Genet Genomics*, 2020, 295: 705–716.
- [24] Wang R J, Gao X F, Yang J, Kong X R. Genome-wide association study to identify favorable SNP allelic variations and candidate

- genes that control the timing of spring bud flush of tea (*Camellia sinensis*) using SLAF-seq. *J Agric Food Chem*, 2019, 67: 10380–10391.
- [25] Fang K X, Xia Z Q, Li H J, Jiang X H, Qin D D, Wang Q S, Wang Q, Pan C D, Li B, Wu H L. Genome-wide association analysis identified molecular markers associated with important tea flavor-related metabolites. *Hortic Res*, 2021, 8: 42.
- [26] 王让剑, 苏德森, 吴建衍, 黄崇耀, 陈立松. 超高效液相色谱-串联质谱法测定茶树新梢中两种香叶醇糖苷含量. *茶叶学报*, 2020, 61(3): 114–119.
Wang R J, Su D S, Wu J Y, Huang C Y, Chen L S. UHPLC MS/MS determination of geraniol glycosides in tea shoots. *Acta Tea Sin*, 2020, 61(3): 114–119 (in Chinese with English abstract).
- [27] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 2009, 25: 1754–1760.
- [28] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytisky A, Garimella K, Altshuler D, Gabriel S, Daly M, Depristo A. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, 2010, 20: 1297–1303.
- [29] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*, 2009, 25: 2078–2079.
- [30] Alexander D H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, 2009, 19: 1655–1664.
- [31] Alkes L P, Nick J P, Robert M P, Michael E W, Nancy A S, David R. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, 2006, 38: 904–909.
- [32] Purcell S, Neale B, Todd-brown K, Thomas L, Ferreira M, Bender D, Maller J, Sklar P, Bakker P, Daly M J. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 2007, 81: 559–575.
- [33] Bradbury P J, Zhang Z W, Kroon D E, Casstevens T M, Ramdoss Y, Buckler E S. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 2007, 23: 2633–2635.
- [34] Robinson M D, McCarthy D J, Smyth G K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 2010, 26: 139–140.
- [35] 王愿, 王晓坤, 戈海曼, 杨磊. 拟南芥富含亮氨酸重复序列类受体激酶 AtLRR78A 的定位及其分选序列研究. *植物生理学报*, 2017, 53: 477–486.
Wang Y, Wang X K, Ge H M, Yang L. The localization and trafficking mechanism of AtLRR78A, a leucine-rich repeat receptor-like kinase (LRR-RLK) in *Arabidopsis*. *J Plant Physiol*, 2017, 53: 477–486 (in Chinese with English abstract).
- [36] Klauser D, Desurmont G A, Glauser G, Vallat A, Flury P, Boller T, Turlings T C J, Bartels S. The *Arabidopsis* Pep-PEPR system is induced by herbivore feeding and contributes to JA-mediated plant defence against herbivory. *J Exp Bot*, 2015, 66: 5327–5336.
- [37] Gou X, Li J. Paired receptor and coreceptor kinases perceive extracellular signals to control plant development. *Plant Physiol*, 2020, 182: 1667–1681.
- [38] Peng H, Zhang Q, Li Y D, Lei C L, Zhai Y, Sun X H, Sun D Y, Sun Y, Lu T G. A putative leucine-rich repeat receptor kinase, OsBRR1, is involved in rice blast resistance. *Planta*, 2009, 230: 377–385.
- [39] Hu L, Ye M, Kuai P, Ye M, Erb M, Liu Y. OsLRRRLK1, an early responsive leucine-rich repeat receptor-like kinase, initiates rice defense responses against a chewing herbivore. *New Phytol*, 2018, 219: 1097–1111.
- [40] Dure L. A repeating 11-mer amino acid motif and plant desiccation. *Plant J*, 1993, 3: 363–369.
- [41] Jung E H, Jung H W, Lee S C, Sang W H, Heu S, Hwang B K. Identification of a novel pathogen-induced gene encoding a leucine-rich repeat protein expressed in phloem cells of *Capsicum annuum*. *Biochim Biophys Acta*, 2004, 1676: 211–222.
- [42] Hasegawa P M, Brsecan R A, Zhu J K, Bohnert H J. Plant cellular and molecular responses to high salinity. *Annu Rev Plant Biol*, 2000, 51: 463–499.
- [43] Xie Y R, Ruarung Y, Chen Z Y, Brown R L, Cleveland T E. ZmGns, a maize class I β -1,3-glucanase, is induced by biotic stresses and possesses strong antimicrobial activity. *J Integr Plant Biol*, 2015, 57: 271–283.
- [44] McFadden H G, Chapple R, Feyter R D E, Dennis E. Expression of pathogenesis-related genes in cotton stem in response to infection by *Verticillium dahliae*. *Physiol Mol Plant Pathol*, 2001, 58: 119–131.
- [45] Jongedijk E, Tigelaar H, Vanroekel J S C, Bresvloemans S A, Dekker I, Vandenelzen P J M, Cornelissen B J C, Melchers L S. Synergistic activity of chitinases and β -1,3-glucanases enhances fungal resistance in transgenic tomato plants. *Euphytica*, 1995, 85: 173–180.
- [46] Jones J D G, Dangl J L. The plant immune system. *Nature*, 2006, 444: 323–329.
- [47] Yuan X, Wang Z Y, Huang J Z, Xuan H, Gao Z Y. Phospholipidase D δ negatively regulates the function of resistance to *Pseudomonas syringae* pv. *Maculicola* 1 (RPM1). *Front Plant Sci*, 2019, 9: 1991.
- [48] Aharoni A, Jongsma M A, Bouwmeester H J. Volatile science? Metabolic engineering of terpenoids in plant. *Trends Plant Sci*, 2005, 10: 594–602.