

基于转录组分析白术倍半萜生物合成的相关基因

杨国^{1, #}, 李鸿慧^{1, #}, 金叶飞¹, 董丽佳¹, 莫亿伟^{1, *}, 罗洁^{2, *}

¹绍兴文理学院生命科学学院, 浙江绍兴312000

²绍兴文理学院元培学院, 浙江绍兴312000

摘要: 本研究通过GC-MS分析了白术挥发油中倍半萜的种类, 同时利用Illumina高通量测序技术对白术叶、茎、块根和花的混合样品进行了测序, 最后对白术萜类骨架合成酶以及倍半萜合成酶有关基因的表达进行了实时荧光定量PCR (qRT-PCR)分析。研究结果显示, 白术倍半萜成分主要为苍术酮和8,9-去氢新异长叶烯, 转录组测序共获得了23 099 928条有效短读序, 组装后获得100 037条Unigene, Q30值为93.48%, N50长度为1 027.2 bp。将拼装所得到的Unigene序列利用BLAST工具分别与 NR、Swiss-Prot、KEGG、KOG、COG和GO数据库进行比对, 共有54 471条Unigene序列被注释, 占总序列数的54.45%。对注释得到的Unigene进行KEGG代谢通路分析, 共涉及133个KEGG代谢通路, 其中81条Unigene参与了萜类骨架生物合成, 18条Unigene参与了倍半萜生物合成。利用MISA软件发现12 124个SSR, 其中单碱基重复SSR共6 028个, 双碱基重复SSR共3 866个。本研究对6个倍半萜合成相关的基因进行qRT-PCR分析, 发现其中4个在白术块根中的表达量显著高于叶、茎和花中的表达量。

关键词: 白术; 挥发油; 倍半萜; 转录组测序; SSR; qRT-PCR

白术(*Atractylodes macrocephalae*)是菊科苍术属多年生草本植物, 其根茎干燥后可入药, 为常用的补益类中药, 在中药药方中有“十药九术”之称, 具有健脾益气、燥湿利水、止汗、安胎等的功效(段启等2008)。现代研究表明白术具有降血糖、抗菌消炎、抗衰老、增强免疫功能及抗肿瘤等作用(Li等2007)。白术为浙江著名的道地药材, 其野生资源极为稀少, 目前绝大部分的药材来自安徽、河北和浙江等省的栽培品种。白术含有倍半萜类挥发油物质(苍术酮和白术内酯)、白术多糖等多种有效药用成分(王瑶等2012), Ji等(2014)发现白术多糖能增强巨噬细胞的吞噬作用来增强机体免疫力和改善心脏功能等, 此外, 陆家佳(2016)研究发现苍术酮对肝癌等5种癌症细胞有抑制作用, 刘昶等(2005)也发现了白术内酯I能抑制癌细胞的生长, 是较有前景的抗肿瘤恶病质药物。

目前国内外对白术的研究主要集中在其化学成分(Peng等2011)、药理作用(Ji等2014)、种植技术(朱玉球等2006; 杨国等2018)和临床应用(刘昶等2005)等方面。孟庆蕊(2015)建立了农杆菌介导的白术遗传转化体系, 获得了抗病的转基因白术植株。但白术重要功能基因以及有效成分生物合成途径的基因调控方面鲜少研究, 基因组和转录

组信息严重匮乏, 使白术在基因水平上的改良和深入开发遇到了瓶颈。转录组测序技术通过将样本中提取的总体RNA (狭义上主要指mRNA)反转录成cDNA, 构建cDNA文库后进行高通量测序来确定样品中整体转录组的表达情况。获得的转录组信息为该物种的发育、次生代谢、抗逆等后续研究提供了基因资源和基础(陈昊和谭晓风 2014)。这对于一些非模式生物的研究有所助益, 已经运用于丹参、党参、三七和金银花等药用植物的研究中(Hua等2011; Gao等2015; Luo等2011; Zhang等2016)。

本研究利用Illumina HiSeq 2000高通量测序平台对白术全株进行转录组测序, 获得白术的转录组信息, 为探究白术的次生代谢、有效成分的生物合成调控、品种改良等方面的研究提供遗传信息及理论参考。

收稿 2019-02-18 修定 2019-10-20

资助 国家自然科学基金(31700476)、浙江省教育厅项目(Y2019-41025)和绍兴市教学建设与改革项目(SXSKG2018069)。

并列第一作者

* 共同通讯作者: 罗洁(lj26wxy@163.com)、莫亿伟(ywmo@163.com)。

1 材料与方法

1.1 材料

材料来自于浙江省绍兴市新昌县双彩乡白术种植基地,经绍兴文理学院汤访评副教授鉴定为菊科植物白术(*Atractylodes macrocephalae* Koidz)。选取一年生白术植株的叶、茎、花和块根部分,液氮速冻保存于 -80°C 冰箱备用。

1.2 白术各组织挥发油含量测定

将白术块根干燥至恒重,粉碎,过60目筛,称取样品10 g,按照药典附录XD挥发油测定法提取挥发油,提取5 h。

GC-MS条件为,色谱柱: HP-5M (30 mm $\times$ 0.25 μm); 升温程序: 以 $2^{\circ}\text{C}\cdot\text{min}^{-1}$ 从 70°C 升到 150°C ,保留3 min,再以 $5^{\circ}\text{C}\cdot\text{min}^{-1}$ 升到 150°C ,保留2 min,最后以 $8^{\circ}\text{C}\cdot\text{min}^{-1}$ 升到 240°C ; 载气流速: 0.8 mL $\cdot\text{min}^{-1}$; 进样口温度: 250°C ; 进样量1 μL 。质量扫描范围: 40~50 amu; 离子源: EI; 电子轰击能量: 70 eV; 倍增电压: 1 200 V; 离子源温度: 230°C ; 接口温度: 150°C 。

1.3 总RNA的提取和文库的构建

采用RNAqueous[®] Total RNA Isolation Kit[™]试剂盒进行白术的叶、茎、花、块根样品的总RNA提取。样品检测合格后,将白术叶、茎、花及块根RNA等量混合成一个样品,用带有Oligo(dT)的磁珠富集白术的mRNA; 接着将mRNA进行随机打断; 以mRNA为模板,用随机引物反转录合成第一条cDNA链,然后以第一条cDNA为模板合成互补的第二链,形成双链cDNA,接着对cDNA进行纯化、末端修复、连接测序接头,最后进行片段大小选择,通过PCR富集得到cDNA文库。

1.4 转录组测序、组装及功能注释

文库检查合格后,使用Illumina HiSeq 2000高通量测序平台对cDNA文库进行测序,转录组测序委托南京集思慧远生物技术有限公司完成。对原始数据进行过滤,去除其中的接头序列及低质量Reads获得有效数据。利用Trinity组装软件(Grabherr等2011)对数据进行组装,获得质量较高的Unigene。

得到的转录组测序文库后首先对其进行质量评估,评估合格后使用BLAST软件(Altschul等1997)

将Unigene序列与NR、Swiss-Prot、GO、COG、KOG、KEGG数据库比对,预测完Unigene的氨基酸序列之后使用HMMER软件(Zhang和Wood 2003)与Pfam数据库比对,获得Unigene的注释信息。

1.5 SSR特征检测及基因表达定量分析

利用MISA (MIcroSAtellite identification tool)软件对筛选得到的1 kb以上的Unigene做SSR分析,再采用Bowtie软件(Langmead等2009)将测序得到的Reads与Unigene库进行比对,根据比对结果,结合RSEM软件(Li和Dewey 2011)进行表达量水平估计。利用FPKM值表示对应Unigene的表达丰度。

1.6 白术不同组织中倍半萜合成途径相关基因qRT-PCR分析

分别以白术叶、茎、花和块根RNA为模板,使用PrimeScript[™] RT reagent Kit with gDNA Eraser (Perfect Real Time)试剂盒(TaKaRa, 日本)反转录合成cDNA第一链,产物放置于 -20°C 冰箱中保存备用。采用的反应体系为: 模板RNA 500 ng、5 $\times$ gDNA Eraser Buffer 2 μL 、gDNA Eraser 1 μL 。根据本实验室建立的白术转录组数据库,筛选了内参基因ubiquitin以及倍半萜合成相关基因(表1),qRT-PCR反应在Eppendorf AG 22331 Hamburg real time PCR仪(Eppendorf, 德国)上进行,反应体系(20 μL)为: cDNA模板2 μL 、上游引物1 μL 、下游引物1 μL 、SYBR[®] Premix Ex Taq[™] II (Tli RNaseHPlus) (2 $\times$) 10 μL 和ddH₂O 6 μL 。所用程序为: 95°C 30 s; 95°C 5 s, 55°C 30 s, 72°C 1 min, 40个循环。

2 实验结果

2.1 白术块根挥发油含量测定

从表2可知,白术块根中的倍半萜类物质占了整个挥发油成分的41.38%,共检测到倍半萜物质17种,其中含量最高的是苍术酮,占整个挥发油含量的20.5%;其次是8,9-去氢新异长叶烯,含量为5.6%,白术内酯I含量为4.5%,白术内酯II含量为4.1%;大根香叶烯-A、芹烷二烯酮和2-(2,5-二甲氧基苯基)环己酮等含量均高于1%。

2.2 转录组数据初步分析

利用Illumina HiSeq 2000高通量测序平台进

表1 qRT-PCR所用引物
Table 1 Primers for qRT-PCR

基因	正向引物(5'→3')	反向引物(5'→3')
内参 <i>ubiquitin</i>	ACCTTGTCCTCCGTCTG	AACCTCTGCTGGTCTGG
<i>HMGR1</i>	TCGACACCCTTGCGGTGTG	GTTCATCCCCATCGCGTCCC
<i>HMGR2</i>	CCCACCTCGGAGGAAGACGA	CTCTGAAGAGCCTGCCGTCG
<i>DXR</i>	GGCCCCGTCTTGATCTCTGC	GCGTACTCACGAGCCCACAA
<i>FDPS</i>	GTCAGTGGGTCGAGCGGATG	ACCAAGGGCACACGCAAGAA
<i>GAS</i>	AAGCCTCTACGAATGTGCC	GTGTCGTCAACGCGTGCTTC
<i>QHS1</i>	GCTTCTGTGCCATTGCAGG	ACGTGCCAAATTGATCACCCG

表2 白术块根倍半萜化合物
Table 2 Sesquiterpenes from rhizomes of *A. macrocephala*

编号	化合物名称	分子式	分子量	相对含量/%	保留时间/s
1	苍术酮	C ₁₅ H ₂₀ O	216	20.49	11.45
2	8,9-去氢新异长叶烯	C ₁₅ H ₂₂	202	5.66	14.15
3	白术内酯I	C ₁₅ H ₁₈ O ₂	230	4.54	15.21
4	白术内酯II	C ₁₅ H ₂₀ O ₂	232	4.13	15.66
5	芹烷二烯酮	C ₁₅ H ₂₂ O	218	1.62	12.29
6	2-(2,5-二甲氧苯基)环己酮	C ₁₄ H ₁₆ O ₃	232	1.12	15.57
7	大根香叶烯-A	C ₁₅ H ₂₄	204	1.10	11.54
8	11-isopropylidenetricyclo[4.3.1.1(2,5)]undec-3-en-10-one	C ₁₅ H ₂₂	202	0.55	14.60
9	甲氧桂乙酯	C ₁₄ H ₁₈ O ₄	250	0.50	14.66
10	β-vatirenene	C ₁₅ H ₂₂	202	0.41	10.10
11	脱氢香橙烯	C ₁₅ H ₂₂	202	0.32	14.95
12	异喇叭烯	C ₁₅ H ₂₄	204	0.25	9.99
13	γ-榄香烯	C ₁₅ H ₂₄	204	0.21	10.27
14	9,10-脱氢异长叶烯	C ₁₅ H ₂₂	202	0.19	10.35
15	棕榈酸甲酯	C ₁₇ H ₃₄ O ₂	270	0.14	14.08
16	9-甲基-s-八氢苯丙烯	C ₁₅ H ₂₀	200	0.12	13.48
17	β-石竹烯	C ₁₅ H ₂₄	204	0.06	9.76

行测序, 共检测得到23 099 928个Clean reads片段, 共获得总核苷酸数6.9 G, GC含量为45.52%, Q30值为93.48%。经过*de novo*拼接后, 获得Transcript有200 070条, 总长度为195.05 Mb, N50长度为1 095 bp; Unigene共100 037条, 总长度82.32 Mb, N50长度为1 027.2 bp, 组装完整性均较高。

在总Unigene中, 长度为300~500 bp的Unigene数量为35 679条(35.67%); 长度为500~1 000 bp的Unigene最多, 有39 973条(39.96%); 长度1 000~2 000 bp的有19 269条(19.26%); 长度大于2 000 bp的共有5 116条(5.11%), 可见, Unigene整体长度分布较均匀。

2.3 Unigene的功能注释

将100 037条Unigene与NR、GO、COG、KOG、KEGG等数据库进行比对, 共有54 471条Unigene序列被注释, 占总序列数的54.45%。在NR数据库中注释数量最多有50 467个Unigene获得同源匹配信息, 注释比例50.45%, 是所有数据库中注释数量最多的; 30 805条Unigene在GO数据库中获得注释, 占30.79%; 其中11 837条Unigene在COG数据库中获得注释, 注释比例为11.83%; 13 028条Unigene在KEGG数据库中获得注释, 占13.02%。

2.4 Unigene的COG分类

将白术叶、茎、块根混合样品的转录组Uni-

gene片段与COG数据库进行比对,发现共有共有11 837个Unigene被注释到COG分类中(图1)。白术转录组中的Unigene根据功能大致可分为25类,并对每一类的基因数量进行了统计。结果(图1)显示,Unigene的COG功能种类比较全面,涉及到大多数的生命活动,仅作为一般功能预测的基因数量最多,有1 836条,占15.51%;其次是基因的复制、重组、修复等,涉及的基因为1 661条,占14.03%,其他类别基因的数量也不尽相同,所有Unigene在胞外结构和核结构两类中没有得到注释。

2.5 Unigene的GO功能分类

Gene ontology (GO)在生物信息学领域内广泛应用于对基因和蛋白质功能进行限定和描述,共有30 805条Unigene在GO数据库中得到注释,按功能被分为细胞组分、分子功能、生物学过程三大类。图2显示,其中5 222条Unigene在细胞组分这一类中得到注释,占在GO数据库注释Unigene总量的16.95%;18 334条在分子功能一类中得到注释,占59.52%;7 249条在生物学过程中得到注释,占23.53%。这三大功能又被细分为51个功能组,其中细胞组分功能组中,细胞要素和细胞被注释的最多,分别为41.50% (12 784个)和41.24% (12 704个);

分子功能部分中,结合功能和催化活性功能所占比例最大,分别为53.64% (17 058个)和 52.5% (16 524个);而生物过程最主要的为代谢过程和细胞过程,依次为71.07% (21 895个)和59.20% (18 236个)。

2.6 Unigene的KEGG注释

结果(图3)显示,13 028条Unigene被注释到KEGG数据库里,白术转录组中的数据分成五大类133种代谢通路,五大类分别为细胞过程、环境信息过程、遗传信息过程、代谢和组织系统。其中跟代谢相关的Unigene有5 681条,所占比例最大;在遗传信息过程一类中,有4 530个Unigene得到注释;环境信息过程中有550个;组织系统407个;细胞过程852个。白术的有效成分有倍半萜类成分、多糖类成分以及氨基酸等。根据注释结果对白术有效成分相关的代谢产物合成路径进行分析,有305个功能基因序列映射到萜类化合物和聚酮类化合物的代谢途径中,其中萜类骨架生物合成途径有137个Unigene,倍半萜及三萜类生物合成途径相关的51个;1 254个功能基因序列映射到氨基酸代谢途径中;1 680个功能基因序列映射到碳水化合物代谢途径中;183个映射到多糖生物合成和代谢中;327个映射到其他次级代谢产物的生物合成中。

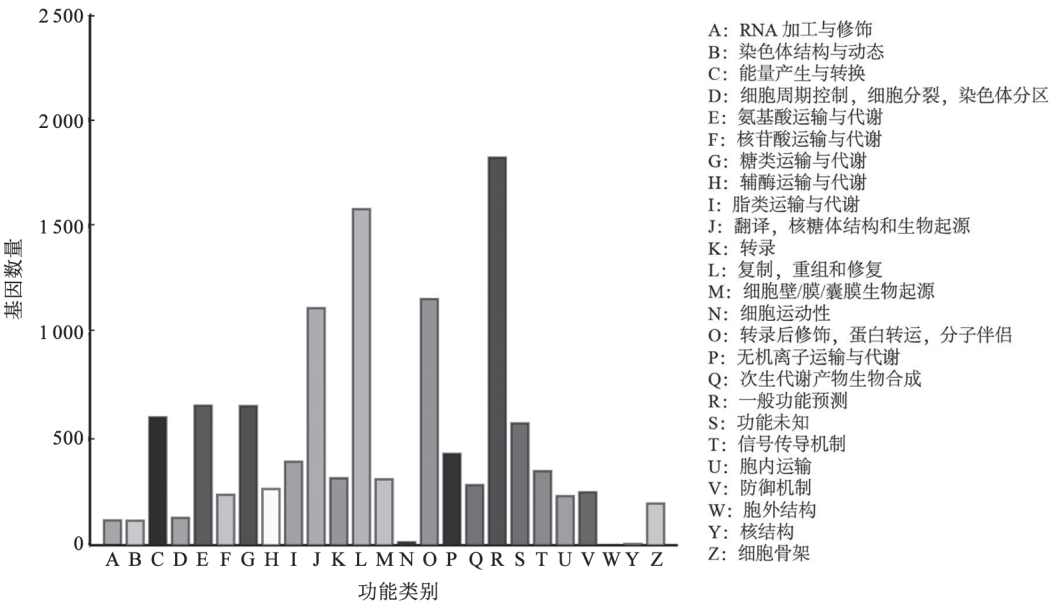


图1 COG功能基因分类

Fig.1 COG functional categories

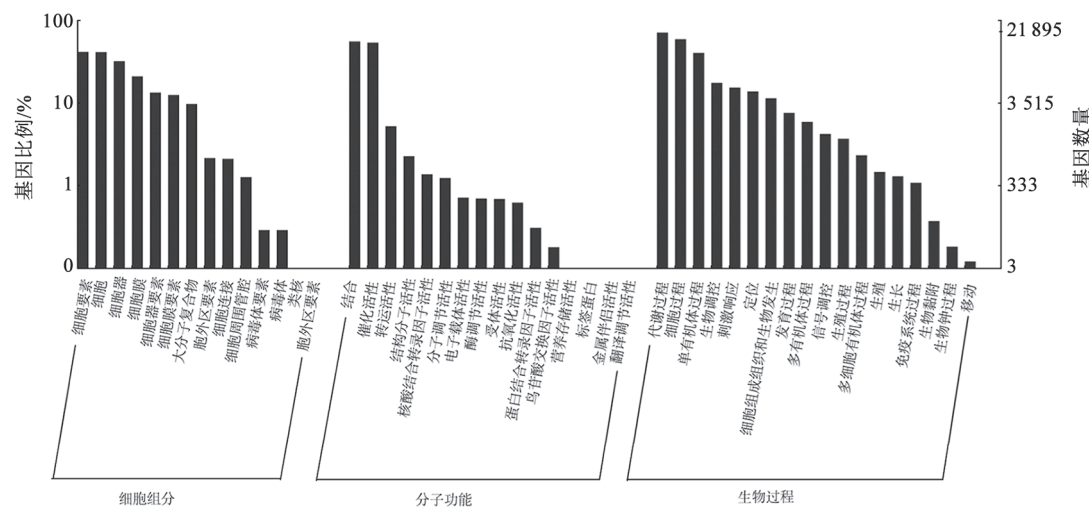


图2 GO功能基因分类
Fig.2 GO functional categories

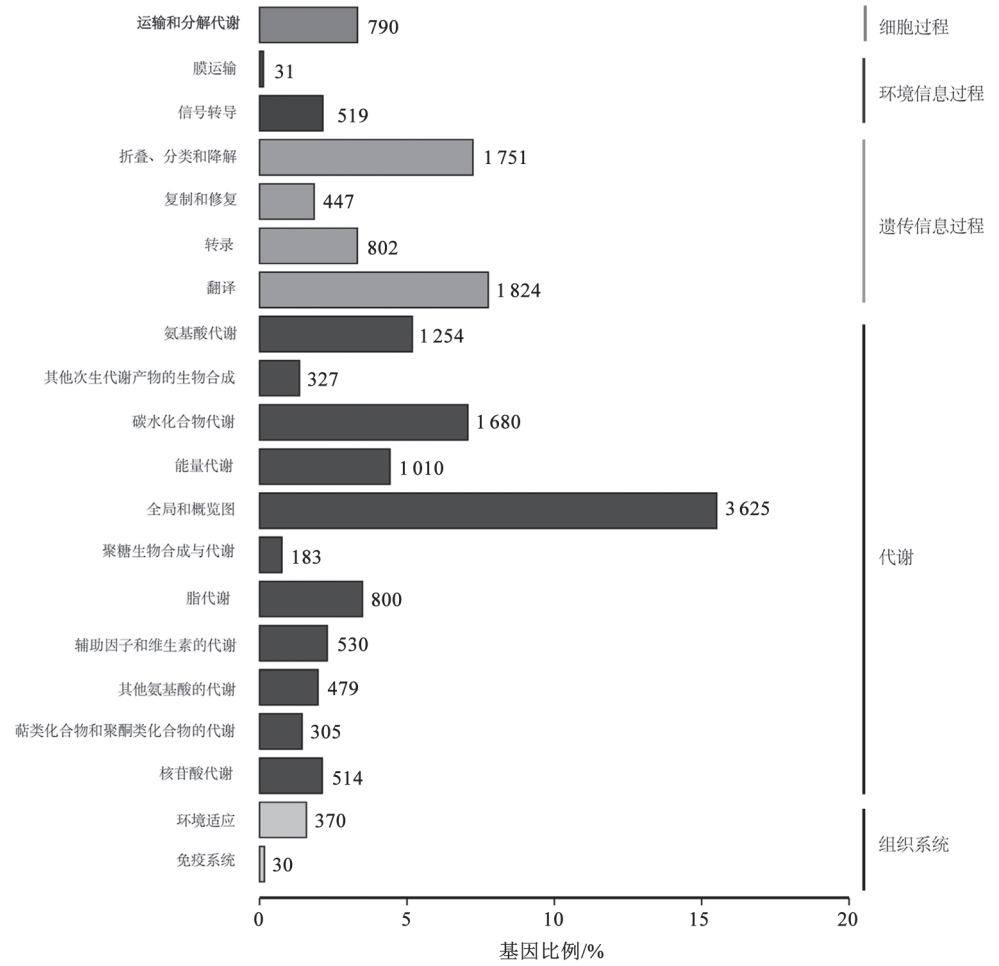


图3 Unigene'sKEGG注释与分类
Fig.3 KEGG annotation and classification of unigenes

2.7 白术倍半萜生物合成途径及相关酶基因表达量分析

通过KEGG代谢通路分析, 结合萜类目前已知的生物合成途径, 白术萜类物质骨架生物合成途径主要有两条, 包括位于细胞质的MVA途径和位于质体中的MEP途径(图4)。在MVA途径中相关酶有6个, 与这6个酶合成相关的Unigene有32个; 在MEP途径中相关酶有6个, 与这6个酶合成相关的Unigene有30个; 法呢基焦磷酸(farnesyl-PP, FPP)是各种萜类物质合成的重要前体物质, FDPS是FPP合成最重要的酶, FDPS合成相关的Unigene共3个; 倍半萜合成酶有7个, 相关Unigene共18个。细胞色素P450酶和过氧化物酶在各种萜类衍生物的生成中起着重要作用, 结果检测到细胞色素P450酶相关的Unigene共有310条。

白术萜类化合物骨架生物合成相关酶有15个

(表3), 与这15个酶合成相关的Unigene有81个, 其中DXPS和ISPG合成相关基因数量最多, 各有10个, 其中ACAT合成相关基因*DN42341_c0_g3*表达量最高, RPKM值为263.15; FDPS合成相关的基因共3个, 其中*DN23232_c2_g1*的表达量最高; 其中倍半萜合成酶分别为FLDH、AFS1、NES1、GERD、TPS1、GAS和QHS1, 其中QHS1相关基因最多, 共5条, TPS1相关基因*DN24041_c0_g7*表达量最高, RPKM值为125.84。

2.8 转录组数据SSR特征分析

对白术进行微卫星位点分析, 共有34 476个1 kb以上的序列被评估, 识别到的SSR总数为12 124, 其中检测到单碱基重复SSR有6 028个, 双碱基重复SSR有3 866个。根据碱基类型的不同, 对六大类型的SSR再进一步做分类型数量统计(图5), 以A/T重复的SSR数量为5 558, 占了单碱基数量的

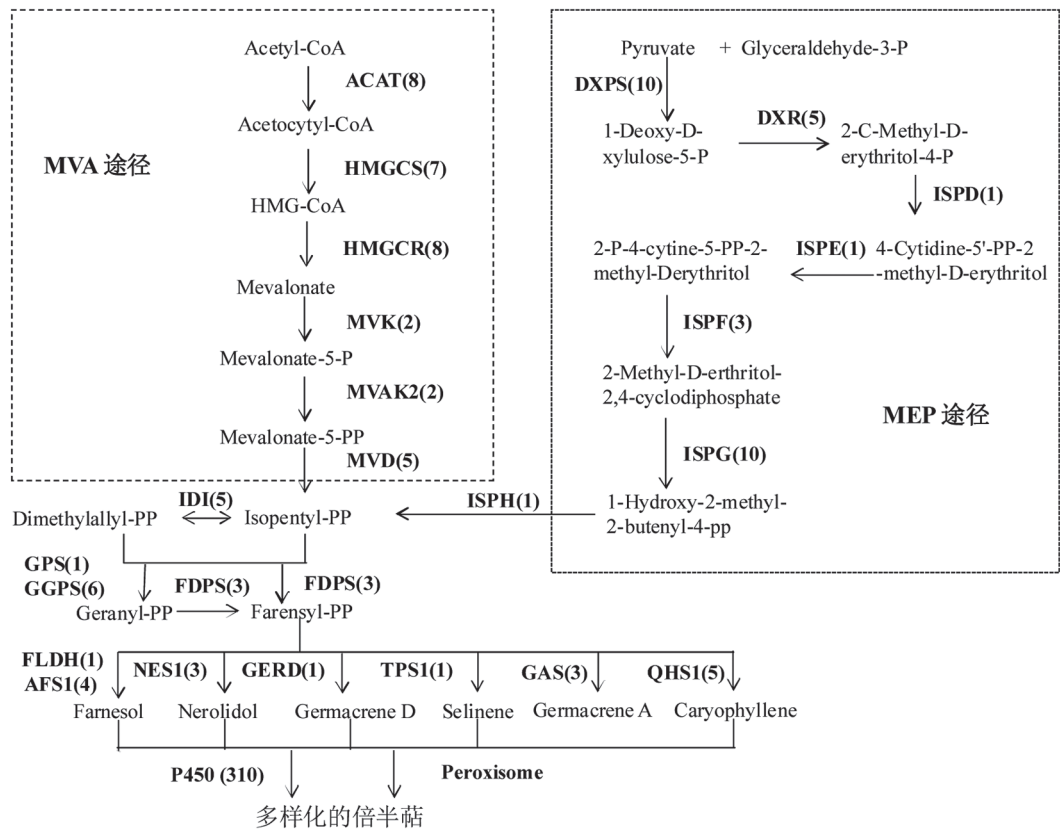


图4 白术的倍半萜类生物合成途径

Fig.4 Sesquiterpene biosynthesis pathway of *A. macrocephala*

括号中的数字代表每个酶相关Unigene的数量; MVA: mevalonate; MEP: 2-C-methyl-d-erythritol-4-phosphate。

表3 白术中萜类骨架和倍半萜合成相关基因

Table 3 Genes related to tepenoids biosynthesis of *A. Macrocephalae*

萜类骨架和倍半萜合成相关酶	EC编号	Unigene名称	RPKM值
酰基转移酶(acetyl-CoA C-acetyltransferase, ACAT)	2.3.1.9	DN42341_c0_g3	263.15
羟基甲基戊二酰-CoA还原酶 (hydroxymethylglutaryl-CoA reductase, HMGCR)	1.1.1.34	DN42341_c0_g2	234.01
1-脱氧-D-木酮糖-5-磷酸还原异构酶 (1-deoxy-D-xylulose-5-phosphate reductoisomerase, DXR)	1.1.1.267	DN20548_c0_g2	51.16
法呢基焦磷酸合酶(farnesyl diphosphate synthase, FDPS)	2.5.1.10	DN23232_c2_g1	127.11
甲羟戊酸激酶(mevalonate kinase, MVK)	2.7.1.36	DN36507_c1_g2	18.27
4-二磷酸胞苷-2C-甲基-D-赤藓糖醇激酶 (4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, ISPE)	2.7.1.148	DN20171_c1_g1	20.22
焦磷酸甲羟戊酸脱羧酶(diphosphomevalonate decarboxylase, MVD)	4.1.1.33	DN33486_c4_g1	70.24
羟甲基戊二酰-CoA合酶 (hydroxymethylglutaryl-CoA synthase, HMGCS)	2.3.3.10	DN20923_c0_g2	102.76
1-脱氧木酮糖-5-磷酸合成酶 (1-deoxyxylulose-5-phosphate synthase, DXPS)	2.2.1.7	DN38629_c1_g2	129.47
2-C-甲基-D-赤藓糖醇2,4-环磷酸合成酶 (2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, ISPF)	4.6.1.12	DN32344_c2_g2	92.22
异戊烯二磷酸δ异构酶 (isopentenyl-diphosphate delta-isomerase, IDI)	5.3.3.2	DN23551_c1_g1	104.21
(E)-4-羟基-3-甲基丁-2-烯基二磷酸合酶 [(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase, ISPG]	1.17.7.1	DN30195_c0_g1	124.53
4-羟基-3-甲基丁-2-烯基-1-二磷酸还原酶 (4-hydroxy-3-methylbut-2-enyl diphosphate reductase, ISPH)	1.17.7.4	DN32151_c7_g2	169.36
香叶基香叶基二磷酸合酶(geranylgeranyl diphosphate synthase, GGPS)	2.5.1.29	DN27118_c0_g1	38.23
香叶基二磷酸合酶(geranyl diphosphate synthase, GPS)	2.5.1.1	DN38721_c0_g1	13.49
NAD+依赖法尼醇脱氢酶 (NAD+-dependent farnesol dehydrogenase, FLDH)	1.1.1.354	DN14600_c0_g2	8.77
阿尔法法尼烯合酶(alpha-farnesene synthase, AFS1)	4.2.3.46	DN20061_c0_g2	36.81
(3S,6E)-橙花叔醇合成酶[(3S,6E)-nerolidol synthase, NES1]	4.2.3.48	DN21947_c2_g1	13.79
大根香叶烯合酶[(-)-germacrene D synthase, GERD]	4.2.3.75	DN22443_c2_g1	25.20
瓦伦烯/7-阿尔法蛇床烯合酶 (valencene/7-epi-alpha-selinene synthase, TPS1)	4.2.3.73, 4.2.3.86	DN24041_c0_g7	125.84
大根香叶烯-A合酶(germacrene-A synthase, GAS)	4.2.3.23	DN27600_c0_g3	16.45
β-石竹烯合酶[(-)-beta-caryophyllene synthase, QHS1]	4.2.3.57	DN41749_c1_g4	31.57

绝大部分, C/G类型仅有470个; 双碱基重复中以AG/CT为重复单位的SSR有2 070个, 占双碱基重复数量的53.54%, 以CG/CG重复的仅有4个。

以SSR基序重复数量为分类依据, 对六大类型进行分类统计。统计结果(表4)显示, 单碱基重复SSR中, 以9~12个基序重复为1单元的SSR占了一半以上; 在双碱基重复SSR中, 以5~8个基序重复为一单元的SSR占一半以上; 三碱基重复的SSR中, 以5~8个基序重复为一单元的SSR占了绝大多数。四碱基、五碱基、六碱基重复SSR均是5~8个基序重复构成的。

2.9 白术不同器官组织的倍半萜合成相关基因的qRT-PCR分析

为了研究倍半萜合成相关基因在不同器官中表达量的差别, 本研究对参与倍半萜生物合成途径的6个基因进行了qRT-PCR分析, 选择了以下基因: MVA途径中关键酶HMGCR的基因(DN42341_c0_g3、DN42341_c0_g2), MEP途径中关键酶DXR的基因(DN20548_c0_g2), 各类倍半萜合成的前体物质相关酶FDPS的基因(DN23232_c2_g1), 2个倍半萜合成酶基因, 分别是GAS的基因(DN27600_

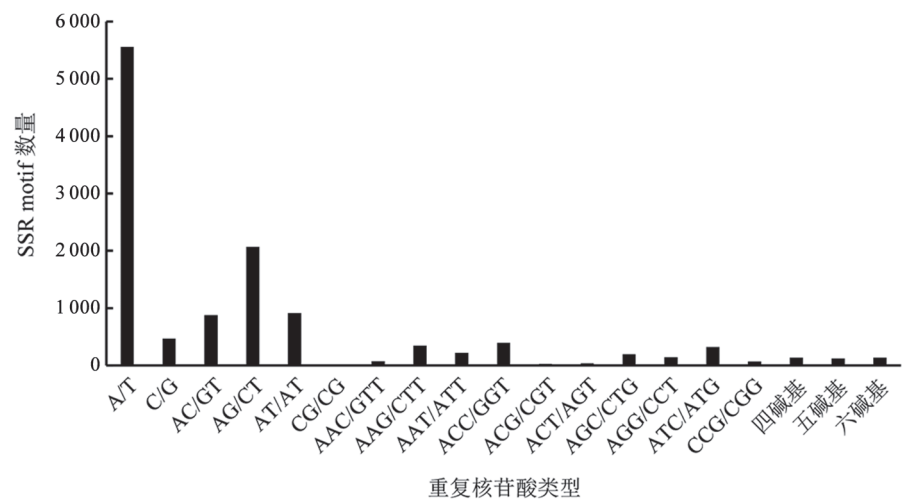


图5 SSR分类的数量统计
Fig.5 Quantity statistics of SSR classification

表4 SSR重复单元数量统计
Table 4 Quantity statistics of SSR repeating units

重复SSR基序类型	单碱基数量	双碱基数量	三碱基数量	四碱基数量	五碱基数量	六碱基数量
2~8个基序重复	0	2 350	1 750	138	120	136
9~12个基序重复	4 560	800	50	0	0	0
13~16个基序重复	900	420	30	0	0	0
17~20个基序重复	350	180	0	0	0	0
21~24个基序重复	110	115	0	0	0	0

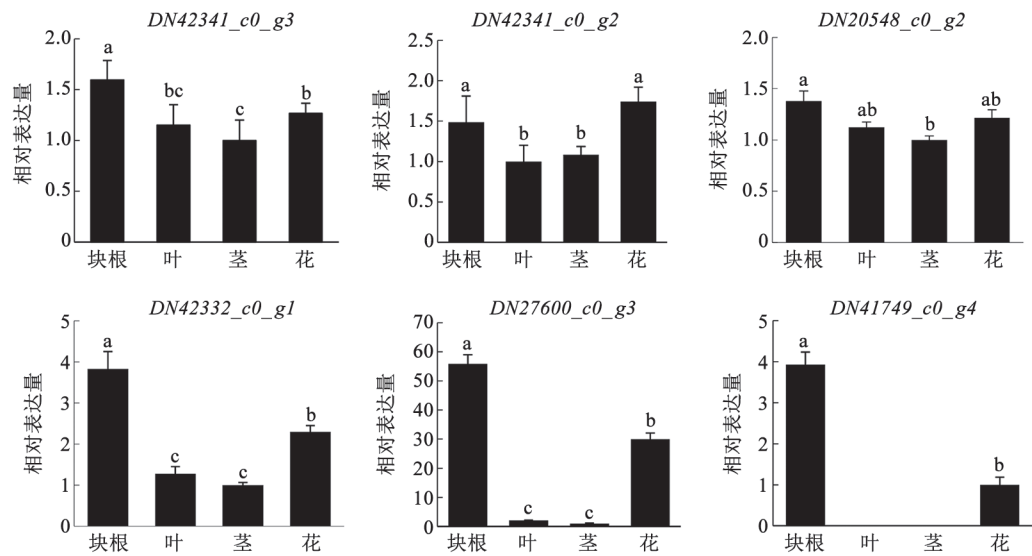


图6 白术倍半萜合成途径中相关酶的基因在不同组织器官中的表达量
Fig.6 Gene expression of related enzymes in sesquiterpene synthesis pathway of *A. macrocephala* in different tissues and organs
不同小写字母表示差异显著($P<0.05$)。

c0_g3)和QHS1的基因(*DN41749_c1_g4*)。结果如图6所示, *DN42341_c0_g3*在块根中表达量最高, 和其他三个组织有显著差异; *DN42341_c0_g2*在花和块根中表达量相对较高, 显著高于叶片和茎中的表达量; *DN20548_c0_g2*在块根中表达量较高, 与茎中的表达量有显著差异; *DN23232_c2_g1*在块根中表达量最高, 在其他三个组织中有显著差异; GAS的基因(*DN27600_c0_g3*)和QHS1的基因(*DN41749_c1_g4*)在白术块根中的表达量相对较高, 其次是花, 而叶片和茎中的表达量接近零。

3 讨论

萜类物质是植物次生代谢物中数量最多且结构最丰富的一类物质(Gershenzon和Dudareva 2007), 萜类化合物起源于乙酰辅酶A (Acetyl-CoA)或丙酮酸(pyruvate)和3-磷酸甘油醛(glyceraldehyde-3-P), 通过细胞质的MVA途径和质体中的MEP途径多步合成法呢基焦磷酸(FPP) (Kirby和Keasling 2009), FPP在不同的萜烯合酶的作用下合成单萜类、倍半萜类、二萜、三萜等化合物(Xia等2012)。本研究通过GC-MS分析表明苍术酮和白术内酯等倍半萜类物质是白术挥发油成分中主要的组成成分, 该研究与前人的研究较一致(陆家佳2016)。

本研究首次利用高通量测序技术对白术叶、茎、块根、花混合样品进行了转录组测序, 获得了大量的转录本信息, 结果共得到100 037条Unigene, 其N50长度为1 027.2 bp, 组装完整性均较高, 这与珠子参、华细辛等药用植物转录组测序结果较相似(张绍鹏等2015; 林懋怡等2017)。将Unigene与COG、GO、KOG、KEGG等数据库进行比对注释, 共有54 471条Unigene序列被注释到以上数据库之一, 占总序列数的54.45%。本研究通过KEGG代谢通路分析, 结合已知的萜类合成途径, 明确了白术萜类基本骨架及倍半萜的生物合成过程关键酶的相关基因, 白术萜类化合物骨架生物合成相关酶有15个, 与这15个酶合成相关的Unigene有81个, HMGCR和DXR分别是MVA途径和MEP途径中最重要限速酶(Cordoba等2009; Schaller等1995), FPP是各种萜类物质合成的重要前体物质, FDPS是FPP合成最重要的酶, 本研究发现HMGCR

合成相关的基因8个, DXR合成相关的基因5个, FDPS合成相关的基因共3个, 同时通过qRT-PCR分析发现HMGCR和DXR合成基因在白术各个组织中的表达量并无差异, 但FDPS合成基因在块根中表达量显著高于其他组织。倍半萜及三萜生物合成单基因簇56个, 其中倍半萜合成酶有7种, 倍半萜合成酶相关的Unigene有18个。前人在同属植物苍术的转录组学研究(Ahmed等2016; Chen等2017)中发现苍术倍半萜合成酶也包含了AFS1、GERA、GERD和TPS1等, 我们的研究与其较一致。qRT-PCR分析发现GAS以及QHS1合成相关的基因在块根中表达量显著高于其他组织。以往的研究表明细胞色素P450氧化酶在各种萜类衍生物的生成中起着重要作用(Mizutani和Ohta 2010)。含氧倍半萜类是特别重要的活性药物成分之一(Zhou等2015)。我们的研究也表明了细胞色素P450酶和过氧化物酶不仅参与多种非氧化倍半萜的生成, 而且还参与了含氧的倍半萜的生成, 对白术倍半萜类多样性的形成起着重要作用。这些结果将有助于进一步研究白术的生长发育和次生代谢产物的生物合成过程。

SSR标记技术具有丰富的多态性、共显性遗传、重复性好和操作简单等优点, 在动植物的分子标记遗传连锁图谱构建、种质资源评估、基因定位、分子标记辅助选择、品种鉴定等领域都有很好的应用价值(Chapman等2009; Neeraja等2007)。传统的SSR标记的开发方法不但价格贵而且费时费力, 随着高通量测序技术的发展, 本研究利用转录组数据对白术进行SSR位点分析, 发现单碱基重复SSR有6 028个, 主要以9~12个基序进行重复, 双碱基重复SSR有3 866个, 主要以5~8个基序进行重复。利用转录组数据开发SSR标记已在许多物种中开始使用(Guo等2011; Kaur等2011; 李海波等2017)。我们的研究为白术相关基因的功能研究以及白术分子育种等提供重要参考。

参考文献(References)

- Ahmed S, Zhan CS, Yang YY, et al (2016). The transcript profile of a traditional Chinese medicine, *Atractylodes lancea*, revealing its sesquiterpenoid biosynthesis of the major active components. PLoS One, 11 (3): e0151975

- Altschul SF, Madden TL, Schäffer A, et al (1997). Gapped blast and psi-blast: A new generation of protein database search programs. *Nucleic Acids Res*, 25 (17): 3389–3402
- Chapman MA, Hvala J, Strever J, et al (2009). Development, polymorphism, and cross-taxon utility of EST-SSR markers from safflower (*Carthamus tinctorius* L.). *Theor Appl Genet*, 120 (1): 85–91
- Chen F, Wei YX, Zhang JM, et al (2017). Transcriptomics analysis investigates sesquiterpenoids accumulation pattern in different tissues of *Atractylodes lancea* (Thunb.) DC. plantlet. *Plant Cell Tiss Org Cult*, 130: 73–90
- Chen H, Tan XF (2014). Excavation of genic resources based on next generation sequencing technologies. *Plant Physiol J*, 50 (8): 1089–1095 (in Chinese with English abstract) [陈昊, 谭晓风(2014). 基于第二代测序技术的基因资源挖掘. *植物生理学报*, 50 (8): 1089–1095]
- Cordoba E, Salmi M, Leon P (2009). Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J Exp Bot*, 60 (10): 2933–2943
- Duan Q, Xu DJ, Liu CX, et al (2008). Research progress of *Atractylodes macrocephala*. *Chin Tradit Herbal Drugs*, 39 (5): 800–802 (in Chinese) [段启, 许冬谨, 刘传祥等 (2008). 白术的研究进展. *中草药*, 39 (5): 800–802]
- Gao JP, Wang D, Cao LY, et al (2015). Transcriptome sequencing of *Codonopsis pilosula* and identification of candidate genes involved in polysaccharide biosynthesis. *PLoS One*, 10 (2): e0117342
- Gershenzon J, Dudareva N (2007). The function of terpene natural products in the natural world. *Nat Chem Biol*, 3 (7): 408–414
- Grabherr MG, Haas BJ, Yassour M, et al (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol*, 29 (7): 644–652
- Guo SG, Liu JG, Zheng Y, et al (2011). Characterization of transcriptome dynamics during watermelon fruit development: sequencing, assembly, annotation and gene expression profiles. *BMC Genomics*, 12 (1): 454
- Hua WP, Zhang Y, Song J, et al (2011). De novo transcriptome sequencing in *Salvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients. *Genomics*, 98 (4): 272–279
- Ji GQ, Chen RQ, Zheng JX (2014). Macrophage activation by polysaccharides from *Atractylodes macrocephala* Koidz through the nuclear factor- κ B pathway. *Pharm Biol*, 53 (4): 1–6
- Kaur S, Cogan NO, Pembleton LW, et al (2011). Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. *BMC Genomics*, 12 (1): 265
- Kirby J, Keasling JD (2009). Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annu Rev Plant Biol*, 60: 335–355
- Langmead B, Trapnell C, Pop M, et al (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10 (3): R25
- Li B, Dewey CN (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12 (1): 323
- Li CQ, He LC, Dong HY, et al (2007). Screening for the anti-inflammatory activity of fractions and compounds from *Atractylodes macrocephala* koidz. *J Ethnopharmacol*, 114 (2): 212–217
- Li HB, Wang S, Ding HM, et al (2017). Development of EST-SSR molecular markers based on transcriptome sequencing of *Camellia oleifera*. *Plant Physiol J*, 53 (7): 1267–1278 (in Chinese with English abstract) [李海波, 王珊, 丁红梅等(2017). 普通油茶转录组EST-SSR分子标记开发. *植物生理学报*, 53 (7): 1267–1278]
- Lin MY, Niu H, Liu JJ, et al (2017). Identification of genes related to biosynthesis of methyleugenol in *Asarum sieboldii* by transcriptome sequence analysis. *Chin Tradit Herbal Drugs*, 48 (15): 3160–3167 (in Chinese with English abstract) [林懋怡, 牛卉, 刘晋杰等(2017). 基于转录组分析华细辛甲基丁香酚生物合成途径的相关基因. *中草药*, 48 (15): 3160–3167]
- Liu Y, Ye F, Qiu GQ, et al (2005). Effects of lactone I from *Atractylodes macrocephala* Koidz on cytokines and proteolysis-inducing factors in cachectic cancer patients. *J First Mil Med Univ*, 25 (10): 1308–1311 (in Chinese with English abstract) [刘映, 叶峰, 邱根全等(2005). 白术内酯I对肿瘤恶病质患者细胞因子和肿瘤代谢因子的影响. *南方医科大学学报*, 25 (10): 1308–1311]
- Lu JJ (2016). A GC-MS Analysis of volatile components of *Atractylodes macrocephala* Koidz and their research of inhibitory activity on five tumor cells. *J Strait Pharm*, 54 (4): 618–626 (in Chinese with English abstract) [陆家佳 (2016). 白术挥发性成分GC-MS分析及对五种肿瘤细胞抑制活性研究. *海峡药学*, 28 (6): 28–31]
- Luo HM, Sun C, Sun YZ, et al (2011). Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. *BMC Genomics*, 12 (Suppl 5): S5
- Meng QR (2015). Establishment of regeneration system and transformation of *Atractylodes macrocephala* Koidz. (dissertation). Wuhan: Central China Agricultural University (in Chinese with English abstract) [孟庆蕊(2015). 白术再生体系建立与农杆菌介导的遗传转化(学位论文). 武汉: 华中农业大学]
- Mizutani M, Ohta D (2010). Diversification of P450 genes during land plant evolution. *Annu Rev Plant Biol*, 61 (1): 291–315

- Neeraja CN, Maghirang-Rodriguez R, Pamplona A, et al (2007). A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. *Theor Appl Genet*, 115 (6): 767–776
- Peng W, Han T, Xin WB, et al (2011). Comparative research of chemical constituents and bioactivities between petroleum ether extracts of the aerial part and the rhizome of *Atractylodes macrocephala*. *Med Chem Res*, 20: 146–151
- Schaller H, Grausem B, Benveniste P, et al (1995). Expression of the *Hevea brasiliensis* (H.B.K.) Mull. Arg. 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1 in tobacco results in sterol overproduction. *Plant Physiol*, 109 (3): 761–770
- Wang Y, Tian W, Wang SL, et al (2012). Quality evaluation on rhizomes of *Atractylodes macrocephala* from different producing areas in Zhejiang Province. *Chin Tradit Herbal Drugs*, 43 (8): 1615–1620 (in Chinese with English abstract) [王瑶, 田薇, 王舒琳等(2012). 浙江不同产区白术质量评价研究. *中草药*, 43 (8): 1615–1620]
- Xia JH, Zhang SD, Li YL, et al (2012). Sesquiterpenoids and triterpenoids from *Abies holophylla* and their bioactivities. *Phytochemistry*, 74 (1): 178–184
- Yang G, Lu K, Zhu GL, et al (2018). Effect of arbuscular mycorrhizal fungus *Glomus mosseae* on photosynthesis and antioxidation enzyme activities in *Atractylodes macrocephala* seedlings under copper stress. *Plant Physiol J*, 54 (4): 618–626 (in Chinese with English abstract) [杨国, 卢可, 朱高樑等(2018). 丛枝菌根真菌摩西球囊霉对铜胁迫下白术幼苗光合特性及抗氧化酶活性的影响. *植物生理学报*, 54 (4): 618–626]
- Zhang LB, Long Y, Fu CH, et al (2016). Different gene expression patterns between leaves and flowers in *Lonicera japonica* revealed by transcriptome analysis. *Front Plant Sci*, 7: 637
- Zhang SP, Jin J, Hu BX, et al (2015). Transcriptome profiling and analysis of *Panax japonicus* var. *Major*. *Chin J Chin Mater Med*, 40 (11): 2084–2089 (in Chinese with English abstract) [张绍鹏, 金健, 胡炳雄等(2015). 珍稀药用植物珠子参的转录组测序及分析. *中国中药杂志*, 40 (11): 2084–2089]
- Zhang Z, Wood WI (2003). A profile hidden Markov model for signal peptides generated by HMMER. *Bioinformatics*, 19 (2): 307–308
- Zhou JY, Yuan J, Li X, et al (2015). Endophytic bacterium-triggered reactive oxygen species directly increase oxygenous sesquiterpenoid content and diversity in *Atractylodes lancea*. *Appl Environ Microb*, 82 (5): 1577–1585
- Zhu YQ, Xia GH, Fang HG, et al (2006). Study on tissue culture and rapid propagation of *Atractylodes macrocephala*. *J Chin Mater Med*, 29 (3): 212–213 (in Chinese with English abstract) [朱玉球, 夏国华, 方慧刚等(2006). 白术组培快繁技术. *中药材*, 29 (3): 212–213]

Analysis of genes related to biosynthesis of sesquiterpene in *Atractylodes macrocephala* by transcriptome

YANG Guo^{1, #}, LI Hong-Hui^{1, #}, JIN Ye-Fei¹, DONG Li-Jia¹, MO Yi-Wei^{1, *}, LUO Jie^{2, *}

¹Academy of Life Sciences, Shaoxing University, Shaoxing, Zhejiang 312000, China

²Yuanpei College, Shaoxing University, Shaoxing, Zhejiang 312000, China

Abstract: In this study, the composition of sesquiterpenes in the volatile oil of *Atractylodes macrocephala* was analyzed by GC-MS. Meanwhile, Illumina HiSeq sequencing technology was used to sequence the transcriptome of the mixed samples included leaves, stems, roots, and flowers from *A. macrocephala*. Finally, the related genes of terpenoid skeleton and sesquiterpene synthesis-related enzymes in *A. macrocephala* were analyzed by real-time fluorescence quantitative PCR (qRT-PCR). The results show that the sesquiterpenes of *A. macrocephala* were mainly atractylone and 8,9-dehydroneoisolongifolene. A total of 23 099 928 valid short sequences were obtained, 100 037 unigenes and 93.48% Q30 were obtained, the length of N50 is 1 027.2 bp. The assembled unigenes sequences were compared with the NR, Swiss-Prot, KEGG, KOG, COG, and GO databases using the BLAST tool. A total of 54 471 unigene sequences were annotated, accounting for 54.45% of the total sequences. Analyze the KEGG metabolic pathway of the annotated unigenes, involving a total of 133 KEGG metabolic pathways, of which 81 unigenes participate in the terpene skeleton biosynthesis, and 18 unigenes participate in the sesquiterpene biosynthesis. Using the MISA software, 12 124 SSRs were found, of which 6 028 were single-base repeat SSRs, and 3 866 were double-base repeat SSRs. In this study, qRT-PCR analysis was performed on six genes related to sesquiterpene synthesis, and it was found that the expression levels of four genes in roots were significantly higher than those in leaves, stems, and flowers.

Key words: *Atractylodes macrocephala*; volatile oil; sesquiterpenes; RNA-seq; SSR; qRT-PCR

Received 2019-02-18 Accepted 2019-10-20

The work was supported by the National Natural Science Foundation of China (31700476), the Education Department Project of Zhejiang Province (Y201941025), and the Shaoxing Teaching Construction and Reform Project (SXSXG2018069).

#Co-first authors.

*Co-corresponding authors: Luo J (lj26wlxy@163.com), Mo YW (ywmo@163.com).