

基于遗传-化学-生态特征的西洋参品质生态型研究

孙 晓^{1,2}, 钱秋玉³, 郑司浩^{1,2}, 陈海梅¹, 黄林芳^{1,2*}

(1. 中国医学科学院、北京协和医学院药用植物研究所, 国家中医药管理局中药资源保护重点研究室, 北京 100193;
2. 中药资源教育部工程研究中心, 北京 100193; 3. 解放军总医院第七医学中心 263 临床部, 北京 101149)

摘要: 我国不同生态型的同种药材品质差异显著, 本课题组前期发现国产西洋参根据人参皂苷的特征存在关外型与关内型两大生态型, 然而引起西洋参品质变异的遗传生态机制尚不清楚。本文基于遗传-化学-生态策略, 运用转录组学技术、HPLC 技术, 并进行西洋参转录组数据、人参皂苷含量及环境气候生态因子的相关性分析。转录组结果显示人参皂苷生物合成关键基因如羟甲基戊二酰辅酶 A 还原酶 (hydroxymethylglutaryl-CoA reductase, HMGR)、 β -氨基酸合成酶 (beta-amyrin synthase, AS) 和法呢基焦磷酸合成酶 (farnesyl diphosphate synthase, FPS) 等在关内型均显著下调。HPLC 结果表明关外型西洋参中人参皂苷含量高于关内型西洋参。相关性分析结果显示西洋参人参皂苷的含量与甲羟戊酸二磷酸脱羧酶 (mevalonate kinase, MK)、羟甲基戊二酰辅酶 A 合酶 (hydroxymethylglutaryl-CoA synthase, HMGS)、HMGR 和 AS 等人参皂苷生物合成关键基因的表达成正相关, 而与糖基转移酶 GT 的表达成负相关, 西洋参的人参皂苷含量与两大生态型所处环境的温度、日照等气候因子成负相关, 与水分成正相关。本文为阐释西洋参两大生态型品质变异的形成机制提供了新思路, 也为道地药材品质生态学理论研究提供科学依据。

关键词: 西洋参; 中药品质生态学; 转录组; 人参皂苷; 道地药材

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2019)09-1695-11

Quality ecotype of *Panax quinquefolium* L. based on heredity-chemistry-ecology characteristics

SUN Xiao^{1,2}, QIAN Qiu-yu³, ZHENG Si-hao^{1,2}, CHEN Hai-mei¹, HUANG Lin-fang^{1,2*}

(1. Key Lab of Chinese Medicine Resources Conservation, State Administration of Traditional Chinese Medicine of China, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China; 2. Engineering Research Center of Chinese Medicine Resource, Ministry of Education, Beijing 100193, China; 3. The 263rd Clinical Department, General Hospital of Beijing Military Area Command of PLA, Beijing 101149, China)

Abstract: Medicinal materials in China differ in quality by different ecological types. Our research group found that there were two ecotypes of domestic *Panax quinquefolium* L. according to the characteristics of ginsenosides, inside versus outside Shanhaiguan. The genetic and ecological mechanisms of quality variation of *Panax quinquefolium* L. is unknown. Based on the genetic-chemical-ecological strategy, transcriptome and HPLC technology were used for comprehensive correlation analyses of transcriptomic data, ginsenoside content and environmental climate ecological factors. The transcriptomic results showed that key genes of ginsenoside biosynthesis, such as HMGR, AS and FPS, were significantly down-regulated in the inside Shanhaiguan ecotype. HPLC results showed that the quality of outside Shanhaiguan ecotype *Panax quinquefolium* L. was higher than that of the inside ecotype, with the content of ginsenosides in outside *Panax quinquefolium* L. was higher than that of inside ecotype except

收稿日期: 2019-05-20; 修回日期: 2019-06-18.

基金项目: 国家自然科学基金资助项目 (81274013, 81473315); 中国医学科学院医学与健康科技创新工程项目 (2016-I2M-3-015).

*通讯作者 Tel: 86-10-57833197, Fax: 86-10-57833197, E-mail: lfhuang@implad.ac.cn

DOI: 10.16438/j.0513-4870.2019-0394

Rb₂. Correlation analyses revealed that content of *Panax quinquefolium* L. ginsenoside is positively related to the expression levels of ginsenoside biosynthesis key genes (*MK*, *HMGS*, *HMGR*, and *AS*), and negatively related to the expression of glycosyl transferase (*GT*). The content of ginsenosides is negative related with climate factors, such as temperature, sunshine, and is positively related with moisture in both ecological environments. This study has provided a new mechanistic insight into the quality variations of two ecotypes for *Panax quinquefolium* L. and established a scientific basis for studying the ecological factors for the quality of traditional Chinese medicine.

Key words: *Panax quinquefolium* L.; quality ecology on chinese medicines; transcriptome; ginsenoside; daodi herbs

西洋参为五加科人参属植物西洋参 (*Panax quinquefolium* L.) 的干燥根, 具有补气养阴、清热生津之功效。西洋参原产于北美, 因其化学成分丰富, 生物活性广泛, 作为保健食品补充剂得到全球认可^[1]。西洋参主要活性成分是人参皂苷类, 因此被广泛地用于抗肿瘤、抗应激、抗衰老、抗疲劳、心脏保护和保肝属性^[2,3]。人参皂苷根据其结构分为原人参二醇 (如 Rb₁、Rb₂、Re、Rd、F₂、Rg₃、Rh₂ 等)、原人参三醇 (如 Re、Rg₁、Rg₂、Rf、Rh₁ 等) 和齐墩果酸 (如 Ro)^[4], 生物体内的合成途径为甲羟戊酸代谢途径 (mevalonic acid pathway, MVA), 主要的关键酶基因有糖基转移酶 (glycosyltransferase, GT)、角鲨烯环氧化酶 (squalene monooxygenase, SE) 等^[5]。本课题组^[6]前期基于人参皂苷成分和气候特征, 应用生物统计、数量分类等分析方法对中国产西洋参进行了生态型划分, 将其分为 Rb₁-Re 山海关外型和 Rg₂-Rd 山海关内型两大化学生态型, 然而缺乏从遗传分子角度的基础性研究。

中药品质生态是研究中药品质及与生态环境关系及机制的关键^[7]。我国幅员辽阔, 拥有复杂的地理地貌条件和多样的气候环境, 这些均极大地影响中药材的品质, 形成了中药材特有的多产地、多道地现象。陈士林^[8]提出: “生态型是道地药材形成的生物学实质”, 从生物与生态学角度阐释道地药材形成的科学内涵。认为生态气候因子通过影响药材生长的环境, 进而影响药材的化学成分含量及遗传物质基础, 并对重楼^[9]、银杏^[10]、川贝母^[11]等大宗道地药材进行了生态适宜性的区划研究。然而相对于水稻等农业经济作物, 中药材的生态型研究相对薄弱, 不同生态型品质差异亟待研究。课题组在多年研究道地性理论及前人工作的基础上, 创新性地提出中药道地性理论研究新体系, 即遗传-化学-生态相关性的综合分析体系。运用 UPLC-QTOF/MS 与转录组和 DNA 条形码等技术, 结合生态气候因子及土壤因子数据综合分析评价道地药材的形成机制, 为深入研究中草药道地性理论提供新思路。该体系已成功应用于肉苁蓉^[12]和黄花蒿^[13]等大宗道地药材。如课题组前期对黄花蒿进行品质变异分析, 发现

秦岭-淮河分界线以南地区黄花蒿的青蒿素含量较高, DNA 条形码分子标记分析显示黄花蒿种质资源存在显著的遗传差异和丰富的多样性, 且遗传多样性与地域相关, GMPGIS 技术预测了湖南武陵山区、广西等优质黄花蒿的生态适宜区^[13]。应用超高效液相色谱法测定山西、内蒙古及甘肃三大主产地的黄芪中的毛蕊异黄酮苷等 6 种化学成分, 运用 DNA 条形码分子标记技术对黄芪进行分析^[14], 同时采用电感耦合等离子体原子发射光谱法 (ICP-AES) 测定黄芪及其根系土壤中 19 中无机元素含量^[15], 通过遗传-化学-生态相关性的综合分析体系探讨了黄芪品质变异的原因及其形成机制。

本研究通过对西洋参山海关内型与山海关外型两大生态型进行转录组测序, 分析比较其转录组差异, 以及人参皂苷合成途径关键酶基因的表达差异, 并结合化学成分含量及气候数据, 通过对遗传、化学及生态的相关性分析探讨影响不同生态型西洋参化学品质变异的潜在分子机制。

材料与方法

材料 本实验材料四年生西洋参样品于 2012 年采自吉林长白山新房子镇北港参场 (关外型: DB) 与山东威海文登张家产镇 (关内型: SD), 采挖后用水洗净, 切断后迅速用锡箔纸包好放入液氮中, 并尽快转移至 -80 °C 冰箱保存。西洋参样品由中国医学科学院药用植物研究所林余霖教授鉴定, 凭证标本保存于中国医学科学院药用植物研究所标本馆。

RNA 提取及检测 应用百泰克植物通用总 RNA 提取试剂盒提取总 RNA。按说明书对西洋参山东及东北样品进行总 RNA 提取。提取后的 RNA 进行琼脂糖凝胶电泳分析 RNA 降解程度, 并运用 Nanodrop (2000, 赛默飞) 检测 RNA 的纯度, 运用 Agilent 2100 生物分析仪 (安捷伦) 精确检测 RNA 的完整性。

文库构建及测序 样品检测合格后, 用带有 Oligo (dT) 的磁珠富集真核生物 mRNA。随后加入破碎缓冲液将 mRNA 打断成短片段, 以 mRNA 为模板, 用六碱

Table 1 Information of *Panax quinquefolium* L. samples. All samples are roots with a growth age of 4 years

Ecotype	Sample ID	Sampling site	Longitude/°	Latitude/°
Outside Shanhaiguan	DB1	New House Town North Port Participation, Changbai Mountain, Jilin	41.5836207800	127.3892081200
	DB2	New House Town North Port Participation, Changbai Mountain, Jilin	41.5861100000	127.3952900000
	DB3	New House Town North Port Participation, Changbai Mountain, Jilin	41.5918778844	127.4018935661
Inside Shanhaiguan	SD1	Zhang Jiachan Town, Shandong	37.0999326500	122.0990977200
	SD2	Zhang Jiachan Town, Shandong	37.1007800000	122.1041300000
	SD3	Zhang Jiachan Town, Shandong	37.1064385767	122.1107613345

基随机引物 (random hexamers) 合成 cDNA 第一链, 然后加入缓冲液、dNTPs 和 DNA polymerase I 合成 cDNA 第二链, 随后利用 AMPure XP beads (Beckman) 纯化双链 cDNA。纯化的双链 cDNA 再进行末端修复、加 A 尾并连接测序接头, 然后用 AMPure XP beads 进行片段大小选择, 最后进行 PCR 富集得到最终的 cDNA 文库。文库构建完成后, 先使用 Qubit2.0 荧光定量仪 (赛默飞) 进行初步定量, 稀释文库至 $1 \text{ ng} \cdot \mu\text{L}^{-1}$, 随后使用 Agilent 2100 对文库的 insert size 进行检测, insert size 符合预期后, 使用 Q-PCR 方法对文库的有效浓度进行准确定量 (文库有效浓度 $>2 \text{ nmol} \cdot \text{L}^{-1}$), 以保证文库质量。库检合格后, 把不同文库按照有效浓度及目标下机数据量的需求合并后进行 Illumina HiSeq 测序。

转录组分析 西洋参无参考基因组, 分析转录组数据时, 先将转录组测得序列拼接成转录本, 以此为参考序列进行后续数据分析。首先对测得数据进行检测, 并去掉质量较差的 reads。应用 Trinity 软件进行转录组拼接, 对拼接后的基因进行功能注释 [参考数据库包括 Nr、Nt、Pfam、KOG/COG、Swiss-Prot、KEGG (Kyoto encyclopedia of genes and genomes)、GO (gene ontology) 等]。对基因进行 GO 注释后, 可将注释成功的基因 GO 三个大类 (BP: Biological process, CC: Cellular component, MF: Molecular function) 的下一层级进行分类, 将 KOG 注释成功的基因按 KOG 的 group 进行分类, 对基因做 KO 注释后, 可根据它们参与的 KEGG 代谢通路进行分类。按 NR、SwissProt、KEGG GENES 的优先级顺序将 unigene 与以上数据库做 blastx 比对, 根据最佳比对结果确定 unigene 基因的 ORF 读码框, 然后根据标准遗传密码子表确定其 CDS 及编码的氨基酸序列; 将与以上数据库都比对不上的 unigene 用 estscan (3.0.3) 软件预测其 CDS 序列。

以 Trinity 拼接得到的转录组作为参考序列 (ref), 采用 RSEM 软件 (bowtie 参数 mismatch 2) 将每个样品的 clean reads 对 ref 做 mapping。RSEM 对 bowtie 的比对结果进行统计, 进一步得到每个样品比对到每个基因上的 read count 数目, 并将之进行 RPKM 转换。RPKM 密度分布能整体检查样品的基因表达模式, 对不同

实验条件下的基因 RPKM 密度分布进行比较能整体上检查不同实验条件之间的 RPKM 分布的情况。两大生态型西洋参基因表达量差异表达的输入数据为基因表达水平分析中得到的 read count 数据。对于有生物学重复的样品, 采用 DESeq 进行分析, 筛选阈值为 $\text{padj} < 0.05$, 该分析方法基于的模型是负二项分布, 第 i 个基因在第 j 个样本中的 read count 值为 K_{ij} , 则有: $K_{ij} \sim \text{NB}(\mu_{ij}, \sigma_{ij}^2)$ 。

火山图可以直观展现 p-value 与 $\log_2(\text{fold change})$ 的关系。差异基因聚类分析用于判断不同实验条件下差异基因表达量的聚类模式。每个比较组合都会得到一个差异基因集, 将所有比较组合的差异基因集的并集在每个实验组/样品中的 RPKM 值用于层次聚类分析。差异基因 GO 富集分析与 KEGG 富集分析可更好地对差异基因的功能进行研究。

人参皂苷含量测定 运用 HPLC 技术对关内型与关外型两大生态型的西洋参进行 7 种人参皂苷 (Rg_1 、 Re 、 Rb_1 、 Ro 、 Rb_2 、 Rd 、 F_2) 含量测定。色谱条件为 Waters C_{18} 色谱柱 ($150 \text{ mm} \times 3.9 \text{ mm}$, $4.6 \mu\text{m}$), 流动相为乙腈与 0.1% 甲酸水, 色谱条件为: 0~15 min, 19% A; 15~20 min, 19%~21% A; 20~22 min, 21%~26% A; 22~40 min, 26%~31% A; 40~45 min, 31% A; 45~57 min, 31%~43% A; 57~60 min, 43%~45% A。流速为 $1 \text{ mL} \cdot \text{min}^{-1}$, 紫外检测波长为 203 nm, 进样量为 $10 \mu\text{L}$ 。方法学考察参考课题组前期实验工作^[16]。

相关性分析 本研究从中国气象数据网 (<http://data.cma.cn/>) 收集整理西洋参两大生态型附近的 6 个气象站 (山东惠民 54 725、潍坊 54 843、成山头 54 776, 吉林长春 54 161、延吉 54 292、临江站 54 374) 1951~2013 年间 24 个气候因子数据, 作为气候因子数据用于后续相关性分析。本研究结合西洋参不同生态型样品差异表达基因及人参皂苷生物合成途径中的关键酶基因表达差异, 找出与人参皂苷合成相关的关键酶基因群。运用各基因表达的 RPKM 值, 结合 7 种人参皂苷化学成分含量以及两产地 24 个生态气候因子数据, 应用 SPSS 19.0 统计分析软件进行 Pearson 相关性分析。

结果与分析

1 西洋参两大生态型7种人参皂苷类成分含量测定

对西洋参中7种主要皂苷类成分进行HPLC含量测定。含量测定结果显示(图1a), 关外型与关内型西洋参样品7种主要皂苷类成分含量存在差异, 关外生态型西洋参的人参皂苷含量较关内生态型西洋参高。PCA分析结果(图1b)也表明两大生态型的西洋参样品差异显著, 明显地被聚成了两类。

2 西洋参两大生态型的转录组测序结果分析

对西洋参转录组数据进行测序错误率分布检查、GC含量分布检查, 并去除低质量的reads, 对原始reads进行过滤, 以得到clean reads(表2)。采用Illumina HiSeq™ 2000高通量测序技术对东北(DB)和山东(SD)2种不同生态型的西洋参测序, 分别获得9.47G和9.48G的原始数据。2组样品分别产生98 526 747和98 840 270条raw reads, 经过质量评估和数据过滤, 去除低质量和带有接头的序列, 分别得到94 713 125和94 822 607条clean reads。Q20(Phred质量值大于20的碱基数占总碱基数clean data的比例)分别为98.26%和98.33%。这些序列经过组装, 总共获得70 793条Unigenes。从Unigene的长度分布情况(图2)可知, Unigene主要集中在200~500 bp区间, 其Unigene

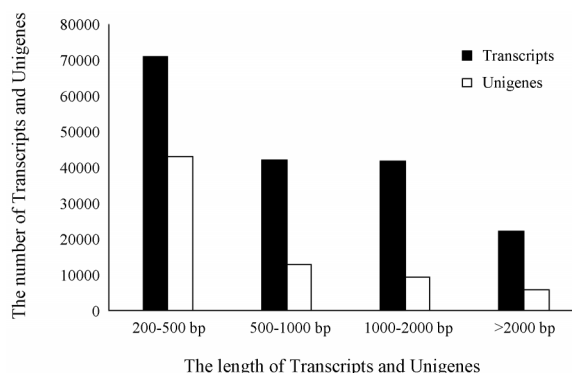


Figure 2 Transcript sequence of *Panax quinquefolium* L. Transcript sequence and Unigene sequence length distribution

数量为43 043, 占总数的60.80%; 其次集中在500~1 000 bp, Unigene数量为12 701个, 占总数的17.94%; 位于1 000~2 000 bp的Unigene数量为9 364个, 占总数的13.23%; 大于2 000 bp的Unigene数量为5 685个, 占总数的8.03%。

3 西洋参两大生态型的基因功能注释

BLASTx同源性比对结果显示(表3), 转录组所产生的70 793条Unigenes中, 共有34 050(48.09%)条Unigenes至少在一个数据库中得到注释。其中, 注释到Nr数据库的有28 580条, 注释到Nt数据库的有15 264

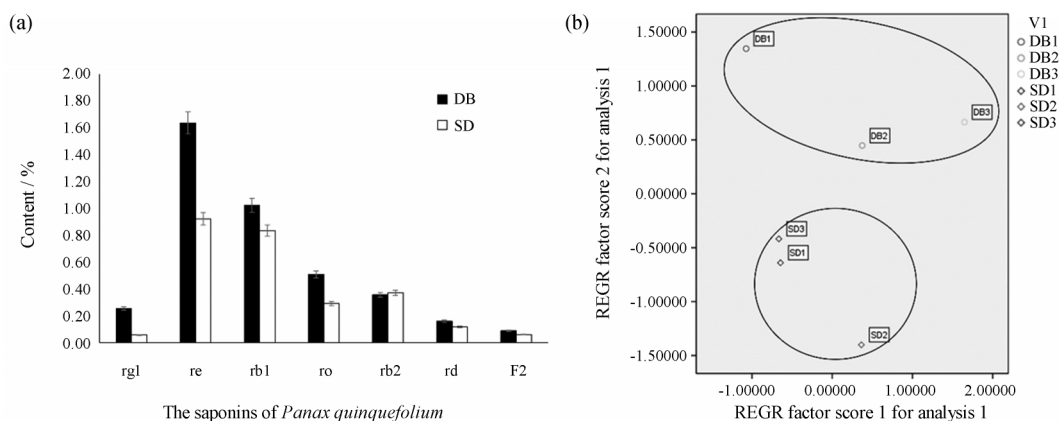


Figure 1 Analysis of saponins in two ecotypes of *Panax quinquefolium* L. (a) Content of 7 saponins (b) PCA analysis of saponins

Table 2 *Panax quinquefolium* L. transcriptome data information

Sample	Raw reads	Clean reads	Clean bases	Error/%	Q20/%	Q30/%	GC/%
SD1_1	34893434	33390538	3.34G	0.03	98.60	94.94	45.11
SD1_2	34893434	33390538	3.34G	0.03	97.92	93.75	45.16
SD2_1	31758271	30599780	3.06G	0.03	98.62	94.95	45.46
SD2_2	31758271	30599780	3.06G	0.03	97.85	93.57	45.52
SD3_1	32188565	30832289	3.08G	0.03	98.65	95.08	45.10
SD3_2	32188565	30832289	3.08G	0.03	97.91	93.75	45.15
DB1_1	32292768	31196124	3.12G	0.02	98.73	95.29	44.35
DB1_2	32292768	31196124	3.12G	0.03	98.06	94.11	44.39
DB2_1	33767990	32488000	3.25G	0.03	98.66	95.08	44.92
DB2_2	33767990	32488000	3.25G	0.03	97.97	93.87	44.98
DB3_1	32465989	31029001	3.1G	0.03	98.65	95.08	44.85
DB3_2	32465989	31029001	3.1G	0.03	97.92	93.78	44.90

Table 3 *Panax quinquefolium* L. transcriptome gene annotation results

Database	The number of the Unigenes	Percent/%
Annotated in NR	28 580	40.37
Annotated in NT	15 264	21.56
Annotated in KO	9 804	13.84
Annotated in SwissProt	21 205	29.95
Annotated in Pfam	22 901	32.34
Annotated in GO	28 873	40.78
Annotated in COG/KOG	11 529	16.28
Annotated in all Databases	4 380	6.18
Annotated in at least one Database	34 050	48.09
Total Unigenes	70 793	100

条, 注释到 KEGG 数据库的有 9 804 条, 注释到 GO、KOG 数据库的分别为 28 873、11 529 条, 其中有 4 380

条被这 7 个数据库同时注释。

将所有 Unigenes 与 GO 数据库比对 (图 3a), 有 28 873 条 Unigenes 被分成三大类, 即“生物学过程 (biological process)”、“细胞组分 (cellular component)”和“分子功能 (molecular function)”。“生物学过程”有最多的四级功能分类, 共有 1 331 条 Unigenes。在“生物学过程”分类中, 注释最多的是“细胞组分的组织和生物合成 (cellular process)”, 其次为“代谢途径 (metabolic process)”, 推测这些途径与代谢产物的合成有关。共有 387 条 Unigenes 注释到“细胞组分”分类中, 其中占比例最多的是“细胞 (cell)”与“细胞区域 (cell part)”, “细胞器 (organelle)”其次。仅有 172 条 Unigenes 注释到“分子功能”分类, 其中“连接 (Binding)”占比例最大, 其次为“催化反应 (catalytic activity)”。

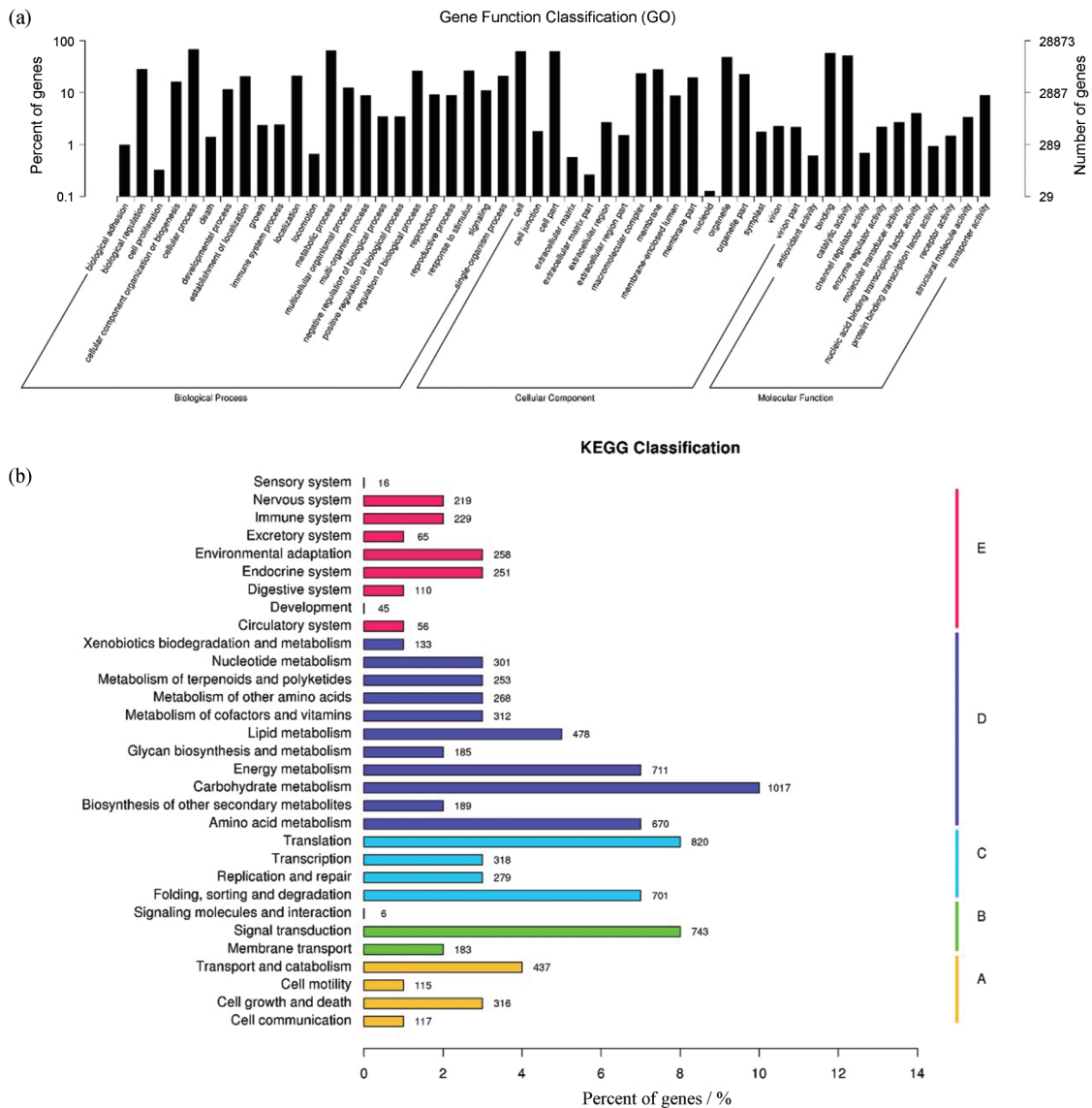


Figure 3 Functional annotation of two ecological *Panax quinquefolium* L. transcriptome genes. (a) GO classification map (b) KEGG classification map. A: Cellular processes; B: Environmental information processing; C: Genetic information processing; D: Metabolism; E: Organismal systems

根据Unigene的KEGG注释,共有9 804个Unigene被注释,被注释到的通路共有31条。Unigene参与的KEGG代谢通路可分为5个分支(图3b):有机系统(organismal systems)、代谢(metabolism)、遗传信息处理(genetic information processing)、环境信息处理(environmental information processing)、细胞过程(cellular processes)。对西洋参的转录组数据进行代谢途径分析,比对到了254个KEGG通路。“碳水化合物代谢(carbohydrate metabolism)”分类拥有最多的Unigenes。“代谢”中,大多数代谢通路为能量、氨基酸、脂质等代谢产物。有6个与人参皂苷生物合成相关的通

路(表4),包括倍半萜和三萜生物合成(sesquiterpenoid and triterpenoid biosynthesis)、萜类骨架生物合成(terpenoid backbone biosynthesis)、类固醇生物合成(steroid biosynthesis)等。其中,有1条涉及到角鲨烯合成酶(SS)、3条涉及角鲨烯环氧酶(SE)、3条涉及磷酸甲羟戊酸激酶(PMK)等。

4 西洋参两大生态型差异表达基因的筛选及功能注释

对两种不同生态型的西洋参进行差异筛选(图4a),在关内型与关外型西洋参中共检测到339个差异表达基因,其中上调基因79个,下调基因260个。这些

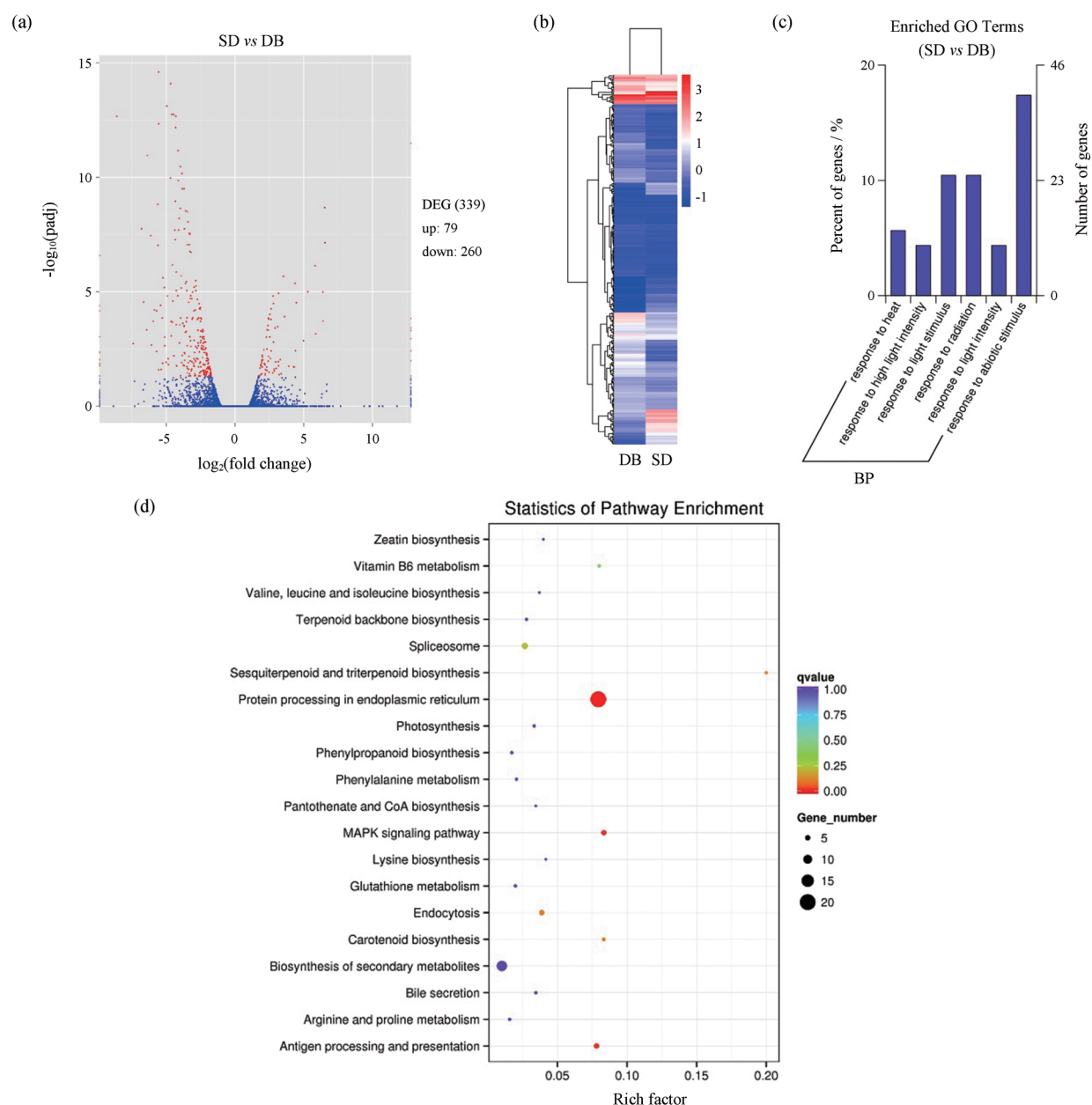


Figure 4 Analysis of differentially expressed genes between two major ecological *Panax quinquefolium* L. a: Volcano map; b: Hierarchical clustering map; c: GO enrichment analysis of DEGs; d: KEGG enrichment analysis of DEGs

Table 4 Two major ecological *Panax quinquefolium* L. ginseng saponins biosynthesis pathway-related KEGG pathway

Ko Number	KEGG pathway	Key unigene	The number of the unigenes
ko00909	Sesquiterpenoid and triterpenoid biosynthesis	comp75083_c0, comp65271_c0, comp76120_c0, comp59474_c0, comp69688_c0, comp70679_c0	10
ko00900	Terpenoid backbone biosynthesis	comp58220_c0, comp58220_c1, comp77697_c0, comp76561_c0, comp69124_c0, comp65081_c0, comp76232_c1, comp67432_c0, comp65471_c0, comp72369_c0	72
ko04146	Peroxisome	comp69588_c0	108
ko00280	Valine, leucine and isoleucine degradation	comp67540_c0	77
ko00100	Steroid biosynthesis	comp70267_c0, comp61865_c0	24
ko02020	Two-component system	comp53463_c0, comp53463_c1, comp63076_c0	122

差异基因中, 共有 267 个有注释信息, 其他基因可能为西洋参所特有基因。两大生态型西洋参的差异基因表达层次聚类 (图 4b), 其颜色相近聚类区内的基因表达模式相近, 表明关内与关外生态型的西洋参具有更多相似的功能或参与调控同一条代谢通路的基因。同时也显示关内与关外生态型西洋参基因的表达模式不尽相同。

通过 GOseq 对两种不同生态型的西洋参基因进行 GO 富集分析 (图 4c)。1 792 条 Unigenes 中有 1 153 条 Unigenes 被富集为“生物过程 (biological process)”。关内型西洋参与关外型西洋参相比, 富集度最高的基因均涉及“对非生物刺激的反应 (response to abiotic stimulus)”、“光刺激反应 (response to light stimulus)”、“应对辐射 (response to radiation)”等, 表明两大生态型西洋参的基因在与胁迫刺激相关中表达差异显著。

对两大生态型西洋参差异基因进行 KEGG pathway 富集分析 (图 4d), 涉及人参皂苷生物合成的代谢通路有 4 条, 包括“倍半萜和三萜类生物合成途径 (sesquiterpenoid and triterpenoid biosynthesis, ko00909)”、“萜类骨架生物合成 (terpenoid backbone biosynthesis, ko00900)”、“次生代谢产物生物合成 (biosynthesis of secondary metabolites, ko01110)”和“缬氨酸、亮氨酸和异亮氨酸降解 (valine, leucine and isoleucine biosynthesis, ko00280)”。其中有 2 条 Unigenes 被注释到“倍半萜和三萜类生物合成途径”、2 条 Unigenes 被注释到“萜类骨架生物合成”、12 条 Unigenes 被注释到“次生代谢产物生物合成”以及 1 条 Unigenes 被注释到“缬氨酸、亮氨酸和异亮氨酸降解”。

5 人参皂苷生物合成相关基因挖掘

人参皂苷属于三萜类, 是由苷元和糖相连而成的糖苷类化合物。结合 KEGG 注释情况与文献^[17], 鉴定了两大生态型西洋参人参皂苷生物合成的关键基因群 (表 5), 如角鲨烯合成酶 (SS)、角鲨烯环氧酶 (SE) 和磷酸甲羟戊酸激酶 (PMK) 等。同时绘制了可能的生

Table 5 Key genetic groups of two major ecological *Panax quinquefolium* L. ginseng saponins biosynthesis pathways

Gene name	EC	Gene ID
Farnesyl-diphosphate farnesyltransferase (SS)	2.5.1.21	comp75083_c0
Squalene monooxygenase (SE)	1.14.13.17	comp65271_c0 comp76322_c0 comp76120_c0
Phosphomevalonate kinase (PMK)	2.7.4.2	comp58220_c0 comp58220_c1 comp77697_c0
Diphosphomevalonate decarboxylase (MVD)	4.1.1.33	comp76561_c0
Mevalonate kinase (MK)	2.7.1.36	comp69588_c0
Isopentenyl-diphosphate delta-isomerase (IDI)	5.3.3.2	comp69124_c0
Hydroxymethylglutaryl-CoA synthase (HMGS)	2.3.3.10	comp67540_c0
Hydroxymethylglutaryl-CoA reductase (HMGR)	1.1.1.34	comp65081_c0 comp76232_c1 comp67432_c0 comp65471_c0 comp72369_c0
Farnesyl diphosphate synthase (FPS)	2.5.1.10	comp70267_c0 comp61865_c0
Cycloartenol synthase (CAS)	5.4.99.8	comp59474_c0 comp69688_c0 comp70679_c0
Beta-amyrin synthase (AS)	5.4.99.39	comp59474_c0 comp69688_c0 comp70679_c0
Acetyl-CoA C-acetyltransferase (AACT)	2.3.1.9	comp53463_c0
Dolichyl-diphosphooligosaccharide-protein glycosyltransferase (GT)	-	comp53463_c1 comp63076_c0 comp70364_c0 comp60837_c0 comp75027_c0 comp64229_c0

物合成通路图 (图 5)。从通路图中可看到两大生态型之间关键基因表达量的差异情况, 通路中除了糖基转移酶 (GT) 外, 大部分关键基因表达量均是关外型高于关内型的。

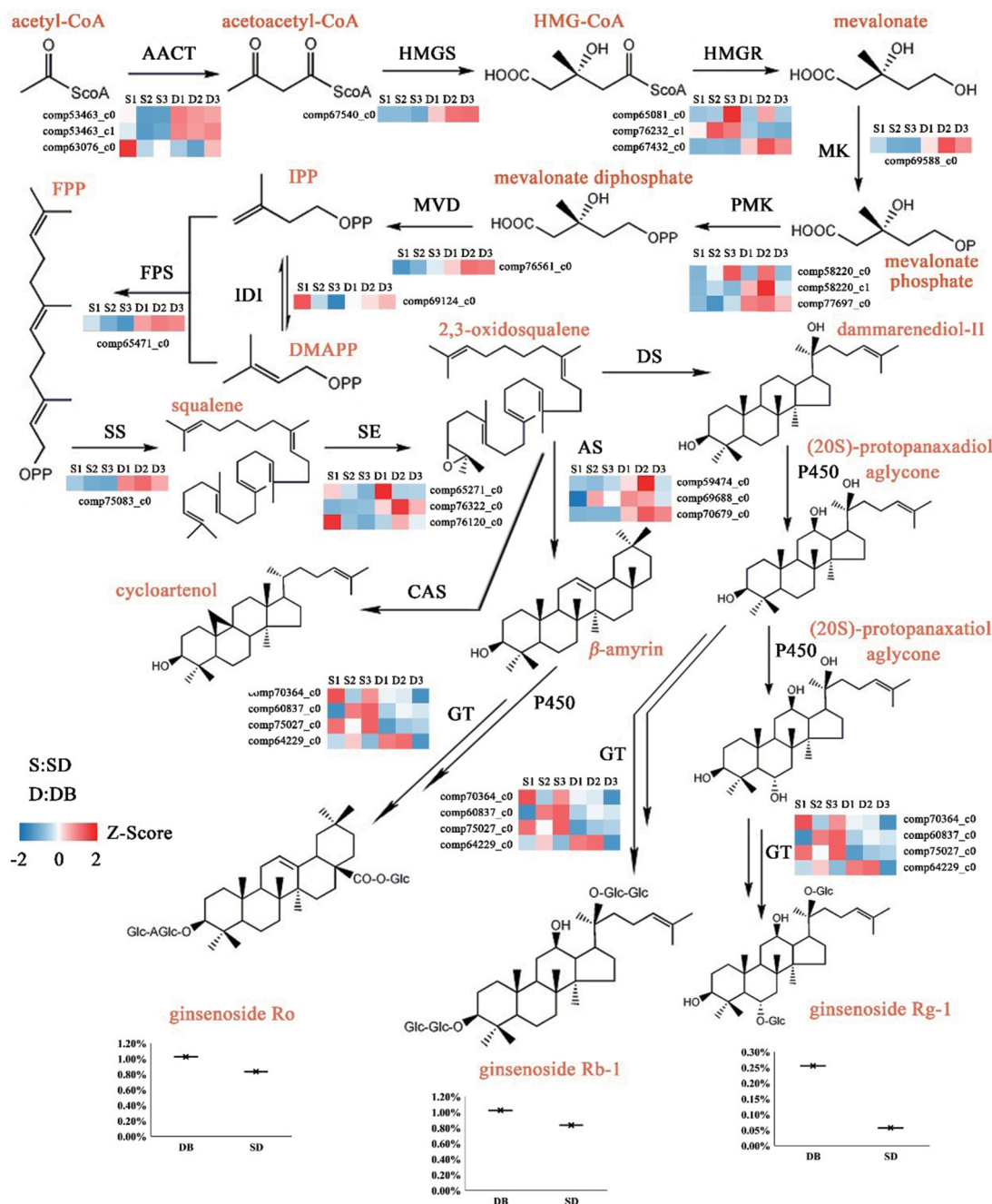


Figure 5 Biosynthesis metabolic pathway of two important *Panax quinquefolium* L. ginseng saponins

进一步研究发现,两个被注释到“倍半萜和三萜类生物合成途径”的差异基因分别为角鲨烯环氧酶 squalene monooxygenase (SE, EC: 1.14.13.17, comp70679_c0) 和 β -氨基酸合成酶 beta-amyrin synthase (AS, EC: 5.4.99.39, comp76322_c0), 且这两个基因在关内型西洋参中显著下调。被注释到“萜类骨架生物合成”的两个差异基因分别为法呢基焦磷酸合成酶 farnesyl diphosphate synthase (FPS, EC: 2.5.1.10, comp72369_c0), 羟甲基戊二酰辅酶 A 还原酶 hydroxymethylglutaryl-CoA reductase (HMGR, EC: 1.1.1.34, comp67432_c0)。

同样的, 这些基因在关内型西洋参中也是显著下调的。

6 西洋参生态气候因子分析

从中国气象数据网下载东北与山东两产地西洋参样品采集地附近的气象站数据, 收集 1951~2013 年间 24 个气候因子的数据 (表 6), 包括极大风速、极大风速的风向、极端最低本站气压、极端最低气温、极端最高本站气压、最小相对湿度、蒸发量 (小型)、极端最高气温、最大日降水量、降水量、平均本站气压、最大风速、最大风速之风向、平均风速、平均气温、平均气温距平、平均水汽压、平均相对湿度、平均最低气温、平均最高

Table 6 Ecological factors of two major ecological American ginseng samples. 54725: Shandong Huimin; 54843: Shandong Weifang; 54776: Shandong Chengshantou; 54161: Jilin Changchan; 54292: Jilin Yanji; 54374: Jilin Linjiang

Item	Ecological factor					
	54725	54843	54776	54161	54292	54374
Extreme wind speed /m·s ⁻¹	219.454 5	230.937 5	331.612 2	235.916 7	268.714 3	178.888 9
Extreme wind Direction	8.727 3	10.225 8	9.574 5	11.416 7	13.833 3	10.555 6
Extreme minimum station pressure /hPa	9 916.119 0	9 890.071 4	9 859.771 4	9 615.767 4	9 673.750 0	9 524.275 0
Extreme minimum temperature /°C	-155.269 8	-148.555 6	-96.016 4	-289.682 5	-268.852 5	-304.721 3
Extreme maximum station pressure /hPa	10 899.761 9	10 381.285 7	10 338.571 4	10 113.767 4	10 149.150 0	9 987.150 0
Minimum relative humidity /%	7.454 5	7.425 5	11.904 8	7.208 3	5.913 0	9.431 8
Evaporation capacity (small) /mm	18 323.627 5	19 336.120 0	14 437.016 9	16 610.941 2	13 218.959 2	10 701.571 4
Extreme maximum temperature /°C	377.015 9	378.714 3	297.450 0	335.825 4	345.868 9	336.163 9
Maximum daily precipitation /mm	833.730 2	1 234.841 3	1 066.596 8	684.193 5	493.465 5	647.508 2
Precipitation /mm	5 760.206 3	6 172.142 9	7 154.983 9	5 877.682 5	5 199.541 0	8 323.901 6
Mean station pressure /hPa	10 156.222 2	10 120.111 1	10 114.322 0	9 868.111 1	9 928.366 7	9 764.508 5
Maximum wind speed /m·s ⁻¹	163.062 5	153.500 0	252.022 7	209.254 5	172.195 1	107.791 7
Maximum wind direction	7.022 7	9.975 0	10.952 4	11.019 6	13.029 4	9.933 3
Average wind velocity /°C	29.428 6	32.952 4	63.161 3	38.666 7	25.475 4	12.409 8
Average temperature /°C	125.444 4	124.888 9	113.573 8	54.793 7	53.262 3	51.508 2
Average temperature anomaly /°C	-0.603 2	-0.142 9	-0.459 0	-2.269 8	-0.786 9	-1.541 0
Mean vapor pressure	119.206 3	118.904 8	120.774 2	83.158 7	82.573 8	87.147 5
Average relative humidity /%	65.730 2	66.555 6	74.129 0	63.666 7	65.065 6	70.082 0
Mean minimum temperature /°C	74.000 0	74.349 2	91.311 5	4.460 3	-3.836 1	-3.163 9
Mean maximum temperature /°C	186.984 1	186.984 1	140.983 3	111.285 7	121.377 0	121.366 7
Daily number of days precipitation ≥ 0.1 mm	70.063 5	78.047 6	89.225 8	100.825 4	102.508 2	138.803 3
Percentage of sunshine /%	58.096 8	58.152 5	55.660 7	58.698 4	52.233 3	49.883 3
Sunshine duration /h	25 751.000 0	25 764.288 1	24 714.232 1	26 071.142 9	23 200.933 3	22 166.233 3

气温、日降水量≥0.1 mm 日数、日照百分率、日照时数。数据下载后, 去掉未检测数据及超出正常范围的数值, 整理以备后续相关性分析用。

7 相关性分析

基于遗传-化学-生态相关性分析的研究新体系, 对两大生态型西洋参转录组的人参皂苷合成相关 24 个差异基因的 RPKM 值、7 种人参皂苷成分含量, 以及 24 个气候因子数据进行相关性分析, 以分析不同生态型西洋参品质差异的分子与生态机制 (图 6)。西洋参人参皂苷的含量与大部分的人参皂苷体内生物合成途径上的关键酶基因的表达成正相关, 而与主要气候因子成负相关。人参皂苷 Rg₁ 的含量与关键基因 MK、HMGS、HMGR (comp67432_c0) 和 AS 的表达量成极强正相关, 与差异关键基因 SS、SE、MVD、FPS 成

强正相关, 而与 HMGR (comp76232_c1) 成强负相关。人参皂苷 Re 与差异关键基因 AS 的表达量成极强正相关, 与 SE、PMK (comp58220_c1) 基因的表达量成强正相关。人参皂苷 Rb₂ 的含量与差异关键基因 HMGR (comp65081_c0) 成极强正相关, 与 PMK (comp58220_c0) 基因的表达量成强正相关。人参皂苷的含量与日降水量≥0.1 mm 日数、极大风速的风向成正相关, 与其他气候因子主要成负相关。其中, 人参皂苷 Rg₁ 的含量与日降水量≥0.1 mm 日数成强正相关, 而与极端最低气温、平均最低气温、平均水汽压、平均本站气压、极端最低本站气压、平均气温、日照百分率及日照时数成强负相关。人参皂苷 Re 的含量与极大风速的风向成强正相关。人参皂苷 Ro 的含量与平均气温及平均最高气温呈强负相关。人参皂苷 Rd 的含量与最大风速成强

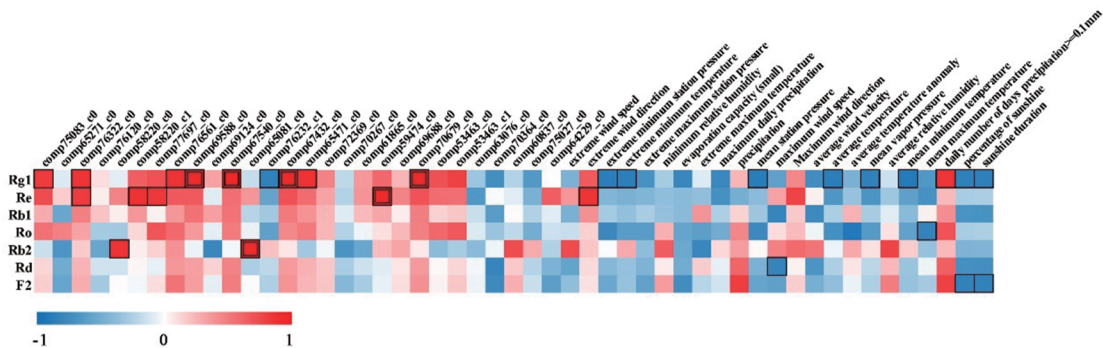


Figure 6 Correlation analysis based on genetic-chemical-ecological characteristics

负相关。人参皂苷 F2 的含量与日照百分率、日照时数成强负相关。由此可得, 人参皂苷的含量与大部分的生物合成途径上的关键酶基因的表达成正相关, 与关键基因 *HMGR* (comp76232_c1) 成负相关。同时人参皂苷的含量与气候因子降水量成正相关, 而与温度、气压及日照等成负相关。

讨论

本研究对西洋参关内型与关外型两大生态型的样品进行基于遗传-化学-生态的综合性相关性分析, 对两大生态型差异基因表达进行比较, 结合两大生态型 7 种皂苷类成分含量测定分析及 24 个气候因子数据, 分析西洋参不同生态型样品基因表达差异, 找出 24 个人参皂苷生物合成途径相关的差异关键酶基因, 并结合其表达量 (RPKM 值)、气候因子与人参皂苷含量进行相关性分析, 为阐明西洋参道地性的形成机制提供参考。

西洋参因其广泛应用及显著的药理活性而受到全球关注, 人参皂苷作为其主要活性成分更是研究的热点。现代药理学研究表明, 人参皂苷具有较好的抗肿瘤、抗炎、抗氧化和抑制细胞凋亡等药理活性^[18]。如 Rg_1 、 Rb_1 具有抗衰老活性, Rg_3 、 Rh_2 具有抗肿瘤活性等^[19]。*HMGR* 被公认为人参皂苷生物合成途径中的第一个限速酶, 是萜类合成过程中最先起作用的关键酶, 通过影响人参皂苷前体 IPP 和 DMAPP 的产量而影响人参皂苷的生物合成^[20]。*FPS* 在人参皂苷的生物合成中起着重要作用^[21]。*SS*^[22]和 *SE*^[23]是催化甾醇和三萜类化合物合成的关键酶基因, 对人参皂苷的含量有重要影响。由氧化鲨烯环化酶 (oxidosqualene cyclase, OSC) 催化的 2,3-氧化鲨烯环化也是三萜皂苷和甾醇生物合成分支形成的关键位点^[24]。*GT* 催化人参皂苷苷元的糖基化, 是人参皂苷生物合成途径中最下游的步骤。糖基化可以增加人参皂苷的稳定性和水溶性, 从而增加了人参皂苷的多样性^[25]。由图 1a 可知, 关外型西洋参的人参皂苷除 Rb_2 外, 其余如 Rg_1 、 Ro 、 Rb_1 等的含量均高于关内型人参皂苷含量。转录组分析结果表明, 两大生态型西洋参的基因在表达模式上不尽相同, 且关内型与关外型相比, 人参皂苷生物合成关键基因如 *SE*、*AS*、*FPS* 和 *HMGR* 等均在关内型显著下调, 与人参皂苷含量差异是一致。图 5 结果显示大多人参皂苷生物合成关键基因如 *AACT*、*HMGR*、*HMGS*、*MK*、*PMK*、*MVD*、*FPS*、*SS*、*SE*、*AS* 等, 在关外型中表达量较高, 而在关内型却较低。值得注意的是, 下游基因糖基转移酶 (*GT*) 在关外型西洋参中表达量较低, 却在关内型中表达量较高。

中药材品质变异是物种对不同生态环境长期适应与自然选择的结果。遗传因素是中药材品质变异的生物学实质, 环境因素是中药材品质变异的生态学实质^[8]。基于遗传-化学-生态相关性分析, 结果表明西洋参人参皂苷的含量与 *AACT*、*HMGR*、*HMGS*、*MK*、*PMK*、*MVD*、*FPS*、*SS*、*SE*、*AS* 等人参皂苷生物合成关键基因的表达成正相关, 而与糖基转移酶 *GT* 的表达成负相关。在药用植物中气候是直接影响药用植物的生长、发育和繁殖的关键因素^[26]。温度、光照和降水等是影响气候较大的因素。相关性分析结果显示, 西洋参的人参皂苷含量与两大生态型所处环境的温度、日照等气候因子成负相关, 与水分成正相关。关外地区属于温带湿润、半湿润大陆性季风气候, 地处长白山脉, 海拔较高且植被丰富, 所以气温较低, 蒸发较弱, 降水量不多, 但湿度仍较高。而关内地区则属于暖温带季风气候且临近海洋^[6]。该结果表明关外型西洋参的人参皂苷含量高于关内型, 可能与关外型所处环境温度更低、日照更少, 湿度较高有关。综上, 关外型西洋参品质较关内型西洋参高, 其中人参皂苷生物合成关键基因在关外型的表达量较高及关外型气温低、日照少和温度高等可能是品质差异的原因。研究结果为科学合理阐释西洋参品质差异成因提供参考, 为中药材品质生态学理论研究提供科学依据, 进一步为西洋参分子育种的优质品种研究奠定基础。

References

- [1] Cheng YQ, Wang CX, Xu T, et al. Progress in quality evaluation of American ginseng medicinal materials [J]. Shandong Sci (山东科学), 2018, 4: 39-43,49.
- [2] Lee SH, Jung BH, Kim SY, et al. The antistress effect of ginseng total saponin and ginsenoside Rg_3 and Rb_1 evaluated by brain polyamine level under immobilization stress [J]. Pharmacol Res, 2006, 54: 46-49.
- [3] Lee J, Jung E, Lee J, et al. *Panax ginseng* induces human Type I collagen synthesis through activation of Smad signaling [J]. J Ethnopharmacol, 2007, 109: 29-34.
- [4] Chen QQ, Wu QQ, Gao XJ, et al. Research progress on fungal biotransformation of ginseng saponins in American ginseng [J]. Food Res Devel (食品研究与开发), 2019, 40: 220-224.
- [5] Wang J, Li J, Li J, et al. Transcriptome profiling shows gene regulation patterns in ginsenoside pathway in response to methyl jasmonate in *Panax quinquefolium* adventitious root [J]. Sci Rep, 2016, 6: 37263.
- [6] Huang LF, Suo FM, Song JY, et al. Quality variation and ecotype division of *Panax quinquefolium* in China [J]. Acta Pharm Sin (药学报), 2013, 48: 580-589.
- [7] Huang LF, Chen SL. Quality ecology of traditional Chinese

- medicine: an emerging interdisciplinary subject [J]. *Chin J Exp Tradit Med Form* (中国实验方剂学杂志), 2017, 23: 1-11.
- [8] Huang LF, Fu J, Chen SL. Academic discussion on ecological variation of Chinese medicinal materials [J]. *Chin Tradit Herb Drugs* (中草药), 2012, 43: 1249-1258.
- [9] Cheng RZ, Wu ML, Shen L, et al. Analysis of ecological suitability of global producing area of traditional Chinese medicine [J]. *Chin J Exp Tradit Med Form* (中国实验方剂学杂志), 2017, 23: 19-24.
- [10] Xu J, Shen L, Wang Y, et al. Analysis of global ecological suitable production area of *Ginkgo biloba* based on GMPGIS [J]. *World Chin Med* (世界中医药), 2017, 12: 969-973.
- [11] Jiang XY, Sun HB, Qin JH, et al. Study on the functional zoning of *Fritillaria cirrhosa* based on growth suitability and quality suitability [J]. *China J Chin Mater Med* (中国中药杂志), 2016, 41: 3194-3201.
- [12] Zheng SH, Jiang X, Huang LF, et al. Chemical and genetic discrimination of *Cistanches Herba* based on UPLCQTOF/MS and DNA barcoding [J]. *PLoS One*, 2014, 9: e98061.
- [13] Lin L, Josef BA, Bing L, et al. Three-dimensional evaluation on ecotypic diversity of traditional Chinese medicine: a case study of *Artemisia annua* L. [J]. *Front Plant Sci*, 2017, 8: 1225.
- [14] Li L, Zheng SH, Brinckmann JA, et al. Chemical and genetic diversity of *Astragalus mongholicus* grown in different ecoclimatic regions [J]. *PLoS One*, 2017, 12: e0184791.
- [15] Li L, Zheng SH, Yang QZ, et al. Distinguishing *Astragalus mongholicus* and its planting soil samples from different regions by ICP-AES [J]. *Molecules*. 2016, 21: 482.
- [16] Wang ZH. Study on the Quality and Origin Variation of American Ginseng and Screening of EGFR Kinase Inhibitors for 49 Medicinal Extracts (西洋参品质和产地变异研究及49种药材提取物EGFR激酶抑制剂筛选) [D]. Beijing: Peking Union Medical College, 2015.
- [17] Yang JL, Gao LL, Zhu P. Advances in the biosynthesis research of ginsenosides [J]. *Acta Pharm Sin* (药学报), 2013, 48: 170-178.
- [18] He DT, Wang B, Chen JM. Research progress on pharmacological effects of ginsenoside [J]. *J Liaoning Univ Tradit Chin Med* (辽宁中医药大学学报), 2012, 14: 118-121.
- [19] Christensen JC. Ginsenosides chemistry, biosynthesis, analysis, and potential health effects [J]. *Adv Food Nutr Res*, 2009, 55: 1-99.
- [20] Wu Q, Sun C, Chen SL. Identification and expression analysis of a 3-hydroxy-3-methylglutaryl coenzyme A reductase gene from American ginseng [J]. *Plant OMICS: J Plant Mol Biol Omics*, 2012, 5: 414-420.
- [21] Kim OT, Bang KH, Jung SJ, et al. Molecular characterization of ginseng farnesyl diphosphate synthase gene and its up-regulation by methyl jasmonate [J]. *Biol Plant*, 2010, 54: 47-53.
- [22] Lee MH, Jeong JH, Seo JW, et al. Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* overexpressing squalene synthase gene [J]. *Plant Cell Physiol*, 2004, 45: 976-984.
- [23] Han JY, In JG, Kwon YS, et al. Regulation of ginsenoside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in *Panax ginseng* [J]. *Phytochemistry*, 2010, 71: 36-46.
- [24] Lee MH, Han JY, Kim HJ, et al. Dammarenediol-II production confers TMV Tolerance in transgenic tobacco expressing *Panax ginseng* dammarenediol-II synthase [J]. *Plant Cell Physiol*, 2012, 53: 173-182.
- [25] Yue CJ, Zhong JJ. Purification and characterization of UDPG: ginsenoside Rd glucosyltransferase from suspended cells of *Panax notoginseng* [J]. *Proc Biochem*, 2005, 40: 3742-3748.
- [26] Xing N, Liu LL, Yin SG, et al. The association between property of Chinese medicinal herbs and chemical components, ecological environment [J]. *J Chin Med Mater* (中药材), 2011, 34: 324-325.