

基于转录组测序的白鼠尾草 2-ODD 基因家族鉴定与表达分析

秦晓静¹, 胡佳栋¹, 房振西¹, 陈万生^{1,2}, 李 卿^{2*}

(1. 上海中医药大学中药研究所, 中药资源与生物技术中心, 上海 201203; 2. 中国人民解放军海军军医大学第二附属医院, 药剂科, 上海 200003)

摘要: α -酮戊二酸 (2OG)/Fe(II) 依赖双加氧酶 (2-ODD) 在植物初生和次生代谢过程中发挥着重要作用。本研究基于高通量测序技术平台 Illumina NovaSeq 6000 进行了白鼠尾草 (*Salvia apiana* Jepson) 转录组测序, 对测序结果进行 *de novo* 拼接后, 共获得 38 534 个独立基因 (unigenes)。将获得的独立基因进行基因功能注释, 有 29 982 个独立基因获得功能注释。通过生物信息学方法在白鼠尾草转录组数据中进行 2-ODD 基因家族鉴定, 并对鉴定得到的基因序列进行序列特征分析, 包括序列的同源性、理化特征、信号肽、跨膜结构域、亚细胞定位、二级结构和三级结构预测等, 另对它们的进化关系及表达模式进行分析。结果显示, 从白鼠尾草转录组中共鉴定得到 39 个具有完整序列结构的 2-ODD 家族基因, 这些基因编码蛋白质的平均长度为 320 个氨基酸, 分子质量约为 36.00 kDa, 多为亲水性蛋白。系统进化分析将白鼠尾草 2-ODD 基因分为多个亚家族, 这些基因在白鼠尾草不同部位均有表达, 多数基因在根中的表达量明显高于叶中的表达量。本研究可为白鼠尾草 2-ODD 基因的深入研究奠定基础。

关键词: 白鼠尾草; 转录组; α -酮戊二酸/Fe(II) 依赖双加氧酶; 生物信息学; 表达分析

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2022)12-3675-11

Transcriptome-wide identification and expression analysis of 2-ODD gene family in *Salvia apiana* Jepson

QIN Xiao-jing¹, HU Jia-dong¹, FANG Zhen-xi¹, CHEN Wan-sheng^{1,2}, LI Qing^{2*}

(1. Research and Development Center of Chinese Medicine Resources and Biotechnology, Institute of Chinese Materia Medica, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China; 2. Department of Pharmacy, Second Affiliated Hospital of Naval Medical University, Shanghai 200003, China)

Abstract: 2-Oxoglutarate/Fe(II)-dependent dioxygenases (2-ODD) play an important role in plant primary and secondary metabolism. Based on the high-throughput sequencing platform Illumina NovaSeq 6000, the transcriptome of *Salvia apiana* Jepson was sequenced, and the obtained reads were *de novo* assembled. A total of 38 534 unigenes were obtained from the transcriptome. The assembled unigenes were annotated and 29 982 unigenes were given functional annotations. The 2-ODD genes were identified from the assembled *S. apiana* transcriptome database by bioinformatics methods, and the genes were analyzed, including the homology of the sequences, physicochemical characteristics, signal peptides, transmembrane domains, subcellular localization, secondary structure and tertiary structure, etc. The evolutionary relationships and the expression patterns of the identified 2-ODD genes were also analyzed. 39 full-length 2-ODD genes were identified from the transcriptome of *S. apiana*. The average length of these 2-ODD encoding proteins was 320 amino acids, the molecular weight was about 36.00 kDa, and most of them were hydrophilic proteins. Phylogenetic analysis divided these 2-ODD genes into several subfamilies. Gene expression analysis indicated that the 2-ODD genes were expressed in different parts of *S. apiana*, and the

收稿日期: 2022-05-26; 修回日期: 2022-06-15.

基金项目: 国家自然科学基金项目 (32070327, 31770329); 国家重点研发计划项目 (2019YFC1711100).

*通讯作者 Tel: 86-21-51322403, E-mail: qli@smmu.edu.cn

DOI: 10.16438/j.0513-4870.2022-0648

expression level of most genes was much higher in roots than that in leaves. This study can lay a foundation for further study of 2-*ODD* genes in *S. apiana*.

Key words: *Salvia apiana* Jepson; transcriptome; 2-oxoglutarate/Fe(II)-dependent dioxygenase; bioinformatics; expression analysis

白鼠尾草 (*Salvia apiana* Jepson) 是唇形科鼠尾草属多年生草本植物, 原产于墨西哥北部和美国南部地区, 我国有小部分人工种植区域。在原产地, 白鼠尾草作为丘马什人 (Chumash) 的食物添加剂和传统药物, 具有悠久的药用历史^[1], 我国尚无白鼠尾草的药用记录。白鼠尾草的种子具有清洁眼睛的作用; 其叶能用于治疗感冒, 具有清洁头发、防止头发变白和消除体味等作用; 其烟雾具有净化居所、驱虫灭菌的功效, 能用于治疗鼻炎、鼻窦炎, 可缓解鼻塞、多涕等症状, 还能用于热灸治疗, 缓解经期痉挛和一些更年期症状^[2]。现代药理学研究表明, 白鼠尾草具有抗菌、抗炎、抗氧化、抗肿瘤和镇痛等药理学活性^[3]。天然产物化学研究发现白鼠尾草主要含有酚酸类和大量二萜类化合物^[3,4]。本课题组前期研究显示, 白鼠尾草具有与同属植物——我国传统中药丹参 (*Salvia miltiorrhiza* Bge.) 类似的化合物及生物合成途径, 特别是以丹参酮 IIA 为代表的松香烷类二萜化合物的生物合成^[5]。探究白鼠尾草中丹参酮 IIA 的分子调控机制, 可将白鼠尾草培育成丹参酮 IIA 提取的替代植物, 达到丹参药用资源保护的目。Song 等^[6]发现丹参中的 *Sm2-ODD14* 基因在丹参酮 IIA 生物合成过程中发挥了重要作用, 能将隐丹参酮直接转化为丹参酮 IIA。

α -酮戊二酸 (2OG)/Fe(II) 依赖双加氧酶 [2-oxoglutarate/Fe(II)-dependent dioxygenases, 2-*ODD*] 广泛存在于自然界, 在微生物、植物和动物中均有报道^[7]。植物中的 2-*ODD* 基因家族成员数量庞大, 不仅广泛参与植物的生长发育过程, 还在次生代谢产物的生物合成过程中发挥着重要作用^[8]。根据序列相似性, 2-*ODD* 基因家族中的成员可分为三大类: DOXA、DOXB 和 DOXC。DOXA 亚家族成员由大肠杆菌中的 DNA 修复蛋白 Alkb 进化而来^[9], 主要参与初生代谢过程; DOXB 亚家族成员相对保守, 编码脯氨酰 4-羟化酶 (prolyl 4-hydroxylase, P4Hs), 主要参与多肽链的翻译后修饰, 对植物信号肽类激素和细胞壁形成意义重大^[10]。DOXC 亚家族成员复杂, 该亚家族成员在不同植物体内的功能和数量差异悬殊, 具有明显的物种间特异性; 该亚家族成员主要参与植物的次生代谢过程, 包括萜类、生物碱类、植物激素、黄酮类和酚酸类等物质的生物合成, 绝大多数参与植物次生代谢的 2-*ODD* 基因都可以归

为这一亚家族^[9]。

2-*ODD* 基因广泛存在, 且作用重要, 目前对白鼠尾草 2-*ODD* 基因的相关研究尚未见报道。由于白鼠尾草缺少基因组信息, 因此本研究基于转录组测序, 对白鼠尾草 2-*ODD* 基因家族进行了鉴定, 并对其表达模式进行了分析, 以期对白鼠尾草 2-*ODD* 基因的功能研究及其在白鼠尾草次生代谢途径中的作用提供参考。

材料与方法

植物材料 植物材料种植于上海中医药大学药圃, 经中国科学院昆明植物研究所向春雷研究员鉴定为白鼠尾草 (*Salvia apiana* Jepson)。为了获得白鼠尾草 *de novo* 拼接的转录组序列, 采集 6 月龄 (2020 年 5~11 月) 白鼠尾草全株, 提取总 RNA 后, 用于转录组测序。为了获得白鼠尾草地上与地下部分的差异表达基因, 分别采集根和叶 (3 个生物学重复), $-80\text{ }^{\circ}\text{C}$ 保存, 用于表达谱测序。

总 RNA 提取、cDNA 文库构建与转录组测序 用 TRIzol[®] 试剂 (美国 Invitrogen 公司) 提取白鼠尾草不同部位的总 RNA, 并用 DNase I 试剂 [宝日医生物技术 (北京) 有限公司] 处理总 RNA。用 Agilent 2100 Bioanalyzer 生物分析仪 (美国安捷伦公司) 进行样品 RNA 完整性测定, RNA 纯度和含量通过 NanoDrop ND-2000 紫外分光光度计 (美国赛默飞公司) 测定。选取 $A_{260}/A_{280} = 1.8\sim 2.2$ 、 $A_{260}/A_{230} \geq 2.0$ 、 $RIN \geq 8.0$ 、 $28S:18S \geq 1.0$, 且 RNA 含量 $> 1\text{ }\mu\text{g}$ 的样品进行后续分析。

使用 Illumina TruSeq RNA Sample Prep Kit (美国 Illumina 公司) 试剂盒制备 cDNA 文库, 由上海美吉生物医药科技有限公司完成转录组和表达谱测序, 测序平台为 Illumina NovaSeq 6000, 采用 $2\times 150\text{ bp}$ 双端测序。

白鼠尾草转录组拼接、注释与表达量计算 使用 fastp 软件及默认参数对测序所得的原始数据 (raw reads) 进行过滤和质控, 通过 Trinity 软件对去掉低质量数据后的待分析数据 (clean reads) 进行 *de novo* 拼接得到转录组数据。表达量数据以转录组数据为模板进行拼接。拼接得到的独立基因 (unigenes) 通过 BLASTX 算法分别与 NCBI 非冗余蛋白数据库 (non-redundant protein sequence database, Nr)、蛋白质数据库 (uni-

Prot)、京都基因与基因组数据库 (Kyoto encyclopedia of genes and genomes, KEGG) 和蛋白质直系同源数据库 (cluster of orthologous groups, COG) 进行比对, 得到白鼠尾草独立基因的蛋白功能注释结果, 并对独立基因进行基因本体数据库 (gene ontology, GO) 功能分类统计。通过 HMMER v3.1b2 进行独立基因蛋白结构域搜索。基因表达量通过 TopHat2 和 Cufflinks 软件计算, 用 TPM 值 (transcripts per million) 表示。

白鼠尾草 2-ODD 基因的寻找与确认 根据文献^[11]报道, 从拟南芥数据库 (TAIR, <https://www.arabidopsis.org/>) 中下载拟南芥的 130 个 2-ODD 基因作为参考序列, 通过 TBLASTN 算法, 在白鼠尾草转录组中进行同源基因比对 ($E\text{-value} = 1.0 \times 10^{-5}$)。比对得到的候选序列, 用 TBtools 软件^[12]预测开放阅读框 (open reading frame, ORF), 将 ORF 翻译成氨基酸序列后, 通过隐马尔可夫模型 (hidden Markov model, HMM) 搜索是否具有 2-ODD 基因的特征结构域 2OG-FeII_Oxy motif (PF03171)。筛除不具有 2OG-FeII_Oxy 结构域的片段序列后, 将余下的候选序列分别在 NCBI 数据库中进行 BLASTP 比对, 进一步删除非 2-ODD 基因序列或片段序列。

白鼠尾草 2-ODD 基因的生物信息学分析 利用各种生物信息学分析软件和在线网站对白鼠尾草 2-ODD 基因编码蛋白序列进行分析。白鼠尾草 2-ODD 基因的分类通过与已知的同属植物丹参序列进行同源比对获得^[11]; 白鼠尾草 2-ODD 基因编码蛋白的理化性质通过 ProtParam (<http://web.expasy.org/protparam/>) 网站进行预测; 信号肽预测在 SignalP-5.0 (<https://services.healthtech.dtu.dk/service.php?SignalP-5.0>) 网站上进行; 使用 TMHMM-2.0 (<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>) 网站进行跨膜结构预测; 亚细胞定位在 subCELLular LOcalization predictor (<http://cello.life.nctu.edu.tw>) 网站上完成; 通过 SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) 网站进行 2-ODD 蛋白的二级结构预测; 2-ODD 蛋白的高级结构同源建模在 Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>)^[13] 网站上进行。

白鼠尾草 2-ODD 基因家族系统进化分析 为了获得白鼠尾草 2-ODD 基因家族中各成员的系统进化关系, 结合 132 个丹参 2-ODD 基因, 利用 MEGA 11 内置的 ClustalW 软件对以上基因的蛋白序列进行多序列比对, 通过 MEGA 11 软件中的邻位连接法 (neighbor-joining, NJ) 构建系统进化树, 用 bootstrap 值检验进化树的拓扑结构, 设置参数为: 重复抽样次数设为 1 000, Gaps/Missing Data Treatment 选择 pairwise deletion。

白鼠尾草 2-ODD 基因的特异性表达 根据白鼠尾草表达谱数据, 找到每个 2-ODD 基因的表达量, 用 TBtools 软件^[12]绘制热图, 使用 0~1 (也称 min-max) 标准化方式, 使基因在不同样品中的表达量最高值为 1, 最低值为 0, 比较 2-ODD 基因在白鼠尾草根与叶中的表达差异。

结果

1 白鼠尾草转录组和表达谱的测序拼接结果

白鼠尾草经过 Illumina 测序平台的高通量测序和拼接后, 共获得 29 895 709 bp 碱基, 38 534 个独立基因。独立基因的长度范围居于 201~6 392 bp 之间, 平均长度为 775.83 bp, 中位长度为 539 bp, $N_{50} = 1 203$ bp, 独立基因的 GC 含量为 44%。

白鼠尾草根和叶两个部位 3 个生物学重复共 6 个样本表达谱的测序数据如表 1 所示, 每个样本测序得到的待分析数据 (去除原始测序数据中的接头污染数据、重复冗余数据和低质量数据后) 占原始数据的比例均在 98.85% 以上, 表明此次 RNA-seq 高通量测序获得的数据质量较优。

2 白鼠尾草转录组注释结果

利用 BLASTX 将拼接得到的 38 534 个独立基因分别与 Nr、UniProt、Pfam、KO (KEGG ontology)、COG 和 GO 数据库 ($E\text{-value} < 1.0 \times 10^{-5}$) 进行比对, 结果显示, 分别有 29 592、21 369、20 224、13 420、25 895 和 25 833 个独立基因比对上 Nr、UniProt、Pfam、KO、COG 和 GO 数据库, 共有 29 982 个独立基因获得功能注释, 占总独立基因的 77.81%。其中有 9 646 个独立基因在所有数据库中都有注释结果。

在 Nr 注释结果中, 与一串红 (*Salvia splendens*) 比

Table 1 RNA-seq information of *S. apiana*

Sample ID	Raw read/bp	Clean read/bp	Clean data rate/%	Aligned read/bp	Aligned rate/%
SAAPI_leaf1	46 755 056	46 218 706	98.85	37 886 090	81.97
SAAPI_leaf2	48 498 062	47 995 282	98.96	39 332 488	81.95
SAAPI_leaf3	47 191 744	46 695 850	98.95	38 295 088	82.01
SAAPI_root1	51 096 844	50 734 930	99.29	37 580 986	74.07
SAAPI_root2	50 934 458	50 484 354	99.12	37 257 044	73.80
SAAPI_root3	50 185 068	49 727 078	99.09	36 960 912	74.33

对上的基因最多,共有 21 934 条注释,占 Nr 注释结果的 74.12%;其次为芝麻 (*Sesamum indicum*),有 1 930 条注释,占注释结果的 6.52%。此注释结果提示,在已知基因组信息的物种中,白鼠尾草与同属植物一串红的序列相似性最大。

3 白鼠尾草转录组独立基因的 COG 注释及分类

白鼠尾草独立基因与 COG 数据库比对结果显示:共有 25 895 个独立基因被归纳到 23 个不同功能的分类当中。其中,有 13 922 个独立基因的功能未知,占总功能序列的 53.76%,在所有功能基因中占比最高;然后依次为蛋白质翻译后修饰-蛋白质周转-分子伴侣 (1 819 个独立基因, 7.02%)、转录 (1 700 个独立基因, 6.57%) 和信号转导机制 (1 276 个独立基因, 4.93%)。只有极少数序列归纳到细胞能动性 (11 个独立基因, 0.04%) 和核结构 (1 个独立基因) 中 (图 1)。

4 白鼠尾草转录组独立基因的 GO 注释及分类

分析白鼠尾草转录组独立基因的 GO 功能注释结果显示:共有 25 833 个独立基因 (总独立基因的 67.04%) 获得功能注释,分布在 52 个亚类中。其中有 15 433、18 936 和 21 129 个独立基因分别注释到生物过程、细胞组成和分子功能中 (图 2)。在生物过程的 23 个功能亚类中,细胞过程和代谢过程占比较多,分别为 40.44% 和 35.21%。在细胞组成的 14 个功能亚类中,以细胞组分占比最多,为 50.63%,其次为膜部件和细胞器,分别占 38.41% 和 29.18%。在分子功能所包括的 15 个亚类中,捆绑和催化活性占比最多,分别为

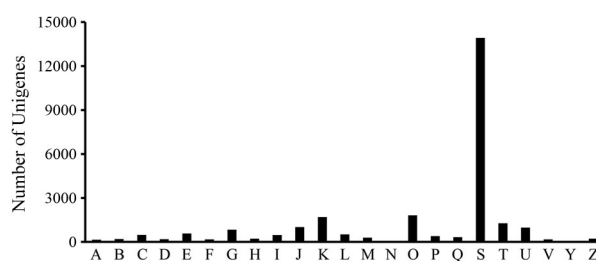


Figure 1 Cluster of orthologous groups (COG) classification of unigenes in *S. apiana*. A: RNA processing and modification; B: Chromatin structure and dynamics; C: Energy production and conversion; D: Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure and biogenesis; K: Transcription; L: Replication, recombination and repair; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Post-translational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms; Y: Nuclear structure; Z: Cytoskeleton

53.02% 和 50.06%。负荷衔接子活性占比最少,只注释上 3 个独立基因,占比仅为 0.01%。

5 白鼠尾草转录组 KEGG 代谢通路分析

白鼠尾草转录组测序结果通过比对 KEGG 数据库

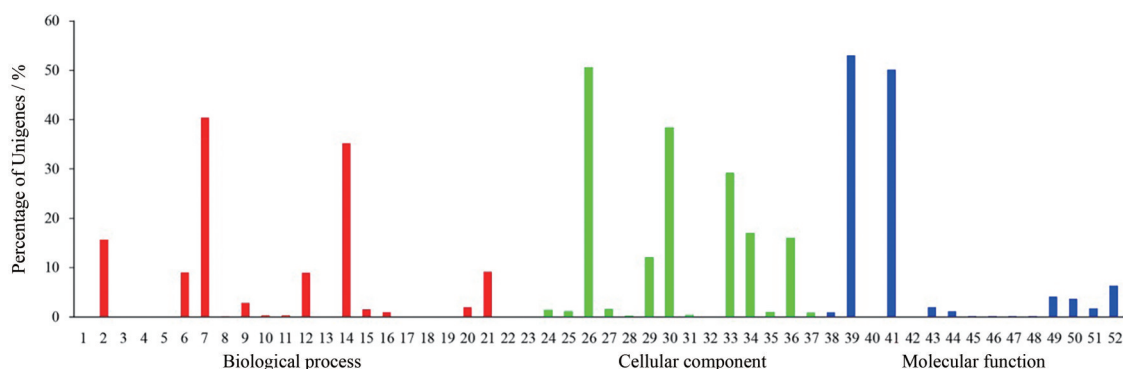


Figure 2 Gene ontology (GO) classification of unigenes in *S. apiana*. 1: Biological adhesion; 2: Biological regulation; 3: Carbon utilization; 4: Cell killing; 5: Cell population proliferation; 6: Cellular component organization or biogenesis; 7: Cellular process; 8: Detoxification; 9: Developmental process; 10: Growth; 11: Immune system process; 12: Localization; 13: Locomotion; 14: Metabolic process; 15: Multicellular organismal process; 16: Multi-organism process; 17: Nitrogen utilization; 18: Pigmentation; 19: Reproduction; 20: Reproductive process; 21: Response to stimulus; 22: Rhythmic process; 23: Signaling; 24: Cell; 25: Cell junction; 26: Cell part; 27: Extracellular region; 28: Extracellular region part; 29: Membrane; 30: Membrane part; 31: Membrane-enclosed lumen; 32: Nucleoid; 33: Organelle; 34: Organelle part; 35: Other organism part; 36: Protein-containing complex; 37: Supramolecular complex; 38: Antioxidant activity; 39: Binding; 40: Cargo adaptor activity; 41: Catalytic activity; 42: Molecular carrier activity; 43: Molecular function regulator; 44: Molecular transducer activity; 45: Nutrient reservoir activity; 46: Protein folding chaperone; 47: Protein tag; 48: Small molecule sensor activity; 49: Structural molecule activity; 50: Transcription regulator activity; 51: Translation regulator activity; 52: Transporter activity

显示: 共有 13 420 个独立基因获得了注释, 占总独立基因的 34.83%, 分布在 20 条代谢通路上 (图 3)。按照获得注释的独立基因数目进行排序, 翻译 (1 390 个独立基因, 10.36%) 通路所获注释最多, 然后依次是糖代谢 (953 个独立基因, 7.10%), 折叠、分类和降解 (804 个独立基因, 5.99%), 运输和分解代谢 (613 个独立基因, 4.57%), 以及信号传导 (592 个独立基因, 4.41%) 等通路。另外, 注释到萜类和聚酮类代谢以及其他次生代谢产物生物合成通路的独立基因各有 291 个, 分别占 KEGG 代谢通路注释结果的 2.17%。此次获得注释的代谢通路几乎代表了白鼠尾草发育的所有代谢过程, 为进一步开展白鼠尾草功能基因组学的研究提供了丰富的数据支撑。

6 白鼠尾草 2-ODD 基因的确证及序列组成分析

通过拟南芥 2-ODD 基因与白鼠尾草转录组的同源比对, 以及 2-ODD 基因特征结构域 (PF03171) 寻找, 共找到 39 条具有完整序列结构的白鼠尾草 2-ODD 基因 (*Sa2-ODD*), 它们与同属植物丹参已知 2-ODD 基因 (*Sm2-ODD*) 的对应及分组关系, 如表 2 所示。

对这 39 条完整的白鼠尾草 2-ODD 基因的编码蛋白进行序列组成成分和理化性质分析 (表 3), 发现它们编码氨基酸的数量从 228aa (*Sa2-ODD24*) 到 377aa (*Sa2-ODD21*) 不等, 分子质量约为 36.00 kDa, 最大为 42.87 kDa (*Sa2-ODD17*), 最小为 24.94 kDa (*Sa2-ODD24*)。这些 2-ODD 基因编码蛋白的理论等电点除 *Sa2-ODD32* ($pI = 7.05$)、*Sa2-ODD5* ($pI = 7.25$)、*Sa2-ODD3* ($pI = 8.19$) 和 *Sa2-ODD2* ($pI = 8.19$) 外, 均小于 7, 提示 *Sa2-ODD* 多为酸性蛋白。每个蛋白的酸碱氨基酸比例与理论等电

点对应, 除 *Sa2-ODD32*、*Sa2-ODD5*、*Sa2-ODD3* 和 *Sa2-ODD2* 之外, 其余蛋白所含碱性氨基酸比例均低于酸性氨基酸比例, 进一步提示大多数 *Sa2-ODD* 为酸性蛋白。对 39 条 *Sa2-ODD* 蛋白稳定系数的预测结果显示, 不稳定系数大于 40 的有 25 条, 不稳定系数小于 40 的为 14 条, 提示这些 *Sa2-ODD* 中, 不稳定蛋白居多。这些 *Sa2-ODD* 除 *Sa2-ODD24* 的总平均亲水性大于 0, 为 0.011, 其余总平均亲水性均为负值, 提示 *Sa2-ODD* 多为亲水性蛋白。

对这 39 条 *Sa2-ODD* 基因进行氨基酸序列相似性比较, 发现氨基酸序列间的两两 (pairwise) 相似度从 16.52% (*Sa2-ODD36* 与 *Sa2-ODD37*) 到 96.86% (*Sa2-ODD6* 与 *Sa2-ODD16*) 不等, 提示序列间变异较大, 与文献^[14,15]报道一致。

7 白鼠尾草 2-ODD 基因编码蛋白的信号肽、亚细胞定位和跨膜结构域预测

信号肽预测显示, 除 *Sa2-ODD1* 和 *Sa2-ODD4* 具有信号肽外, 其余 *Sa2-ODD* 都不具有信号肽。

亚细胞定位预测显示, *Sa2-ODD1* 和 *Sa2-ODD4* 都定位在细胞外, 与信号肽预测结果一致; *Sa2-ODD7*、*Sa2-ODD9*、*Sa2-ODD11*、*Sa2-ODD12*、*Sa2-ODD13*、*Sa2-ODD16*、*Sa2-ODD17*、*Sa2-ODD18*、*Sa2-ODD19*、*Sa2-ODD20*、*Sa2-ODD21*、*Sa2-ODD22*、*Sa2-ODD23*、*Sa2-ODD24*、*Sa2-ODD26*、*Sa2-ODD27*、*Sa2-ODD28*、*Sa2-ODD29*、*Sa2-ODD31*、*Sa2-ODD32*、*Sa2-ODD34*、*Sa2-ODD35*、*Sa2-ODD36* 和 *Sa2-ODD38* 都定位在细胞质内; *Sa2-ODD30* 和 *Sa2-ODD33* 定位在细胞核内; *Sa2-ODD8*、*Sa2-ODD10* 和 *Sa2-ODD39* 定位在细胞质和

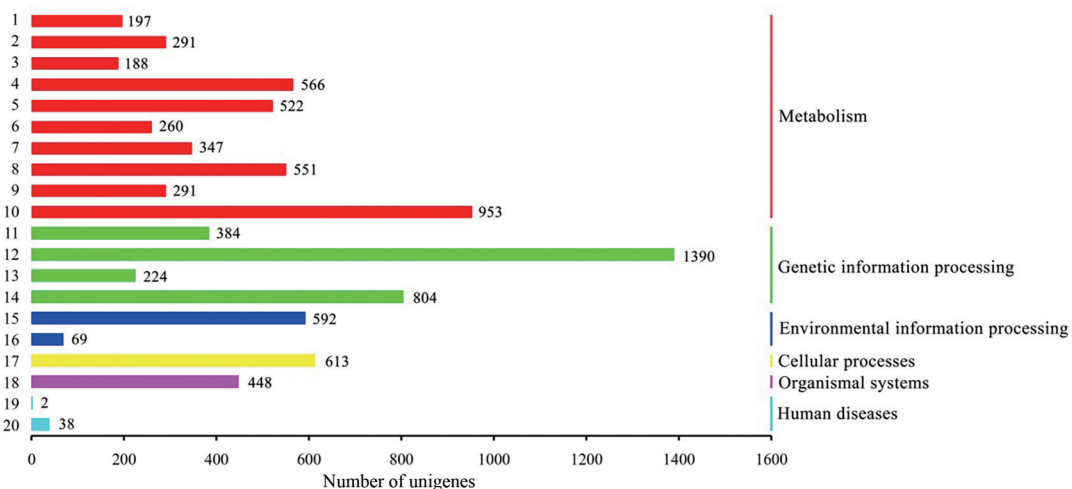


Figure 3 KEGG classification of unigenes in *S. apiana*. 1: Nucleotide metabolism; 2: Metabolism of terpenoids and polyketides; 3: Glycan biosynthesis and metabolism; 4: Lipid metabolism; 5: Energy metabolism; 6: Metabolism of other amino acids; 7: Metabolism of cofactors and vitamins; 8: Amino acid metabolism; 9: Biosynthesis of other secondary metabolites; 10: Carbohydrate metabolism; 11: Transcription; 12: Translation; 13: Replication and repair; 14: Folding, sorting and degradation; 15: Signal transduction; 16: Membrane transport; 17: Transport and catabolism; 18: Environmental adaptation; 19: Drug resistance: antimicrobial; 20: Endocrine and metabolic disease

Table 2 The identification of homologous 2-oxoglutarate/Fe(II)-dependent dioxygenases (2-*ODD*) genes between *S. apiana* and *S. miltiorrhiza*

Designate name	Transcript ID	<i>Sm2-ODD</i> ID	Identity/%	E-value	Bit score	Class	Clade
<i>Sa2-ODD1</i>	TRINITY_DN16972_c0_g1	SMil_00004726	91.089	0	583	DOXB	P4H
<i>Sa2-ODD2</i>	TRINITY_DN18814_c0_g1	SMil_00018996	95.848	0	506	DOXB	P4H
<i>Sa2-ODD3</i>	TRINITY_DN3023_c0_g1	SMil_00013255	86.293	0	519	DOXB	P4H
<i>Sa2-ODD4</i>	TRINITY_DN3151_c0_g1	SMil_00001764	94.539	0	551	DOXB	P4H
<i>Sa2-ODD5</i>	TRINITY_DN6900_c0_g1	SMil_00001015	95.862	0	550	DOXB	P4H
<i>Sa2-ODD6</i>	TRINITY_DN9762_c0_g1	SMil_00006860	94.521	0	526	DOXB	P4H
<i>Sa2-ODD7</i>	TRINITY_DN1005_c0_g1	SMil_00011064	94.286	0	593	DOXC	ACO
<i>Sa2-ODD8</i>	TRINITY_DN10817_c0_g1	SMil_00006040	90.698	0	651	DOXC	ACO
<i>Sa2-ODD9</i>	TRINITY_DN11493_c0_g1	SMil_00022077	83.271	6.06E-149	412	DOXC	ACO
<i>Sa2-ODD10</i>	TRINITY_DN12053_c0_g1	SMil_00019259	79.048	0	528	DOXC	AOP
<i>Sa2-ODD11</i>	TRINITY_DN17768_c0_g1	SMil_00000574	33.333	7.34E-07	42	DOXC	AOP
<i>Sa2-ODD12</i>	TRINITY_DN19158_c0_g1	SMil_00000574	87.047	1.58E-130	363	DOXC	AOP
<i>Sa2-ODD13</i>	TRINITY_DN3380_c0_g1	SMil_00013636	91.290	0	558	DOXC	AOP
<i>Sa2-ODD14</i>	TRINITY_DN3583_c0_g1	SMil_00016764	49.138	9.13E-83	245	DOXC	AOP
<i>Sa2-ODD15</i>	TRINITY_DN5191_c0_g1	SMil_00016763	84.076	0	553	DOXC	AOP
<i>Sa2-ODD16</i>	TRINITY_DN6073_c0_g1	SMil_00016764	87.342	0	572	DOXC	AOP
<i>Sa2-ODD17</i>	TRINITY_DN10852_c0_g1	SMil_00013081	22.179	1.31E-12	60.5	DOXC	CODM/NCS
<i>Sa2-ODD18</i>	TRINITY_DN2211_c0_g2	SMil_00024282	86.813	2.08E-180	493	DOXC	CODM/NCS
<i>Sa2-ODD19</i>	TRINITY_DN22361_c0_g1	SMil_00012259	71.273	1.54E-151	422	DOXC	CODM/NCS
<i>Sa2-ODD20</i>	TRINITY_DN2908_c0_g1	SMil_00016116	56.475	1.01E-111	321	DOXC	D4H/GSLOH/BX6
<i>Sa2-ODD21</i>	TRINITY_DN30979_c0_g1	SMil_00018669	86.096	0	637	DOXC	D4H/GSLOH/BX6
<i>Sa2-ODD22</i>	TRINITY_DN4864_c0_g1	SMil_00016116	81.167	0	620	DOXC	D4H/GSLOH/BX6
<i>Sa2-ODD23</i>	TRINITY_DN9009_c0_g1	SMil_00012033	90.437	0	686	DOXC	D4H/GSLOH/BX6
<i>Sa2-ODD24</i>	TRINITY_DN1201_c0_g1	SMil_00023724	88.073	2.06E-142	393	DOXC	DAO
<i>Sa2-ODD25</i>	TRINITY_DN23679_c0_g1	SMil_00023724	54.110	2.66E-115	328	DOXC	DAO
<i>Sa2-ODD26</i>	TRINITY_DN1185_c0_g1	SMil_00013745	92.717	0	667	DOXC	F3H
<i>Sa2-ODD27</i>	TRINITY_DN1324_c0_g1	SMil_00022938	72.590	8.89E-179	491	DOXC	F3H
<i>Sa2-ODD28</i>	TRINITY_DN4831_c0_g2	SMil_00028133	87.330	6.70E-146	402	DOXC	F3H
<i>Sa2-ODD29</i>	TRINITY_DN8474_c0_g2	SMil_00005410	81.873	0	557	DOXC	F3H
<i>Sa2-ODD30</i>	TRINITY_DN4585_c0_g1	SMil_00008090	86.737	0	672	DOXC	GA20ox
<i>Sa2-ODD31</i>	TRINITY_DN13439_c0_g1	SMil_00004141	81.040	4.08E-176	484	DOXC	GA2ox
<i>Sa2-ODD32</i>	TRINITY_DN8276_c0_g1	SMil_00017774	91.212	0	617	DOXC	GA2ox
<i>Sa2-ODD33</i>	TRINITY_DN15665_c0_g1	SMil_00027988	78.402	0	570	DOXC	H6H
<i>Sa2-ODD34</i>	TRINITY_DN17826_c0_g1	SMil_00013059	72.271	0	530	DOXC	S3H
<i>Sa2-ODD35</i>	TRINITY_DN16851_c0_g1	SMil_00020342	91.117	0	638	DOXC	Unknown
<i>Sa2-ODD36</i>	TRINITY_DN4633_c0_g1	SMil_00009012	89.773	0	676	DOXC	Unknown
<i>Sa2-ODD37</i>	TRINITY_DN4633_c0_g2	SMil_00009013	82.173	0	587	DOXC	Unknown
<i>Sa2-ODD38</i>	TRINITY_DN5898_c0_g1	SMil_00018124	87.805	0	557	DOXC	Unknown
<i>Sa2-ODD39</i>	TRINITY_DN7999_c0_g1	SMil_00021185	87.069	0	619	DOXC	Unknown

细胞核内; Sa2-ODD14和Sa2-ODD15定位在细胞质和细胞质膜上; Sa2-ODD6定位在细胞质和细胞质膜上,还穿过细胞膜定位在细胞外; Sa2-ODD2定位在线粒体上; Sa2-ODD3定位在细胞质膜和细胞外; Sa2-ODD5定位在细胞外和线粒体上; Sa2-ODD25定位在细胞质和叶绿体上; Sa2-ODD37定位在细胞外和细胞质中。

跨膜结构域预测显示, Sa2-ODD2、Sa2-ODD3、Sa2-ODD5、Sa2-ODD6、Sa2-ODD14和Sa2-ODD17具有跨膜结构域。

8 白鼠尾草 2-*ODD* 基因编码蛋白的二级和三级结构预测

利用在线分析软件 SOPMA 对白鼠尾草 2-*ODD* 基因的蛋白序列进行二级结构预测 (表 4), 结果显示白鼠尾草 2-*ODD* 基因的蛋白序列由 4 个结构组件组成, 以

无规则卷曲为主导, α -螺旋次之, 延伸链居第三, β -折叠所占比例最少。其中 Sa2-ODD7、Sa2-ODD18、Sa2-ODD19、Sa2-ODD20 和 Sa2-ODD21 中, α -螺旋所占比例最多; Sa2-ODD6 中, 延伸链所占比例大于 α -螺旋。

白鼠尾草 2-*ODD* 基因编码蛋白的三级结构预测采用 Phyre2 线串法的从头计算模式完成, 所有三级结构都在 > 90% 的置信度下建模, 结果如图 4 所示。2-*ODD* 基因编码蛋白的功能域中包含 1 个扭曲的双股 β 螺旋核心折叠结构域, 包括 8 条 β -折叠股, 每 2 条连续的 β -折叠股之间以反向平行方式排列, 最终 4 对 β -折叠股形成 1 个扁桶状的 DSBH 结构^[8]。

9 白鼠尾草 2-*ODD* 基因的系统进化分析

为了研究白鼠尾草 2-*ODD* 基因家族的系统进化关系, 使用 MEGA 11 软件的 Neighbor-joining 算法, 用

Table 3 Composition and physicochemical character analysis of Sa2-ODDs

ID	Length /aa	Molecular weight/Da	Theoretical pI	Negatively charged residue/aa	Positively charged residue/aa	Formula	Instability index	Aliphatic index	Grand average of hydropathicity
Sa2-ODD1	302	33 464.88	5.89	43	36	C ₁₄₈₂ H ₂₂₈₅ N ₄₀₁ O ₄₅₀ S ₁₇	36.67	69.80	-0.367
Sa2-ODD2	289	32 322.24	9.16	33	40	C ₁₄₄₃ H ₂₂₉₅ N ₃₉₉ O ₄₁₉ S ₁₂	45.87	86.99	-0.323
Sa2-ODD3	289	32 272.65	8.19	33	35	C ₁₄₄₄ H ₂₂₄₀ N ₃₉₄ O ₄₂₇ S ₁₀	40.45	77.58	-0.312
Sa2-ODD4	292	32 521.71	6.04	41	38	C ₁₄₃₈ H ₂₂₃₁ N ₃₈₉ O ₄₄₄ S ₁₄	45.52	70.48	-0.473
Sa2-ODD5	289	32 575.86	7.25	35	35	C ₁₄₅₆ H ₂₂₅₇ N ₄₀₃ O ₄₃₂ S ₈	43.50	80.97	-0.424
Sa2-ODD6	293	32 824.28	5.92	36	33	C ₁₄₇₃ H ₂₂₈₁ N ₃₉₃ O ₄₃₈ S ₁₀	37.82	76.25	-0.398
Sa2-ODD7	314	35 624.86	5.06	50	37	C ₁₆₀₇ H ₂₄₉₆ N ₄₁₀ O ₄₇₄ S ₁₅	34.40	84.81	-0.291
Sa2-ODD8	343	38 476.58	6.07	41	36	C ₁₇₁₅ H ₂₆₅₉ N ₄₆₇ O ₅₁₅ S ₁₃	44.49	80.35	-0.429
Sa2-ODD9	267	30 277.43	6.37	34	33	C ₁₃₅₄ H ₂₁₀₅ N ₃₆₇ O ₄₀₂ S ₁₀	41.36	79.59	-0.456
Sa2-ODD10	312	35 418.39	5.91	40	32	C ₁₆₀₆ H ₂₄₈₂ N ₄₁₂ O ₄₇₃ S ₉	42.20	83.97	-0.402
Sa2-ODD11	362	40 338.28	6.47	43	41	C ₁₇₉₁ H ₂₈₁₂ N ₄₉₄ O ₅₂₅ S ₂₁	55.41	77.82	-0.331
Sa2-ODD12	318	35 878.55	5.18	39	25	C ₁₆₁₆ H ₂₄₄₂ N ₄₁₄ O ₄₈₃ S ₁₅	37.37	77.23	-0.227
Sa2-ODD13	311	35 470.34	5.91	45	35	C ₁₆₀₅ H ₂₄₅₈ N ₄₁₆ O ₄₇₁ S ₁₁	37.75	81.16	-0.412
Sa2-ODD14	290	32 938.85	5.64	33	26	C ₁₄₉₆ H ₂₂₇₉ N ₃₆₉ O ₄₃₃ S ₁₈	35.05	82.59	-0.050
Sa2-ODD15	313	35 137.79	5.09	40	26	C ₁₅₇₇ H ₂₄₁₈ N ₄₁₀ O ₄₇₅ S ₁₃	41.78	82.52	-0.191
Sa2-ODD16	315	35 559.41	5.78	39	31	C ₁₅₉₉ H ₂₄₄₅ N ₄₁₇ O ₄₇₃ S ₁₅	43.35	75.21	-0.302
Sa2-ODD17	376	42 871.21	6.43	47	45	C ₁₉₃₅ H ₃₀₀₂ N ₅₁₆ O ₅₅₄ S ₁₆	49.08	81.38	-0.334
Sa2-ODD18	278	31 511.15	5.27	40	33	C ₁₄₀₉ H ₂₂₂₁ N ₃₇₁ O ₄₂₁ S ₁₃	30.95	88.38	-0.364
Sa2-ODD19	282	31 504.65	5.28	37	29	C ₁₄₃₀ H ₂₂₆₈ N ₃₆₄ O ₄₁₀ S ₁₂	48.17	101.03	-0.009
Sa2-ODD20	293	32 774.56	5.32	40	31	C ₁₄₆₂ H ₂₂₉₅ N ₃₉₉ O ₄₂₉ S ₁₄	42.40	90.78	-0.194
Sa2-ODD21	377	42 030.58	4.88	58	39	C ₁₈₄₉ H ₂₉₄₅ N ₅₁₁ O ₅₈₀ S ₁₃	33.76	94.38	-0.279
Sa2-ODD22	362	40 831.32	5.09	51	37	C ₁₈₂₅ H ₂₈₄₇ N ₄₉₃ O ₅₄₉ S ₁₁	41.03	90.99	-0.246
Sa2-ODD23	365	41 216.31	5.55	50	40	C ₁₈₃₆ H ₂₈₇₉ N ₄₉₅ O ₅₄₁ S ₂₁	51.46	83.86	-0.284
Sa2-ODD24	228	24 942.55	4.62	28	18	C ₁₁₁₅ H ₁₇₄₂ N ₂₈₈ O ₃₃₆ S ₁₂	42.22	92.32	-0.011
Sa2-ODD25	302	33 354.98	5.41	37	28	C ₁₄₈₃ H ₂₃₁₅ N ₄₀₅ O ₄₄₃ S ₁₄	41.23	86.56	-0.112
Sa2-ODD26	349	40 006.28	5.40	55	42	C ₁₈₁₁ H ₂₇₇₅ N ₄₇₃ O ₅₃₆ S ₈	37.16	79.60	-0.529
Sa2-ODD27	295	33 722.49	5.28	40	32	C ₁₅₂₂ H ₂₃₄₀ N ₃₉₆ O ₄₄₅ S ₁₃	42.15	84.27	-0.324
Sa2-ODD28	310	34 936.57	5.34	43	32	C ₁₅₈₄ H ₂₄₁₉ N ₄₁₇ O ₄₆₃ S ₇	44.03	79.64	-0.396
Sa2-ODD29	332	37 073.20	5.26	47	37	C ₁₆₅₂ H ₂₅₆₅ N ₄₄₇ O ₄₉₂ S ₁₆	31.72	82.65	-0.308
Sa2-ODD30	374	42 197.10	6.75	39	37	C ₁₈₉₅ H ₂₉₀₃ N ₅₁₅ O ₅₄₅ S ₁₈	38.67	75.67	-0.372
Sa2-ODD31	316	35 724.25	5.38	44	34	C ₁₅₈₇ H ₂₄₄₀ N ₄₂₈ O ₄₈₅ S ₁₄	49.87	74.34	-0.341
Sa2-ODD32	329	36 561.87	7.05	39	39	C ₁₆₄₀ H ₂₅₆₈ N ₄₄₀ O ₄₈₁ S ₁₃	33.92	85.35	-0.233
Sa2-ODD33	337	38 324.80	6.17	33	26	C ₁₇₁₄ H ₂₆₄₃ N ₄₆₅ O ₄₉₇ S ₁₉	55.17	80.14	-0.343
Sa2-ODD34	336	37 950.17	5.76	44	37	C ₁₆₉₉ H ₂₆₃₃ N ₄₅₃ O ₅₀₆ S ₁₄	42.16	82.56	-0.385
Sa2-ODD35	348	38 962.94	5.31	45	31	C ₁₇₃₃ H ₂₆₈₉ N ₄₇₅ O ₅₂₇ S ₁₁	47.67	86.09	-0.305
Sa2-ODD36	360	40 427.78	5.50	45	34	C ₁₈₀₉ H ₂₇₇₈ N ₄₈₄ O ₅₄₀ S ₁₅	35.29	78.33	-0.280
Sa2-ODD37	353	39 373.90	5.95	40	33	C ₁₇₇₂ H ₂₇₅₁ N ₄₇₁ O ₅₂₁ S ₁₂	35.84	89.24	-0.148
Sa2-ODD38	344	37 940.01	5.20	45	32	C ₁₆₈₀ H ₂₆₃₁ N ₄₆₃ O ₅₁₁ S ₁₄	44.73	84.48	-0.285
Sa2-ODD39	349	38 728.01	5.81	39	30	C ₁₇₂₁ H ₂₆₉₂ N ₄₇₄ O ₅₁₆ S ₁₄	54.80	89.74	-0.194

氨基酸序列构建 39 个白鼠尾草 2-ODD 基因与 132 个丹参 2-ODD 基因的系统进化树。根据丹参 2-ODD 基因的分组规则^[11], 171 个 2-ODD 基因被分成了 3 大类 15 组 (图 5), 其中红色区域为 DOXA 亚家族 (XV), 成员最少; 黄色区域为 DOXB 亚家族 (XIV), 成员居中; 蓝色区域为 DOXC 亚家族 (I-XIII), 绝大多数 2-ODD 基因都属于 DOXC 类。DOXB 类的成员, 都属于 P4H (prolyl 4-hydroxylase, 脯氨酸 4-羟化酶, XIV)。DOXC 类分成了 13 个小组, 分别是功能未知的 I 组, S3H (salicylic acid 3-hydroxylase, 水杨酸 3-羟化酶, II)、H6H (hyoscyamine 6 β -hydroxylase, 莨菪碱 6 β -羟化酶, III)、ACO (1-aminocyclopropane carboxylic acid oxidase, 1-氨基环丙烷羧酸氧化酶, IV)、FLS/ANS (flavonol synthase/

anthocyanin synthase, 黄酮醇合成酶/花青素合成酶, V)、CODM/NCS (codeine O-demethylase/norcochlorine synthase, 可待因-O-脱甲基酶/去甲乌药碱合成酶, VI)、D4H/GSLOH/BX6 (desacetoxyvindoline-4-hydroxylase/GSLOH/BENZOXAZINLESS6, VII)、GA2ox (gibberellin 2-oxidase, 赤霉素 2-氧化酶, VIII)、GA3ox (gibberellin 3 β -oxidase, 赤霉素 3 β -氧化酶, IX)、GA20ox (gibberellin 20-oxidase, 赤霉素 20-氧化酶, X)、DAO (dioxygenase for auxin oxidation, XI)、AOP (glucosinolate biosynthesis, 葡萄糖异硫氰酸合成酶, XII) 和 F3H (flavanone 3-hydroxylase, 黄烷酮 3-羟化酶, XIII)。根据 DOXC 亚家族的成员组成, 可以推测该亚家族的成员可能参与了植物黄酮、生物碱和生长素类物质的代谢

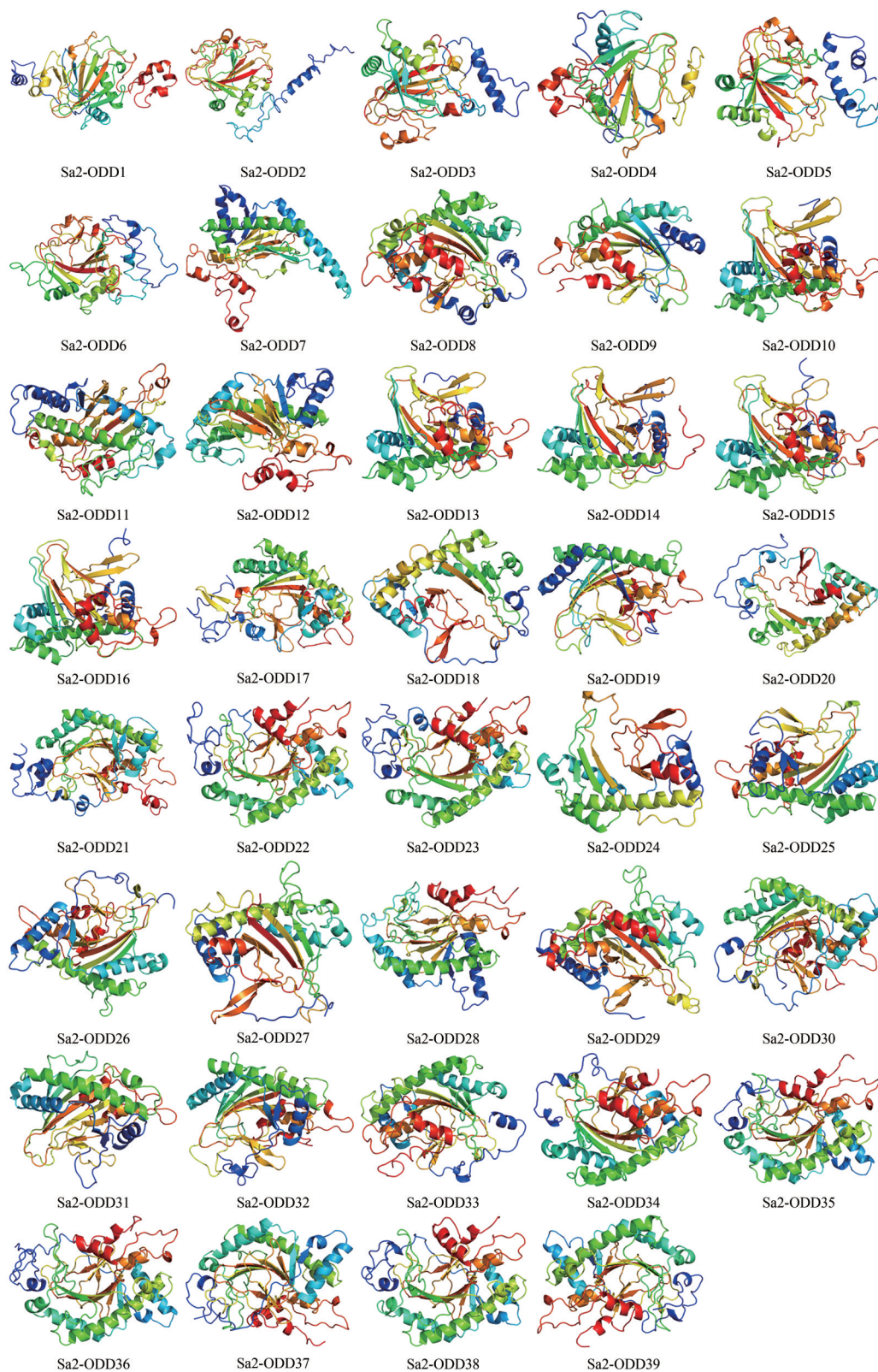


Figure 4 Tertiary structure of Sa2-ODDs

过程。

10 白鼠尾草 2-ODD 基因的表达分析

为了研究白鼠尾草 2-ODD 家族基因在地上和地下部位的表达模式, 利用高通量测序技术, 对白鼠尾草

的根和叶进行了表达谱测序。基于表达谱数据, 对白鼠尾草 2-ODD 基因的表达模式进行分析 (图 6), 发现绝大多数 2-ODD 基因在根中的表达量明显高于叶中的表达量; *Sa2-ODD7*、*Sa2-ODD8*、*Sa2-ODD9*、*Sa2-*

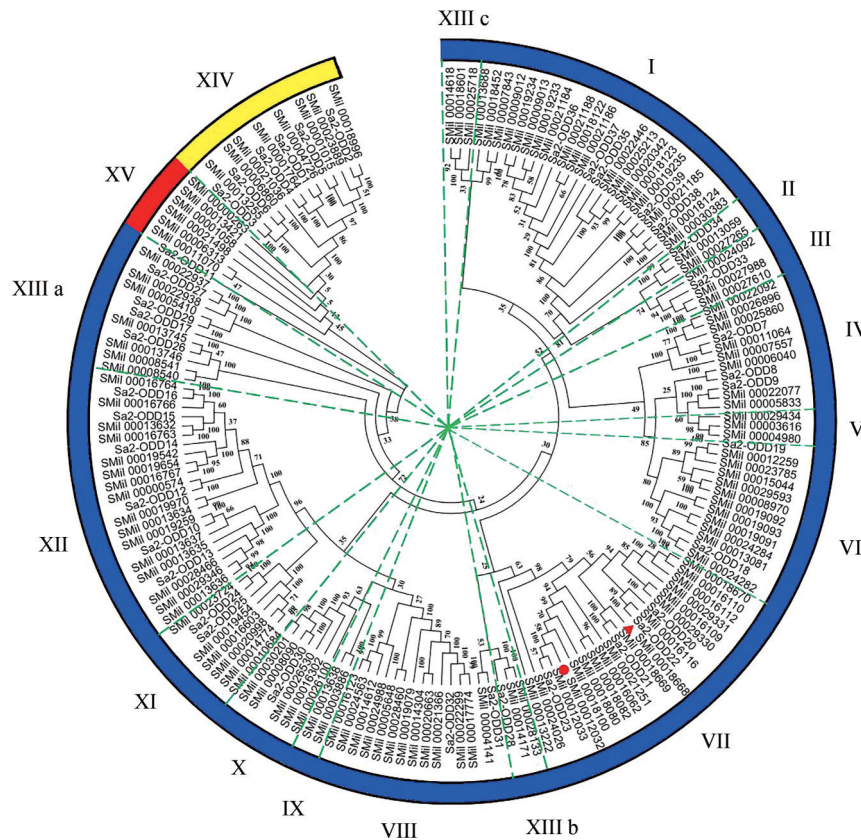


Figure 5 Phylogenetic tree (NJ) of Sa2-ODDs and Sm2-ODDs. A total of 171 2-ODD proteins from *S. apiana* and *S. multiorrhiza* are clustered into 3 classes. Class DOXA (XV): The red region; Class DOXB (XIV): The yellow region; and Class DOXC (I-XIII): The blue region. Bootstrap values are presented for all branches. The red dot indicates Sm2OGD5 (SMil_00012032) mentioned in Xu's paper^[11] and the red triangle indicates Sm2-ODD14 (SMil_00018668) mentioned in Song's paper^[6]

ODD22、Sa2-ODD29、Sa2-ODD32 和 Sa2-ODD38 基因在叶中表达量较高; Sa2-ODD11、Sa2-ODD23、Sa2-ODD27 和 Sa2-ODD34 基因在根和叶中的表达量相当, 相较于其他基因, 这 4 个基因的表达量相对偏低。以上基因表达特征结果显示, Sa2-ODD 基因家族中的成员可能参与了白鼠尾草不同代谢产物的调控, 并且其功能存在分化。

讨论

2-ODD 是一个庞大的基因家族, 拟南芥 (*Arabidopsis thaliana*) 和丹参基因组中都包含超过 130 条 2-ODD 基因^[11]。该家族成员在翻译后修饰、去甲基化、植物激素合成和特殊次生代谢产物生物合成过程中发挥着重要作用^[8]。本研究从白鼠尾草转录组数据库中筛选获得 39 条具有完整结构的 2-ODD 基因家族成员, 其中 6 条属于 DOXB 亚家族, 提示它们可能与白鼠尾草响应信号肽类激素以及细胞壁形成有关; 33 条属于 DOXC 亚家族, 提示白鼠尾草中的大部分 2-ODD 基因参与了该植物的次生代谢过程, 且可能参与了黄酮、生

物碱和生长素类物质的代谢过程; 暂未发现 DOXA 亚家族成员, 这可能与转录组基因信息不如基因组信息全面有关。GO 注释结果显示, 6 条 DOXB 亚家族成员中的 3 条 (Sa2-ODD1、Sa2-ODD2 和 Sa2-ODD5) 与细胞组分相关; 33 条 DOXC 亚家族成员中, 有 13 条与分子功能相关, 4 条与生物过程相关。这些结果进一步加强了白鼠尾草 2-ODD 基因功能的推测。

基于生物信息学方法, 对白鼠尾草中鉴定得到的 2-ODD 基因家族成员进行理化特性和序列特征分析。理化特性分析结果显示, 该家族基因编码氨基酸数量平均为 320 个氨基酸, 多为酸性和亲水性蛋白, 并且大多数基因所编码的蛋白定位在细胞质中, 这与大多数文献报道结果一致^[16,17]。序列相似性比较显示, 白鼠尾草 2-ODD 基因家族成员的序列保守性为 16.52%~96.86%, 序列间差异较大, 这也与 2-ODD 基因文献^[14,15]报道一致。以上结果表明, 不同物种中的 2-ODD 基因在理化特性上具有较高的相似性, 但在基因序列上存在较大差异。

系统进化分析结果显示, F3H 相关的基因被分成

Table 4 The proportion of main components in the secondary structure of Sa2-ODDs

ID	Random coil/%	Alpha helix/%	Extended strand/%	Beta turn/%
Sa2-ODD1	54.64	21.52	18.87	4.97
Sa2-ODD2	49.13	22.84	21.80	6.23
Sa2-ODD3	51.21	22.15	20.42	6.23
Sa2-ODD4	52.40	23.97	18.49	5.14
Sa2-ODD5	50.17	24.22	19.03	6.57
Sa2-ODD6	49.15	21.84	22.53	6.48
Sa2-ODD7	31.53	43.95	17.83	6.69
Sa2-ODD8	41.40	34.69	18.37	5.54
Sa2-ODD9	40.45	31.46	20.60	7.49
Sa2-ODD10	42.31	37.18	15.71	4.81
Sa2-ODD11	42.27	34.81	14.92	8.01
Sa2-ODD12	39.94	36.16	17.92	5.97
Sa2-ODD13	40.51	37.62	17.04	4.82
Sa2-ODD14	40.00	37.24	16.90	5.86
Sa2-ODD15	41.21	34.82	17.57	6.39
Sa2-ODD16	43.17	35.56	15.87	5.40
Sa2-ODD17	42.02	38.56	15.43	3.99
Sa2-ODD18	36.69	38.85	19.42	5.04
Sa2-ODD19	35.46	40.43	17.73	6.38
Sa2-ODD20	38.57	38.91	18.09	4.44
Sa2-ODD21	35.28	41.11	17.77	5.84
Sa2-ODD22	38.12	37.85	17.40	6.63
Sa2-ODD23	38.90	38.08	17.81	5.21
Sa2-ODD24	43.42	35.53	15.79	5.26
Sa2-ODD25	42.05	35.43	16.89	5.63
Sa2-ODD26	40.97	36.10	18.34	4.58
Sa2-ODD27	41.69	34.92	18.31	5.08
Sa2-ODD28	42.90	30.97	19.35	6.77
Sa2-ODD29	45.48	35.54	15.96	3.01
Sa2-ODD30	41.71	34.22	17.65	6.42
Sa2-ODD31	41.14	32.28	20.25	6.33
Sa2-ODD32	42.86	34.35	17.02	5.78
Sa2-ODD33	41.25	35.91	17.51	5.34
Sa2-ODD34	40.48	35.71	17.56	6.25
Sa2-ODD35	40.52	37.36	15.23	6.90
Sa2-ODD36	41.11	35.56	18.06	5.28
Sa2-ODD37	37.96	36.54	18.98	6.52
Sa2-ODD38	43.02	34.59	17.44	4.94
Sa2-ODD39	40.11	36.39	17.48	6.02

了3组, 分别在进化树的不同位置, 这与文献^[11]报道中的聚类结果有所差异, 这可能与序列相似度以及参与进化树构建的基因数目不同有关。

通过表达谱分析发现, Sa2-ODD基因在根和叶中均有表达, 大多数Sa2-ODD基因(28个)在根中表达量较高, 这与白鼠尾草根中丹参酮类物质积累较多的分布模式一致^[5], 与丹参中丹参酮类物质的积累方式也相符, 进一步提示Sa2-ODD基因可能参与丹参酮类物质的合成。刘文超等^[14]曾克隆得到丹参中的Sm2-ODD1基因, 该基因在丹参各个器官中都有表达, 但表达水平具有组织特异性, 根中表达量最高, 叶中表达量最低。Xu等^[11]发现了丹参中与丹参酮IIA合成相关的一条

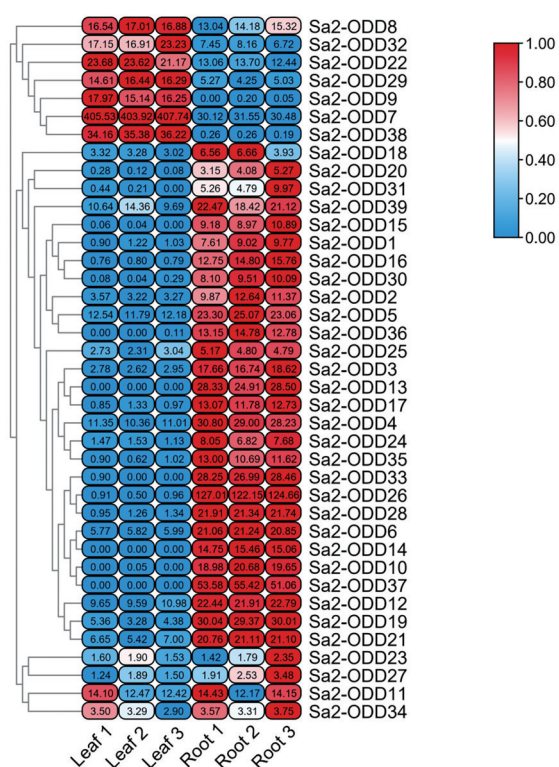


Figure 6 The expression of Sa2-ODDs in leaves and roots [the data in the rectangular box is the transcripts per million (TPM) value of each gene]

Sm2-ODD基因(SMil_00012032, 图5红色圆形标记), 通过进化树比较, 发现白鼠尾草中与该基因最接近的是Sa2-ODD23基因, 但是Sa2-ODD23基因在白鼠尾草根和叶中的表达量均较低, 且根和叶中的表达量没有差异(图6)。Song等^[6]通过进一步研究丹参中的2-ODD基因, 没有找到与SMil_00012032相近的基因, 研究者发现了丹参中与丹参酮IIA合成直接相关的Sm2-ODD14基因(SMil_00018668, 图5红色三角形标记), 该基因具有催化呋喃环芳构化的作用, 能将隐丹参酮直接转变为丹参酮IIA, 且该基因在丹参根中的表达量明显高于叶中的表达量, 与丹参酮IIA的分布规律一致。通过进化树分析, 发现白鼠尾草中与该基因最接近的是Sa2-ODD21基因, Sa2-ODD21基因的表达模式与Sm2-ODD14一致, 均是根中的表达量明显高于叶中的表达量(图6), 而且白鼠尾草中丹参酮IIA的积累也是地下部分明显多于地上部分^[5]。由此推测, Sa2-ODD21基因可能与白鼠尾草中丹参酮IIA的生物合成有关, 值得进一步研究关注。综上所述, 本研究可为深入挖掘白鼠尾草2-ODD基因的功能奠定基础。

作者贡献: 秦晓静负责数据分析和文章撰写; 胡佳栋负责样品采集和转录组测序; 房振西参与数据分析; 陈万生负责实验统筹和论文审阅; 李卿负责实验设计、指导和论文

修改。

利益冲突: 所有作者均声明不存在利益冲突。

References

- [1] Krol A, Kokotkiewicz A, Luczkiewicz M. White sage (*Salvia apiana*) -a ritual and medicinal plant of the chaparral: plant characteristics in comparison with other *Salvia* species [J]. *Planta Med*, 2022, 88: 604-627.
- [2] Service USDoANRC [OL]. <http://plant-materials.nrcs.usda.gov/intranet/pfs.html>.
- [3] Afonso AF, Pereira OR, Fernandes ÂSF, et al. The health-benefits and phytochemical profile of *Salvia apiana* and *Salvia farinacea* var. *Victoria Blue* Decoctions [J]. *Antioxidants*, 2019, 8: 241.
- [4] Srivedavyasari R, Hayes T, Ross SA. Phytochemical and biological evaluation of *Salvia apiana* [J]. *Nat Prod Res*, 2017, 31: 2058-2061.
- [5] Hu J, Wang F, Liang F, et al. Identification of abietane-type diterpenoids and phenolic acids biosynthesis genes in *Salvia apiana* Jepson through full-length transcriptomic and metabolomic profiling [J]. *Front Plant Sci*, 2022, 13: 919025.
- [6] Song J, Fang X, Li C, et al. A 2-oxoglutarate-dependent dioxygenase converts dihydrofuran to furan in *Salvia* diterpenoids [J]. *Plant Physiol*, 2022, 188: 1496-1506.
- [7] Carolis ED, Luca VD. 2-Oxoglutarate-dependent dioxygenase and related enzymes: biochemical characterization [J]. *Phytochemistry*, 1994, 36: 1093-1107.
- [8] Wang X, Wang B, Li H, et al. Research progress of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant [J]. *Plant Physiol J* (植物生理学报), 2021, 57: 2077-2090.
- [9] Kawai Y, Ono E, Mizutani M. Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants [J]. *Plant J*, 2014, 78: 328-343.
- [10] Nadi R, Mateo-Bonmati E, Juan-Vicente L, et al. The 2OGD superfamily: emerging functions in plant epigenetics and hormone metabolism [J]. *Mol Plant*, 2018, 11: 1222-1224.
- [11] Xu Z, Song J. The 2-oxoglutarate-dependent dioxygenase superfamily participates in tanshinone production in *Salvia miltiorrhiza* [J]. *J Exp Bot*, 2017, 68: 2299-2308.
- [12] Chen C, Chen H, Zhang Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data [J]. *Mol Plant*, 2020, 13: 1194-1202.
- [13] Kelley LA, Mezulis S, Yates CM, et al. The Phyre2 web portal for protein modeling, prediction and analysis [J]. *Nat Protoc*, 2015, 10: 845-858.
- [14] Liu W, Wang D, Wang Z, et al. Cloning and expression analysis of 2-oxoglutarate-dependent dioxygenase from *Salvia miltiorrhiza* Bunge [J]. *Acta Bot Boreali-Occid Sin* (西北植物学报), 2012, 32: 1289-1294.
- [15] Huang W, Yang X, Zhang C, et al. Cloning and sequence analysis of flavonol synthase gene from tobacco (*Nicotiana tabacum* L.) [J]. *Plant Physiol Commun* (植物生理学通讯), 2006, 42: 1059-1062.
- [16] Ma X, Zhu M, Wang N, et al. Identification and expression analysis of 2-*ODD* gene family involved in flavonoid biosynthesis pathway in safflower [J]. *Acta Pharm Sin* (药学学报), 2021, 56: 2858-2866.
- [17] Wei S, Zhang W, Fu R, et al. Genome-wide characterization of 2-oxoglutarate and Fe(II)-dependent dioxygenase family genes in tomato during growth cycle and their roles in metabolism [J]. *BMC Genomics*, 2021, 22: 126.