



分子植物育种
Molecular Plant Breeding
ISSN 1672-416X, CN 46-1068/S

《分子植物育种》网络首发论文

题目： 利用 RNA-seq 筛选调控当归抽薹主要候选基因
作者： 郭李智，宗渊，席杏媛，李建民，刘宝龙
网络首发日期： 2020-03-06
引用格式： 郭李智，宗渊，席杏媛，李建民，刘宝龙. 利用 RNA-seq 筛选调控当归抽薹主要候选基因. 分子植物育种.
<http://kns.cnki.net/kcms/detail/46.1068.S.20200306.1454.008.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

研究报告

Research Report

利用 RNA-seq 筛选调控当归抽薹主要候选基因

郭李智¹ 宗渊² 席杏媛² 李建民^{1*} 刘宝龙^{2*}

1 青海师范大学, 西宁, 810008; 2 中国科学院西北高原生物研究所, 青海省作物分子育种重点实验室, 西宁, 810001

*通信作者, beyond_3862740@163.com; bliliu@nwipb.cas.cn

摘要 当归为伞形科当归属多年生草本药用植物, 提前抽薹严重影响当归产量, 而当归抽薹分子遗传机理尚不清楚。本研究利用 RNA-seq 技术对连续五代抽薹当归的叶片、茎段、种子、根进行测序, 对所有组装后的 Unigenes 进行功能注释, 寻找调控当归抽薹主要候选基因。结果表明: 在当归的叶片、根、种子、茎段转录本中分别得到 7.1 GB、7.63 GB、7.23 GB、6.59 GB 数据, 通过 Nr、Nt、Swissprot、KEGG、KOG、Interpro、GO、Intersection、Overall 等蛋白数据库进行蛋白质预测, 分别预测了 71 024、59 931、49 694、51 832、55 320、46 908、35 598、17 133、79 161 个蛋白。所有转录本中 80.13% 预测蛋白与胡萝卜同源。在 KEGG 通路中共有 2 666 个差异表达的基因参与植物信号传导; 与拟南芥、胡萝卜、芥菜中参与调控开花相关基因比对, 共筛选出 56 条候选基因, 将 56 条候选基因在 NCBI 中进行核酸比对, 得到与直接控制抽薹 AP1 转录因子同源的 Unigene45683_All、激活 AP1 转录因子的整合子 FT 和 FD 同源的 Unigene11264_All、参与拟南芥光周期调控基因 CO (CONSTANS) 的同源基因 Unigene4344_All、在芥菜中促进抽薹整合因子 SOC1 的表达的 AGL24 或直接作用于抽薹决定基因 LFY 的同源基因 Unigene46836_All、Unigene7820_All、CL10006.Contig2_All。因此, 上述候选基因很有可能参与调控当归抽薹相关分子机制。本研究筛选到与当归抽薹相关的候选基因, 有助于了解当归抽薹相关分子机制, 为挖掘控制当归抽薹相关基因提供理论依据。

关键词 当归(*Angelica sinensis*), 抽薹, RNA-Seq, 同源基因

RNA-seq Screening to Control the Main Candidate Genes for Bolting of *Angelica sinensis*

Guo Lizhi¹ Zong Yuan² Xi Xingyuan² Li Jianmin^{1*} Liu Baolong^{2*}

1 Qinghai Normal University, Xining, 810008; 2 Key Laboratory of Crop Molecular Breeding, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810001

* Corresponding authors, beyond_3862740@163.com; bliu@nwipb.cas.cn

Abstracts Angelica belongs to perennial herbaceous medicinal plants, and early bolting seriously affects the yield of Angelica, and the molecular genetic mechanism of Angelica bolting is still unclear. This study used RNA-seq technology to sequence the leaves, stem segments, seeds, and roots of *Angelica sinensis*, and annotated all the assembled Unigenes to find the main candidate genes for regulating *Angelica sinensis* bolting, and performed functional analysis of candidate genes. The results showed that 7.1, 7.63, 7.23, and 6.59 GB of data were obtained in the transcription of leaves, roots, seeds, and stems of Angelica, respectively. Protein prediction through protein databases such as Nr, nt, Swisstorm, KEGG, KOG, Interpro, GO, Intersection, Overall, etc.. 71 024, 59 931, 49 694, 51 832, 55 320, 46 908, 35 598, 17 133, 79 161 proteins were predicted, respectively. 80.13 % of the predicted proteins in all transcripts are homologous to carrots. A total of 2,666 genes expressed differently in the KEGG pathway are involved in plant signaling; A total of 56 candidate genes were screened and 56 candidate genes were compared with the genes involved in the regulation of flowering in Arabidopsis, carrots, and mustard, and 56 candidate genes were compared with nucleic acids in NCBI. Unigene45683_All, Unigene11264_All, which is homologous to the direct control of bolting AP1 transcription factor, the integrin FT which activates AP1 transcription factor and the FD homologous, and the homologous gene Unigene4344_All that participates in the optical cycle regulatory gene CO (CONSTANS) are obtained. Promoted in mustard The expression of AGL24 by bolting integration factor SOC1 can also directly affect the homologous genes of bolting gene LFY, Unigene468836_All, Unigene7820_All, and CL10006. Contig 2_All. Therefore, the above candidate genes are likely to participate in the regulation of Angelica pull-related molecular mechanisms. In this study, the candidate genes associated with bolting of *Angelica sinensis* were selected, which is helpful to understand the bolting related molecular mechanism of *Angelica sinensis*, and provide theoretical basis for excavating and controlling bolting related genes of *Angelica sinensis*.

Keywords Angelica (*Angelica sinensis*), Bolting, RNA-Seq, Homologous genes

当归(*Angelica sinensis* (Oliv.) Diels)是伞形科多年生草本植物,高 0.4~1 m, 花期 6~7 月, 果期 7~9 月。当归主要分布在甘肃、云南、四川、青海、贵州等地,主产于甘肃省岷县、漳县一带,年产量占全国总产量的 90% 以上。当归始载于《神农本草经》,以干燥根入药,具有活血补血,调经止痛和润肠通便等多种功效(中华人民共和国药典委员会, 2005,中华人民共和国药典(2005 年版一部), pp.89)。现代药物化学

和药理学研究表明,当归化学成分主要为水溶成分(张丽筑等, 2009)和挥发油(孙敏等, 2019)两大部分, 包含了苯酚类化合物、多糖、有机酸以及丰富的氨基酸和微量元素等多种成分, 具有广泛的药理作用, 对人体的神经系统、免疫系统、循环系统、呼吸道系统均产生影响。近年来, 随着现代医学的发展和人们对天然药物的需求量增加, 中药行业进入全面快速发展的新时期。其中, 有关当归新产品的开发及应用市场不断增大, 当归已经广泛用于美容保健剂、饮料、调味品、化工原料等制造业的生产(李少锋和孙朝辉, 2005; 杨妍等, 2006)。

当归已经成为甘肃省支柱产业之一, 由于在生产上缺乏科学规范化栽培规程, 导致药材产量和质量不稳定(韩静, 2018)。其中, 抽薹是影响当归产量与质量的最大因素, 龚成文等(2016)通过调查表明, 一般年份当归抽薹率在 20%~30%, 遇到高温干旱年份, 抽薹率高达 60%~80%, 严重影响经济效益。抽薹是植株营养生长向生殖生长转化的形态学指标, 由于十字花科植物产品器官具有多样性, 对于食用花茎、花蕾、角果的作物来说, 适当促进抽薹是重要的; 而对于食用对象为变态根、茎、叶等营养器官的作物来说, 防止先期抽薹是保障作物高产稳产的前提(汪精磊等, 2015)。提早抽薹的当归大部分营养成分用于抽薹结实, 根部不在膨大呈肉质, 而是呈柴性, 无油气, 不能入药, 经济价值大大降低(蔡年辉等, 2016)。目前, 研究人员分别从生境因子、营养物质、育苗技术、栽培模式等方面对当归抽薹进行研究, 比如, 杨永林等(2014)通过“梨归蒜”复合种植模式、惠娜娜等(2015)利用外源激素对当归茎叶进行处理、金素奎(2013)通过对内源激素和蛋白磷酸化的调节等方面分别对当归抽薹问题进行研究, 效果均不太明显。

植物抽薹受内源信号和外部环境共同作用。近年来, 通过基因分离等手段, 模式植物拟南芥开花诱导过程逐渐清晰, 筛选出包括光周期途径 CO、赤霉素途径 LFY、和春化途径 FRI、FLC、VRN 等多个控制开花关键基因。其中 FLC 是春化途径的主要调控基因, 通过抑制成花信号的整合因子 SOC1 和 FT, 从而控制下游的 LFY 和 AP1 等开花决定基因, 最终调控开花过程。光周期途径中 CO 基因是光周期调控的关键基因, 其直接作用于成花整合因子 SOC1。赤霉素具有调控开花的作用, 赤霉素通过激活开花决定基因 LFY 从而启动成花。综上所述, FLC 是自主和春化途径的关键基因, 光周期途径通过 CO 间接作用于 FT, 最终通过促进或抑制开花决定基因 AP1 或 LFY, 从而决定植物是否开花。

转录组测序(RNA-seq)已经发展为一种低成本高效益的工具, 具有通量高、成本低、不受基因组序列信息限制等优势, 是现今分子生物学研究中最活跃的领域之一(赵振宇等, 2017)。本实验对当归的叶片、

根、种子、茎段进行 RNA-Seq 测序，通过与胡萝卜中控制抽薹相关基因进行同源比对，可以快速筛选出当归不同组织中与抽薹相关候选基因。

1 结果与分析

1.1 序列组装及编码区预测

当归样品共测得 190 388 594 个 Reads，不同组织中的 Q20/Q30 均大于 94%，测序质量高，可以用于后续分析(表 1)。叶片共获得 7.1 Gb 数据，测得 47338454 个 Reads，Q20/Q30 百分比分别为 98.23%和 94.88%；根部共获得 6.59 Gb 数据，测得 43 944 790 个 Reads，Q20/Q30 百分比分别为 98.28%和 94.99%；种子共获得 7.23 Gb 数据，测得 48 214 288 个 Reads，Q20/Q30 百分比分别为 98.05%和 94.47%；根中共获得 7.63 Gb 数据，测得 50 891 062 个 Reads，Q20/Q30 百分比分别为 94.47%和 94.59%。

表 1 转录组测序数据统计

Table 1 Data Statistics on Transcription Group Sequencing

类别	叶	根	种子	茎
Classification	Leaf	Root	Seed	Stem
原始总读数	47 338 454	43 944 790	48 214 288	50 891 062
Total Raw Reads				
清除后总读数	47 336 254	43 942 788	48 212 110	50 888 708
Total Clean Reads				
清除后总核苷酸长度(nt)	7 100 768 100	6 591 718 500	7 232 143 200	7 633 659 300
Total Clean Nucleotides (nt)				
Q20 (%)	98.23	98.28	98.05	94.47
Q30 (%)	94.88	94.99	94.47	94.59
N (%)	0.00	0.00	0.00	0.00
GC (%)	39.26	42.1	39.33	39.04
清除后总数据(Gb)	7.1	6.59	7.23	7.63
Total clean bases (Gb)				

利用 Trinity 软件生成 366 738 个 Contigs，通过鉴定得到 113 573 个 Unigenes，平均长度为 1 068 nt，N50 长度为 1 802 nt (图 1)。

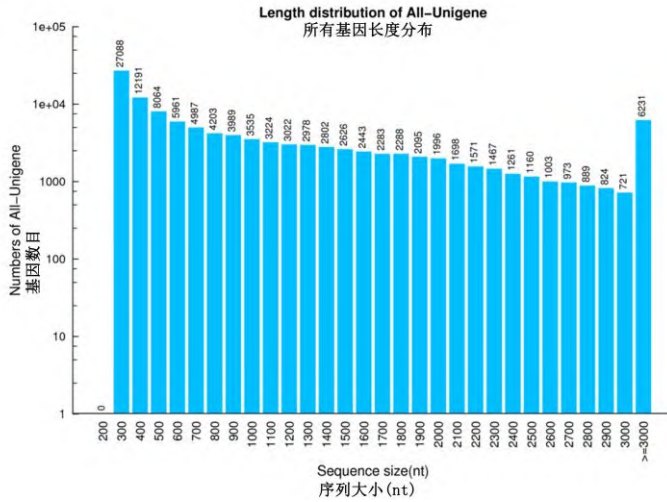


图 1 当归转录本中所有基因长度分布

Figure 1 Distribution of lengths of the assembled unigenes in Angelica

1.2 功能注释

将拼接的 Unigenes 通过与 Nr (NCBI 蛋白质数据库)、Nt (非冗余核苷酸数据库)、Swissprot (蛋白质序列数据库)、KEGG 数据库、KOG(真核生物直系同源数据库)、Interpro 蛋白数据库、GO (基因本体论数据库)、Intersection、Overall 等蛋白数据库进行比对, 分别预测了 71 024、59 931、49 694、51 832、55 320、46 908、35 598、17 133、79 161 个蛋白, 分别占总量的 62.54%、52.75%、43.76%、45.64%、48.71%、41.30%、31.34%、15.09%、69.70% (表 2)。

表 2 基因功能注释统计

Table2 Summary statistics of functional annotation for Unigenes

数据库	数量	百分比(%)
Databases	Number	Percentage (%)
Nr	71 024	62.54
Nt	59 931	52.75
Swissprot	49 694	43.76
KEGG	51 832	45.64
KOG	55 320	48.71
Interpro	46 908	41.30
GO	35 598	31.34
Intersection	17 133	15.09
Overall	79 161	69.70

总计

113 573

100

Total

对预测蛋白进行同源比对后发现，当归转录本中 80.13% 预测蛋白与胡萝卜同源(图 2)，剩余 19.87% 预测蛋白与其他物种同源。其中胡萝卜和当归都属于伞形科植物，而转录本中大部分蛋白与该科具有高度同源性(图 2)。

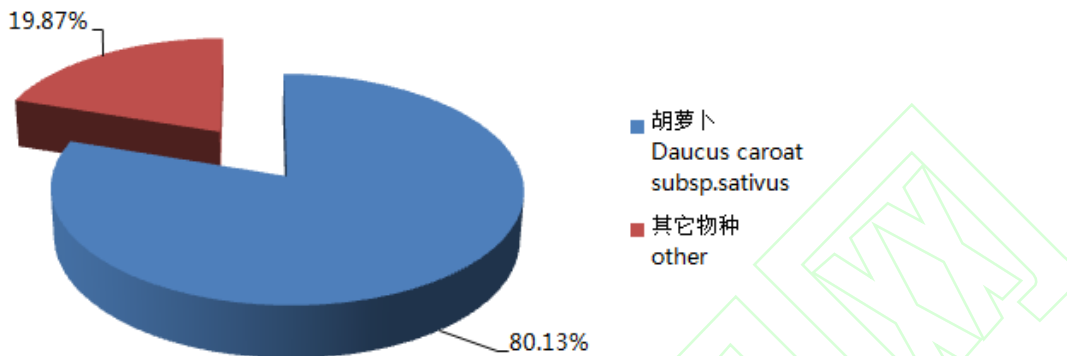


图 2 预测蛋白与同源蛋白物种分布

Figure 2 The species distributions of the homologous proteins of predicted proteins

1.3 Unigenes 的 KOG 分类

KOG 功能共注释到 55 320 个 Unigenes (表 3)，被注释的 Unigenes 涵盖了所有生命活动相关。其中参与“信号转导机制”(Signal transduction mechanisms)的 Unigenes 有 7 062 个，所占比例为 12.76%。

表 3 当归的 KOG 功能分类及数量

Table 3 Classification and quantity of KOG functions in Angelica

序号	功能	数量	序号	功能	数量
Number	Function	Quantity	Number	Function	Quantity
1	RNA 的处理与修饰 RNA processing and modification	2 506	14	细胞能动 Cell activation	49
2	染色体结构与动力学 Chromatin structure and dynamics	860	15	转录后修饰, 蛋白质转换 Posttranslational modification, protein turnover, chaperones	5 633
3	能量产生与转化 Energy production and conversion	2 209	16	无机盐离子运输与代谢 Inorganic ion transport and metabolism	1 452
4	细胞周期调控 Cell cycle control	1 404	17	次生代谢产物的合成运输分解 Secondary metabolites biosynthesis,	1 867

				transport and catabolism	
5	氨基酸的转运与代谢	2 036	18	基因功能预测	14 114
	Amino acid transport and metabolism			Gene function prediction	
6	核酸的转运与代谢	620	19	未知功能	4 962
	Nucleotide transport and metabolism			Function unknown	
7	碳水化合物的运输与代谢	3 104	20	信号转导机制	7 062
	Carbohydrate transport and metabolism			Signal transduction mechanisms	
8	辅酶运输与代谢	523	21	细胞内运输分泌	2 588
	Coenzyme transport and metabolism			Intracellular transport secretion	
9	脂质的运输与代谢	2 609	22	防御机制	947
	Transport and metabolism of lipids			Defensive mechanism	
10	翻译, 核糖体结构	3 051	23	细胞外结构	381
	Translation, ribosomal structure			Extracellular structure	
11	转录	4 603	24	核结构	257
	Transcription			Nuclear structure	
12	复制, 重组与修复	1 547	25	细胞骨架	1 226
	Replication, recombination and repair			Cytoskeleton	
13	细胞壁, 膜信使	1 283			
	Cell wal, membrane Messenger				

1.4 当归不同组织差异表达基因的 GO 功能分类和 KEGG 通路分布

将本次测序获得的 Unigenes 通过与 KEGG 数据库比对, 由 FPKM 值确定差异表达基因, 按照错配率 (FDR)≤0.001, |log₂ Ratio|>1 计算, 得到以叶片为参照, 鉴定出 10 091 个基因在当归根中表达水平较高, 另有 20 809 个基因在当归叶中表达水平较高(图 A); 鉴定出 14 281 个基因在当归种子中表达水平较高, 另有 6 388 个基因在当归叶中表达水平较高(图 B); 鉴定出 12 292 个基因在当归茎中表达水平较高, 另有 7 429 个基因在当归叶中表达水平较高(图 C)。以根为参照, 鉴定出 27 622 个基因在当归种子中表达水平较高, 另有 6 854 个基因在当归根中表达水平较高(图 D); 鉴定出 24 780 个基因在当归茎中表达水平较高, 另有 7 668 个基因在当归根中表达水平较高(图 E); 以种子为参照, 鉴定出 8 468 个基因在当归茎中表达水平较高, 另有 11 735 个基因在当归种子中表达水平较高(图 F)。

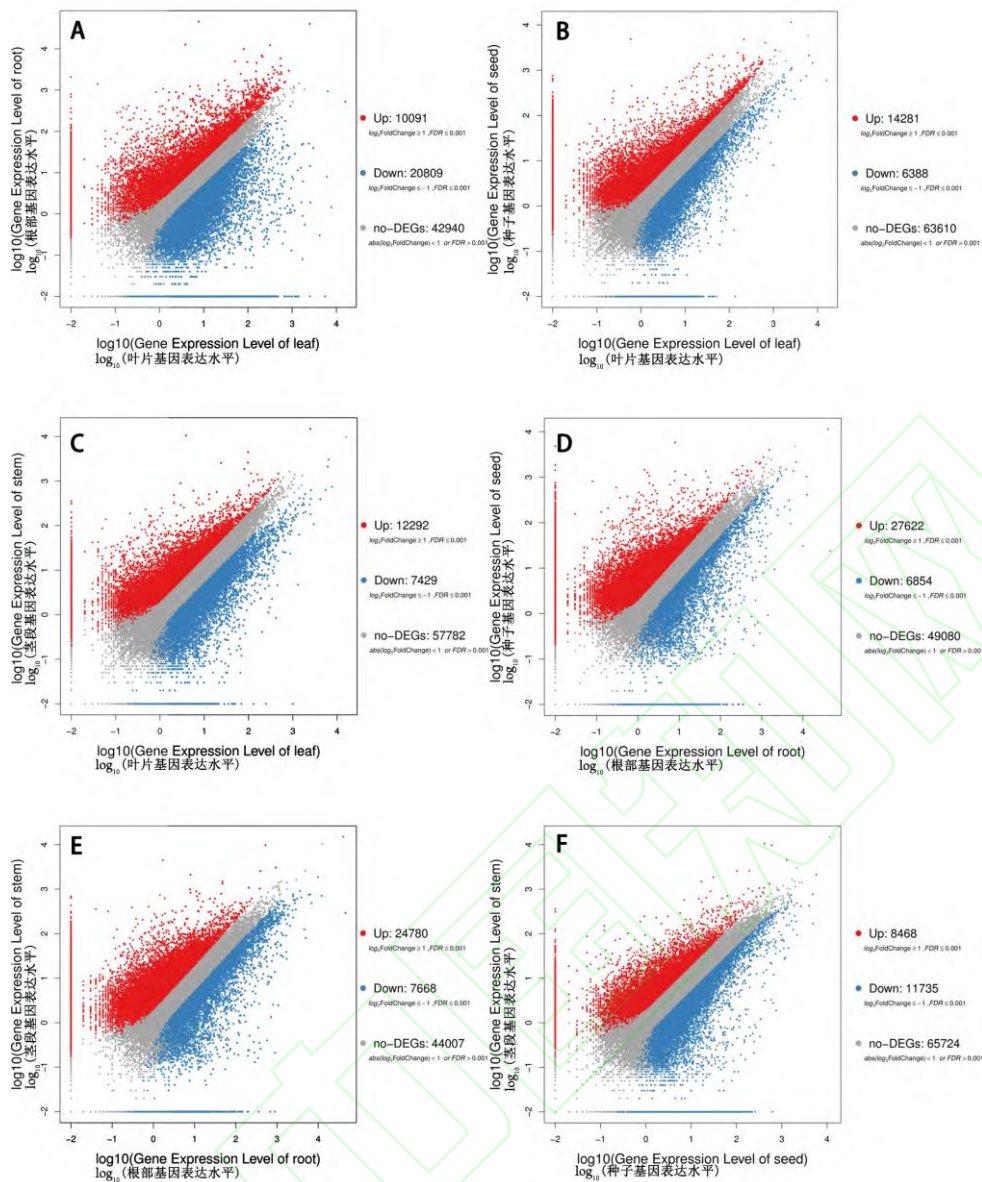


图3 当归不同组织差异表达基因统计

注: 图中横坐标, 纵坐标分别为当归不同组织中的基因表达水平计算值; A: 根部与叶片基因表达差异个数统计; B: 种子和叶片基因差异表达个数统计; C: 茎段和叶片基因差异表达个数统计; D: 种子和根基因差异表达个数统计; E: 茎段和根部基因差异表达个数统计; F: 茎和种子基因差异表达个数统计; 所有基因分为三类, 红色代表上调; 蓝色代表下调; 灰色表示该基因在两个部位中无差异表达

Figure 3 The statistics of differentially expressed gene in Angelica Different tissues

Note: The transverse coordinates and longitudinal coordinates in the figure are calculated as the gene expression levels in different tissues of Angelica; A: Shows the difference in gene expression between root and leaf; B: Statistics on the number of gene expression differences between seeds and leaves; C: Shows the number of gene expression of stem and leaf; D: Shows the number of different expression of seed and root genes; E: Statistics on the number of different expression of stem and root genes; F: Statistics on the number of gene expression differences between stem and seed; All genes are divided into three categories, red representing an

increase; Blue representing downward adjustment, Gray indicates that there is no difference in the expression of the gene in two parts

Gene ontology(GO)主要是对基因和蛋白质进行功能描述，主要类别有细胞成分、分子功能和生物过程。其中 104 623 个 Unigenes 得到 GO 功能注释其中，参与细胞成分有 68 469 个 Unigenes，占 60.28%；分子功能有 39 977 个 Unigenes，占 35.20%；注释到生物过程的有 46 828 个 Unigenes，占 41.23% (图 4)。

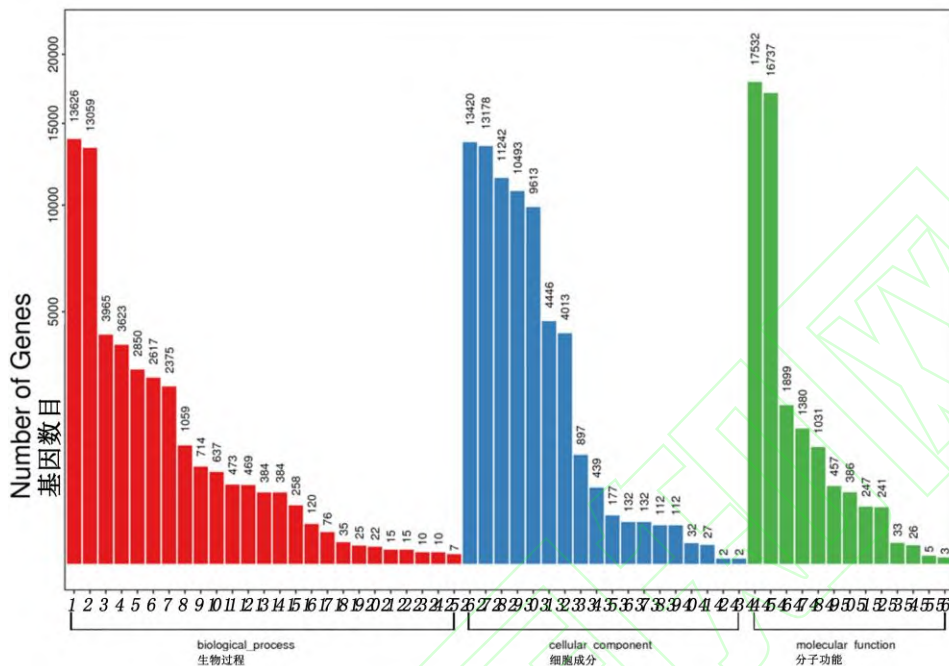


图 4 当归 Unigene 的 GO 功能分类

注: 1: 细胞过程; 2: 代谢过程; 3: 生物调节; 4: 生物过程调节; 5: 局部化; 6: 应对刺激; 7: 细胞成分组织; 8: 信号; 9: 发育过程; 10: 多细胞组织过程; 11: 生物工程正调控; 12: 生物过程负调控; 13: 繁殖; 14: 生殖过程; 15: 多生物过程; 16: 生长; 17: 免疫系统过程; 18: 细胞增殖; 19: 解毒; 20: 节律过程; 21: 生物附着; 22: 运动; 23: 碳利用率; 24: 色素淀积; 25: 氮利用率; 26: 细胞; 27: 细胞成分; 28: 膜; 29: 膜组分; 30: 细胞器; 31: 大分子配合物; 32: 细胞器组分; 33: 膜封闭腔; 34: 细胞外区; 35: 超分子配合物; 36: 细胞连接; 37: 共质体; 38: 病毒; 39: 病毒组分; 40: 细胞外区组分; 41: 类核; 42: 其他生物体; 43: 其他生物体组分; 44: 粘合物; 45: 催化活性; 46: 转运活性; 47: 分子结构活性; 48: 转录因子活性; 49: 分子功能调节剂; 50: 信号传感器活性; 51: 分子传感器活性; 52: 抗氧化活性; 53: 营养库活动; 54: 分子载体活动; 55: 蛋白质标记; 56: 毒素活动

Figure 4 Angelica Unigene's GO function classification

Note: 1: Cellular process; 2: Metabolic process; 3: Biological regulation; 4: Regulation of Biological process; 5: Localization; 6: Response to stimulus; 7: Cellar component organization; 8: Signaling; 9: Developmental process; 10: Multicellular tissue process; 11: Positive regular of biological process; 12: Negative regular of biological process; 13: Reproduction; 14: Reproductive process; 15: Multi-organism processes; 16: Growth; 17: immune system process; 18: Cell proliferation; 19: Detoxification; 20: Rhythm process; 21: Biological adhesion; 22: Locomotion; 23: Carbon utilization; 24: Pigmentation; 25: Nitrogen utilization; 26: Cell; 27: Cell part; 28: Membrane; 29: Membrane part; 30: Organelles; 31: Macromolecular complex; 32: Organelle part; 33: Membrane enclosed lumen;

34: Extracellular region; 35: Supermolecular complex; 36: Cell junction; 37: Symplast; 38: Virion; 39: Virion part; 40: Extracellular region part; 41: Nucleoid; 42: Other organism; 43: Other organism part; 44: Binding; 45: Catalytic activity; 46: Transporter activity; 47: Structural molecular activity; 48: Transcription regulator activity; 49: Molecular functional regulators; 50: Signal transducer activity; 51: Molecular transducer activity; 52: Antioxidant activity; 53: Nutrient reservoir activity; 54: Molecular carrier activity; 55: Protein tag; 56: Toxin activity

KEGG (Kyoto Encyclopedia of Gene and Genomes)通过整合基因组、化学分子和生化系统等方面的数据，系统分析基因产物功能的数据库。根据功能可将这些 KEGG 注释的 Unigenes 分为细胞过程、环境信息通路、遗传信息处理、新陈代谢和生物有机系统 5 个类别。抽薹开花属于环境信息通路中的信号转导途径，其中差异表达的 2 666 个 Unigenes 参与信号转导(图 5)。

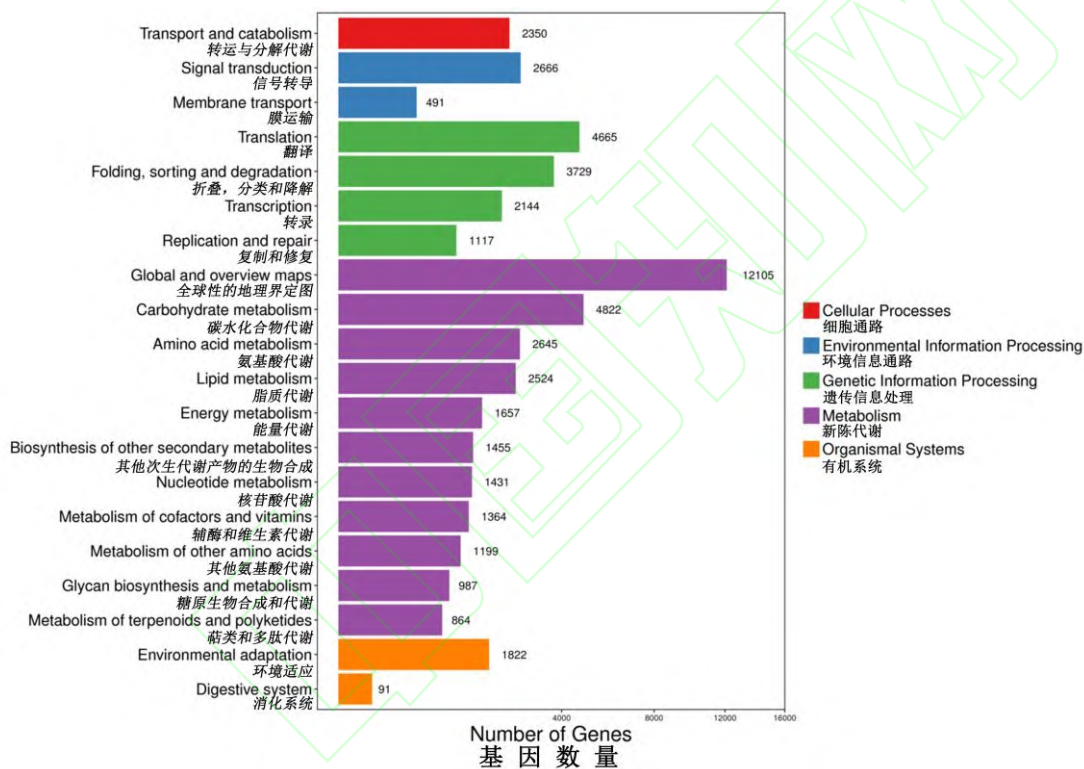


图 5 差异基因在 KEGG 通路统计

Figure 5 Differential genes are counted in the KEGG pathway

1.5 同源基因筛选

十字花科植物拟南芥抽薹诱导主要受光周期调控、春化途径、自主途径和 GA 信号转导途径调控。通过 BLASTx 在美国转录组数据库比对，参照抽薹相关 KEGG 途径进行分析，筛选出 56 条候选基因。分析 56 条候选基因的表达规律，发现控制当归抽薹的基因在当归种子高表达，在叶片、根和茎中表达差异不显著(表 4)。

表 4 当归中与抽薹相关候选基因的 RPKM 值

Table 4 The differential expression of gene relative to flowering in Angelica

基因 Gene	KO	基因编号 Gene ID	不同组织表达量				差异表达是否明显 Differential expression
			Different tissue expression				
			叶 Leaf	根 Root	种子 Seed	茎 Stem	
AP1	K09264	CL8072.Contig3_All	13.47	34.65	20.66	21.07	否 NO
		CL8072.Contig2_All	2.16	0	2.15	2.7	否 NO
		CL8072.Contig1_All	2.7	4.79	2.79	7.22	否 NO
		CL9349.Contig2_All	1.67	0	21.52	15.98	否 NO
		CL9349.Contig1_All	1.74	0	17.48	29.25	否 NO
		Unigene7820_All	0.41	0	39.97	1.06	否 NO
		Unigene45683_All	0	0	202.03	18.35	是 YES
		Unigene21987_All	0.09	0.37	1.51	10	否 NO
		Unigene8770_All	0	0	173.59	16.89	是 YES
		CL1578.Contig3_All	0	0	1.23	0	否 NO
		CL1578.Contig2_All	0	0	9.53	0	否 NO
		CL1578.Contig1_All	0	0	26.43	0.26	否 No
CO	K12135	CL11021.Contig1_All	15.79	0	0.57	0.77	否

		CL7524.Contig2_All	38.65	9.51	1.21	17.28	否 No
		CL7524.Contig1_All	6.02	10.24	5.62	18.17	否 No
		CL3664.Contig2_All	32.14	25.57	0.76	45	否 NO
		CL3664.Contig1_All	37.92	40.33	1.54	46.63	否 NO
AGL24	K09260	CL8626.Contig2_All	29.34	144.67	5.15	13.84	否 No
		CL8626.Contig1_All	20.86	105.43	10.72	3.29	是 YES
		Unigene46836_All	0	0	109.88	0.08	是 YES
		CL627.Contig1_All	0	0	18.17	0.07	否 No
	K09264	Unigene7820_All	0.41	0	39.97	1.06	否 No
		CL8978.Contig1_All	3.42	3.87	15.68	23.86	否 No
		CL10006.Contig2_All	0	0	85.86	0.14	否 No
		CL10006.Contig1_All	0	0	5.11	0	否 No
		CL9349.Contig2_All	1.67	0	21.52	15.98	否 No
		CL9349.Contig1_All	1.74	0	17.48	29.25	否 No
		Unigene14361_All	0	0	0.9	8.27	否 No
		Unigene45930_All	0	0	24.74	0.13	否

							No
		Unigene45683_All	0	0	202.03	18.35	否
							No
SVP	K09260	CL8626.Contig2_All	29.34	144.67	5.15	13.84	否
							No
		CL8626.Contig1_All	20.86	105.43	10.72	3.29	否
							No
		CL10016.Contig3_All	26.88	4.01	0.71	1.61	否
							No
		CL10016.Contig2_All	4.54	0.16	0.18	0.06	否
							No
		CL10016.Contig1_All	4.05	19.29	0	0.6	否
							No
	K09264	Unigene7820_All	0.41	0	39.97	1.06	否
							No
		CL1578.Contig3_All	0	0	1.23	0	否
							No
		CL1578.Contig2_All	0	0	9.53	0	否
							No
		CL1578.Contig1_All	0	0	26.43	0.26	否
							No
		CL8978.Contig1_All	3.42	3.87	15.68	23.86	否
							No
FT	K16223	Unigene23117_All	0	0	3.6	2.18	否
							No

目前，模式植物拟南芥的抽薹调控机制相对清晰，与抽薹相关的结构基因受 AP1 (APETALA2) 转录因子调控，参照伞形科植物胡萝卜中调控花期的决定基因 AP1，筛选出同源基因有：CL8072.Contig3_All、CL8072.Contig2_All、CL8072.Contig1_All、CL9349.Contig2_All、CL9349.Contig1_All、Unigene7820_All、Unigene45683_All、Unigene21987_All、Unigene8770_All、CL1578.Contig3_All、CL1578.Contig2_All、CL1578.Contig1_All，它们编码 AP1 转录因子从而促进植物抽

薹。比较同源基因的表达水平，AP1 转录因子 Unigene45683_All、Unigene8770_All 在叶和根中均没有表达，与茎相比，在种子中明显分别上调 11.0、10.28 倍，其他基因表达水平无差异或下调。抽薹决定基因 AP1 属于多重调控位点，受到 LFY 和 FT-FD 复合物的双重调控，以拟南芥控制抽薹途径为参照，信号途径汇聚在整合子 FT 和 FD 上来发起抽薹过渡，对 FT 与特异分生组织 bZIP 类转录因子 Flowering locus D(FD) 进行分析，筛选出同源基因有：CL11965.Contig2_All、CL11965.Contig1_All、CL7527.Contig1_All、Unigene11264_All、Unigene23117_All，它们激活 AP1 转录因子从而促进成花并启动花发育过程。在拟南芥中，FLC 属于自主和春化途径中的抑制基因，直接作用于成花信号的整合因子 SOC1 和 FT，通过与胡萝卜中 FLC 同源比较，筛选出同源基因有：Unigene2473_All、CL7801.Contig2_All、CL7801.Contig1_All。CO (CONSTANS)是光周期途径中的关键基因，其表达可直接诱导成花整合因子 FT 上调表达，FT 蛋白作为成花素有叶片转移到茎尖分生组织促进抽薹，通过与拟南芥光周期调控基因 CO 同源比对，筛选出同源基因有:CL11021.Contig1_All、Unigene4344_All；参照十字花科植物芥菜抽薹调控蛋白 AGAMOUS-LIKE24 (AGL24)与 SOC1 基因相互作用的结果，AGL24 能够促进抽薹整合因子 SOC1 的表达，也可直接作用于抽薹决定基因 LFY，通过与芥菜中 AGL24 进行同源比较，筛选出同源基因有：CL8626.Contig2_All、CL8626.Contig1_All、Unigene46836_All、CL627.Contig1_All、Unigene7820_All、CL8978.Contig1_All、CL10006.Contig2_All、CL10006.Contig1_All、CL9349.Contig2_All、CL9349.Contig1_All、Unigene14361_All、Unigene45930_All、Unigene45683_All。此外，SVP 蛋白通过结合 FT 和 SOC1 的 CArG 基序也能调控 AP1 表达，筛选出同源基因有:CL11021.Contig2_All、CL8626.Contig2_All、CL8626.Contig1_All。

2 讨论

当归作为我国名贵中药材之一，药效广泛。早期抽薹严重影响当归产量与质量，有关当归抽薹分子机理方面还未见报道。对于缺乏基因组信息的物种，转录组测序可快速获得各种植物类型转录本，已成为非模式植物功能基因挖掘有效工具之一。毛笈华(2013)利用 RNA-seq 获得胡萝卜转基因数据库，通过与拟南芥成花调控基因进行比对，筛选出包括春化途径上的 FRI、FLC，光周期途径关键基因 GI、COL，成花信号整合因子 SOC1 等(毛笈华, 2013)。本研究采用 RNA-seq 技术对当归四种组织进行转录组测序，得到了 190 388 594 个 Reads，进一步鉴定得到 113 573 个 Unigenes，构建了当归的 cDNA 文库，很有可能找出控制当归抽薹相关基因，为当归抽薹分子机理提供可靠依据。

当归 GO 分析结果显示, 41.23% 的 Unigenes 参与生物学途径, KEGG 数据库中有 2 666 个 Unigenes 参与信号转导。信号转导属于生物学途径中的一种, 其中控制抽薹途径的变化比其他次生代谢途径更明显。参照伞形科植物胡萝卜抽薹途径决定基因 AP1, 确定其是否与胡萝卜控制抽薹基因同源。经比对后发现, Unigene45683_All 与胡萝卜 MCM1 基因同源性最高, Unigene45683_All 与胡萝卜 AP1 基因同源。且该 Unigene 在种子中表达量明显高于其他组织中, 因此, Unigene45683_All 很有可能是控制当归抽薹的转录因子; 以此类推, 通过与激活 AP1 转录因子的整合子 FT 和 FD 同源比对, 发现 Unigene11264_All 与 FD 基因同源, 并且只在种子中表达, 很有可能是激活 AP1 转录因子从而促进成花并启动花发育的候选基因; 参照十字花科植物拟南芥抽薹抑制基因 FLC, 得到同源基因 Unigene2473_All、CL7801.Contig2_All、CL7801.Contig1_All, 但差异表达均不明显, 因此, 基因 FLC 有可能没有参与当归抽薹调控。参照拟南芥光周期途径中的关键基因 CO (CONSTANS), 得到同源基因 CL11021.Contig1_All、Unigene4344_All, 其中 Unigene4344_All 在种子中表达明显高于在其它组织中表达水平, 因此, Unigene4344_All 有可能是光周期控制基因 CO 的同源基因; 十字花科植物芥菜中 AGL24 能够促进抽薹整合因子 SOC1 的表达, 其同源基因 Unigene46836_All、Unigene7820_All、CL10006.Contig2_All 差异表达比较明显, 很有可能也能够促进抽薹整合因子 SOC1 的表达, 进而控制抽薹。

3 材料与方法

3.1 材料

供试验的当归来自于中科院西北高原生物研究所, 连续五代种植, 抽薹率达到 80%~90%。将抽薹当归留种直播于培养皿中, 暗培养 20 d, 待植株萌发后 3 d, 移栽至花盆中, 待植株长至 20 cm 时, 从当归植株上分别截取茎段、叶、根于离心管中, 快速置于液氮桶中, 用于提取 RNA。利用天根总 RNA 提取试剂盒分别提取当归幼苗的根、茎、叶和采集的种子进行总 RNA 提取, 利用琼脂糖凝胶电泳检测后, 用核酸检测仪(赛默飞, 美国)测定浓度。

3.2 文库构建、测序及组装

根据北京诺禾生物科技有限公司样品制备方法创建当归 cDNA 文库。将制备好的当归 cDNA 文库放入 Illumina HiSeq 2000 仪器进行测序。将测得的原始读序进行过滤, 去除低质量、接头序列从而完成读数。然后由 Trinity 进行组装, 得到当归转录本。

3.3 基因功能预测与差异表达基因筛选

将组装的 Unigenes 分别提交 NR、KEGG 和 GO 数据库进行功能预测，然后把当归不同组织中所有差异显著的 Unigenes 定位到 GO 和 KEGG 途径数据库中，对单个 Unigene 分析寻找存在显著性富集的通路注释。收集 KEGG 途径中与控制抽薹相关基因，通过比较当归种子中与抽薹相关的单基因长度与参考基因长度，探讨分离抽薹相关基因的可能性。通过搜集十字花科植物拟南芥和芥菜、伞形科植物胡萝卜开花调控途径上的 8 个已知基因 AP1 (K09264)、CO (K12135)、FLC (K09264)、FT (K16223)、FD (K11450)、SVP (K09260)、SOC1 (K09264)、AGL24 (K09260)，根据这些基因名称在 NCBI 上下载抽薹相关基因 CDS 序列，以之为参考序列在美国转录组数据库中进行 BLASTn 比对。以上获得的所有候选序列再利用 BLASTx 进行比对确认，具有抽薹基因相关结构域的为该基因在当归中的同源基因。

作者贡献

郭李智是本研究的实验设计和实验研究和的执行人，完成数据分析和论文初稿的写作；席杏园参与实验设计和试验结果分析；宗渊是项目的构思者及负责人，指导实验设计和数据分析；刘宝龙和李建民完成指导论文写作和修改。全体作者都阅读并同意最终的文本。

致谢

本研究由青海省植物基因与基因组学研究创新服务平台项目(2018-ZT-T08)资助。

参考文献

- Cai N.H., Deng L.L., and Xu Y.L., 2016, Transcriptome analysis of Yunnan pine based on high-throughput sequencing, *Zhiwu Yanjiu (Journal of Plant Research)*, (1): 75-83 (蔡年辉, 邓丽丽, 许玉兰, 2016, 基于高通量测序的云南松转录组分析, 植物研究, (1): 75-83)
- Gong C.W., Deng L.L., and Xu Y.L., 2016, Discussion on early moss extraction of *Angelica sinensis*, *Zhiwu Yanjiu (Plant Research)*, (1): 75-83 (龚成文, 邓丽丽, 许玉兰, 2016, 当归提早抽薹的问题商榷, 植物研究, (1): 75-83)
- Han J., 2018, Research on key techniques of standardized cultivation of *Angelica sinensis*, Thesis for M.S., Gansu agricultural university, Supervisor: Chen H., pp. 1-4 (韩静, 2018, 当归规范化栽培关键技术研究, 硕士学位论文, 甘肃农业大学, 导师: 陈桓, pp.1-4)

- Hui N.N., Wang L., and Li J.P., 2015, Effects of 8 exogenous hormones on the convulsion and yield of *Angelica sinensis*, *Gansu Nongye Keji (Journal of Gansu Agricultural Sciences)*, (12): 27-30 (惠娜娜, 王立, 李继平, 2015, 8 种外源激素对当归抽薹及产量的影响, *甘肃农业科技*, (12): 27-30)
- Jin S.K., 2013, Study on the physiological mechanism and preventive measures of *Angelica* in early twitching, Thesis for M.S., Lanzhou University, Supervisor: Pu T.L., pp.38-39 (金素奎, 2013, 当归提前抽薹的生理机制与防止措施研究, 硕士学位论文, 兰州大学, 导师: 浦铜良, pp.38-39)
- Li SH.F., and Sun Z.H., 2005, High-yield and high-efficiency cultivation technology of *Angelica sinensis*, *Yunnan Nongye (Yunnan Agriculture)*, (4): 10 (李少锋, 孙朝辉, 2005, 当归高产高效栽培技术, *云南农业*, (4): 10)
- Mao J.H., 2013, Using RNA-seq Technology to Excavate Carrot Bolting and Flowering Related Genes, Thesis for M.S., Chinese Academy of Agricultural Sciences, Supervisor: Zhuang Y.F., pp.47 (毛笈华, 2013, 利用 RNA-seq 技术挖掘胡萝卜抽薹开花相关基因, 硕士学位论文, 中国农业科学院, 导师: 庄云飞, pp.47)
- Sun M., Ma Q.L., and Liu F.L., 2019, New development of volatile oil from *Angelica sinensis*, *Shijie Zuixin Yixue Xinxi Wenzhai (World Latest Medical Information Digest)*, (2): 56-58 (孙敏, 马清林, 刘峰林, 2019, 当归挥发油研究新进展, *世界最新医学信息文摘*, (2): 56-58)
- Wang J.L., Li X.X., and Qiu Y., 2015, Advances in regulation mechanism and molecular breeding of flowering traits of cruciferous vegetables, *Zhiwu Yichuan Ziyuan Xuebao (Journal of Plant Genetic Resources)*, 16 (6): 1283-1284 (汪精磊, 李锡香, 邱杨, 2015, 十字花科蔬菜抽薹抽薹性状的调控机理和分子育种研究进展, *植物遗传资源学报*, 16(6): 1283-1284)
- Yang Y., Dou X.C., and Zhang Z.Z., 2006, Economic benefit analysis of *Angelica sinensis* in min county, *Gansu Nongye Daxue Xuebao (Journal of Gansu Agricultural University)*, 41(4): 109-112 (杨妍, 窦学诚, 张正卓, 2006, 岷县当归经济效益分析, *甘肃农业大学学报*, 41(4): 109-112)
- Yang Y.L., Yang X.Q., and Du W.K., 2014, Experimental study on compound planting of "Pear Guizhi", *Anhui Nongxue Tongbao (Anhui Agricultural Science Bulletin)*, 20(13): 110-111 (杨永林, 杨秀青, 杜伟坤, 2014, “梨归蒜”复合种植试验研究, *安徽农学通报*, 20(13): 110-111)
- Zhang L.Z., Han Y.L., and Hao X.Y., 2009, Study on chemical constituents of water extract of *Angelica sinensis*, *Guizhou Kaixue (Guizhou Science)*, (6): 28-30 (张丽筑, 韩应莲, 郝小燕, 2009, 当归水提物化学成分的研究, *贵州科学*, (6): 28-30)

Zhao Z.Y., Wang S.Y., and Guo F.G., 2017, Transcriptome sequencing and its application in medicinal plants, *Jiyinzuxue Yu Yingyong Shengwuxue (Genomics and Applied Biology)*, 35(2): 820-825 (赵振宇, 王仕玉, 郭凤根, 2017, 转录组测序及其在药用植物上的应用, *基因组学与应用生物学*, 35(2): 820-825)

