

晚香玉鳞茎的全长转录组分析

姜福星¹, 吴菁熙¹, 朱 泉¹, 黄远祥², 周 鹏², 高 平², 李 西¹

(¹四川农业大学风景园林学院, 成都 611130; ²四川天艺生态园林股份有限公司, 成都 611130)

摘要: 晚香玉为石蒜科的多年生鳞茎花卉, 是著名的香花植物, 具有较高的观赏、经济和药用价值, 鳞茎是其主要的繁殖器官和药用器官, 为深入探析其生长发育的机制和次生代谢途径, 利用高通量测序PacBio Sequel平台, 以其鳞茎为材料, 使用单分子长读测序技术(SMART, single molecule long reading sequencing technology)对晚香玉鳞茎进行三代全长转录组测序及分析。平台共获得508012条高质量环形一致序列(CCS, circular consensus)和402032条全长非嵌合序列(FLNC, full-length non-chimeric read), 对全长非嵌合序列进行聚类得到高质量一致序列共137215个, 经质控和去冗余后共获得81719个转录本(Unigene)和36215个完整CDS区。获得的所有转录本经过NR、SwissProt、KOG、KEGG、GO等数据库进行注释和功能分类, 结果共有64362个单基因被注释, NR注释的基因数量最多, 为63538个, 占98.72%; KEGG注释的基因最少, 58878个基因被注释到148条途径, 代谢途径分布的基因最多(11112)。同时分析出41240个SSR、15835个lncRNA和4903个转录因子, RLK转录因子家族的基因最多; 发现同源盒转录因子可能参与了晚香玉鳞茎的发育过程, 并对其中的*PtKNOX1L13*基因进行了克隆, 发现其在鳞茎的起始、膨大和成熟3个时期表达量均比较高, 瞬时过量表达表明其能促进珠芽的发育, 可能是参与鳞茎形成和发育的重要功能基因。本研究为深入研究晚香玉的重要功能基因及种质资源创新提供了参考和依据。

关键词: 晚香玉; 鳞茎; 全长转录组; 种质创新

Full-length Transcriptome Analysis of Bulbs of *Polianthes tuberosa* L.

JIANG Fuxing¹, WU Jingxi¹, ZHU Xiao¹, HUANG Yuanxiang², ZHOU Peng², GAO Ping², LI Xi¹

(¹School of Landscape Architecture, Sichuan Agricultural University, Chengdu 611130;

²Sichuan Tianyi Ecological Landscape Co., Ltd, Chengdu 611130)

Abstract: *Polianthes tuberosa* L. is a perennial bulbous flower of the Amaryllidaceae family. It is a famous fragrant flower plant with high ornamental, economic and medicinal values. Bulbs are their main reproductive and medicinal organs. In order to further explore its growth and development mechanism and secondary metabolic pathways, we use the high-throughput sequencing PacBio Sequel platform to make use of its bulbs as materials, the single molecule long reading sequencing technology (SMART) was used to sequence and analyze the full-length transcriptome of *Agave amica* bulbs for three generations. The platform obtained a total of 508012 high-quality CCS reads and 402032 full-length non chimeric sequences (FLNC), and clustered the full-length non chimeric sequences to obtain a total of 137215 high-quality consistent sequences. After quality control and redundancy removal, a total of 81719 transcripts (Unigene) and 36215 complete CDS regions were obtained. All transcripts obtained were annotated and functionally classified through databases such as NR, SwissProt, KOG, KEGG GO, etc. The result showed that a total of 64362 single genes were annotated, with NR having the highest annotation count of 63538, accounting for 98.72%, while KEGG had the lowest number of annotated genes, with 58878 genes annotated to 148 pathways. And 41240 SSRs, 15835 lncRNAs and 4903 transcription factors were analyzed. RLK transcription factor family has the most genes. The discovery of homologous box transcription factors family may be involved in the development process of tuber, *PtKNOX1L13* gene was

收稿日期: 2023-12-11 网络出版日期: 2024-07-17

URL: <https://doi.org/10.13430/j.cnki.jpgr.20231211001>

第一作者研究方向为园林植物遗传育种, E-mail: bnjfx@126.com

基金项目: 成都市高等院所人才创新服务项目(2021-RC03-00001-CG)

Foundation project: Chengdu Higher Education Institute Talent Innovation Service Project(2021-RC03-00001-CG)

cloned, its expression levels were relatively high during the initiation, enlargement, and maturity stages of tubers, transient overexpression indicates that it can promote the development of bulbils, so that it may be an important functional gene involved in bulb formation and development. This study provides reference and basis for in-depth research and development of molecular markers, important functional genes and germplasm resources innovation of *Polianthes tuberosa* L.

Key words: *Polianthes tuberosa* L.; bulb; full-length transcriptome; germplasm innovation

晚香玉(*Polianthes tuberosa* L.)为多年生鳞茎花卉,传统上为石蒜科,近年来将其归到天门冬科龙舌兰属(*Agave* L.)植物,命名为*Agave amica* Medik,原产于墨西哥和南美洲等地区,后在世界各地广泛栽培,也成为印度和法国等国家花卉产业的主要支柱和经济收入的重要来源^[1]。晚香玉翠叶素茎,碧玉秀荣,幽香四溢,观赏价值较高,可作为盆花、切花和地被植物栽培,尤其是花朵可提取香精油,可用于化工和养生保健等领域;叶片和鳞茎中含有类黄酮等次生代谢成分,味甘淡凉,清热解毒、主治疮痍肿毒,药用价值也较高。国内外均认为,晚香玉是极具发展潜力的花卉^[2-3]。

选育优良新品种是提高经济效益和高质量发展的有效途径,然而现有的晚香玉主要以杂交育种和诱变育种为主,存在着周期长和效益低等问题^[4],生物技术育种成为新的重要手段^[5-6]。目前,晚香玉的基因组信息未知,因此可以利用转录组测序技术来获知其遗传信息。

晚香玉的花香非常引人瞩目,最先用二代转录组对晚香玉开花和花发育等进行了分析,但是主要是采用Illumina HiSeq™ 2000和4000等二代测序平台,往往存在读序很短、拼接易出错、不能准确组装出完整的转录本及获得目的基因的全长等问题,不能识别超家族基因、同源基因等表达的转录本,难以得到更完整更准确的信息^[7-10]。随着单分子测序等第三代测序技术的出现,在转录组学研究及获取物种的遗传信息等方面表现出独到的优势,基于PacBio SMART单分子实时测序技术的全长转录组测序无须打断RNA片段,直接反转录得到全长cDNA,而且能够超长读取(平均10 kb)包含了单条完整转录本的序列信息,后期分析无需组装,即可得到较准确的全长的基因信息^[11],但是应用于晚香玉的研究仍尚未见报道。

晚香玉能绽放出美丽、芳香的花朵,关键在于地下优质的鳞茎。从仔球的繁育、休眠与萌发、花芽的分化、开花的数量与质量到病虫害防治等种植的关键环节,都是围绕着鳞茎而展开的^[1,10,12]。同

时鳞茎也能生产生物碱和龙舌兰素等重要药用成分^[13-14],更是进行物理化学诱变和分子育种的主要材料^[4,6]。由此可见鳞茎对晚香玉的研发具有极其重要的价值与意义,然而对其形成发育等方面分子机理的研究却鲜见报道,成为制约科研和生产的瓶颈问题^[4,15]。

KNOX(*KNOTTED1-like HOMEODOMAIN*)基因是最早发现的调控植物生长发育的同源盒转录因子,主要包括两大类型,第一大类型包括*STM*、*KNAT1/BP*、*KNAT2*和*KNAT6*等成员,主要在分生组织中表达,在调控植物干细胞活性、器官再生发育中发挥着重要作用。第二大类型包括*KNAT7*、*KNAT5*、*KNAT4*和*KNAT3*等成员,表达范围更加广泛,在所有的器官和组织中均有表达^[16]。已有报道表明*KNOX*基因家族能促进拟南芥的种子等器官的成熟^[16],也参与了马铃薯的球茎、蕨藜苜蓿的共生根瘤及龙舌兰和百合的珠芽(小鳞茎)等地下器官的形成过程^[17-19],因此*KNOX*基因家族对于晚香玉的鳞茎等器官的形成发育也可能具有重要的价值与意义,但是在晚香玉的鳞茎等器官中的研究尚未见到报道。

因此本研究采用更先进的PacBio测序平台对晚香玉鳞茎进行三代全长转录组测序分析,对转录本进行注释、转录因子和SSR标记等进行研究,并对鳞茎发育相关功能基因进行了初步分析,为探讨晚香玉鳞茎形成的分子机理提供参考,也为其种质资源创新及新品种培育提供帮助。

1 材料与方法

1.1 试验材料

选择课题组培育的盆栽晚香玉的栽培品种灵感,以其鳞茎为材料,将鳞茎从土壤中挖出,用清水冲洗,去除上面附着的泥土和杂质,按照起始、膨大和成熟将鳞茎发育分为3个不同时期(图1),挑选每个时期具有代表性的3~6个不同的个体作为后续试验的样本,立即在液氮冷冻,存储于-80℃冰箱。



图1 晚香玉鳞茎发育的3个不同时期

Fig.1 Three different stages of bulb development in *Polianthes tuberosa*

1.2 RNA提取、建库测序与拼接组装

使用植物RNA提取试剂盒(大连宝生物工程公司),分别提取晚香玉鳞茎发育3个不同时期的总RNA,用琼脂糖凝胶电泳和酶标仪检测所提取的RNA的完整性和质量,选择OD值(260nm/230nm)>1.8的3个不同时期的总RNA等量混合后作为一个样本,交由北京百迈克生物公司进行三代全长转录组测序。采用PacBio三代测序仪对晚香玉的鳞茎cDNA进行测序,检测后高质量的RNA使用NEBNext Single Cell/Low Input cDNA Synthesis & Amplification Module合成mRNA的全长cDNA,PCR扩增全长cDNA。对全长cDNA进行损伤修复/末端修复;连接SMRT哑铃型接头以构建文库。文库构建完成后,对文库质量进行检测,质量达到要求后方可上机测序,库检合格后,按照目标下机数据量,使用PacBio测序仪进行全长转录组测序,获得原始数据^[20-21]。

对获得的原始数据进行处理,首先获得subreads序列,进行校正后获得环形一致序列(Consensus, circular consensus)。再将序列分为全长序列(FL, full-length)、全长非嵌合序列(FLNC, full-length non-chimeric read)与非全长序列(nFL, non-full-length read),然后将全长非嵌合序列中相似的序列(即同一转录本的多个拷贝)进行聚类,得到质量较高的一致性序列即同一的转录本,然后将冗余序列去除,获得非冗余的转录本序列,即Unigene。采用TransDecoder软件进行编码区序列预测及Pfam蛋白结构域分析和功能注释^[20-21]。

1.3 功能注释及转录因子分析

使用DIAMOND软件将得到的非冗余转录本序列与NR、Swissprot、COG、KOG、KEGG等数据库比对。InterProScan利用InterPr整合的数据库分析非冗余转录本序列的GO结果,预测非冗余转录本

的氨基酸序列后,使用HMMER软件与Pfam数据库比对,获得转录本的注释信息^[20-21]。

使用iTAK软件进行植物转录因子预测,用动物转录因子数据库animalTFDB2.0对转录组中类似的动物转录因子进行鉴定,并对不同类型的转录因子进行统计^[20-21]。

1.4 微卫星(SSR)分析及非编码LncRNA及其靶向转录本预测

采用MISA(Microsatellite identification tool)对转录本序列进行分析,鉴定不同类型的SSR,并进行统计分析^[20-21]。

LncRNA与能编码蛋白质的RNA结合而起作用,因此采用CPC、CNCL、CPAT和Pfam等最广泛的编码潜能分析方法对转录本进行LncRNA的预测,把有编码潜能的转录本过滤去除,从而得到非编码的LncRNA,并取其交集作为最终预测结果;根据其mRNA进行碱基互补配对产生作用的规则,利用LnTar靶基因预测工具进行靶基因预测^[20-21]。

1.5 鳞茎发育相关的转录因子挖掘及同源盒*PtKNOX1L13*转录因子的克隆

1.5.1 鳞茎发育相关的转录因子挖掘 根据已有的研究, *WUSCHEL* 和 *KNOX* (*KNOTTED1-like HOMEODOMAIN*)等同源盒转录因子在调控植物干细胞活性和器官发育中具有重要功能和作用^[16],本研究结合晚香玉鳞茎三代全长的转录组数据,尤其是其中的转录因子和基因家族分析结果,逐条逐个搜索、查找和统计可能参与鳞茎的再生、发育和形成相关的转录因子。

1.5.2 *PtKNOX1L13*基因的克隆及生物信息学分析 根据晚香玉鳞茎三代全长转录组数据,查找目的基因的全长序列,并设计引物(表1),进行PCR扩增、电泳、凝胶回收、连接T载体、转化大肠杆菌后,

交由上海生工测序,并采用DNAMAN和MEGAX等软件进行氨基酸保守序列分析和系统发育树的构建,通过Prabi(<https://npsa-prabi.ibcp.fr/>)在线网站预测和分析二级结构模型^[22]。PCR扩增体系为:

cDNA 2 μ L,正、反向引物各 1.5 μ L, 2 \times Flash PCR Master Mix (康为世纪) 25 μ L, ddH₂O 20 μ L; 反应程序: 95 $^{\circ}$ C 预变性 5min; 95 $^{\circ}$ C 变性 20 s, 63 $^{\circ}$ C 退火 30 s, 72 $^{\circ}$ C 延伸 45 s, 35个循环; 72 $^{\circ}$ C 延伸 1 min。

表1 引物名称和序列

Table 1 Primers name and sequence

引物名称 Primer name	引物序列(5'→3') Primer sequence(5'→3')	用途 Usage
<i>PtKNIL13F</i>	ATGTCTTACCACTCCCACATCGACCTC	全长克隆
<i>PtKNIL13R</i>	TCACCTTTTTCGCTTGCTCTTAACTGC	
<i>PtKN130F</i>	GGTCCCACCTGGCTCAACAG	相对定量
<i>PtKN130R</i>	CCATGATGGAGTCGCCGGAG	
<i>PtGAPDH-180-F</i>	GCTGCTAAGGCTGTTGGGAAAG	内参引物
<i>PtGAPDH-180-R</i>	TCCCTTCATGCTCCCTCAG	
<i>PtKNIL13F</i>	cagctcggtaccgggatccATGTCTTACCACTCCCACAT	表达载体构建
<i>PtKNIL13R</i>	catgtcgactctagaggatccCCTTTTTCGCTTGCTCTTAACTGC	

小写字母表示植物表达载体上的序列

Lowercase letters represent sequences on plant expression vectors

1.5.3 *PtKNOX1L13* 基因的实时荧光定量 PCR (real-time PCR) 利用软件 Primer-premier 5 (<http://frodo.wi.mit.edu/primer5/>) 设计 real-time PCR 引物(表1)。扩增产物的长度范围在 100~250 bp。反应体系和条件参照 SYBR Premix Ex TaqTM kit 说明书,以 *GAPDH* 基因作为晚香玉表达分析的内参,设置重复 3 次。通过分析熔解曲线对扩增的特异性进行评估,使用 $2^{-\Delta\Delta Ct}$ 法来确定基因的相对表达量^[23]。

1.5.4 *PtKNOX1L13* 基因的瞬时转化 将目的基因连接到植物表达载体 *pCYH05251* 上(表1),将空载体和目的基因载体均用冻融法导入根瘤农杆菌 *EHA105* 中,对白花虎眼万年青的离体叶片进行异源瞬时转化^[24-26]。试验分为两组,以空载体为对照,以目的基因载体为处理,每组处理 30 个叶片,并进行 3 次生物学重复。在转化 48~72 h 后,每组随机抽取 6 个叶片,提取 RNA 进行相对荧光定量表达分析^[24-26],同时将材料在组培室内 25 $^{\circ}$ C 培养 15~20 d 后,观察分析叶片切口产生珠芽的情况,用数码相机进行拍照分析。

2 结果与分析

2.1 晚香玉鳞茎全长转录组数据分析

晚香玉鳞茎的发育过程可分为 3 个时期,第 1 个时期是起始期,刚刚开始启动,为白色的小嫩芽;第 2 个时期是膨大期,逐渐膨大变绿,棕褐色的外皮

也慢慢开始形成;第 3 个时期是成熟期,稍肥大的圆锥形的鳞茎基本形成,并外被多重外皮(图1)。利用单分子长读数测序技术(SMRT, single-molecule long read sequencing technology),Pacbio 高通量测序平台对 cDNA 文库进行测序共获得 49.69 Gb 数据量,根据 full passes 大于 3 且序列准确性大于 0.9 的原则,从原始序列中提取环形一致序列(CCS),并对 CCS 序列进行矫正,CCS 序列数为 508012 条,CCS 的碱基总数为 887701911 个,CCS 序列平均长度为 1747 个碱基,508012 条 CCS 序列中得到全长非嵌合序列 402032 条,将全长非嵌合序列中的相似序列(即同一转录本的多个拷贝)进行聚类得到 137247 条一致的转录本序列(Consensus isoform),进行合并整理后获得 81719 条非冗余转录本序列(Unigene),采用 TransDecoder 软件进行转录本序列的开放阅读框(ORF, open reading frame)及其编码氨基酸的预测,共获得 ORF 有 67972 条,其中完整 ORF 36215 条。

2.2 晚香玉鳞茎全长转录组功能注释

2.2.1 转录本 NR 注释 利用 NCBI 中非冗余蛋白质数据库,包含了各个蛋白质数据库及蛋白质数据信息,通过序列比对寻找同源物种,NR 比对结果同源物种分布如下:芦笋最多(39226 条,占被注释基因总数的 61.74%),其次为油棕(4304 条,占被注释基因总数的 6.77%),再次为海枣(3379 条,占被注释基因总数的 5.32%),粗柄象脚蕉最少(484 条,占被

注释基因总数的0.76%),其他物种有12169条,占被注释基因总数的19.15%。

2.2.2 转录本GO分类 通过GO数据库比对分析,共有203412个基因被注释,分为三大类:生物过程、细胞组分和分子功能。生物过程的基因最多,其次为分子功能,细胞组分最少。在第一大类生物过程中,细胞过程(Cellular process)最多,其次为代

谢过程(Metabolic process),再次为细胞调控(Biological regulation);在第二大类分子功能中,结合(Binding)最多,其次为催化活性(Catalytic activity),再次为运输活性(Transporter activity);第三大类细胞组分中,细胞结构体最多(Cellular anatomical entity),其次为细胞内(Intracellular),再次为蛋白复合体(Protein-containing complex)(图2)。

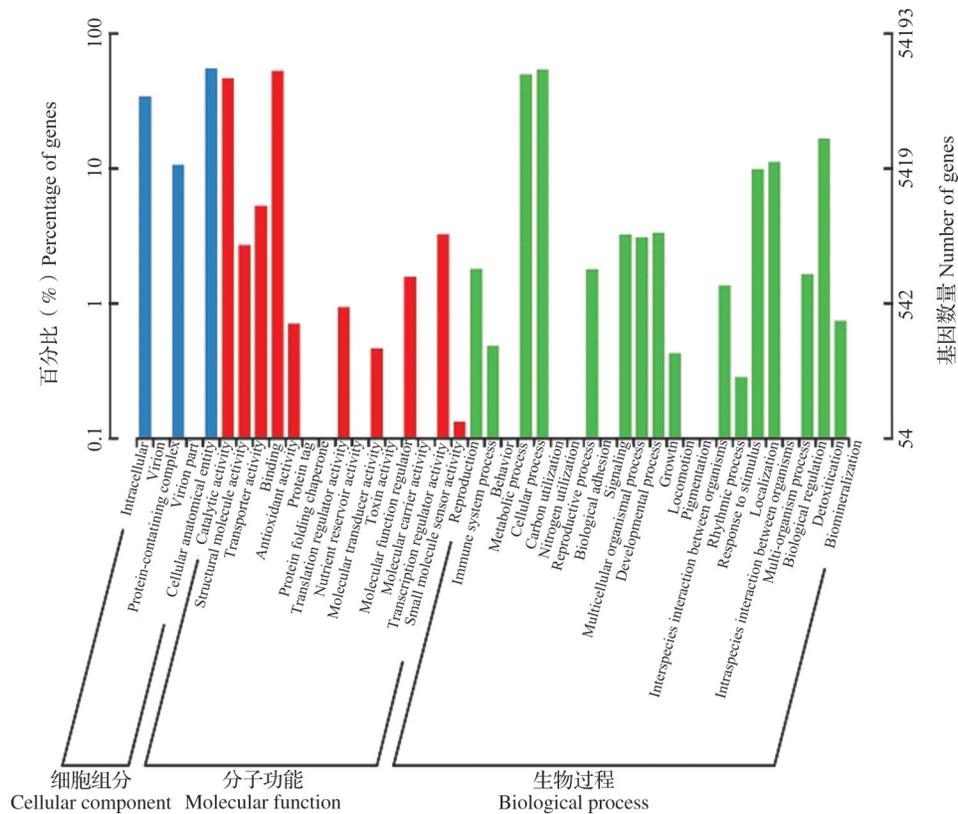


图2 GO功能注释

Fig.2 GO function annotation

2.2.3 COG分析 共有22676个基因被COG注释,共分为25类,其中最多的是翻译后修饰、蛋白转换和伴侣(Posttranslational modification, protein turnover, chaperones),其次是糖类运输与代谢(Carbohydrate transport and metabolism),再次为一般功能预测(General function prediction only),最少的是RNA加工与修饰(RNA processing and modification),仅包括4个基因,还有未知功能基因348个。1268个基因与次生代谢合成、运输和分解有关,这些基因可能与晚香玉鳞茎中次生代谢产物和药用成分的生物合成、运输、积累和分解等重要作用有关(图3)。

2.2.4 KEGG通路分析 通过KEGG数据库比对分析,共有58878个基因被注释到了139条途径,其中代谢途径(Metabolic pathways)分布的基因最多,其次是次生代谢产物生物合成(Biosynthesis of

secondary metabolites)、碳代谢(Carbon metabolism)、植物激素信号途径(Plant hormone signal transduction)、植物病原菌互作(Plant-pathogen interaction)和氨基酸合成(Biosynthesis of amino acids)等(表2)。另有基本转录因子(Basal transcription factors)198个,可能参与了代谢和次生代谢产物的调控过程。

与次生代谢密切相关的还有丁酸甲酯(Butanoate metabolism, 代谢通路中基因数量为130)、二苯乙烯及二芳基庚酸类和姜酚合成路径(Stilbenoid, diarylheptanoid and gingerol biosynthesis, 代谢通路中基因数量为134)、萜类相关的代谢途径(萜类骨架、单萜类合成、二萜合成及倍半萜和三萜类合成等, Terpenoid backbone biosynthesis, monoterpene biosynthesis, diterpenoid biosynthesis, sesquiterpene and triterpene biosynthesis, and so on, 代谢通路中

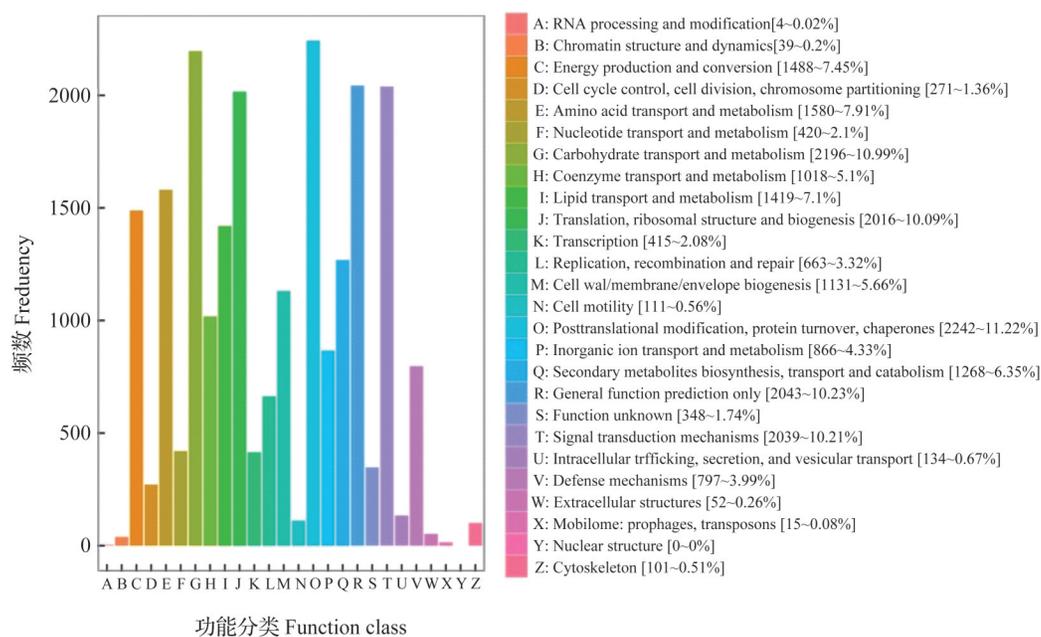


图3 COG功能分类

Fig.3 COG function annotation

表2 KEGG通路富集

Table 2 KEGG pathway enrichment

序号 No.	途径 Pathway	途径编码 Pathway ID	注释基因数量 Number of annotation genes	比例(%) Percent
1	代谢途径	ko01100	11112	18.87
2	次生代谢途径	Ko01110	6172	10.48
3	碳代谢	ko01200	1730	2.94
4	植物激素信号途径	ko04075	1347	2.29
5	植物病原菌互作	ko04626	1323	2.25
6	内质网蛋白加工	ko04141	1305	2.22
7	氨基酸合成	ko01230	1277	2.17
8	核糖体	ko03010	1232	2.09
9	MAPK 信号途径	ko04016	963	1.64
10	RNA 运输	ko03013	949	1.61
11	内吞作用	ko04144	897	1.52
12	泛素介导的蛋白质降解	ko04120	823	1.40
13	氧化磷酸化	ko00190	773	1.31
14	氨糖和核酸糖的代谢	ko00520	666	1.13
15	苯丙酸合成	ko00940	653	1.11
16	mRNA 监控途径	ko03015	646	1.10
17	乙醛酸及二羧酸代谢	ko00630	629	1.07
18	丙酮酸盐	ko00620	590	1.00
19	碳固定和光合作用	ko00710	588	0.99
20	RNA 降解	ko03018	509	0.86

基因数量为 290)、生物碱有关的代谢途径(吲哚生物碱、异喹啉生物碱以及托烷哌啶和吡啶类生物碱等, Indole alkaloid biosynthesis, isoquinoline alkaloid biosynthesis, tropane, piperidine and pyridine alkaloid biosynthesis, and so on, 代谢通路中基因数量为 348)、油菜素甾醇(Brassinosteroid biosynthesis, 代谢通路中基因数量为 65)、苯并噁唑啉酮(Benzoxazinoid biosynthesis, 代谢通路中基因数量为 27)等药用成分的代谢途径;这与已有报道中,晚香玉的鳞茎中含有单糖(Monosaccharides)、螺甾皂苷(Spirostanol saponin)、类固醇糖苷类(Steroid glycosides)、呋甾烷醇类(Furostanols)、呋甾烷醇皂苷(Furostanol saponins)及生物碱(Alkaloid)等次生代谢成分,具有抗氧化、抗癌杀菌等功能^[1-3]相互验证,在医药化工等方面有潜在的开发价值。

2.3 转录组中转录因子分析

约有 4903 个转录因子参与了晚香玉鳞茎及其次生代谢产物的形成过程,其中类受体激酶 *RLK* (receptor-like kinase) 基因家族最多,为 1049 个。在类受体激酶 *RLK* 中的 *RLK Pelle DLSV* 家族有 173 个,其次为 *bHLH* 169 个,再次为 *bZIP* 146 个,也包括 *MYB*、*WRKY*、*C2H2* 和 *NAC* 等基因家族,最少的是 *SET* 基因 68 个、*CAMK_CAMKL-CHK1* 转录因子 66 个和 *PHD* 基因 65 个(图 4)。

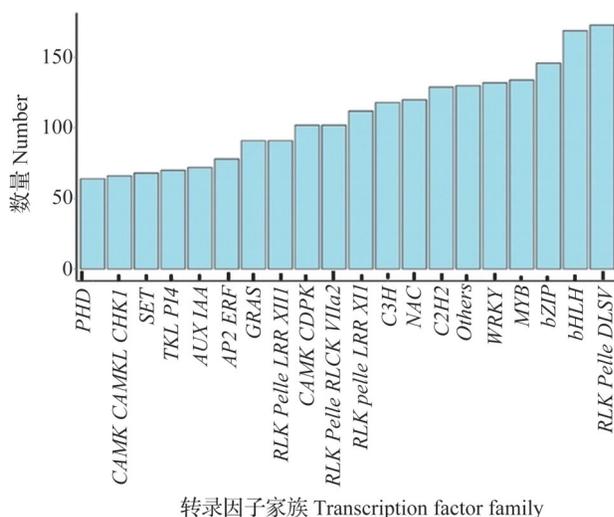


图 4 转录因子家族分类和数量

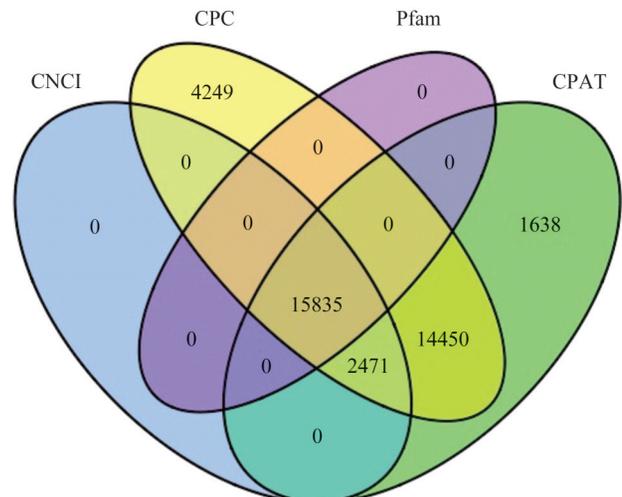
Fig.4 Types and quantities of transcription factors family

2.4 转录组中微卫星 SSRs 位点的分布

对含有 130666731 个碱基的 52370 条序列进行检索,共获得 41240 条 SSR 序列,含有 SSR 的转录本数量为 34322 条,含有 1 个以上 SSR 的序列有 9296 条。从数量和分布密度(每 Mb 数量)来看,单碱基最多最密有 19777 个,其次为双碱基 13741 个,再次为三碱基 7072 个,最少的是四碱基、六碱基和五碱基,分别为 366 个、146 个和 138 个,含有复合型 SSR 6918 个。

2.5 长链非编码 RNA

用 4 种分析方法预测 LncRNA,得到的结果取交集,结果共获得 15835 个 LncRNA 序列,其中有 13406 个与 266746 个转录本匹配,平均每个 LncRNA 与 19.91 个转录本产生互作(图 5)。



数字是将以上 4 种分析软件鉴定得到的非编码转录本采取逐级筛选的方法得到的交集,即共有的转录本的数量

The number is the intersection obtained by using a stepwise screening method to identify non coding transcripts identified by the above four analysis software, that is, the number of shared transcripts

图 5 根据 4 种筛选方法预测的 LncRNA 的维恩图

Fig.5 Venn diagram of predicted LncRNA by four kinds of methods

2.6 鳞茎发育相关的转录因子的挖掘

根据晚香玉鳞茎三代全长转录组数据,并结合转录因子和不同基因家族的分析结果,鳞茎形成发育相关的主要转录因子有 *MYB*、*AP2*、*GRAS* 和 *WOX* (Homeodomain) 等基因家族(表 3)。

表3 鳞茎发育相关的转录因子的挖掘

Table 3 Exploration of transcription factors related to bulb development

序号 No.	转录因子家族 Transcription factor	主要功能基因 Main functional genes	主要功能 Main function	转录本 数量 Number of transcripts
1	<i>MYB</i>	<i>MYB115</i>	器官再生发育、次生代谢及抗逆等	176
2	<i>AP2</i>	<i>ERF</i> (Ethylene response factor)/ <i>BABYBOOM</i>	器官再生发育及抗逆等	90
3	<i>HOX</i>	<i>BELL</i>	叶片等侧生器官发育	49
4	<i>GRAS</i>	<i>GRAS</i>	赤霉素响应因子	96
5	<i>AT hook motif</i>	AT-hook motif nuclear-localized protein	器官发育	2
6	Homeobox KN domain	<i>KNOX</i>	器官发育	125
7	Homeobox associated leucine zipper	Homeobox-leucine zipper protein <i>HOX</i>	器官发育	50
8	Homeodomain	<i>WOX</i>	器官发育	168

2.7 鳞茎发育相关的同源盒 *PtKNOX1L13* 转录因子的克隆及功能分析

2.7.1 *PtKNOX1L13* 基因的克隆 从晚香玉鳞茎三代全长转录组数据中发现了1个编号为Ptbulbtranscripti-125603的转录本,该转录本全称为*KNOTTED1-LIKE HOMEBOX13*,可能是*KNOX*基因在晚香玉中的同源基因,经过初步分析将其命名为*PtKNOX1L13*基因,该目的基因全长有1798个碱基,包括5'端非编码区178个碱基,3'端非编码区453个碱基,及中间完整编码区1167个碱基,共编码388个氨基酸,在NCBI上Genbank提交,获得登录号(The GenBank accession number PP066245)。将其转录本编号在长链非编码RNA中进行搜索发现其与编号为Ptbulb_transcript_49465、Ptbulb_transcript_82205和Ptbulb_transcript_54645等LncRNA存在互作,可能受到上述非编码RNA的调控。

2.7.2 *PtKNOX1L13* 氨基酸保守区域分析 经过DNAMAN软件分析发现*PtKNOX1L13*基因包括*KNOXI*、*KNOXII*、*ELK*和*HD*等*KNOX*基因家族典型的保守区域^[16],属于*KNOX*基因家族的同源异形盒转录因子(图6)。

2.7.3 *PtKNOX1L13* 系统发育树构建 系统发育树可分为两大类群,第一大类群主要是单子叶植物*KNIL13*基因组成,先由单子叶植物的晚香玉*PtKNIL13*基因和油棕*EgKNIL13*基因,与双子叶植物的蒺藜苜蓿*MtKNOX3*基因、拟南芥*AtKNAT3*以及棉花*GnKNL1*基因聚合成为一个分支,说明其亲缘进化关系较近,功能上也可能有相似的地方,然后海枣、菠萝、生姜等单子叶植物的*KNIL13*基因也

陆续融入这个分支中,组成第一大类群;第二类群主要是双子叶植物的*KNIL13*基因,主要包括梅花、白梨和欧洲桉木等;这与传统的植物分类和亲缘进化演变基本一致(图7)。

2.7.4 *PtKNOX1L13* 基因的相对定量表达分析 对目的基因在叶片、花梗(花茎)、花芽(花茎顶端分生组织)、花蕾(花苞)、初开(花朵刚刚开放)、盛开(花朵完全开放)、起始鳞茎、膨大鳞茎和成熟鳞茎等器官中的表达量进行分析,结果表明,*PtKNOX1L13*基因在鳞茎发育的3个时期(起始、膨大和成熟)的表达量均比较高,尤其是在成熟的鳞茎中表达量最高,在盛开的花朵和花梗等其他较成熟的器官中表达量也较高,在叶片中的表达量最低(图8)。

2.7.5 *PtKNOX1L13* 基因的功能分析 为了更好地分析*PtKNOX1L13*基因是否参与了鳞茎的形成,将其连接到植物表达载体*pCYH05251*上并采用瞬时过量表达的方法来验证其功能。因晚香玉鳞茎很难进行转化分析,而与其亲缘进化关系较近的且同为天门冬科虎眼万年青属(春慵花属)的白花虎眼万年青(*Ornithogalum thyrsoides*)叶片切口较易产生珠芽(小鳞茎),便于进行观察分析,因此对其离体叶片进行瞬时转化^[24-26]。在48~72 h后,分别取对照组和处理组的叶片进行荧光定量分析表明,处理组的基因表达量急剧升高,而对照组完全为零,有极其显著的差异(图9);再于15~20 d后进行观察分析表明,空载处理的对照组多数尚处于启动状态,而瞬时过表达的处理组叶片切口已经产生明显的小绿球(珠芽开始成形)(图9),表明*PtKNOX1L13*基因能促进珠芽和鳞茎的形成和发育。

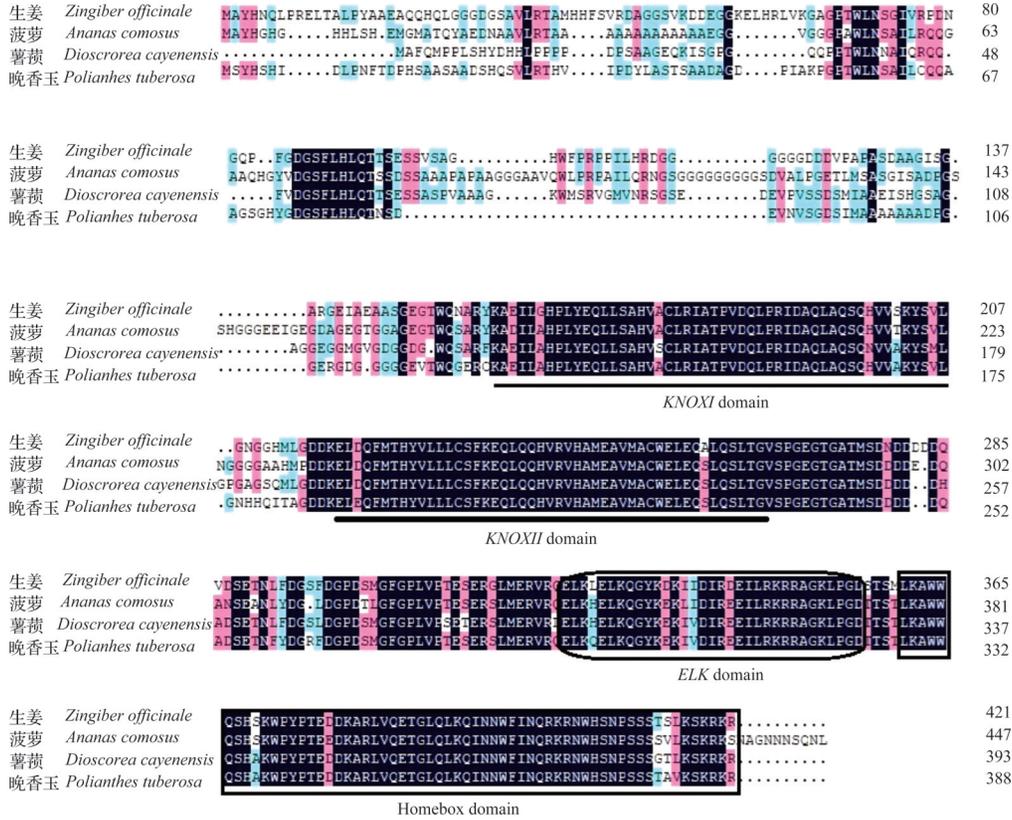


图 6 *PtKNOX1L3* 基因保守区域分析

Fig.6 Sequence alignment analysis of the *PtKNOX1L3* gene

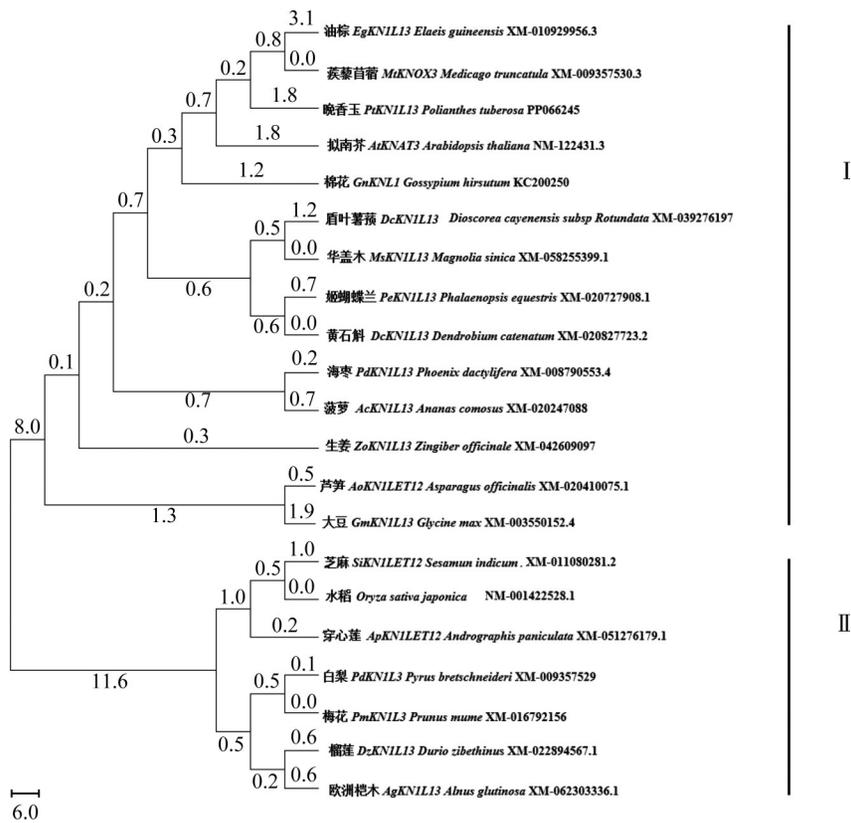


图 7 *PtKNOX1L3* 系统发育树

Fig.7 Phylogenetic of the *PtKNOX1L3* protein

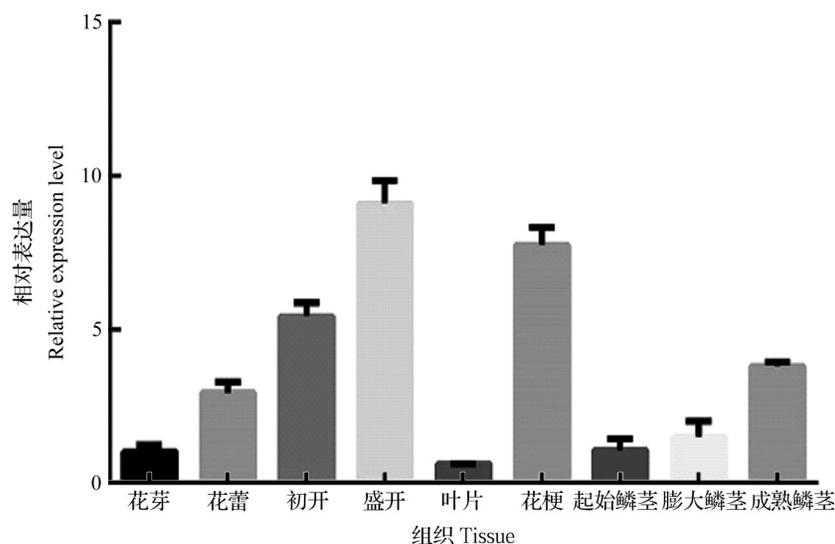
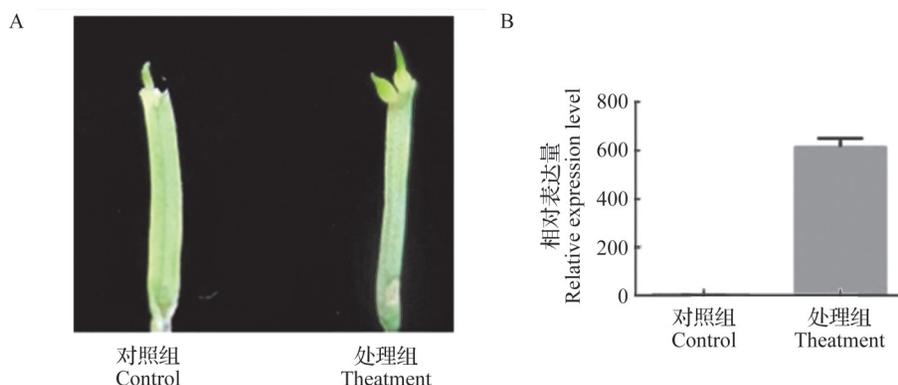


图8 *PtKNIL13* 基因相对定量表达分析
Fig.8 *PtKNIL13* gene relative expression



对照组为 *pCYH05251* 空载体瞬时表达, 处理组为 *PtKNIL13* 的 *pCYH05251* 载体瞬时过表达

Control was transient overexpression with *pCYH05251* empty vector, treated was transient overexpression with *pCYH05251-PtKNIL13*

图9 *PtKNIL13* 基因对白花虎眼万年青叶片切口产生珠芽的影响(A)及相对定量表达分析(B)

Fig.9 The effect of *PtKNIL13* genes on the formation of bulbils in the incision of *Ornithogalum thyrsoides* leaves *in vitro* (A) and relative expression analysis of *PtKNIL13* in *O.thyrsoides* (B)

3 讨论

晚香玉是鳞茎球根花卉中的翘楚, 色香型味俱佳, 是难得的多功能多用途花卉。尽快选育出新品种是提高核心竞争力和创造更高经济效益的重要途径, 然而晚香玉目前仍多采用杂交育种和诱变育种, 存在技术落后、育种周期长和突变位点有限等问题, 不利于新品种的选育和高质量发展^[4]。生物育种越来越成为花卉育种的主要途径^[5-6], 而晚香玉基因组序列未知, 转录组测序成为获取其遗传信息的重要手段。已有研究用二代转录组测序技术对其花香等进行了分析, 但是存在读序短和拼接易出错等问题, 效果并不理想^[6-10]。

“根深才能叶茂, 球好才能花香”, 鳞茎, 也是种

球, 对于种质创新和种植生产的重要价值得到了深刻认识和充分重视^[1]。本研究利用三代高通量测序平台对晚香玉的鳞茎进行了全长转录组分析, 获得 49.69 Gb 数据量, 从原始序列中提取环形一致序列 (CCS), 508012 个 CCS 序列平均长度为 1747 个碱基, 得到全长非嵌合序列 402032 条, 聚类得到 137247 条一致的转录本序列 (Consensus isoform), 合并整理后共获得 81719 条非冗余转录本序列 (Unigene) 和 36215 条完整 ORF, 转录因子 4903 个, SSR 序列 41240 个, LncRNA 序列 15835 个。NR 注释的同源物种分布中, 注释到芦笋中的最多, 占 61.74%, 其次为油棕和海枣, 分别占 6.77% 和 5.32%, 然后为菠萝、小果野蕉、黄石斛、葡萄、野蕉、虎眼万年青和粗柄象脚蕉等物种, 说明晚香玉可能

与这些物种的亲缘进化关系较近,注释到其他物种的约占 19.15%,这与晚香玉二代转录组分析的结果非常相似^[6-10]。

KEGG 分析表明,晚香玉鳞茎中存在吡啶生物碱和油菜素合成等代谢途径,这与晚香玉鳞茎中有多酚、皂苷和甾醇等抗氧化、抗癌杀菌等药用成分^[1-3]比较吻合,充分说明了传统医学中,晚香玉具有清热解毒的功效、能够治疗痈疮肿毒,是有物质基础和分子依据的,其重要成分需要深入分离、验证和研发。

本研究发现约有 4903 个转录因子和 15835 个 LncRNA 序列参与了晚香玉鳞茎及其次生代谢产物的形成过程,类受体激酶 *RLK* (receptor-like kinase) 基因家族最多,为 1049 个,在类受体激酶 *RLK* 中的 *RLK Pelle DLSV* 家族有 173 个,其次为 *bHLH* 有 169 个,再次为 *bZIP* 有 146 个,也包括 *MYB*、*WRKY*、*C2H2* 和 *NAC* 等基因家族,可为挖掘抗逆性、再生发育、次生代谢和药用成分的关键基因奠定基础,为研究鳞茎形成的基因调控网络提供了参考。

将三代全长转录组与相关研究结果进行综合分析,晚香玉的鳞茎形成可能是 *BABYBOOM* (*AP2* transcription factor) 转录因子和同源盒转录因子包括 *WOX* (*WUSCHEL*-related homeobox) 和 *KNOX1* 等基因受到细胞分裂素和生长素的调控使得表达量显著提高,从而激活了晚香玉植株中的植物干细胞活性及形成了干细胞和胚性干细胞团,并在 *GRAS*、*MYB*、*YABBY* 和 *BELL* 等转录因子的作用下,进一步再生、分化成为薄壁细胞、表皮细胞等其他类型的细胞和组织,鳞茎也逐渐发育膨大并最终形成的^[16]。

同源盒转录因子一直被认为是参与鳞茎等器官形成的重要功能基因,已有研究表明,拟南芥的 *AtKNAT3* 和棉花的 *GhKNL1* 等基因的表达量提高可以调控植物细胞中木质化积累、纤维延长和次生壁的形成来促进种子及其纤维等地上器官的成熟^[16],而且紫花苜蓿的 *MtKNOX3* 基因在根瘤菌感染其根系和结节逐渐形成时,表达量显著提高,在其根系中异位表达可产生结节状结构(无根瘤菌侵染),也可促进根瘤结节等地下器官的形成^[17]。

本研究表明, *PtKNOXIL13* 基因与拟南芥的 *AtKNAT3*、棉花的 *GnKNL1* 基因和紫花苜蓿的 *MtKNOX3* 基因的亲缘进化关系非常相近,聚为一类,且 *PtKNOXIL13* 基因不但在晚香玉初开和全开的花器官和花梗等较成熟的地上器官中表达量较

高,在其地下鳞茎的起始、膨大和成熟等 3 个时期表达量均较高,特别是在鳞茎的成熟时期表达量最高;结合对白花虎眼万年青叶片进行瞬时转化,发现其能够促进珠芽(小鳞茎)的形成和发育;说明 *PtKNOXIL13* 基因也有可能是通过参与晚香玉植物细胞的纤维化、木质化及次生细胞壁的产生来促进鳞茎的形成,是调控晚香玉器官发育的重要功能基因。

本研究有助于晚香玉新基因的重要功能基因挖掘,深入探讨鳞茎形成的分子机制和加快晚香玉的繁育、种质创新及产业化开发,也促进了植物器官发育分子生物学的研究。

参考文献

- [1] Sadhukhan R, Chowdhuri T K, Datta S K, Tuberose (*Polyanthes tuberosa* Linn./*Agave amica*) //Datta S K, Gupta Y C. Floriculture and ornamental plants. handbooks of crop diversity: Conservation and use of plant genetic resources. Singapore: Springer, 2021: 307-357
- [2] Rahmatullah R N, Jannat K, Islam M, Rahman T, Jahan R, Rahmatullah M. A short review of *Polianthes tuberosa* L. considered a medicinal plant in Bangladesh. Journal of Medicinal Plants Studies, 2019, 7(1): 1-4
- [3] Copetta, A, Marchioni I, Mascarello C, Pistelli L, Cambournac L, Dimita R, Ruffoni B. *Polianthes Tuberosa* as edible flower: *in vitro* propagation and nutritional properties. International Journal of Food Engineering, 2020, 6(2): 57-62
- [4] Kumar M, Chaudhary V, Kumar M, Sirohi U, Yadav M K. Application of conventional and mutation approaches in genetic improvement of Tuberose (*Polianthes tuberosa* L.): A review on recent development and future perspectives. International Journal of Agriculture and Rural Development, 2021, 14(3): 277-297
- [5] Bautista-Montes E, Hernández-Soriano L, Simpson J. Advances in the micropropagation and genetic transformation of *Agave* Species. Plants (Basel), 2022, 11(13): 1757
- [6] Singh K B M, Madhavan J, Chandra S, Rao U, Mandal P K. In planta transformation of *Polianthes tuberosa* for concomitant knockdown of *flp-1*, *flp-12* and *flp-18* genes induced root-knot nematode resistance. Scientia Horticulturae, 2023(311): 111764
- [7] Fan R, Chen Y, Ye X, Wu J, Lin B, Zhong H. Transcriptome analysis of *Polianthes tuberosa* during floral scent formation. PLoS ONE, 2018, 5, 13(9): e0199261
- [8] Madhavan J, Jayaswal P, Singh K B M, Rao U. Identification of putative flowering genes and transcription factors from flower *de novo* transcriptome dataset of tuberose (*Polianthes tuberosa* L.). Data Brief, 2018, 22(20): 2027-2035
- [9] Madhavan J, Jayaswal P, Singh K B M, Thakur P K, Rao U.

- Identification of SSR and miRNA from transcriptome of tuberose. *Indian Journal of Horticulture*, 2019, 76(1):141-147
- [10] Singh K B M, Jayaswal P, Chandra S, M J, Mandal P K. Comparative transcriptome profiling of *Polianthes tuberosa* during a compatible interaction with root-knot nematode *Meloidogyne incognita*. *Molecular Biology Reports*, 2022, 49(6):4503-4516
- [11] 赵陆滢, 曹绍玉, 龙云树, 张应华, 许俊强. 全长转录组测序在植物中的应用研究进展. *植物遗传资源学报*, 2019, 20(6):1390-1398
- Zhao L Y, Cao S Y, Long Y S, Zhang Y H, Xu J Q. Applications and research progresses of full-length transcriptome sequencing in plants. *Journal of Plant Genetic Resources*, 2019, 20(6):1390-1398
- [12] Zamin M, Muhammad A, Jan I, Ullah H, Shah S, Amin M, Rashid H U. Production of tuberose (*Polianthes tuberosa* L.) as affected by bulb size and planting medium. *Sarhad Journal of Agriculture*, 2020, 36(4):1156-1161
- [13] Barghout N, Chebata N, Moumene S, Khennouf S, Gharbi A, Hadi D E. Antioxidant and antimicrobial effect of alkaloid bulbs extract of *Polianthes tuberosa* L. (Amaryllidaceae) cultivated in Algeria. *Journal of Drug Delivery and Therapeutics*, 2020, 10(4):44-48
- [14] Salomé-Abarca L F, Márquez-López R E, López M G. *Agave amica* a potential model for the study of agavins metabolism. *Scientific Reports*, 2023, 13(1):19888
- [15] Kumar M, Sirohi U, Malik S, Kumar S, Ahirwar G K, Chaudhary V, Yadav M K, Singh J, Kumar A, Pal V, Prakash S. Methods and factors influencing *in vitro* propagation efficiency of ornamental tuberose (*Polianthes* Species): A systematic review of recent developments and future prospects. *Horticulturae*, 2022, 8(11):998
- [16] Jia P, Wang Y, Sharif R, Dong QL, Liu Y, Luan HA, Zhang X M, Guo S P, Qi G H. *KNOTTEDI-like homeobox (KNOX)* transcription factors - Hubs in a plethora of networks: A review. *International Journal of Biological Macromolecules*. 2023, 253(3):126878
- [17] Azarakhsh M, Kirienko A, Zhukov V, Lebedeva M, Dolgikh E, Lutova L. *KNOTTEDI-LIKE HOMEOBOX 3*: A new regulator of symbiotic nodule development. *Journal of Experimental Botany*, 2015, 66(22):7181-7195
- [18] Abraham-Juarez M J, Martinez-Hernandez A, Leyva-Gonzalez M A, Herrera-Estrella L, Simpson J. Class I *KNOX* genes are associated with organogenesis during bulbil formation in *Agave tequilana*. *Journal of Experimental Botany*, 2010, 61(14):4055-4067
- [19] Zhang Y, Zeng Z, Yong Y, Lyu Y. Hormonal regulatory patterns of *LaKNOXs* and *LaBEL1* transcription factors reveal their potential role in stem bulblet formation in LA Hybrid Lily. *International Journal of Molecular Sciences*, 2021, 22(24):13502
- [20] 任毅鹏, 张佳庆, 孙瑜, 吴振峰, 阮吉寿, 贺秉军, 刘国卿, 高山, 卜文俊. 基于 PacBio 平台的全长转录组测序研究. *科学通报*, 2016, 61(11):1250-1254
- Ren Y P, Zhang J Q, Sun Y, Wu Z F, Ruan J S, He B J, Liu G Q, Gao S, Bu W J. Full-length transcriptome sequencing on PacBio platform. *Chinese Science Bulletin*, 2016, 61(11):1250-1254
- [21] 梅瑜, 李向荣, 蔡时可, 顾艳, 王继华. 药食同源植物甘葛藤的全长转录组分析. *华北农学报*, 2021, 36(5):10-17
- Mei Y, Li X R, Cai S K, Gu Y, Wang J H. Full-length transcriptome analysis of a homology of medicine and food of *Pueraria thomsonii*. *Acta Agriculturae Boreali-Sinica*, 2021, 36(5):10-17
- [22] 周雅, 张祥, 赵权, 孙建强, 王晓波, 邱丽娟. 大豆 DMP 基因的生物信息学及基因多样性分析. *植物遗传资源学报*, 2023, 24(6):1744-1754
- Zhou Y, Zhang X, Zhao Q, Sun J Q, Wang X B, Qiu L J. Bioinformatics and gene diversity analysis of the *DMP* genes in soybean. *Journal of Plant Genetic Resources*, 2023, 24(6):1744-1754
- [23] Livak K J, Schmittgen T D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods*, 2001, 25:402-408
- [24] 殷小雨, 何国仁, 毕蒙蒙, 唐玉超, 郝春莲, 渠雨潇, 郝泽慧, 徐雷锋, 胡凤荣, 杨盼盼, 明军. 卷丹侧生器官边界域基因 *LILBD18* 的克隆和功能分析. *园艺学报*, 2023, 50(10):2117-2127
- Yin X Y, He G R, Bi M M, Tang Y C, Hao C L, Qu Y X, Hao Z H, Xu L F, Hu F R, Yang P P, Ming J. Cloning and functional analysis of *LiLBD18* in *Lilium lancifolium*. *Acta Horticulturae Sinica*, 2023, 50(10):2117-2127
- [25] 姜福星, 陈其兵, 黄远祥, 周鹏. ZL201610431551.2 一种白花虎眼万年青的组织培养快速繁殖方法. 北京: 中华人民共和国国家知识产权局, 2018
- Jiang F X, Chen Q B, Huang Y X, Zhou P. ZL201610431551.2 A rapid propagation method for tissue culture of *Ornithogalum thyrsoides*. Beijing: China National Intellectual Property Administration of the People's Republic of China, 2018
- [26] Tripathi P K, Ayzenshtat D, Kumar M, Zemachn H, Yedidia I, Bocobza S E. An efficient and reproducible *Agrobacterium*-mediated genetic transformation method for the ornamental monocotyledonous plant *Ornithogalum dubium* Hoult. *Plant Growth Regulation*, 2023, 101:201-214