

## 基于番红花转录组的碳氮代谢基因挖掘

周琳, 杨柳燕, 蔡友铭, 王桢, 张永春\*

上海市农业科学院林木果树研究所, 上海市设施园艺技术重点实验室, 上海201403

**摘要:** 番红花球茎是其开花和新球茎形成的基础, 促进番红花种球中营养物质积累、提高番红花繁殖效率是解决国内番红花产量供不应求的有效途径, 而番红花种球生长和繁殖与生长发育期间碳氮代谢途径密切相关, 深入研究番红花生长发育机理, 尤其是碳氮代谢途径是番红花栽培和繁育体系优化的前提。通过对番红花已有转录组数据进行从头组装, 得到62 105条unigene序列。注释结果表明, 仅有55.89%序列能够注释到公共数据库, 注释到GO分类和KEGG注释的unigene分别为36 661和14 185条。通过生物信息学分析, 挖掘番红花多个参与碳氮代谢的重要基因和转录因子, 包含了蔗糖合成酶基因、蔗糖磷酸合成酶基因、磷酸烯醇式丙酮酸羧化酶基因、甘油醛-3-磷酸脱氢酶基因、谷氨酸合成酶基因和碳酸酐酶基因等, 以及转录因子GATA和DOF。研究结果可为番红花生长发育功能基因的发掘与利用研究提供参考。

**关键词:** 番红花; 转录组; 碳氮代谢; 生物信息学分析

番红花(*Crocus sativus*), 又名西红花、藏红花, 为鸢尾科(Iridaceae)番红花属(*Crocus*)多年生球茎花卉, 主要分布于地中海和亚洲西南部地区(Gresta等2008), 且已有数千年的栽培历史。番红花的干燥柱头, 称为藏红花, 既是名贵香料, 广泛应用于高级美容化妆品、食品、染料工业等行业; 在许多国家也是重要药材, 具有凉血解毒、活血化瘀、散郁开结等功效, 并于2001年被国家中医药管理局列为重点发展的中药材品种(姚冲等2013)。随着研究深入, 番红花中含有藏红花素、藏红花醛、藏红花酸等多种重要次生代谢产物, 具有抗肿瘤、防癌、抗氧化和抗诱变作用, 具有广谱抗癌活性且低毒性, 有可能成为未来理想抗癌药物之一(李玲蔚2013), 具有广阔的市场前景。

由于番红花17万朵花才能得到1 kg的柱头; 且因其花期短、采收费时费力、易感染病毒等问题, 导致藏红花生量产量极低。我国先后从日本和德国引种繁殖番红花, 但番红花仅依靠球茎进行自然无性繁殖, 繁殖系数较低, 限制其推广, 最终导致国内藏红花资源严重短缺, 难以满足市场需求, 需每年从国外大量进口藏红花。虽然, 自1987年至今, 国内外已开展大量番红花组织培养探索(王桢等2019), 但均因培养周期过长和成本过高而不能应用于产业。为提高番红花繁殖系数, 国内外在母球茎大小与留芽数目(姚冲等2017)、栽培方式(Gresta等2016)、最佳种植密度和深度(Karra

等2017)、施肥类型和方式(Behnia等1999)等方面进行了大量摸索。然而, 研究结果间存在较大的差异, 例如不同国家、区域报道番红花最佳栽培密度存在较大差异(Karra等2017; Nasseer等2018), 主要原因之一是番红花生长发育相关的分子生物学的研究基础较为薄弱。

利用分子生物学手段对番红花生长发育进行更深层次的研究, 对于指导番红花栽培和繁育具有重要的科学意义。因此, 本研究通过分析已有的番红花球茎、萼片、叶、柱头和雄蕊组织转录组数据(Baba等2015; Jain等2016), 重点挖掘碳氮代谢途径相关基因, 为阐明生长发育提供分子基础。

### 1 材料与方法

#### 1.1 番红花转录组数据过滤

在NCBI序列片段存档数据库(Sequence Read Archive, SRA)中选取番红花(*Crocus sativus* L.)开花期(开花前、开花时、开花后)转录组数据(SRP056059) (Baba等2015), 在高通量基因表达数据库(Gene Expression Omnibus, GEO)中选取番红花球茎、萼片、叶、柱头和雄蕊组织转录组数据(SRP052616) (Jain等2016), 相关信息见表1。7个

收稿 2019-08-08 修定 2019-12-18  
资助 上海市现代农业产业技术体系(沪农科产字2019第8号)和上海市科技兴农成果转化项目(沪农科转字2016第2-4号)。

\* 通讯作者(saasflower@163.com)。

表1 番红花转录组测序数据  
Table 1 RNA-Seq data of *C. sativus*

| 登录号       | SRA编号      | 测序样品类型 | 测序文库          | 测序数据量/Gb | 测序平台                         |
|-----------|------------|--------|---------------|----------|------------------------------|
| SRP056059 | SRR1909702 | 花      | 双末端(pair-end) | 3.3      | Illumina Genome Analyzer IIx |
|           | SRR1909704 | 柱头     |               | 2.6      |                              |
| GSE65103  | SRR1767298 | 球茎     | 双末端(pair-end) | 2.3      | Illumina HiSeq 2000          |
|           | SRR1767300 | 萼片     |               | 3.2      |                              |
|           | SRR1767299 | 叶      |               | 2.5      |                              |
|           | SRR1767302 | 柱头     |               | 3.0      |                              |
|           | SRR1767301 | 雄蕊     |               | 2.6      |                              |

番红花转录组数据组装前先进行数据过滤,即除去含接头、含N比例大于10%,以及低质量的原始读数序列(质量值 $Q \leq 20$ 的碱基数占整个原始读数的40%以上),以获得高质量纯净读数(clean reads)数据。

1.2 转录组数据组装及分析

1.2.1 转录组数据组装

使用短原始读数(reads)组装软件Trinity对clean reads进行从头组装(Grabherr等2011),首先将具有一定长度overlap的reads连成更长的片段,这些通过reads overlap关系得到的不含N的组装片段作为组装出来的unigene。

1.2.2 序列注释功能分类和生物学通路分析

通过blastx将unigene序列比对到蛋白数据库Nr ( $E$ -value值 $<1 \times 10^{-5}$ ),获得相应unigene具有最高序列相似性的蛋白,进而得到该unigene编码蛋白功能注释信息。根据Nr注释信息,使用Blast2GO软件(Conesa等2005)得到unigene的GO (gene ontology)注释信息;随后,用WEGO (Ye等2006)软件对所有unigene做GO功能分类统计。unigene分别与GO、KOG (Karyotic Ortholog Groups)、KEGG (Kyoto Encyclopedia of Genes and Genomes)和Swiss-Prot数据库比对,分析预测基因的功能分类和参与的生物学途径。

1.3 预测编码蛋白框和转录因子

预测编码蛋白框(coding DNA sequence, CDS)参照武红霞等(2016)的方法,即先将unigene序列与4大数据库blastx比对( $E < 10^{-5}$ ),选择比对结果中排列最高的蛋白确定该unigene的编码区序列,并通过编码区序列的翻译得到氨基酸序列;对于未能

比对上的unigene,则通过ESTScan软件(Iseli等1999)预测编码区,得到其编码区核苷酸和氨基酸序列。预测转录因子(transcription factors, TF)参考李依民等(2018)的方法,即将预测的CDS与植物转录因子数据库Plant TFDB进行hmmscan比对,检索转录因子家族和其成员。

2 实验结果

2.1 番红花转录组组装分析

本研究所选7个转录组测序量均在2.0 Gb以上,测序样品涵盖了番红花球茎、萼片、叶和花器官(表1),整合7组转录组序列,过滤去除低质量片段,各样品数据统计如表2所示。番红花各组织碱基质量超过 $Q_{30}$ 的比例均在90%以上,GC含量均在50%左右,说明测序组装效果好,可进行后续分析。

利用Trinity软件对上述获得的clean reads进行de novo组装,共获得62 105条非冗余unigene序列,N50长达1 144 nt,最小长度201 bp,最大长度11 913 bp,平均序列长727 bp。经统计分析发现:unigene在200~3 000 bp区间内均有分布,在200~999 bp长度范围内数量较多;除了大于3 000 bp的1 118条unigene以外,序列数目随序列长度的增加而减少。

2.2 unigene的功能注释

为获得unigene功能信息,将62 105条unigene与Nr等四大数据库进行比对分析,共有34 711条unigene基因(55.89%)被注释,四个数据库条目占得比例如表3所示。其中,在NR数据库中获得注释的unigene数量最多(34 507条,55.56%)。Swiss-Prot和KOG注释的比例分别为37.92%和32.46%;KEGG

表2 番红花转录组测序数据质量分析

Table 2 Quality analysis of transcriptome sequencing data of *C. sativus*

| SRA编号      | 测序样品类型 | raw reads数 | clean reads数 | clean reads比例/% | clean bases数  | GC/%  | 超过Q <sub>30</sub> | 超过Q <sub>30</sub> 比例/% |
|------------|--------|------------|--------------|-----------------|---------------|-------|-------------------|------------------------|
| SRR1909702 | 花      | 75 432 904 | 62 629 064   | 83.03           | 4 429 189 036 | 48.39 | 4 179 529 008     | 94.36                  |
| SRR1909704 | 柱头     | 59 043 670 | 52 157 996   | 88.34           | 3 690 728 400 | 48.83 | 3 486 857 931     | 94.48                  |
| SRR1767298 | 球茎     | 35 932 216 | 35 661 368   | 99.25           | 3 568 957 596 | 46.18 | 3 316 655 813     | 92.93                  |
| SRR1767300 | 萼片     | 48 312 982 | 47 934 316   | 99.22           | 4 792 902 099 | 47.36 | 4 405 631 169     | 91.92                  |
| SRR1767299 | 叶      | 37 650 344 | 37 341 556   | 99.18           | 3 736 576 711 | 47.30 | 3 440 571 623     | 92.08                  |
| SRR1767302 | 柱头     | 45 865 500 | 45 502 630   | 99.21           | 4 562 890 066 | 45.45 | 4 255 565 375     | 93.26                  |
| SRR1767301 | 雄蕊     | 38 535 588 | 38 228 368   | 99.20           | 3 822 092 063 | 47.56 | 3 507 961 204     | 91.78                  |

表3 番红花unigene的注释统计

Table 3 The statistics analysis of the unigene annotation of *C. sativus*

| 数据库        | unigene数目 | 注释比例/% |
|------------|-----------|--------|
| Nr         | 34 507    | 55.56  |
| Swiss-Prot | 23 553    | 37.92  |
| KOG        | 20 159    | 32.46  |
| KEGG       | 14 185    | 22.84  |
| 总数         | 34 711    | 55.89  |

数据库仅有14 185条(22.84%) unigene得到了注释, 注释信息最少。

比对到Nr数据库的34 507条unigene注释同源基因的物种, 在相似序列匹配度较高的物种中, 石刁柏(*Asparagus officinalis*)所占比例最高8 299条(24.05%); 其次为油棕(*Elaeis guineensis*) 4 698条(13.61%)和海枣(*Phoenix dactylifera*) 4 200条(12.17%), 其余匹配物种如凤梨(*Ananas comosus*)、粳稻(亚种) (*Oryza sativa* subsp. *japonica*)、玉米(*Zea mays*)等比例在7%以下。由于缺乏番红花基因组, 仍有部分unigene未能在NR库中匹配到同源序列。

2.3 功能分类

2.3.1 unigene的GO分类

在转录本中, 能够被注释到GO分类的unigene仅有36 611条, 分别注释到生物过程、细胞组分和分子功能3个GO类别的46个亚类(图1)。在第一大类生物过程中, 有 20 个亚类, 主要聚集在代谢过程和细胞过程, 涉及的基因分别有4 228条和3 920条。细胞组分有15个亚类, 其中细胞和细胞片段相关基因丰度最高, 均为2 731条, 其次是细胞器有

2 010条, 此外膜和膜部分的基因较多, 分别为1 522和1 059条。在分子功能方面, 有11个亚类, 其中具有催化活性和结合功能的基因数量较高, 分别为4 004和3 068条, 其他类别基因数目均较少。总体而言, 番红花细胞活动的基因表达丰度较高, 基于7个转录组数据进行碳氮代谢基因挖掘具有一定可行性。

2.3.2 unigene的KOG分类

将番红花unigene与KOG数据库进行比对, 预测unigene的功能并对其进行分类统计。统计结果显示番红花unigene根据其功能大致可以分为25类(图2), 且涉及了大多数的生命活动。有20 159条unigene注释到KOG数据库, 这些unigene中注释到数目较多的组依次是整体功能预测(6 438条, 31.94%), 蛋白质翻译后修饰、折叠及分子伴侣(3 883条, 19.26%), 信号转导机制(3 176条, 15.75%), RNA加工和修饰(1 960条, 9.72%); 此外, 有1 431条(7.10%)功能未知。

2.4 KEGG通路分析

通过KEGG功能注释发现, 番红花不同组织间共有7 860条unigene参与了131条代谢通路; 占较大比例的是代谢相关途径, 有12 635条unigene; 其次是遗传信息处理有3 803条; 而细胞程序、环境信息程序和有机体系统所占比例较小。代谢通路细分后(表4, 排名前30位的代谢通路), 代谢途径参与基因最多, 达2 904条(36.95%); 其次是次生代谢产物生物合成途径, 有1 504条(19.13%) unigene参与。

从番红花不同组织转录组KEGG代谢途径分析结果可见, 有409条unigene直接注释到碳代谢

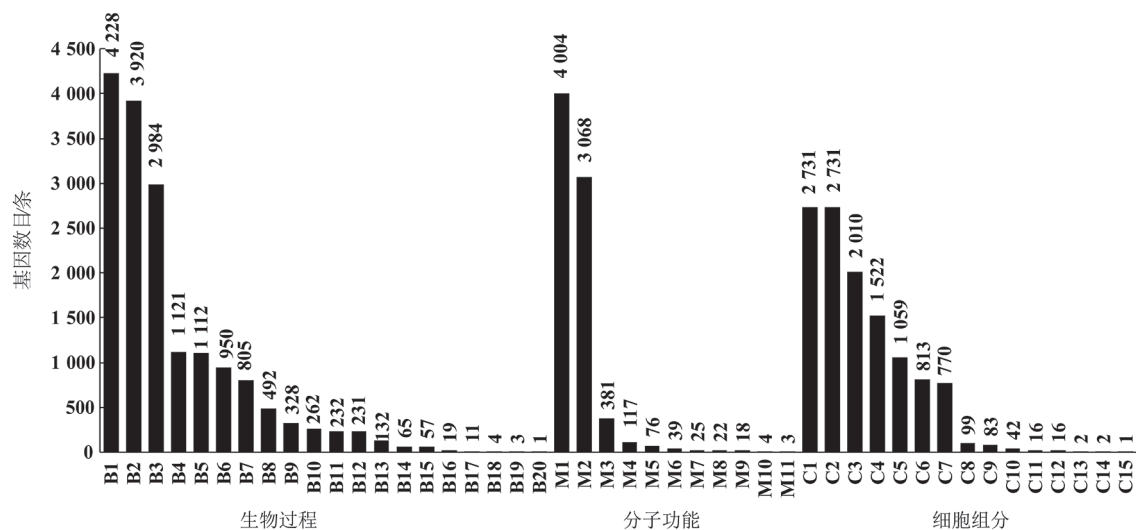


图1 番红花unigene的GO功能分类  
Fig.1 GO function classification of unigene in *C. sativus*

B1: 代谢过程; B2: 细胞过程; B3: 单一生物过程; B4: 生物调节; B5: 定位; B6: 应激反应; B7: 细胞成分组织或生物发生; B8: 发育进程; B9: 多细胞生物过程; B10: 信号; B11: 生殖; B12: 生殖过程; B13: 多有机体过程; B14: 生长; B15: 免疫系统过程; B16: 生物粘附; B17: 节律过程; B18: 排毒; B19: 移动; B20: 细胞杀伤; M1: 催化活性; M2: 结合; M3: 运输活性; M4: 结构分子活性; M5: 核酸结合转录因子; M6: 信号传感器活动; M7: 分子功能调节剂; M8: 抗氧化活性; M9: 分子转导活性; M10: 电子载体; M11: 转录因子活性, 蛋白质结合; C1: 细胞; C2: 细胞组分; C3: 细胞器; C4: 膜; C5: 膜部分; C6: 细胞器部分; C7: 高分子复合物; C8: 细胞连接; C9: 细胞外区域; C10: 膜封闭的管腔; C11: 病毒体; C12: 病毒体部分; C13: 细胞外区域部分; C14: 超分子纤维; C15: 拟核。

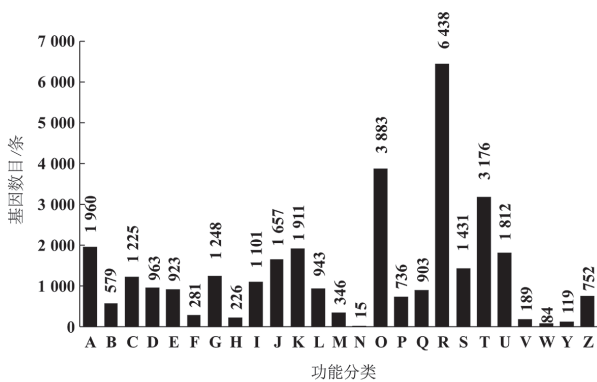


图2 番红花unigene的KOG功能注释分类  
Fig.2 KOG functional annotation classification of unigene in *C. sativus*

A: RNA加工与修饰; B: 染色质结构与动力学; C: 能源产生和转化; D: 细胞周期控制, 细胞分裂, 染色体分离; E: 氨基酸运输与代谢; F: 核苷酸运输与代谢; G: 碳水化合物运输与代谢; H: 辅酶运输与代谢; I: 脂质运输与代谢; J: 翻译, 核糖体结构与生物合成; K: 转录; L: 复制, 重组与修复; M: 细胞壁/膜/胞外被膜; N: 细胞运动; O: 翻译后修饰, 蛋白质折叠, 分子伴侣; P: 无机离子运输与代谢; Q: 次生代谢产物的生物合成, 运输与代谢; R: 整体功能预测; S: 功能未知; T: 信号转导机制; U: 胞内运输, 分泌与囊泡运输; V: 防御机制; W: 胞外结构; Y: 核结构; Z: 细胞骨架。

(ko01200), 注释到氮代谢途径的unigene则略少, 仅有46条; 此外, 有378条unigene注释到淀粉和蔗糖代谢途径(ko00500)、143条unigene注释到光合生物中的碳固定途径(ko00710)、98条unigene注释到光合途径(ko00195), 这些途径均与碳代谢紧密相关。碳氮代谢途径中主要unigene见表5, 包括了磷酸烯醇式丙酮酸羧化酶(K01595)、甘油醛-3-磷酸脱氢酶(K00134)、谷氨酸合成酶(K00264)和碳酸酐酶(K01674)的基因, 除了表中列出的相关基因外, 还有在蔗糖代谢中主要调控作用的蔗糖合成酶基因(K00695)、蔗糖磷酸合成酶基因(K00696)等, 是进一步研究番红花碳氮代谢途径的基础。

2.5 CDS预测

对CDS进行预测能为番红花碳氮代谢功能基因的挖掘利用提供重要参考。按照NR、SwissProt、KEGG和KOG数据库的优先级顺序, 通过blastx比对共获得34 180条unigene的CDS, 与以上蛋白数据库皆比对不上的unigene用软件ESTScan预测后获

表4 番红花unigene的KEGG功能注释分类(前30个)

Table 4 KEGG functional classification of unigene in *C. sativus* (top 30)

| 编号 | 通路           | 通路编码    | 基因数量/条 | 比例/%  | 编号 | 通路            | 通路编码    | 基因数量/条 | 比例/% |
|----|--------------|---------|--------|-------|----|---------------|---------|--------|------|
| 1  | 代谢途径         | ko01100 | 2 904  | 36.95 | 16 | mRNA监测途径      | ko03015 | 281    | 3.58 |
| 2  | 次生代谢物的生物合成   | ko01110 | 1 504  | 19.13 | 17 | RNA降解         | ko03018 | 277    | 3.52 |
| 3  | 抗生素的生物合成     | ko01130 | 671    | 8.54  | 18 | 植物-病原体相互作用    | ko04626 | 269    | 3.42 |
| 4  | 核糖体          | ko03010 | 579    | 7.37  | 19 | 真核生物中的核糖体生物发生 | ko03008 | 234    | 2.98 |
| 5  | 微生物在不同环境中的代谢 | ko01120 | 565    | 7.19  | 20 | 嘌呤代谢          | ko00230 | 232    | 2.95 |
| 6  | 内质网中的蛋白质加工   | ko04141 | 419    | 5.33  | 21 | 氨基糖和核苷酸糖代谢    | ko00520 | 226    | 2.88 |
| 7  | 碳代谢          | ko01200 | 409    | 5.20  | 22 | 半胱氨酸和蛋氨酸代谢    | ko00270 | 196    | 2.49 |
| 8  | 剪接           | ko03040 | 396    | 5.04  | 23 | 嘧啶代谢          | ko00240 | 183    | 2.33 |
| 9  | 淀粉和蔗糖代谢      | ko00500 | 378    | 4.81  | 24 | 糖酵解/糖异生       | ko00010 | 176    | 2.24 |
| 10 | 氨基酸的生物合成     | ko01230 | 374    | 4.76  | 25 | 甘油磷脂代谢        | ko00564 | 171    | 2.18 |
| 11 | RNA转运        | ko03013 | 354    | 4.50  | 26 | 吞噬体           | ko04145 | 163    | 2.07 |
| 12 | 内噬作用         | ko04144 | 348    | 4.43  | 27 | 苯丙烷类生物合成      | ko00940 | 161    | 2.05 |
| 13 | 植物激素信号转导     | ko04075 | 341    | 4.34  | 28 | 丙酮酸盐代谢        | ko00620 | 150    | 1.91 |
| 14 | 氧化磷酸化        | ko00190 | 293    | 3.73  | 29 | 戊糖和葡萄糖醛酸的相互转化 | ko00040 | 149    | 1.90 |
| 15 | 泛素介导的蛋白水解    | ko04120 | 290    | 3.69  | 30 | 光合生物中的碳固定     | ko00710 | 143    | 1.82 |

表5 参与碳氮代谢的部分番红花unigene

Table 5 Part of *C. sativus* unigene involved in carbon and nitrogen metabolism

| 代谢途径         | unigene编码蛋白功能           | unigene数目/条 | 编码     |
|--------------|-------------------------|-------------|--------|
| 碳代谢(ko01200) | 磷酸烯醇式丙酮酸羧化酶             | 25          | K01595 |
|              | 果糖-二磷酸醛缩酶               | 22          | K01623 |
|              | 丙酮酸激酶                   | 21          | K00873 |
|              | 苹果酸酶                    | 12          | K00026 |
|              | 转酮醇酶                    | 12          | K00615 |
|              | 甘油醛-3-磷酸脱氢酶             | 11          | K00134 |
|              | 2-氧代戊二酸脱氢酶E1组分          | 10          | K00164 |
|              | 甘氨酸羟甲基转移酶               | 10          | K00600 |
|              | 半胱氨酸合成酶A                | 10          | K01738 |
|              | 苹果酸脱氢酶(草酰乙酸脱羧酶) (NADP+) | 9           | K00029 |
|              | 6-磷酸果糖激酶1               | 9           | K00850 |
|              | 丙酮酸脱氢酶E2组分(二氢硫辛酸乙酰转移酶)  | 8           | K00627 |
|              | (S)-2-羟基-酸氧化酶           | 8           | K11517 |
|              | 谷氨酸合成酶(NADPH/NADH)      | 8           | K00264 |
|              | 碳酸酐酶                    | 6           | K01673 |
| 氮代谢(ko00910) | 硝酸还原酶[NAD(P)H]          | 6           | K10534 |
|              | 谷氨酸合成酶(铁氧还蛋白)           | 5           | K00284 |
|              | 碳酸酐酶                    | 5           | K01674 |
|              | 谷氨酸脱氢酶[NAD(P)]          | 4           | K00261 |
|              | 铁氧还蛋白-亚硝酸还原酶            | 4           | K00366 |

得5 211个CDS。由图3可见, 通过BLAST获得的CDS序列长度较长,  $\geq 1\ 000$  nt的有7 563条, 占22.13%; 而由ESTScan获得的CDS较短, 主要集中在200~399 nt (共3 512条, 占67.40%)。

2.6 TF分析

将预测的蛋白序列同相应的TF数据库(Plant TFDB)进行hmmscan比对, 共预测出55个TF家族, 共1 070个TF的unigene (图4), 其中, bHLH、MYB-

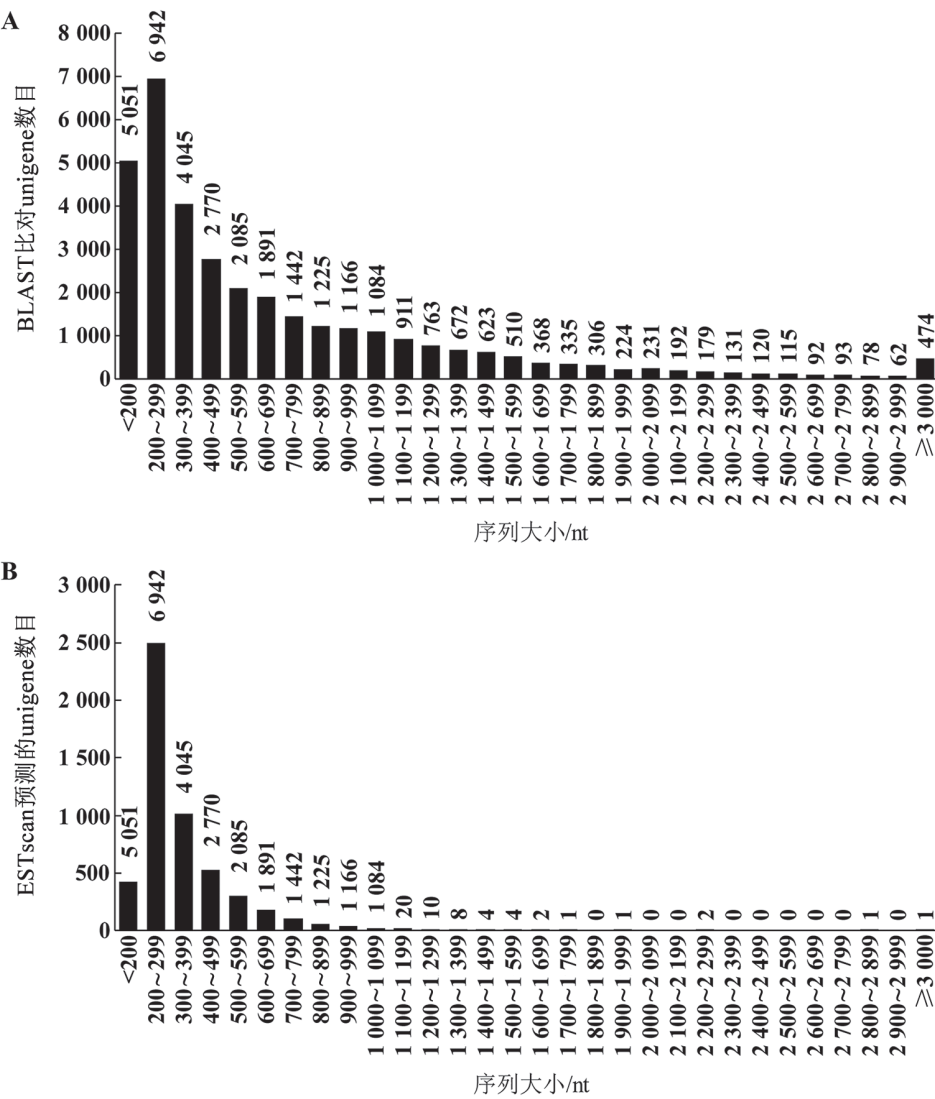


图3 番红花unigene的CDS长度分布图  
Fig.3 CDS length distribution of unigene in *C. sativus*

related、NAC、C2H2、ERF、bZIP以及C3H类TF为多基因家族, 分别占总预测量的7.29%、6.54%、6.26%、6.26%、5.05%、5.05%和4.86%, 而VOZ、STAT、NZZ/SPL、HRT-like、HB-PHD和GRF所占比例较少, 均为0.093%。在55个TF家族中, 与碳氮代谢相关的TF包括: GATA转录因子(17个)和DOF转录因子(18个)等。这些TF的发现将为后续研究番红花生长发育过程中碳氮代谢提供新思路。

3 讨论

随着测序成本的降低, 转录组测序已广泛应用

于药用植物功能基因组研究, 现已解析人参(*Panax ginseng*, Jung等2018)、杜仲(*Eucommia ulmoides*, 李铁柱等2012)、羽叶三七(*Panax pseudoginseng* var. *bipinnatifidus*, 黄勋等2017)等药材的转录组特征, 为阐明中药材种质资源遗传基础、次生代谢途径等奠定重要基础。番红花作为重要药用植物之一, 2015年Baba等首次开展番红花花器官转录组测序, 随后Jain等(2016)进行了球茎、萼片、叶、柱头和雄蕊组织转录组测序, 均关注于次生代谢途径尤其是类胡萝卜素代谢途径。然而, 除了番红花次生代谢途径和药理作用, 国内番红花产量供

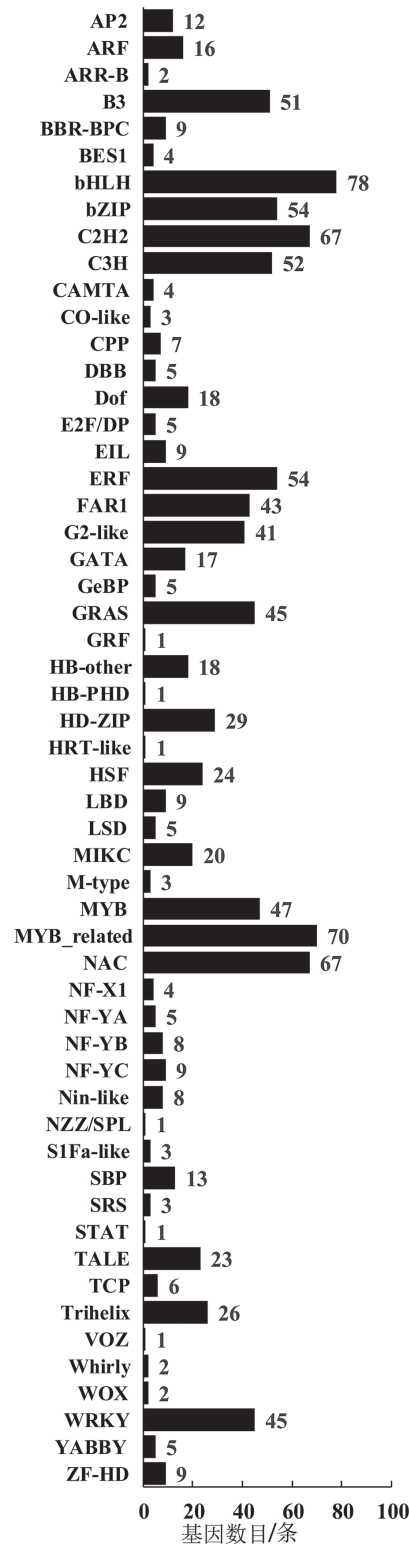


图4 番红花unigene的转录因子家族分类图  
Fig.4 Transcription factor family classification  
of unigene in *C. sativus*

不应求的现状需得到充分重视, 需开展番红花生长发育机理研究, 以此指导栽培方式和繁育体系优化, 从而提高番红花种球繁殖效率和花丝产量。

本文通过生物信息学的方法, 对番红花不同组织的7个转录组进行数据过滤及*de novo*组装、功能注释和分类、代谢通路分析以及重要基因和转录因子的挖掘。7个转录组数据过滤和组装后, 共获得62 105条unigene序列, 平均序列长727 bp, 在200~999 bp长度范围内数量较多。在NR数据库中获得注释的unigene数量最多(34 507条, 55.56%), 同源基因的物种相似序列匹配度较高的物种为石刁柏。与Swiss-Prot数据库相比较, 检测出23 553个蛋白同源序列; 进行GO功能分类分析, 36 611条unigene被注释到生物过程、细胞组分和分子功能3个GO类别的46个亚类, 其中涉及代谢过程和细胞过程的基因较多; 与KOG数据库比对, 发现20 159条unigene被KOG功能注释, 共分为25类; 通过BLAST比对共获得34 180条unigene的CDS; 共预测出55个TF家族。

基于转录组数据, 从中挖掘出碳氮代谢相关基因和转录因子。筛选出的与碳氮代谢相关的基因中, 磷酸烯醇式丙酮酸羧化酶基因在植物细胞中参与光合碳同化等重要代谢途径, 也参与调控种子营养物质合成与代谢过程, 控制糖类物质流向、脂肪酸合成或蛋白质合成途径(魏绍巍和黎茵2011); 谷氨酸合成酶(glutamate synthase, GOGAT)是植物体内氮素同化与循环的关键酶, 高等植物体内95%以上的 $\text{NH}_4^+$ 通过GS/GOGAT循环(无机氮转化为有机氮的第1步, 也是目前为止所发现的无机氮转化成有机氮的主导途径)同化, 而GOGAT是该途径的限速酶(Lea和Mifflin 1974; Lea等1990; Joy等1992), 模式植物GOGAT基因在氮代谢调控中的应用备受关注且已经取得了一定进展(牛超等2018)。转录因子在植物生长发育和形态建成中发挥着重要的调控作用。番红花转录组数据中, 与碳氮代谢密切相关的转录因子有GATA和DOF。GATA转录因子在植物光响应调控、叶绿素合成、细胞分裂素响应以及碳氮代谢等生物学过程中发挥重要作用(袁岐等2017); DOF参与光应答与光周期调控开花、种子发育和萌发、植物激素和

防御应答、碳氮代谢等生物学过程(蔡晓锋等2013);而且在拟南芥和水稻中发现*ZmDof1*在低氮条件下可增加氮同化和促进生长(武姣娜等2018)。除了与碳氮代谢直接密切相关的转录因子以外,也挖掘到与光信号响应(bHLH) (阮林伟等2019)和种子(球)发育(ERF) (曹国丽等2018)密切相关的转录因子。这些基因和转录因子的挖掘,为番红花生生长发育尤其是碳氮代谢途径的研究提供了重要线索和依据。

### 参考文献(References)

- Baba SA, Mohiuddin T, Basu S, et al (2015). Comprehensive transcriptome analysis of *Crocus sativus* for discovery and expression of genes involved in apocarotenoid biosynthesis. *BMC Genomics*, 16: 698
- Behnia MR, Estilai A, Ehdaie B (1999). Application of fertilizers for increased saffron yield. *J Agron Crop Sci*, 182: 9–15
- Cai XF, Zhang YY, Zhang JH, et al (2013). Advances in research on function of the *Dof* gene family in plant. *Plant Physiol J*, 49 (1): 1–12 (in Chinese with English abstract) [蔡晓锋, 张余洋, 张俊红等(2013). 植物*Dof*基因家族功能研究进展. *植物生理学报*, 49 (1): 1–12]
- Cao GL, Zhang YF, Sun YQ, et al (2018). Functional analysis of ERF transcription factor family in taproot development of sugar beet. *Plant Physiol J*, 54 (5): 745–754 (in Chinese with English abstract) [曹国丽, 张永丰, 孙亚卿等(2018). ERF转录因子家族在甜菜块根发育中的功能分析. *植物生理学报*, 54 (5): 745–754]
- Conesa A, Götz S, García-Gómez JM, et al (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21 (18): 3674–3676
- Grabherr MG, Haas BJ, Yassour M, et al (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*, 29 (7): 644–652
- Gresta F, Lombardo GM, Siracusa L, et al (2008). Saffron, an alternative crop for sustainable agricultural systems. A review. *Agron Sustain Devt*, 28 (1): 95–112
- Gresta F, Santonoceto C, Avola G (2016). Crop rotation as an effective strategy for saffron (*Crocus sativus* L.) cultivation. *Sci Hortic*, 211: 34–39
- Huang X, Jin J, Yu J, et al (2017). Transcriptome sequencing of *Panax pseudoginseng* and identification of key enzyme genes in triterpeno saponin biosynthesis. *Genomics Appl Biol*, 36 (6): 2531–2538 (in Chinese with English abstract) [黄勋, 金健, 余洁等(2017). 羽叶三七的转录组测序与三萜皂苷生物合成的关键酶基因的识别. *基因组学与应用生物学*, 36 (6): 2531–2538]
- Iseli C, Jongeneel CV, Bucher P (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol*, 99: 138–148
- Jain M, Srivastava PL, Verma M, et al (2016). *De novo* transcriptome assembly and comprehensive expression profiling in *Crocus sativus* to gain insights into apocarotenoid biosynthesis. *Sci Rep*, 6: 22456
- Joy KW, Blackwell RAYD, Lea PJ (1992). Assimilation of nitrogen in mutants lacking enzymes of the glutamate synthase cycle. *J Exp Bot*, 43 (247): 139–145
- Jung I, Kang H, Kim JU, et al (2018). The mRNA and miRNA transcriptomic landscape of *Panax ginseng* under the high ambient temperature. *BMC Syst Biol*, 12 (Suppl 2): 27
- Karra Y, Boujghagh M, Serghini MA, et al (2017). Effect of planting density on productivity of saffron corms. *Acta Hortic*, 1184: 119–124
- Lea PJ, Mifflin BJ (1974). Alternative route for nitrogen assimilation in higher plants. *Nature*, 251 (5476): 614–616
- Lea PJ, Robinson SA, Stewart GR (1990). The enzymology and metabolism of glutamine, glutamate, and asparagine. *Biochem Plants*, 16: 121–159
- Li LW (2013). Study on biological characteristics and utilization technics of *Crocus* resources (dissertation). Suzhou, Jiangsu: Suzhou University (in Chinese) [李玲蔚(2013). 西红花资源生物学与利用技术的研究(学位论文). 江苏苏州: 苏州大学]
- Li TZ, Du HY, Liu HM, et al (2012). Transcriptome data assembly and gene function annotation of *Eucommia* mature fruits and young fruits. *J Cent South Univ For Technol*, 32 (10): 9–17 (in Chinese with English abstract) [李铁柱, 杜红岩, 刘慧敏等(2012). 杜仲幼果和成熟果实转录组数据组装及基因功能注释. *中南林业科技大学学报*, 32 (10): 9–17]
- Li YM, Zhang HW, Chen Y, et al (2018). High-throughput transcriptome sequencing of roots of *Dictamnus dasycarpus* and data analyses. *Chin Tradit Herbal Drugs*, 49 (21): 4975–4982 (in Chinese with English abstract) [李依民, 张化为, 陈莹等(2018). 白鲜根转录组高通量测序及数据分析. *中草药*, 49 (21): 4975–4982]
- Nasseer S, Nagoo SA, Dar NA, et al (2018). High density plantation in saffron (*Crocus sativus* L.) for achieving higher yields. *Int J Curr Microbiol App Sci*, 7 (5): 748–750
- Niu C, Liu GJ, Qu CP, et al (2018). A review of glutamate synthase gene and its regulatory role in plant nitrogen metabolism. *Jiangsu Agric Sci*, 46 (9): 10–16 (in Chinese) [牛超, 刘关君, 曲春浦等(2018). 谷氨酸合成酶基

- 因及其在植物氮代谢中的调节作用综述. 江苏农业科学, 46 (9): 10–16]
- Ruan LW, Fu Y, Wang MH, et al (2019). Investigation on the function and possible mechanism of light signal in anther development based on transcriptome and protein interaction network analyses. *Plant Physiol J*, 55 (2): 159–168 (in Chinese with English abstract) [阮林伟, 付营, 王梦涵等 (2019). 基于转录组和蛋白互作网络分析光信号在花药发育过程中的功能和可能机制. 植物生理学报, 55 (2): 159–168]
- Wang Z, Zhang YC, Yang LY, et al (2019). Tissue culture of *Crocus sativus* L.: research progress. *Chin Agric Bull*, 35 (12): 100–106 (in Chinese with English abstract) [王桢, 张永春, 杨柳燕等 (2019). 西红花组织培养研究进展. 中国农学通报, 35 (12): 100–106]
- Wei SW, Li Y (2011). Functions of plant phosphoenolpyruvate carboxylase and its applications for genetic engineering. *Chin J Biotech*, 27 (12): 1702–1710 (in Chinese with English abstract) [魏绍巍, 黎茵 (2011). 植物磷酸烯醇式丙酮酸羧化酶的功能及其在基因工程中的应用. 生物工程学报, 27 (12): 1702–1710]
- Wu HX, Xu WT, Luo C, et al (2016) Transcriptome data assembly and gene function annotation of mangofruits. *Chin J Trop Crop*, 37 (11): 2191–2198 (in Chinese with English abstract) [武红霞, 许文天, 罗纯等 (2016). 芒果果实转录组数据组装及基因功能注释. 热带作物学报, 37 (11): 2191–2198]
- Wu JN, Wei XD, Li X, et al (2018) Research progress in nitrogen use efficiency in plants. *Plant Physiol J*, 54 (9): 1401–1408 (in Chinese with English abstract) [武姣娜, 魏晓东, 李霞等 (2018). 植物氮素利用效率的研究进展. 植物生理学报, 54 (9): 1401–1408]
- Yao C, Cao HB, Wang DL, et al (2013). Web of Science-based visual analysis of citations in papers on saffron crocus. *Chin J Med Libr Inf Sci*, 22 (12): 70–78 (in Chinese with English abstract) [姚冲, 曹恒斌, 王大力等 (2013). 基于 Web of Science 的藏红花研究引文可视化分析. 中华医学图书情报杂志, 22 (12): 70–78]
- Yao C, Liu BB, Zhou GF, et al (2017). Advances in research on factors affecting the yield and quality of saffron. *J Chin Med Mater*, 40 (3): 738–743 (in Chinese with English abstract) [姚冲, 刘兵兵, 周桂芬等 (2017). 影响西红花产量和品质的诸因素研究进展. 中药材, 40 (3): 738–743]
- Ye J, Fang L, Zheng H, et al (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res*, 34 (Suppl 2): W293–W297
- Yuan Q, Zhang CL, Zhao TT, et al (2017). Research advances of GATA transcription factor in plant. *Mol Plant Breed*, 15 (5): 1702–1707 (in Chinese with English abstract) [袁岐, 张春利, 赵婷婷等 (2017). 植物中 GATA 转录因子的研究进展. 分子植物育种, 15 (5): 1702–1707]

## Exploring carbon and nitrogen metabolism pathway genes based on transcriptome of *Crocus sativus*

ZHOU Lin, YANG Liuyan, CAI Youming, WANG Zhen, ZHANG Yongchun\*

Forestry and Pomology Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai Key Laboratory of Protected Horticultural Technology, Shanghai 201403, China

**Abstract:** The saffron (*Crocus sativus*) bulb is the basis for flowering and new bulb formation. Promoting the accumulation of nutrients in the saffron bulb and improving the reproductive efficiency of saffron are effective ways to solve the shortage of domestic saffron production. The growth and reproduction of saffron bulbs are closely related to the carbon and nitrogen metabolism pathways during growth and development. Therefore, making a profound study of the mechanism of growth and development of saffron, in particular, the carbon and nitrogen metabolic pathway is the prerequisite for the optimization of saffron cultivation and breeding system. We obtained 62 105 unigene sequences by *de novo* assembly of transcriptome data from saffron. Only 55.89% unigene can be annotated in the public database, and the unigene annotated to GO classification and KEGG annotation were 36 661 and 14 185 respectively. Through bioinformatics analysis, we had extracted multiple important genes and transcription factors involved in carbon and nitrogen metabolism in saffron, including genes encoding sucrose synthase, sucrose phosphate synthase, phosphoenolpyruvate carboxylase, glyceraldehyde 3-phosphate dehydrogenase, glutamate synthase, carbonic anhydrase, and transcription factors GATA and DOF. This research results can provide reference for the discovery and utilization of the growth and development function genes of saffron.

**Key words:** saffron (*Crocus sativus*); transcriptome; carbon and nitrogen metabolism; bioinformatics analysis

Received 2019-08-08 Accepted 2019-12-18

This work was supported by the Agriculture Research System of Shanghai, China (201908) and Shanghai Agricultural Applied Technology Development Program, China (20160204).

\*Corresponding author (saasflower@163.com).