• 药材与资源 •

西红花柱头全长转录组测序及 bZIP 转录因子家族鉴定分析

高广春[1]，李　军[2*]，钱纯节[1]，潘可欣[1]，潘欣怡[1]

1. 嘉兴大学医学院，浙江 嘉兴　314001
2. 嘉兴职业技术学院现代农业学院，浙江 嘉兴　314036

摘　要：目的　深入研究西红花苷生物合成途径，基于西红花 *Crocus sativus* 柱头进行全长转录组测序，挖掘可能参与西红花苷生物合成调控的 bZIP 转录因子基因。**方法**　通过 PacBio Sequel 高通量测序技术对西红花柱头进行全长转录组测序及生物信息学分析，从中鉴定具有完整结构域的全长 CsbZIPs，分析其编码蛋白的理化性质、保守基序、系统进化及互作网络，并筛选可能参与西红花苷生物合成调控的候选基因。**结果**　全长转录组测序共获得 37 732 条 Isoforms 序列，平均长度为 2 358 bp，N50 长度为 3 594 bp；CsbZIPs 蛋白主要为定位到细胞核中的不稳定亲水蛋白，氨基酸数目 57～607 aa，相对分子质量 6 696.55～65 982.04，等电点 5.01～12.16；64 个 CsbZIPs 蛋白都包含典型碱性区域和亮氨酸拉链区域，根据其与拟南芥 AtbZIPs 系统发育关系被分为 11 个亚族，其中 A 亚族含 CsbZIPs 最多；基因表达模式分析及转录组与代谢组联合分析结果表明，E、S 及 A 亚族的 *CsbZIPs* 基因，尤其是 E 亚族的 *CsbZIP44*、*CsbZIP54*、*CsbZIP55*、*CsbZIP62*、*CsbZIP64* 等基因可能参与西红花苷生物合成的转录调控。**结论**　较为全面地揭示了可能参与西红花苷生物合成的 bZIP 转录因子基因，为深入研究西红花苷生物合成调控提供了丰富的数据基础。

关键词：西红花；转录组；bZIP 转录因子；西红花苷；生物合成途径

中图分类号：R286.12　　**文献标志码：**A　　**文章编号：**0253 - 2670(2024)24 - 8526 - 12

DOI: 10.7501/j.issn.0253-2670.2024.24.020

Identification and analysis of bZIP transcription factor family based on full-length transcriptome sequencing in *Crocus sativus* stigmas

GAO Guangchun[1], LI Jun[2], QIAN Chunjie[1], PAN Kexin[1], PAN Xinyi[1]

1. School of Medicine Science, Jiaxing University, Jiaxing 314001, China
2. School of Modern Agriculture, Jiaxing Vocational & Technical College, Jiaxing 314036, China

Abstract: Objective　To investigate the biosynthesis pathway of crocin, identify bZIP transcription factor genes that may be involved in the biosynthesis regulation of crocin, based on the full-length transcriptome sequencing of saffron stigma. **Method**　Sequel high-throughput sequencing technology was used to perform full-length transcriptome sequencing and bioinformatics analysis on the stigma of saffron. Identifying full-length CsbZIPs with complete domains, and analyzing its physicochemical properties, conserved motifs, phylogenetic evolution, and interaction networks of their encoded proteins. Screening candidate genes that may participate in the regulation of saffron biosynthesis. **Results**　A total of 37 732 Isoforms were obtained, with an average length of 2 358 bp and an N50 length of 3 594 bp; The CsbZIP proteins are mainly unstable hydrophilic proteins located in the nucleus, with amino acid number of 57—607 aa, the relative molecular weight of 6 696.55—65 982.04, and isoelectric point of 5.01—12.16; All of the 64 CsbZIP proteins contain typical alkaline regions and leucine zipper regions. Based on the phylogenetic relationship with AtbZIPs, they are divided into 11 subfamilies, in which Group A containing the most CsbZIPs; The results of gene expression pattern analysis and transcriptome metabolome association analysis indicate that *CsbZIP* genes in E, S, and A subgroups, especially *CsbZIP44*, *CsbZIP54*, *CsbZIP55*, *CsbZIP62*, *CsbZIP64* genes in E subgroups, may be involved in the transcriptional regulation of crocin biosynthesis. **Conclusion**　The data mining comprehensively reveals the bZIP transcription factor genes that may be involved in the biosynthesis regulation of crocin, provides a rich data foundation for in-depth research on the biosynthesis regulation of crocin.

Key words: *Crocus sativus* L.; transcriptome; bZIP transcription factor; crocin; biosynthesis pathway

收稿日期：2024-08-05

基金项目：国家自然科学基金项目（81703667）；浙江省基础公益研究计划项目（LGN21C150004）；嘉兴市公益性研究计划项目（2023AY11048）；嘉兴市科技特派员专项（2021K117）

作者简介：高广春，博士，副教授，研究方向为中药资源与药用活性成分的开发。Tel: 13738275846　E-mail: gaogcjx@163.com

***通信作者：**李　军，博士，教授，研究方向为药用植物生物技术。Tel: 13738286846　E-mail: lijun@jxvtc.edu.cn

西红花 *Crocus sativus* L.为鸢尾科西红花属球茎类药用植物，在浙江、江苏、上海、四川等地广泛栽培，其干燥柱头称为藏红花、西红花，既是名贵香料又是一种名贵中药材，西红花苷为中药西红花的主要活性成分，具有活血化瘀、解郁安神、凉血解毒等功效[1-2]。对西红花苷生物合成途径的解析及相关基因的挖掘一直是国内外研究学者广泛关注的热点，目前已完成基因组测序[3-4]，并鉴定出 *CsCCD2*[5]、*CsALDH3*[6]、*CsUGT74AD1*[7]等功能基因。

bZIP（basic leucine zipper）转录因子是植物十大转录因子家族之一，是一类广泛分布于真核生物中数量最大、最保守的基因家族。bZIP 结构域由 60～80 个氨基酸构成，含有 1 个碱性区域和 1 个亮氨酸拉链结构域[8-9]，碱性区域由 16～20 个碱性氨基酸残基组成，含有保守的 N-X7-R/K 基序，具有细胞核定位以及与 DNA 顺式元件特异性结合的作用；亮氨酸拉链区含有 L-X6-L-X6-L 结构，具有二聚化功能，形成二聚体后依赖于碱性区域结合在双链 DNA 分子上发挥作用。Wolfgang 等[10]根据 bZIP 家族的碱性区以及其他保守序列的相似相，将拟南芥 *Arabidopsis thaliana* L. 78 个 bZIP 成员分为 A～K、M 及 S 共计 13 个亚家族。bZIP 转录因子广泛参与植物次生代谢产物合成、植物生长发育以及胁迫响应等生物学过程[8]。bZIP 转录因子通过参与激素信号转导，激活某些次生代谢产物的生物合成通路基因表达来影响次生代谢物的合成。研究发现，VvibZIPC22 调控葡萄 *Vitis vinifera* L.黄酮醇生物合成[11]、DkbZIP5 调控柿子 *Diospyros kaki* T.原花青素生物合成[12]、AabZIP1 调控黄花蒿 *Artemisia annua* L.青蒿素生物合成[13]、MtbZIP17 和 MtbZIP60 调控紫花苜蓿 *Medicago truncatula* L.三萜皂苷生物合成[14]。西红花苷作为药用植物西红花柱头中的一种主要次生代谢产物，哪些 bZIP 转录因子参与西红花苷的生物合成调控及其调控机制尚未报道。

转录组测序是获得植物基因转录水平序列信息及表达丰度的关键技术手段，基于 PacBio RSII 测序平台的第 3 代全长转录组测序技术具有长读、长测序的优势，可直接获得完整的全长转录本，对无参考基因组物种的转录组研究具有非常突出的优势，近年来全长转录组测序技术在地黄 *Rehmannia glutinosa* L.[15]、滇黄精 *Polygonatum kingianum* C.[16]、滇重楼 *Paris polyphylla* S.[17]、多花黄精 *P. cyrtonema* H.[18]、掌叶大黄 *Rheum palmatum* L.[8]等药用植物功能基因挖掘中广泛应用。随着转录组测序技术的发展及日趋成熟，该技术在西红花苷生物合成相关功能基因及转录因子研究方面逐步得到应用[6,19-20]，鉴定出一批 *CYP450*[21]、*Bhlh*[22]、*MYB*[23]类等转录因子基因，然而对西红花 bZIP 转录因子的鉴定和功能研究尚未报道。本研究基于三代全长转录组测序技术，挖掘西红花 bZIP 转录因子基因家族，对该家族成员进行特征分析，同时结合 RNA-seq 及代谢组数据分析 *CsbZIPs* 基因在不同花发育阶段的表达模式、与西红花苷生物合成功能基因的共表达特性及与西红花次生代谢产物合成的相关性，为进一步研究西红花 bZIP 转录因子的生物学功能奠定基础。

1 材料与仪器

1.1 材料

西红花全长转录组测序样本采自于嘉兴职业技术学院西红花种植基地，经成都中医药大学陈江副教授鉴定为西红花 *C. sativus* L.，于盛花期前后采集 5 个不同发育时期（开花前 2 d、开花前 1 d、开花期、开花后 1 d 及开花后 2 d）的柱头（图 1），每个发育时期分别采集来自 10 个球茎的柱头等量混合，设 3 次重复，立即放入液氮中速冻，−80 ℃冰箱冻存备用。

1.2 仪器

ABI StepOne Plus 实时荧光定量 PCR 仪（美国 Thermo Fisher Scientific 公司）；NanoDrop 2000 分光光度计（美国 Thermo Fisher Scientific 公司）；Veriti 96-Well 型梯度 PCR 仪（美国 Applied Biosystems 公



A-开花前 2 d；B-开花前 1 d；C-开花期；D-开花后 1 d；E-开花后 2 d；图 7 同。

A-two days before blooming; B-one day before blooming; C-blooming day; D-one day after blooming; E-two days after blooming, same as Fig. 7.

图 1 不同发育时期西红花

Fig. 1 *Crocus sativus* at different developmental stages

司）；DYCP-31DN 型电泳仪（北京六一生物科技有限公司）。

2 方法

2.1 全长转录组测序及 *CsbZIPs* 基因鉴定

采用 Trizol 法提取柱头总 RNA，通过琼脂糖凝胶电泳及 Nanodrop 2000 进行总 RNA 质量、纯度及浓度检测。总 RNA 检测合格后，等量混合 5 个不同花期柱头总 RNA，委托上海美吉生物医药科技有限公司使用 PacBio 测序仪进行全长转录组测序。

使用 BlastX（http://www.ncbi.nlm.nih.gov/BLAST/）将全长转录组测序所得完整的 isoform 在 NR 数据库（non-redundant database，NR）、Swiss-Prot 蛋白质序列数据库（Swiss-Prot protein sequence database，Swiss-Prot）、GO 数据库（gene ontology，GO）、KOG 数据库（enKaryotic orthologous groups，KOG）、COG 数据库（cluster of orthologous groups of Proteins，COG）、Pfam 数据库（protein family，Pfam）和京都基因与基因组百科全书（Kyoto encyclopedia of genes and genomes，KEGG）数据库进行功能注释分析。将预测的蛋白序列与植物转录因子数据库 Plant TFDB（http://planttfdb.gao-lab.org/）进行 hmmscan 比对，获得转录因子信息并从中筛选 bZIP 转录因子基因，进而通过 Blastp（https://blast.ncbi.nlm.nih.gov/Blast.cgi）比对、ORF Finder 分析及 ExPASy（https://prosite.expasy.org/prosite. html）分析，剔除重复冗余及不完整序列，最终获得西红花全长 *bZIP* 基因。

2.2 蛋白序列特征分析

利用 ProtParam（https://web.expasy.org/ protparam/）对西红花 bZIP 转录因子蛋白理化性质进行分析；WoLF PSORT（https://www.genscript.com/wolf-psort.html）进行亚细胞定位；Jalview (https://www.jalview.org/download）进行多序列比对；MEME（https://meme-suite.org/meme/tools/streme）进行保守基序 Motif 分析。

2.3 系统进化分析

拟南芥 bZIP 家族成员序列来源于 TAIR 拟南芥数据库（https://www.arabidopsis.org/）。使用 MEGA 7.0 软件对西红花和拟南芥 bZIP 家族成员蛋白质序列进行多重序列比对，使用邻接法（neigh-joining，NJ）构建进化树。利用 iTOL（https://itol.embl.de/）对构建的进化树进行美化。

2.4 基因表达模式分析

利用课题组前期西红花不同发育时期柱头转录组测序数据，分析 *CsbZIPs* 转录因子基因在不同发育时期柱头中的表达模式；另外，分析 *CsbZIPs* 转录因子基因与西红花苷生物合成功能基因（*CsCCD2L*、*CsBCH1*、*CsALDH3I1*、*CsUGT74AD1*、*CsLCYB2a*）的共表达特性。

随机选取 5 个 *CsbZIPs* 基因（*CsbZIP35*、*CsbZIP47*、*CsbZIP49*、*CsbZIP51*、*CsbZIP52*），利用实时荧光定量 PCR 验证测序数据的可靠性。选取与 5 个功能基因（*CsBCH1*、*CsCCD2L*、*CsLYCB2a*、*CsUGT74AD1*、*CsALDH3I1*）表达模式相近的 CsbZIPs 基因（*CsbZIP54*、*CsbZIP58*、*CsbZIP56*、*CsbZIP62*、*CsbZIP37*、*CsbZIP44*），利用实时荧光定量 PCR 分析表达一致性，利用 GraphPad Prism9 软件的单因素方差分析（$P<0.05$）进行表达趋势差异显著性分析并进行可视化作图。利用 Primer Premier 5.0 设计 *CsbZIPs* 基因及功能基因特异性引物，以西红花 *Actin* 作为内参基因用于定量分析（表 1）。用 TRIzol 法分别提取开花前 2 d、开花前 1 d、开花期、开花后 1 d 及开花后 2 d（图 1）共 5 个时期的柱头总 RNA，使用反转录试剂盒（PrimeScriptTM RT Master Mix）合成 cDNA 第一链。根据 TB Green Premix ExTaqTM（TakaRa 公司）试剂盒说明书配置反应体系，模板 cDNA 1 μL，总反应体积 10 μL。反应条件为 95 ℃预变性 2 min，95 ℃，10 s，58 ℃、20 s，72 ℃、30 s，循环数为 40 个，72 ℃ 延伸 30 s 收集荧光，并添加熔解曲线。3 次生物学重复，每次生物学重复进行 3 次技术重复，结果用 $2^{-\Delta\Delta Ct}$ 计算基因的相对表达量差异。

2.5 转录组与代谢组联合分析

利用课题组前期西红花不同发育期柱头转录组和代谢组分析数据，在上海美吉生物云平台进行 CsbZIPs 转录因子与西红花苷及生物合成途径中的相关次生代谢产物西红花苷I、西红花苷IV、西红花酸、西红花酸二醛、苦西红花素、脱落酸的联合分析。选择相关系数大于 0.5 的数据进行相关性网络分析，采用 Cytoscape 3.9.1 软件对结果进行可视化分析。

3 结果与分析

3.1 全长转录组数据统计

通过 PacBio Sequel 高通量测序技术对 5 个采样时期的西红花柱头混合样品开展全长转录组测序，共获得 31.25 Gb 原始数据。根据 Full passes≥

表 1 qPCR 引物序列

Table 1 qPCR primer sequences

基因名称	上游引物（5’-3’）	下游引物（5’-3’）
CsbZIP35	GTTTACAGACAACGGAGCAA	TGCCTGGTTATGGTGATAGA
CsbZIP37	AAGAAGCAATGCCTGGAT	CAAACTGGAGCACCGAAT
CsbZIP44	CAGACACCATCAGCGTAC	CATCCCGAAATGAACTATC
CsbZIP47	GGAGGAAACCGAAGAAAC	GTTGGAGGAATGCCGAAA
CsbZIP49	CAGCAGTAGAGGAAGGAAGAGT	TGCCGCCTCCGAATTGTT
CsbZIP51	GCATTCACAGGAAGCCACATAT	CTGACTGCTCGGAATCACTTG
CsbZIP52	CGTCAGCCTTCTAGGAATGTTC	GTCTCGCCTCGGTGTGATT
CsbZIP54	AGGCACAGAGCGAATGCA	CGACAATGCCGATACCTCC
CsbZIP56	TCTCGGGTTGGACAGGGAT	TCGGCAGAACAGGGAACG
CsbZIP58	GTAAGGTGGATGATAGCGAAGT	CCCGACGTTCCAGATGTT
CsbZIP62	CCTCACTAGGCTCCATTCT	TCCCTTCAAACCCATTAC
CsALDH3I1	CCTGCCACATCATCATTG	TGGTGTTAGGTGCTTTGC
CsUGT74AD1	CAGCCTATGAATGCCAAGT	TCCCACCCTCACTAACAGA
CsLCYB2a	ACGAGTTTGCGGAGTTGG	GGCCTTGGCCTTATGGAA
CsCCD2L	ACTATTATTCTTGACATGGAAT	AGGATGGCAGCAATTGTAAT
CsBCH1	TGACGATGTGGAGGAAGAGAA	CAAGCGAAGCGGTAGTAGAC
Cs18S rRNA	ACGAAACCCCGGCGCAGTGGGC	TCGCTACGTTCTTCATCGATGCG

3 且序列准确性大于 0.9 从原始序列中提取环形一致性序列（circular consensus isoforms，CCS），从中提取了 249 154 条 CCS 序列，平均长度 3 169 bp，进一步筛选得到 214 804 条全长非嵌合序列（full-length non-chimeric，FLNC），约占 86.21%。对 214 804 条 FLNC 进行聚类得到 58 904 条一致序列，其中高质量一致序列 57 180 条，通过进一步合并、去冗余，最终得到 37 732 条 Isoforms 序列，平均长度为 2 358 bp，N_{50} 长度为 3 594 bp。

3.2 全长转录本功能注释

37 732 条 Isoforms 中共有 35 669 条在 NR、GO、COG、Pfam、KEGG、KOG 和 SwissProt 数据库获得注释（表 2），在上述数据库中注释的转录本分别有 35 537、26 268、14 939、28 785、17 000、24 842、25 833 条，注释率达 94.53%。以 NR 数据库为基础，对所有 Isoforms 进行序列比对，发现西红花与油棕 *Elaeis guineensis* Jacq.似率最高，达 33.88%；其次为海枣 *Phoenix dactylifera* L.和芭蕉 *Musa acuminata* Colla，相似率分别为 27.71%和 9.01%。24 842 条在 KOG 数据库中注释成功的 Isoforms，大致可分为 25 个功能类型。其中数量较多的类别为翻译后修饰（O）、信号转导机制（T）、转录（K）、细胞内运输（U）、能量的产生和转化（C）和碳水化合物转运和代谢（G），分别有 3 621、2 613、1 536、1 581、1 490 和 1 446 条（图 2）。

表 2 西红花转录本在公共数据库的功能注释

Table 2 Functional annotation of *C. sativus* transcripts in public databases

公共数据库	被注释基因数目/条	比例/%
NR	35 537	94.18
Swiss-Prot	25 833	68.46
GO	26 268	69.62
KOG	24 842	65.84
COG	14 939	39.59
Pfam	28 785	76.29
KEGG	17 000	45.05

26 268 条在 GO 数据库中注释成功的 Isoforms 涉及生物过程（biological process，BP）、分子功能（molecular fuction，MF）和细胞组分（cellular component，CC）3 大类别，进一步分为 51 个亚类（图 3）。在 BP 类别中，细胞过程和代谢过程涉及的转录本较多，分别有 14 363 和 14 142 条；在 MF 类别中，结合和催化活性涉及的转录本数量远高于其他亚类，均在 14 000 条以上；在 CC 类别中，细胞、细胞部分和细胞器涉及到的转录本较多，均在 10 000 条以上。

3.3 转录因子家族预测

基于西红花全长转录组测序数据，将 37 732 条全长转录本序列通过 TF 数据库进行 hmmscan 比对预测，共预测到 1 232 个转录因子，分属于 59 个

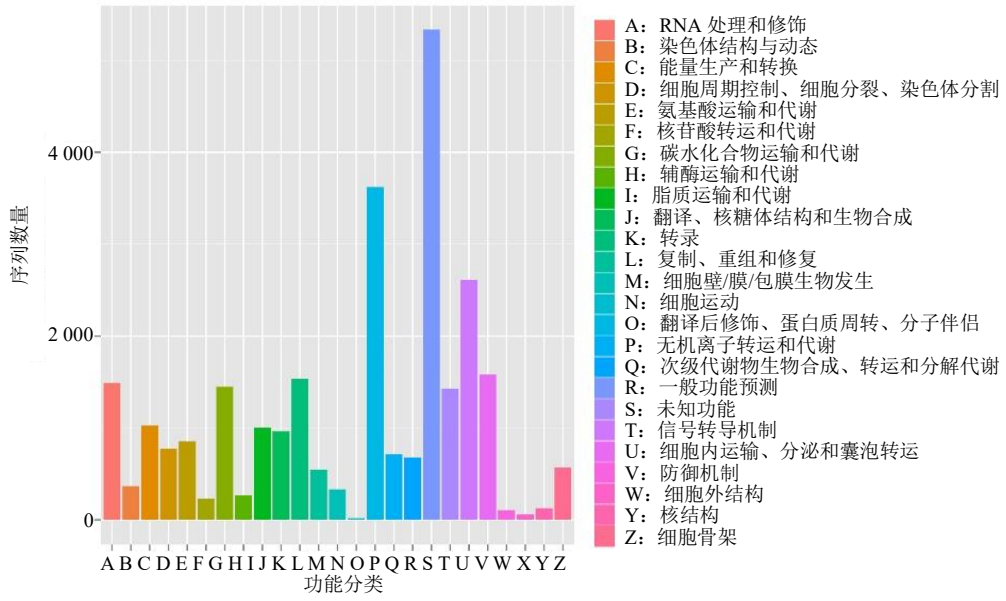


图 2　KOG 功能注释及分类

Fig. 2　KOG functional annotation and classification

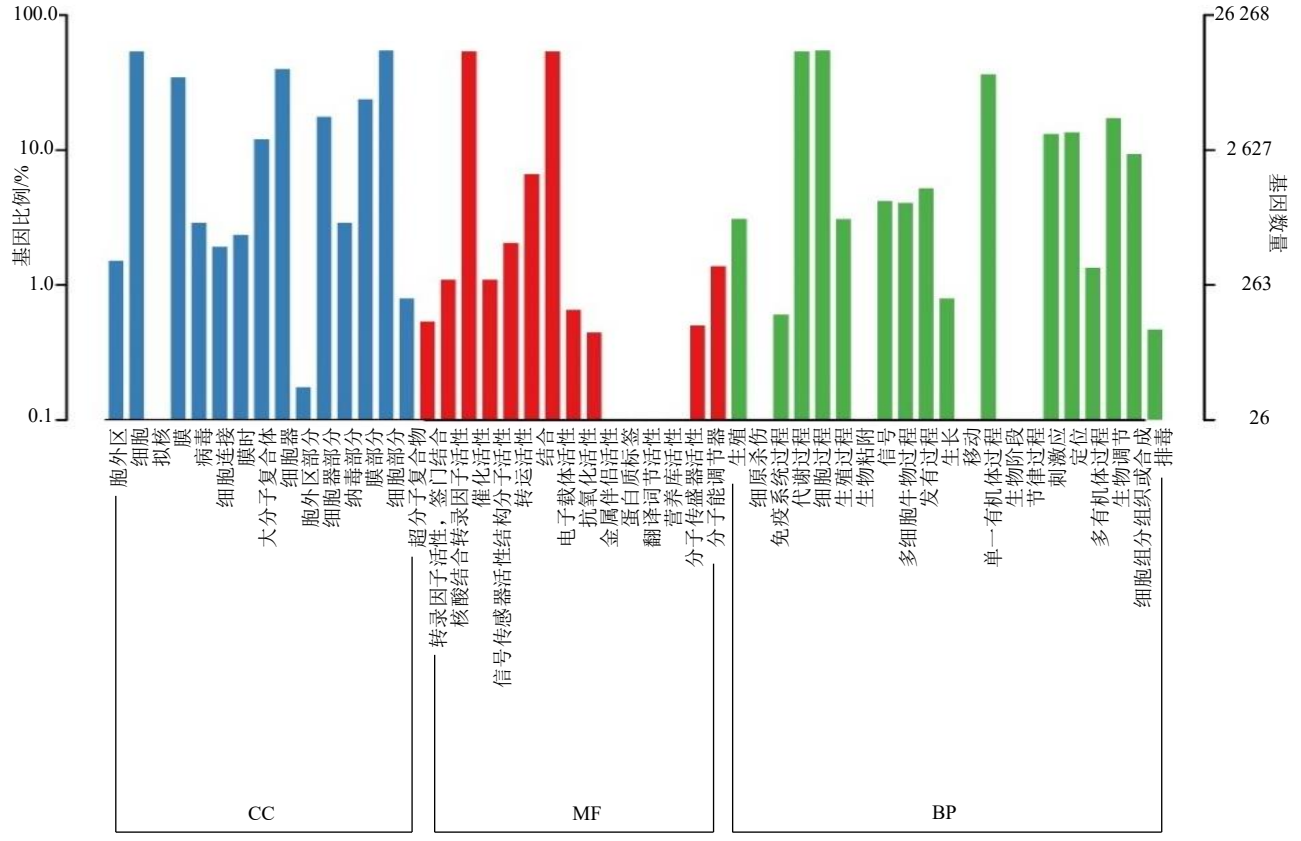


图 3　GO 功能注释

Fig. 3　GO functional annotation

TFs 家族。其中 C3H 转录因子家族的转录本数量最多（132 个），其余依次是 MYB_related（99 个）、bZIP（82 个）、NAC（75 个）、B3-ARF（69 个），另外 BSD、C2C2-YABBY、LIM、LOB、MADS-M、NF-YB、STAT 转录因子家族的转录本数量最少，都只有 1 个。

3.4 *CsbZIPs* 基因鉴定及蛋白理化性质分析

将预测到的 82 个 bZIP 转录因子编码蛋白逐个进行 blastp 比对及 Batch CD-Search，剔除重复冗余及不完整序列，共得到 64 个具有完整结构域的全长 CsbZIPs，分别命名为 CsbZIP1～CsbZIP64（GenBank 登录号 PQ369348～PQ369411）。CsbZIPs 蛋白理化性质分析发现（表 3），*CsbZIPs* 基因编码蛋白的氨基酸残基数目为 57（CsbZIP3）～607（CsbZIP57）aa，相对分子质量 6 696.55～65 982.04，等电点为 5.01（CsbZIP51）～12.16（CsbZIP41），

表 3 CsbZIP 家族成员理化性质和亚细胞定位

Table 3 Physicochemical properties and subcellular localization of CsbZIP family members

转录本名称	GenBank 登录号	氨基酸数/aa	相对分子质量	等电点	不稳定系数	脂肪系数	亲水性	亚细胞定位
CsbZIP1	PQ369348	364	39 837.96	8.81	47.64	65.81	−0.631	细胞核
CsbZIP2	PQ369349	162	17 726.06	9.09	72.35	69.69	−0.561	细胞核
CsbZIP3	PQ369350	57	6 696.55	10.81	74.46	77.02	−1.188	细胞核
CsbZIP4	PQ369351	342	38 165.25	7.00	52.17	81.64	−0.480	细胞核
CsbZIP5	PQ369352	185	20 181.62	6.92	46.13	65.76	−0.882	细胞核
CsbZIP6	PQ369353	432	47 830.80	7.71	55.36	75.73	−0.516	细胞核
CsbZIP7	PQ369354	469	51 574.60	6.13	55.99	70.17	−0.542	细胞核
CsbZIP8	PQ369355	410	43 429.72	5.84	63.11	44.21	−0.931	细胞核
CsbZIP9	PQ369356	179	19 744.37	9.04	71.89	65.79	−0.641	细胞核
CsbZIP10	PQ369357	118	13 538.87	6.92	51.60	59.23	−1.562	细胞核
CsbZIP11	PQ369358	75	8 921.28	9.46	56.87	55.54	−1.326	细胞核
CsbZIP12	PQ369359	264	30 053.69	7.00	63.69	68.97	−0.859	细胞核
CsbZIP13	PQ369360	459	50 578.73	6.34	57.76	73.19	−0.519	细胞核
CsbZIP14	PQ369361	70	8 024.09	6.27	66.49	60.72	−0.994	叶绿体
CsbZIP15	PQ369362	140	15 759.10	9.09	61.80	83.45	−0.512	细胞核
CsbZIP16	PQ369363	142	16 097.46	6.73	62.54	82.98	−0.578	细胞核
CsbZIP17	PQ369364	366	40 829.11	6.63	54.26	78.68	−0.507	细胞核
CsbZIP18	PQ369365	436	48 037.96	6.73	56.97	75.72	−0.474	细胞核
CsbZIP19	PQ369366	295	32 644.66	7.21	57.52	67.38	−0.774	细胞核
CsbZIP20	PQ369367	366	38 946.56	6.68	54.28	72.47	−0.525	细胞核
CsbZIP21	PQ369368	139	15 814.17	10.32	61.77	85.51	−0.562	细胞核
CsbZIP22	PQ369369	219	23 469.31	6.38	50.60	63.53	−0.581	细胞核
CsbZIP23	PQ369370	459	50 649.81	6.34	59.58	73.41	−0.525	细胞核
CsbZIP24	PQ369371	349	37 930.67	7.80	45.96	66.95	−0.607	细胞核
CsbZIP25	PQ369372	242	27 192.66	9.66	50.83	60.41	−0.889	细胞核
CsbZIP26	PQ369373	335	35 797.08	5.24	60.21	66.86	−0.627	细胞核
CsbZIP27	PQ369374	360	39 934.10	8.59	57.25	67.63	−0.696	细胞核
CsbZIP28	PQ369375	536	58 810.25	6.89	71.08	55.68	−0.863	细胞核
CsbZIP29	PQ369376	290	32 444.98	11.41	81.90	71.59	−0.812	细胞核
CsbZIP30	PQ369377	322	34 355.14	6.53	50.34	70.22	−0.736	细胞核
CsbZIP31	PQ369378	208	22 572.47	6.93	46.19	76.76	−0.548	细胞核
CsbZIP32	PQ369379	414	43 543.74	5.95	56.18	44.29	−0.916	细胞核
CsbZIP33	PQ369380	358	39 705.89	8.59	56.11	68.01	−0.691	细胞核
CsbZIP34	PQ369381	372	39 536.53	6.27	54.14	49.19	−0.922	细胞核
CsbZIP35	PQ369382	374	41 539.88	5.84	55.51	58.82	−0.900	细胞核
CsbZIP36	PQ369383	154	17 694.03	9.04	60.51	74.03	−0.948	细胞核
CsbZIP37	PQ369384	137	15 673.01	8.95	55.19	82.63	−0.620	细胞核
CsbZIP38	PQ369385	121	13 780.50	6.50	48.45	73.39	−0.485	线粒体
CsbZIP39	PQ369386	389	41 975.10	8.83	56.49	68.74	−0.597	细胞核
CsbZIP40	PQ369387	162	18 927.30	8.82	71.90	68.52	−0.981	细胞核
CsbZIP41	PQ369388	126	14 526.74	12.16	68.95	94.60	−0.617	细胞核
CsbZIP42	PQ369388	149	17 340.66	9.50	59.52	64.09	−1.029	细胞核
CsbZIP43	PQ369390	277	29 411.29	7.13	58.22	49.35	−0.953	细胞核
CsbZIP44	PQ369391	374	42 485.38	8.68	70.60	62.86	−0.953	细胞核
CsbZIP45	PQ369392	132	15 197.52	11.70	72.45	101.97	−0.569	细胞核
CsbZIP46	PQ369393	334	37 273.32	6.01	47.83	58.83	−0.876	细胞核
CsbZIP47	PQ369394	522	57 613.83	7.24	74.07	54.25	−0.911	细胞核
CsbZIP48	PQ369395	313	34 066.61	6.43	57.56	65.85	−0.791	细胞核
CsbZIP49	PQ369396	345	39 073.94	9.46	62.59	63.91	−0.957	细胞核
CsbZIP50	PQ369397	288	31 148.90	6.66	73.77	67.08	−0.693	细胞核
CsbZIP51	PQ369398	343	37 785.87	5.01	65.43	70.58	−0.666	细胞核
CsbZIP52	PQ369399	603	65 184.25	6.69	46.53	72.77	−0.535	内质网
CsbZIP53	PQ369400	365	38 666.02	6.68	61.99	60.74	−0.624	细胞核
CsbZIP54	PQ369401	265	29 520.82	5.85	62.92	70.34	−0.680	细胞核
CsbZIP55	PQ369402	290	31 949.60	5.78	60.69	71.76	−0.581	细胞核
CsbZIP56	PQ369403	138	15 767.05	9.50	62.54	85.51	−0.536	细胞核
CsbZIP57	PQ369404	607	65 982.04	6.80	50.34	70.00	−0.610	细胞质
CsbZIP58	PQ369405	263	28 825.78	5.12	61.53	66.05	−0.826	细胞核
CsbZIP59	PQ369406	138	15 801.07	7.97	68.30	76.30	−0.636	细胞核
CsbZIP60	PQ369407	174	19 741.27	9.00	58.71	73.45	−0.875	细胞核
CsbZIP61	PQ369408	317	33 851.46	8.00	52.10	66.91	−0.681	细胞核
CsbZIP62	PQ369409	300	33 027.98	5.95	62.33	76.10	−0.628	叶绿体
CsbZIP63	PQ369410	299	32 687.91	7.81	59.89	66.22	−0.733	细胞核
CsbZIP64	PQ369411	262	29 880.47	7.31	66.99	68.13	−0.897	细胞核

说明 CsbZIPs 蛋白主要为中性和碱性蛋白；蛋白不稳定系数最低为 45.96，均为不稳定蛋白；脂肪系数为 44.21（CsbZIP8）～101.97（CsbZIP45）；亲水性均为负值，均为亲水性蛋白。除 CsbZIP14 及 CsbZIP62 定位到叶绿体、CsbZIP38 定位到线粒体、CsbZIP57 定位到细胞质、CsbZIP52 定位到内质网以外，其余 59 个 bZIP 家族蛋白均定位到细胞核。

3.5　CsbZIPs 蛋白结构分析

CsbZIPs 多序列比对结果表明，64 个 CsbZIPs 蛋白都包含碱性区域（N-X7-R/K 结构）和亮氨酸拉链区域（L-X6-L-X6-L 结构）这 2 个典型的 bZIP 家族结构域（图 4）。另外，发现保守区域部分氨基酸被疏水残基取代，其中 CsbZIP49 碱性区域中的 R/K

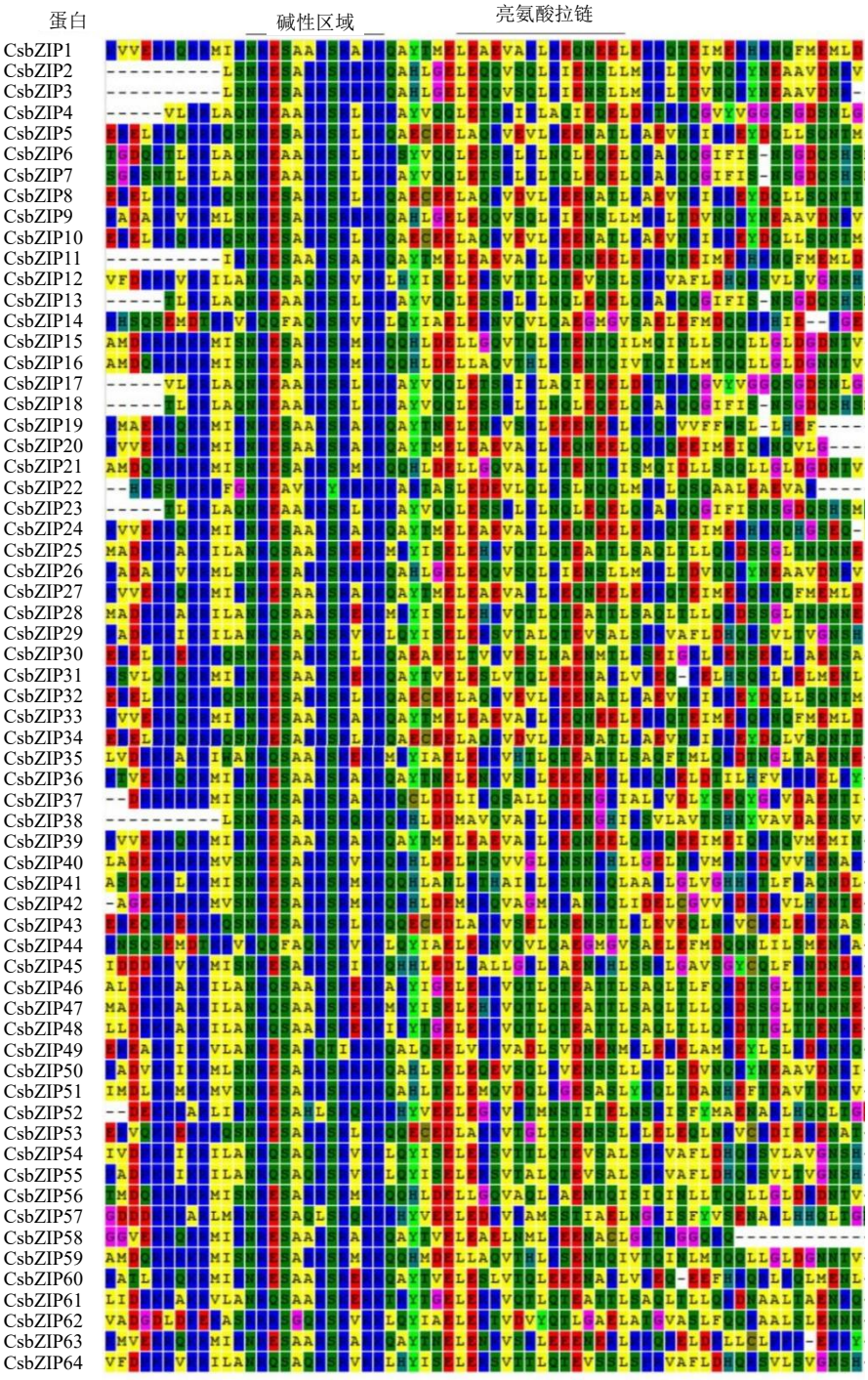
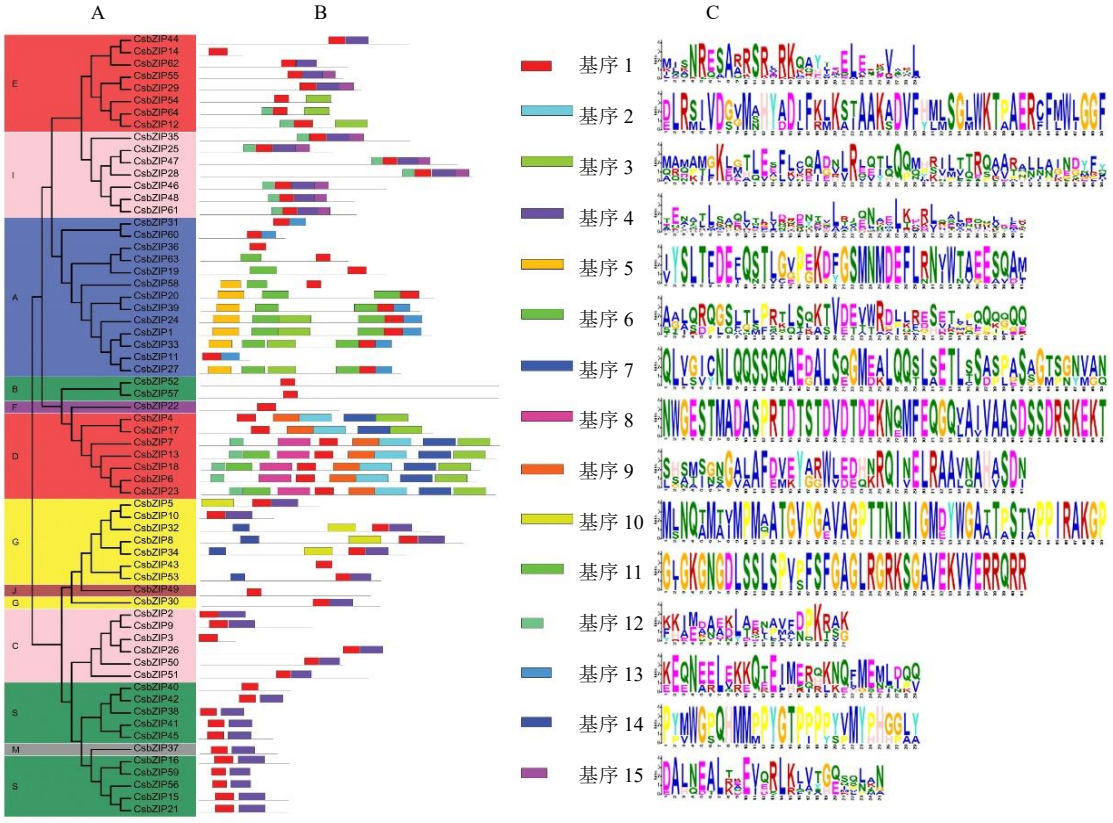


图 4　CsbZIPs 多序列分析

Fig. 4　Sequence alignment of CsbZIPs

被异亮氨酸（I）取代，CsbZIP14 及 CsbZIP44 碱性区域中的 N 被赖氨酸（K）取代，CsbZIP62 碱性区域中的 N 被精氨酸（R）取代；L-X6-L-X6-L 亮氨酸拉链中的部分亮氨酸被蛋氨酸（M）（CsbZIP38、CsbZIP42、CsbZIP49、CsbZIP52、CsbZIP57）、酪氨酸（Y）（CsbZIP62）、缬氨酸（V）（CsbZIP14、CsbZIP44）、异亮氨酸（I）（CsbZIP37、CsbZIP59）等氨基酸取代。

使用 MEME 对 64 个 CsbZIPs 蛋白进行保守基序分析（图 5），共鉴定了 15 个 Motif，所有 CsbZIP 蛋白均含有包含典型碱性区域和亮氨酸拉链区域的 Motif1，Motif2、Motif8 和 Motif9 只存在于 D 亚族，Motif10 仅存在于 G 亚族部分成员中，Motif5 和 Motif13 只存在于 A 亚族。



A-CsbZIP 蛋白系统进化树；B-保守基序分布；C-基序序列组成。

A-phylogenetic relationships of CsbZIP proteins; B-distribution of conserved motifs; C-amino acid sequence of motifs.

图 5　西红花 CsbZIPs 蛋白系统进化、保守基序分布和基序序列组成

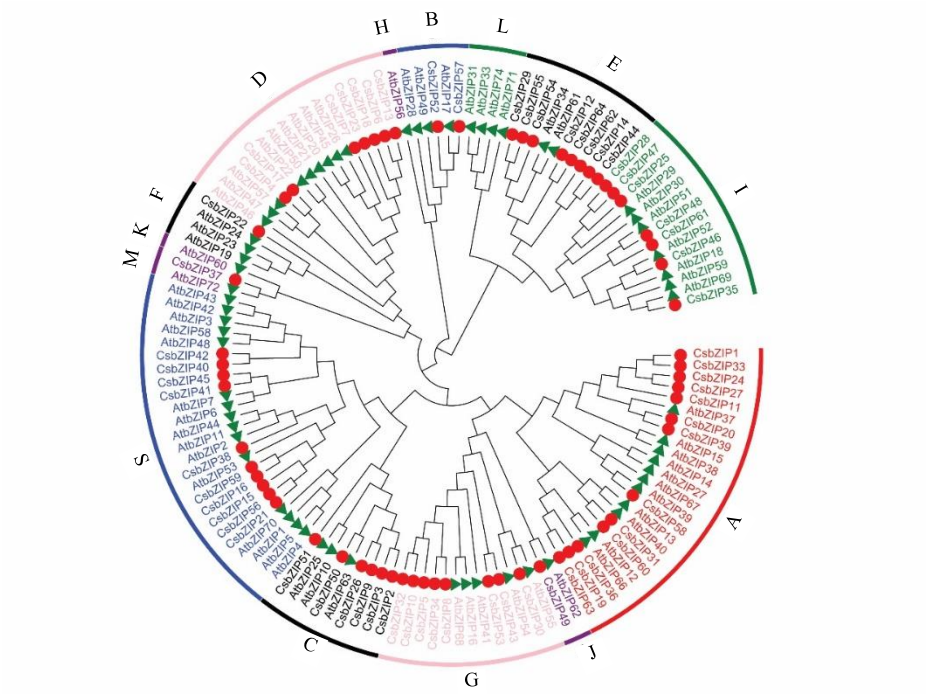
Fig. 5　Phylogenetic relationships distribution and sequence of conserved motif of CsbZIP proteins in *C. sativus*

3.6　CsbZIPs 系统进化分析

使用 MEGA 7.0 软件对 64 个西红花 CsbZIPs 与 66 个拟南芥 AtbZIPs 进行序列同源性比对，采用邻接法构建系统进化树，再通过 iTOL 进行美化处理（图 6）。根据拟南芥 AtbZIPs 的分组规则，将所有 bZIPs 分为 13 个亚族（A～K、M 和 S），除了 K 和 H 亚族无西红花 bZIP 成员以外，其他 11 个亚族都同时含有拟南芥和西红花 bZIP 成员。其中，S 亚族成员最多，包含 10 个 CsbZIP 成员及 15 个 AtbZIP 成员；A 亚族含 CsbZIPs 最多（13 个），该亚族含有 11 个 AtbZIP 成员；K 和 H 亚族均只包含 1 个 AtbZIP 成员。

3.7　*CsbZIPs* 基因表达模式分析

通过课题组前期西红花不同发育期柱头转录组测序数据分析，获得 30 个 *CsbZIPs* 转录因子基因在不同发育期柱头中的表达数据。表达量热图（图 7-a）显示，随着开花期的延续，有 3 个 *CsbZIPs*（*CsbZIP41*、*CsbZIP61*、*CsbZIP43*）表达量逐渐增加，7 个 *CsbZIPs*（*CsbZIP56*、*CsbZIP59*、*CsbZIP64*、*CsbZIP45*、*CsbZIP38*、*CsbZIP46*、*CsbZIP49*）表达量逐渐减小。同一亚族成员基因表达模式相近的有 5 组，包括 I 亚族（*CsbZIP35*、*CsbZIP47*）、S 亚族



A～K、M 和 S 表示 CsbZIP 的亚族分类。

A—K, M, and S represent the subfamily classification of CsbZIP protein members.

图 6　西红花与拟南芥 bZIP 成员系统进化分析

Fig. 6　Phylogenetic relationships of *C. sativus* and *Arabidopsis* bZIP members

（*CsbZIP40*、*CsbZIP41*）、B 亚族（*CsbZIP52*、*CsbZIP57*）、E 亚族（*CsbZIP55*、*CsbZIP62*）和 S 亚族（*CsbZIP38*、*CsbZIP45*）。另外，CsbZIPs 转录因子基因与西红花苷生物合成功能基因的共表达分析发现，*CsbZIP36*、*CsbZIP55*、*CsbZIP56* 及 *CsbZIP62* 与 *CsUGT74AD1* 基因表达模式相近；*CsbZIP54*、*CsbZIP58* 与 *CsBCH1*、*CsCCD2L* 及 *CsLCYB2a* 的基因表达模式相近；*CsbZIP37*、*CsbZIP42* 及 *CsbZIP44* 与 *CsALDH3I1* 基因表达模式相近。

为了进一步验证转录组表达模式分析结果，随机选取 5 个 *CsbZIPs* 基因（*CsbZIP35*、*CsbZIP47*、*CsbZIP49*、*CsbZIP51*、*CsbZIP52*）利用 qRT-PCR 分析了其在 5 个不同发育期柱头中的表达情况（图 7-b）。结果显示，这 5 个基因的表达模式与转录组结果一致，说明转录组数据可靠。另外，利用 qRT-PCR 分析了 5 个功能基因（*CsBCH1*、*CsCCD2L*、*CsLYCB2a*、*CsUGT74AD1*、*CsALDH3I1*）与其表达模式相近的 *CsbZIPs* 基因的表达情况，结果显示 6 个 *CsbZIPs* 基因在 5 个不同发育期柱头中的表达趋势与功能基因的表达趋势无显著差异（图 7-c～e）。

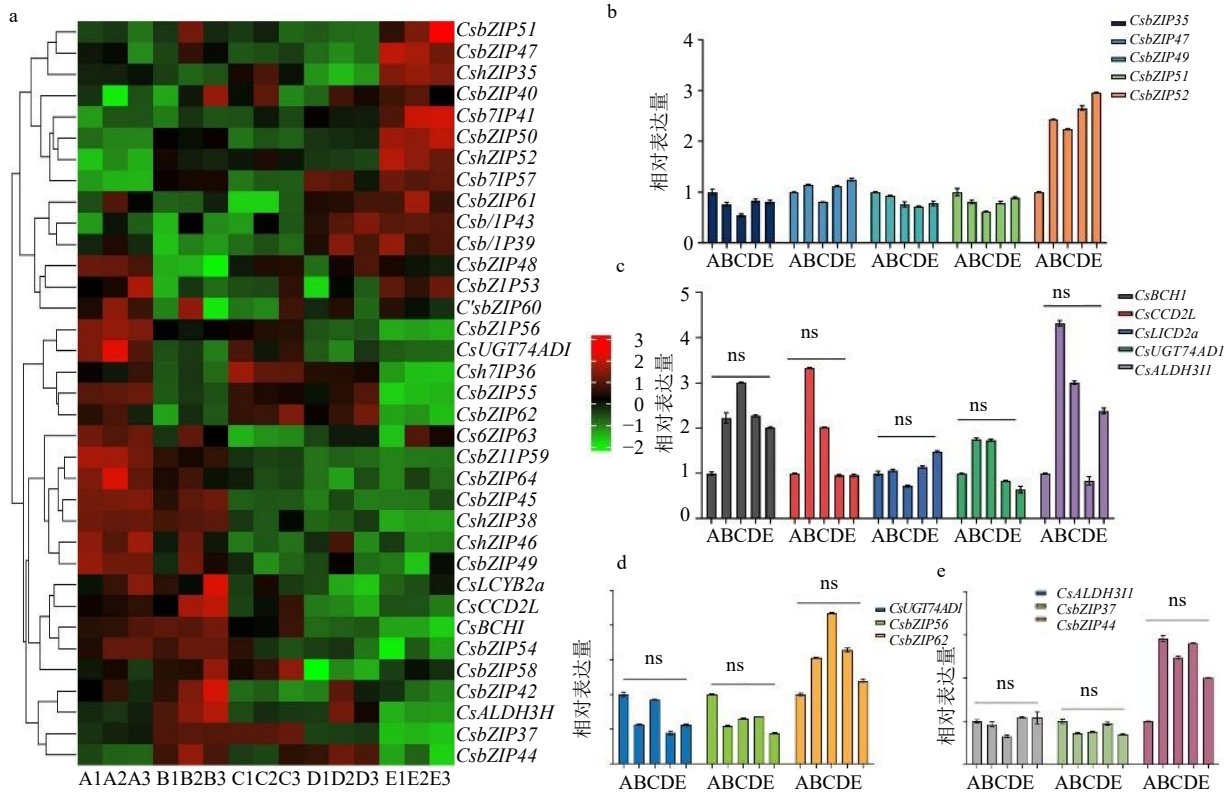
3.8　转录组与代谢组联合分析

通过对不同发育时期柱头的转录组和代谢组联合分析，获得了 *CsbZIPs* 转录因子基因与西红花苷及生物合成相关次生代谢产物的相关性（图 8）。结果发现 13 条 *CsbZIPs* 基因与西红花苷及生物合成相关次生代谢产物的相关性较强，相关系数大于 0.5，其中 12 条相关性差异显著（$P<0.05$）。其中 *CsbZIP38* 与西红花酸呈极显著的负相关，相关系数为−0.775（$P<0.001$），同时它还与苦西红花素、脱落酸和西红花苷 IV 存在显著的负相关性。其次 *CsbZIP49* 与脱落酸也呈现显著的负相关，相关系数为−0.653（$P<0.01$）。另外 *CsbZIP64* 与西红花酸和西红花苷 IV 存在显著的负相关性，*CsbZIP51* 与西红花苷 I 和西红花酸二醛存在显著的正相关性，不与其它代谢物存在明显的相关性。以上结果说明 *CsbZIPs* 基因通过调控功能基因的表达从而影响次生代谢产物在柱头发育过程中的积累。

4　讨论

4.1　西红花全长转录组测序

西红花作为一种名贵中药材，对其主要功能成分西红花苷生物合成途径的研究一直是国内外研

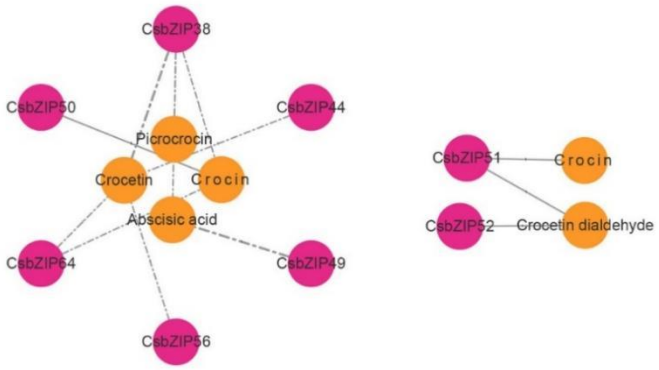


a-基因表达量热图；b～e-qRT-PCR 表达分析；ns 表示差异不显著。

a-Heatmap of gene expression; b—e-qRT-PCR expression analysis; ns expression the difference is not significant.

图 7　西红花 *CsbZIPs* 基因及功能基因在不同发育期表达模式分析

Fig. 7　Gene expression patterns of *CsbZIPs* and functional genes in *C. sativus* at different developmental stages



实线代表正相关，虚线代表负相关；线的宽度代表相关系数，相关系数越大线越粗。

Solid and dashed lines represent positive and negative correlations, respectively; the width of the lines reflects the strength of the correction, the larger the correlation coefficient, the thicker the line.

图 8　西红花 *CsbZIPs* 基因与西红花苷相关代谢产物联合分析

Fig. 8　Correlation analysis of *CsbZIPs* gene and crocin related metabolites in *C. sativus*

究学者广泛关注的热点，从西红花基因组及转录组水平解析西红花苷生物合成及代谢调控机制研究已有多篇报道。Ambardar 等[3]和 Xu 等[4]先后以西红花叶片为材料，开展了西红花基因组测序，获得的基因组大小从 3.5 Gb 增加至 7.59 Gb，为后续功能基因挖掘奠定了良好基础。卢晓慧等[24]开展了西红花柱头 miRNA 测序，鉴定出 164 条保守 miRNA。吴瑞等[25-26]以 MeJA 处理的西红花柱头全长转录组测序数据为基础，对 CsR2R3-MYB 转录因子和 bHLH 转录因子进行了鉴定和系统分析。本研究首

次采用 PacBio Sequel 高通量测序技术对 5 个不同发育期柱头进行了全长转录组测序，获得的转录本平均长度 2 358 bp，N50 长度为 3 594 bp，获得的全长转录本集中在柱头组织中并且覆盖不同发育期，为后期进一步开展西红花功能基因研究提供了丰富的数据基础。

4.2 西红花 bZIP 家族成员分析

bZIP 转录因子广泛参与植物次生代谢产物合成等生物学过程，在葡萄[11]、柿子[12]、黄花蒿[13]、紫花苜蓿[14]等植物的功能性次生代谢产物合成中的作用已被广泛研究。*bZIP* 基因家族数目在不同植物中变化较大，拟南芥中鉴定出 78 个 bZIPs[10]，随着基因组与转录组测序技术的发展，现已在掌叶大黄（63 个）[8]、艾（156 个）、丹参（70 个）[27]、紫花苜蓿（138 个）[28]等多种药用植物中得以广泛鉴定。本研究基于西红花柱头全长转录组数据，鉴定出 64 个 *CsbZIPs* 转录因子基因，在数量上与其他药用植物相比偏少，可能与取样组织单一有关。CsbZIPs 蛋白主要为定位到细胞核中的不稳定亲水蛋白，等电点 5.01～12.16，这与掌叶大黄[8]、鱼腥草[29]及单叶蔷薇[30]等植物中的 bZIP 蛋白相似。系统进化分析表明，CsbZIP 家族成员分布在除 K 和 H 亚族以外的 11 个亚族内，其中含有 CsbZIP 家族成员最多的 2 个亚族为 A 亚族（13 个）和 S 亚族（10 个），这 2 个亚族分别参与 ABA 响应和逆境胁迫的转录调控[31]、植物生殖发育以及响应干旱低温等[32]。

4.3 西红花 *CsbZIPs* 基因参与西红花苷生物合成分析

西红花苷生物合成途径包括 3 部分[1]，上游部分是通过非甲羟戊酸（methylerythritol phosphate，MEP）途径产生牻牛儿基焦磷酸（geranylgeranyl diphosphate，GGPP）；中游部分是通过类胡萝卜素代谢途径，将 GGPP 生成玉米黄质等类胡萝卜素类化合物，关键酶包括八氢番茄红素合成酶（phytoene synthase，PSY）、八氢番茄红素脱氢酶(phytoene desaturase，PDS）、胡萝卜脱氢酶（ζ-carotene desaturase，ZDS）、番茄红素 β-环化酶（lycopene β-cyclase，LCYB）及 β-胡萝卜素羟化酶（β-carotene hydroxylase，BCH）等；下游部分是类胡萝卜素的降解、氧化及糖基化反应途径，将玉米黄质生成西红花酸二醛、西红花酸及西红花苷，关键酶包括类胡萝卜素裂解双加氧酶（carotenoid cleavage dixoygenase，CCD）、乙醛脱氢酶（aldehyde dehydrogenase，ALDH）及 UDP-糖基转移酶（UDP-glucuronosyltransferase，UGT）；以玉米黄质为前体物质，与西红花苷产生竞争性的途径是通过玉米黄质环氧酶（zeaxanthin epoxidase，ZEP）等酶合成脱落酸（abscisic acid，ABA）。结合基因表达模式分析及转录组与代谢组联合分析发现，与西红花苷合成功能基因表达模式相近的基因以及与西红花苷相关次生代谢产物合成相关的基因主要分布在 E 亚族（*CsbZIP44*、*CsbZIP54*、*CsbZIP55*、*CsbZIP62*、*CsbZIP64*）、S 亚族（*CsbZIP38*、*CsbZIP42*、*CsbZIP56*）以及 A 亚族（*CsbZIP36*、*CsbZIP58*）。研究发现 E 亚族的 *AtbZIP34* 基因与花粉萌发和花粉管生长相关[33]。因此，E、S 及 A 亚族的 *CsbZIPs* 基因，尤其是 E 亚族的 *CsbZIP44*、*CsbZIP54*、*CsbZIP55*、*CsbZIP62*、*CsbZIP64* 等基因可能参与西红花苷生物合成的转录调控，其具体功能还有待深入研究。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 浦香东, 徐志超, 高冉冉, 等. 西红花苷生源途径解析及其合成生物学研究进展 [J]. 科学通报, 2021, 66(2): 219-232.

[2] 李军, 吴霁蓁, 高广春, 等. 西红花多柱头突变体 Cs5 的转录组分析 [J]. 江苏农业学报, 2021, 37(6): 1630-1632.

[3] Ambardar S, Vakhlu J, Sowdhamini R. Insights from the analysis of draft genome sequence of *Crocus sativus* L [J]. *Bioinformation*, 2022, 18(1): 1-13.

[4] Xu Z C, Chen S S, Wang Y L, *et al*. *Crocus* genome reveals the evolutionary origin of crocin biosynthesis [J]. *Acta Pharm Sin B*, 2024, 14(4): 1878-1891.

[5] Frusciante S, Diretto G, Bruno M, *et al*. Novel carotenoid cleavage dioxygenase catalyzes the first dedicated step in saffron crocin biosynthesis [J]. *Proc Natl Acad Sci U S A*, 2014, 111(33): 12246-12251.

[6] Tan H X, Chen X H, Liang N, *et al*. Transcriptome analysis reveals novel enzymes for apo-carotenoid biosynthesis in saffron and allows construction of a pathway for crocetin synthesis in yeast [J]. *J Exp Bot*, 2019, 70(18): 4819-4834.

[7] Demurtas O C, Frusciante S, Ferrante P, *et al*. Candidate enzymes for saffron crocin biosynthesis are localized in multiple cellular compartments [J]. *Plant Physiol*, 2018, 177(3): 990-1006.

[8] 唐璟, 杜桥, 胡晓晨, 等. 基于全长转录组测序的掌叶大黄 bZIP 基因家族鉴定分析 [J]. 草业科学, 2024, 41(3): 638-651.

[9] Glover J N, Harrison S C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA [J]. *Nature*, 1995, 373(6511): 257-261.

[10] Wolfgang D L, Snoek B L, Snel B, *et al*. The *Arabidopsis* bZIP transcription factor family-an update [J]. *Curr Opin Plant Biol*, 2018, 45(Pt A): 36-49.

[11] Malacarne G, Coller E, Czemmel S, *et al*. The grapevine VvibZIPC22 transcription factor is involved in the regulation of flavonoid biosynthesis [J]. *J Exp Bot*, 2016, 67(11): 3509-3522.

[12] Akagi T, Katayama-Ikegami A, Kobayashi S, *et al*. Seasonal abscisic acid signal and a basic leucine zipper transcription factor, DkbZIP5, regulate proanthocyanidin biosynthesis in persimmon fruit [J]. *Plant Physiol*, 2012, 158(2): 1089-1102.

[13] Shu G P, Tang Y L, Yuan M Y, *et al*. Molecular insights into AabZIP1-mediated regulation on artemisinin biosynthesis and drought tolerance in *Artemisia annua* [J]. *Acta Pharm Sin B*, 2022, 12(3): 1500-1513.

[14] Ribeiro B, Erffelinck M L, Lacchini E, *et al*. Interference between ER stress-related bZIP-type and jasmonate-inducible bHLH-type transcription factors in the regulation of triterpene saponin biosynthesis in *Medicago truncatula* [J]. *Front Plant Sci*, 2022, 13: 903793.

[15] 王丰青, 杨旭, 左鑫, 等. 地黄全长转录组测序及苯乙醇苷合成途径催化酶基因鉴定 [J]. 药学学报, 2022, 57(3): 831-838.

[16] 米琪, 赵艳丽, 徐萍, 等. 滇黄精全长转录组测序及生物信息学分析 [J/OL]. 药学学报, (2024-03-02), https://doi. org/10.16438/j.0513-4870.2023-1259.

[17] 徐萍, 米琪, 罗文秀, 等. 滇重楼种子的全长转录组测序与休眠基因的挖掘 [J/OL]. 时珍国医国药, (2024-05-06), https://link.cnki.net/urlid/ 42.1436.R.20240319.100.002.

[18] 周先治, 饶宝蓉, 高晖, 等. 基于 DNA 条形码的多花黄精系统发育和变异位点分析研究 [J]. 中草药, 2020, 51(15): 4003-4010.

[19] Hu J, Liu Y P, Tang X H, *et al*. Transcriptome profiling of the flowering transition in saffron (*Crocus sativus* L.) [J]. *Sci Rep*, 2020, 10(1): 9680.

[20] Yue J Y, Wang R, Ma X J, *et al*. Full-length transcriptome sequencing provides insights into the evolution of apocarotenoid biosynthesis in *Crocus sativus* [J]. *Comput Struct Biotechnol J*, 2020, 18: 774-783.

[21] Gao G C, Wu J M, Li B, *et al*. Transcriptomic analysis of saffron at different flowering stages using RNA sequencing uncovers cytochrome P450 genes involved in crocin biosynthesis [J]. *Mol Biol Rep*, 2021, 48(4): 3451-3461.

[22] Ahrazem O, Argandoña J, Fiore A, *et al*. Multi-species transcriptome analyses for the regulation of crocins biosynthesis in *Crocus* [J]. *BMC Genomics*, 2019, 20(1): 320.

[23] Pathak J, Chettry U, Chrungoo N K, *et al*. RNA-Seq analysis reveals the role of MYB12, MYB111 and MBW complex repressors in regulation of flavonoid biosynthesis in stigmas of saffron (*Crocus sativus* L.) flowers [J]. *Acta Physiol Plant*, 2022, 44(4): 42.

[24] 卢晓慧, 朱志明, 岳俊阳. 藏红花柱头小 RNA 高通量测序及生物信息学分析 [J]. 植物生理学报, 2018, 54(5): 827-836.

[25] 吴瑞, 林定, 罗栋, 等. 西红花R2R3-MYB转录因子的鉴定与时空表达分析 [J]. 农业生物技术学报, 2022, 30(3): 457-472.

[26] 吴瑞. 调控西红花苷生物合成的 bHLH 转录因子筛选及克隆分析 [D]. 杭州: 浙江农林大学, 2022.

[27] Zhang Y, Xu Z C, Ji A J, *et al*. Genomic survey of bZIP transcription factor genes related to tanshinone biosynthesis in *Salvia miltiorrhiza* [J]. *Acta Pharm Sin B*, 2018, 8(2): 295-305.

[28] 齐晓, 张正社, 闵学阳, 等. 紫花苜蓿bZIP基因家族的鉴定、进化及表达分析 [J]. 草业科学, 2017, 34(8): 1635-1648.

[29] 蔡伟, 肖敬忠, 余小丽, 等. 鱼腥草*bZIP*基因的鉴定与分析 [J]. 中草药, 2023, 54(13): 4295-4305.

[30] 孙彦琳, 于超, 罗乐, 等. 单叶蔷薇bZIP转录因子家族鉴定与表达分析 [J]. 西北农林科技大学学报 (自然科学版), 2022, 50(6): 82-92.

[31] Fujita Y, Fujita M, Satoh R, *et al*. AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in *Arabidopsis* [J]. *Plant Cell*, 2005, 17(12): 3470-3488.

[32] Llorca C M, Berendzen K W, Malik W A, *et al*. The elucidation of the interactome of 16 *Arabidopsis* bZIP factors reveals three independent functional networks [J]. *PLoS One*, 2015, 10(10): e0139884.

[33] Gibalová A, Renák D, Matczuk K, *et al*. AtbZIP34 is required for *Arabidopsis* pollen wall patterning and the control of several metabolic pathways in developing pollen [J]. *Plant Mol Biol*, 2009, 70(5): 581-601.

[责任编辑 时圣明]