

网络出版时间:2019-01-24 16:29

DOI:10.13207/j.cnki.jnwfufu.2019.08.008

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1390.S.20190124.1627.016.html>

冬瓜高通量转录组测序及分析

叶新如,张前荣,陈敏氢,王 彬,刘建汀,朱海生,温庆放

(福建省农业科学院 作物研究所,福建 福州 350013)

【摘 要】【目的】获得冬瓜转录组序列、遗传变异等信息,从中挖掘冬瓜基因数据及 SSR 分子标记,为冬瓜后续研究提供数据支撑。【方法】以冬瓜嫩叶为材料,利用 Illumina HiSeq™ 2000 技术对冬瓜进行转录组测序,构建数据库从中获得干净序列。经 De novo 拼接组装后,将获得的单基因簇(Unigene)数据在非冗余蛋白数据库(nonredundant protein database,Nr)、蛋白质序列数据库(Swiss Prot protein database,Swiss Prot)、基因本体论数据库(gene ontology,GO)、蛋白质真核同源数据库(eukaryotic orthologous groups,KOG)、东京基因与基因组百科全书(Kyoto encyclopedia of genes and genomes,KEGG)、蛋白质家族域数据库(protein families database,Pfam)6 个公共数据库中进行比对,最终得到冬瓜单基因簇注释信息。利用 MISA 软件对转录组单基因簇进行搜索,获得单基因簇中的 SSR 位点。【结果】从冬瓜嫩叶中得到 62 021 032 条高品质序列,组装后获得 40 611 条单基因簇,平均长度 955 bp。将所有单基因簇在 Nr 和 Swiss Prot 数据库中进行比对,结果分别比对到 27 474 及 19 573 条单基因簇;在 GO 数据库中,所注释到的 10 659 条单基因簇分别匹配到生物功能、分子功能和细胞组分 3 个本体的 47 个功能组中;与 KOG 数据库进行注释比对,根据其功能将注释到的单基因簇划分为 25 类;KEGG 数据库比对注释到 10 799 条冬瓜的单基因簇,可分为 5 个大类、19 个亚类、125 条代谢途径;在 Pfam 数据库中比对到 17 990 条单基因簇,分属于 369 个类群。SSR 位点搜索发现,有 5 086 条单基因簇包含 SSR 序列,获得 5 474 个 SSR 位点。【结论】利用高通量测序获得大量冬瓜转录组信息,有助于从分子水平对冬瓜进行深入研究。

【关键词】 冬瓜;高通量测序;转录组;基因注释

【中图分类号】 S642.301

【文献标志码】 A

【文章编号】 1671-9387(2019)08-0054-11

High-throughput sequencing and analysis of transcriptome of *Benincasa hispida* Cogn

YE Xinru,ZHANG Qianrong,CHEN Mindong,WANG Bin,LIU Jianting,
ZHU Haisheng,WEN Qingfang

(Crops Research Institute,Fujian Academy of Agricultural Sciences,Fuzhou,Fujian 350013,China)

Abstract: 【Objective】The transcriptional sequence,genetic variation and other information of *Benincasa hispida* were obtained and genetic data and SSR markers were extracted to improve researches on *Benincasa hispida*. 【Method】The Illumina HiSeq™ 2000 technology was used for the transcription sequencing of *Benincasa hispida* tender leaves, and a transcriptome database was built to obtain clean reads. The annotation information of the Unigenes was obtained by De novo splicing assembly and compared with public database (Includes Nr,Swiss Prot,GO,KOG,KEGG,and Pfam). MISA software was used to search the transcription group Unigenes and obtain the SSR sites in Unigenes. 【Result】A total of 62 021 032 reads

【收稿日期】 2018-07-13

【基金项目】 福建省属公益类科研院所基本科研专项(2018R1026-2);中央引导地方科技发展专项(2018L3005);福建省农业科学院“青年科技英才百人计划”项目(YC2017-5)

【作者简介】 叶新如(1991—),女,福建福州人,研究实习员,主要从事园艺植物分子生物学研究。E-mail:731122298@qq.com

【通信作者】 朱海生(1978—),男,福建福州人,研究员,博士,主要从事园艺植物生理生化与分子生物学研究。

E-mail:zhs0246@163.com

fragments were generated by sequencing. The reads gave 40 611 Unigenes with an average length of 955 bp. A total of 27 474 and 19 573 Unigenes were annotated against the Nr and Swiss Prot databases, respectively. Gene ontology analysis revealed that annotated 10 659 unigenes were grouped into 47 different categories in terms of biological function, molecular function and cellular component. Unigenes in the transcriptome of *Benincasa hispida* were divided into 25 classes according to the function comparing with the KOG database. Moreover, 10 799 unigenes were annotated to 125 KEGG pathway and broadly divided into 5 categories of 19 branches, and 17 990 unigenes were annotated against Pfam database and formed 369 groups. A total of 5 474 SSRs were identified from the sequence of transcription, distributed in 5 086 unigenes. 【Conclusion】 Large amount of transcriptome information was obtained by high-throughput sequencing, which was helpful for further study of *Benincasa hispida* at molecular level.

Key words: *Benincasa hispida* Cogn; high-throughput sequencing; transcriptome; gene annotation

冬瓜(*Benincasa hispida* Cogn.)原产于中国南部、东南亚以及印度等地,分布于亚洲的热带、亚热带及温带地区^[1],在中国有 1 500 多年的栽培历史。其作为我国重要的蔬菜品种,对调节蔬菜淡季品种稀少、保证蔬菜周年供应有着非常重要的作用。冬瓜产量高,花、叶、果皮、果肉、种子等都具有很好的药食兼用价值^[2],能够用于治疗咳嗽^[3]、消化性溃疡^[4]、高血压、炎症^[3]等。前人对冬瓜的研究局限在栽培技术^[4-5]、药用开发^[6-9]、营养价值^[10-11]、遗传育种^[12-14]等方面,而关于冬瓜基因组学的研究极少,有关冬瓜重要经济性状分子机制的研究进展缓慢。

转录组是特定的细胞或组织在某一发育阶段或功能状态下转录出的所有 RNA 总和^[15]。目前,基于高通量测序的转录组分析成为功能基因研究的主要方法^[16-17]。该技术也被广泛应用于植物上,如拟南芥^[18]、玉米^[19]、香椿^[20]、紫色黄秋葵^[21],但在冬瓜上鲜有关于转录组测序的报道。

本研究以冬瓜嫩叶为材料,利用 Illumina HiSeqTM2000 技术对其进行转录组测序,获得冬瓜转录组序列、遗传变异等信息,从中挖掘冬瓜的基因数据和 SSR 分子标记,以期后续研究提供数据支撑。

1 材料与方法

1.1 试验材料

试验材料为“黑皮冬瓜”,统一播种于福建省农业科学院作物研究所福清市东张镇蔬菜科研基地。采用穴盘育苗,待植株长至三叶一心时,取成型的嫩叶 3 片,等量混样进行转录组测序,测序由广州基迪奥生物有限公司完成。

1.2 冬瓜总 RNA 提取

冬瓜总 RNA 提取采用 Trizol Reagent 法,并对

RNA 纯度、品质进行检测。若提取的 RNA 序列满足如下条件: OD_{260}/OD_{280} 在 1.8 ~ 2.2, $OD_{260}/OD_{230} \geq 2$, $28S/18S \geq 1$, RNA 完整值(RIN) ≥ 7.5 ,则认为 RNA 符合建库测序标准。

1.3 冬瓜转录组测序

提取冬瓜总 RNA 后,用带有 Oligo(dT)的磁珠富集冬瓜 mRNA。以 mRNA 为模板,用六碱基随机引物(random hexamers)合成第一条 cDNA 链;然后加入缓冲液、dNTPs、RNase H 和 DNA polymerase I 合成第二条 cDNA 链;经 QiaQuick PCR 试剂盒纯化及 EB 缓冲液洗脱后做末端修复,加 poly(A)并连接测序接头;最后进行 PCR 扩增,建好的测序文库用 Illumina HiSeqTM2000 进行测序。

1.4 冬瓜转录组序列的拼接与组装

利用 Illumina HiSeqTM2000 测序平台对建立好的冬瓜转录组文库进行测序,得到原始序列(raw reads),其中含有低品质、重复测序以及接头污染的序列,因此必须进行严格过滤,去除低品质序列,获得干净序列(clean reads);统计干净序列的长度、数量、GC 含量等。使用组装软件 Trinity 进行拼接,将有重叠信息(overlap)的序列连接成长片段(contig),通过不断地延伸,拼接组装成转录本(Transcript),然后使用 Tgicl 软件对转录本进行去冗余聚类,最终获得单基因簇(Unigene)。利用 RPKM 法对单基因簇表达量进行计算分析,其计算公式为:

$$RPKM = C \times 10^6 / (NL / 10^3)。$$

式中: C 为比对到基因的 reads 数, N 为比对到所有基因的总 reads 数, L 为基因的碱基数。

1.5 冬瓜单基因簇的基本功能注释

通过 Blastx^[22]将冬瓜的单基因簇序列与公共数据库(E 值 $< 1 \times 10^{-5}$, E 值为单基因簇与数据库中匹配序列为同源序列的假阳性概率)进行比对,得

到与给定单基因簇具有最高序列相似性的蛋白,从而获得该单基因簇的蛋白功能注释信息。公共数据库包括非冗余蛋白数据库(nonredundant protein database,Nr)、蛋白质序列数据库(Swiss Prot protein database, Swiss Prot)、基因本体论数据库(gene ontology,GO)、蛋白质真核同源数据库(eukaryotic orthologous groups,KOG)、东京基因与基因组百科全书(Kyoto encyclopedia of genes and genomes,KEGG)、蛋白质家族域数据库(protein families database,Pfam)。

通过 Blastx 软件将冬瓜单基因簇序列比对至 Nr 数据库,可获得冬瓜单基因簇对应的蛋白编码序列。将冬瓜单基因簇与 Swiss Prot 数据库比对,可确定冬瓜基因簇的功能。利用 Blast2 GO 和 WE GO 软件对冬瓜单基因簇序列进行 GO 注释及功能分类统计,同时与 KOG 数据库比对注释,可统计冬瓜基因功能和分类信息。在 KEGG 数据库中对冬瓜单基因簇进行相似性分析,可获得冬瓜单基因簇代谢通路的注释和预测。利用 Hmmer 3.0 将冬瓜单基因簇与 Pfam 蛋白结构数据库进行比对,可获得冬瓜单基因簇的注释信息。

1.6 冬瓜转录因子分析及 SSR 位点分析

将冬瓜蛋白序列与 TF 数据库(plant TFdb/animal TFdb)进行 hmmscan 比对,可获得相应转录因子家族,以进一步进行转录因子分析。使用软件 MISA 对转录组所有单基因簇中的 SSR 进行搜索,并对 SSR 重复基序进行特征分析,为获得合适的引

物序列提供数据支撑。

2 结果与分析

2.1 冬瓜转录组测序结果

2.1.1 转录组测序及组装后不同长度单基因簇的分布 用高通量测序平台对冬瓜嫩叶测序,共获得 63 525 954 条原始序列,碱基数据量 9 528 893 100 bp(9.5 Gb)。去除低品质的原始序列后,获得 62 021 032 条高品质干净序列,占原始序列的 97.63%。碱基数据量 9 154 299 990 bp(9.1 Gb),G+C 含量为 46.88%,碱基 Q20(即测序错误率小于 1%的比例)和 Q30(即测序错误率小于 0.1%的比例)分别为 98.31%及 94.96%。测序结果表明数据合格,可以用于后续分析。

利用 Trinity 软件对干净序列进行组装,获得 40 611 条单基因簇,删除 7 条比对到核糖体的基因后,对剩余的 40 604 条单基因簇进行统计及注释,G+C 含量为 41.64%,覆盖 50%核苷酸的最大单基因簇长度(N50)为 1 591 bp,最短和最长单基因簇分别为 201 及 13 773 bp,平均长 955 bp,不同长度单基因簇的分布情况见图 1。由表 1 可知,长 200~999 bp 的单基因簇占总数(40 604 条)的 66.96%,长 1 000~1 999 bp 的单基因簇占比为 20.93%,长 2 000~2 999 bp 的单基因簇占比为 7.95%,长度≥3 000 bp 的单基因簇占比为 4.16%。总体而言,随着长度增加,单基因簇数量减少。

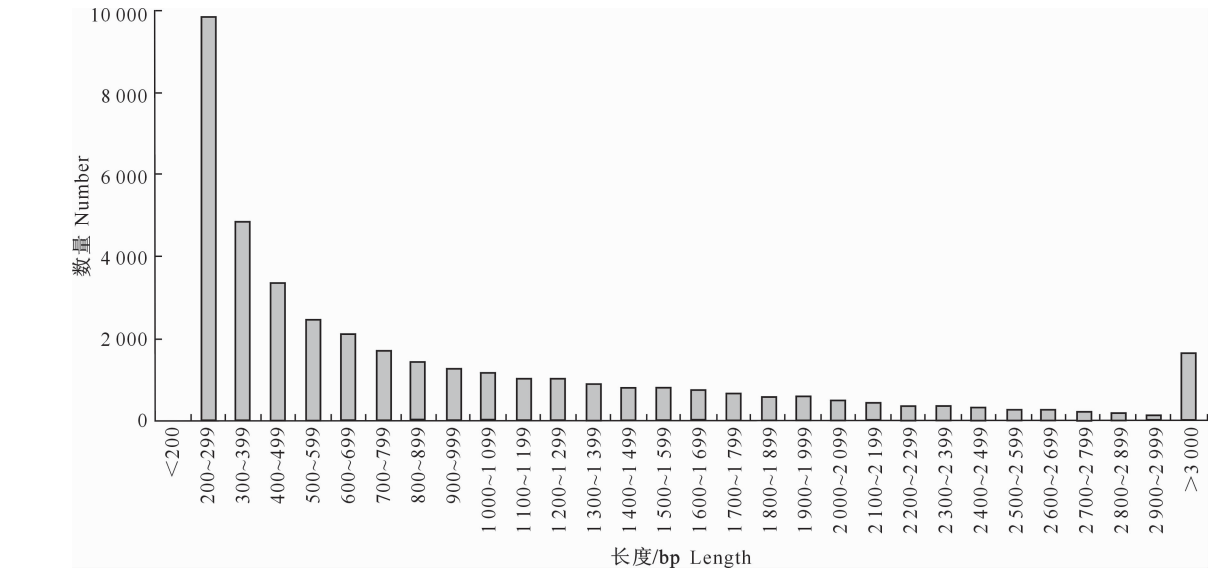


图 1 冬瓜单基因簇的长度分布

Fig. 1 Length distribution of Benincasa hispida unigenes

2.1.2 单基因簇的总体表达量 测序得62 021 032条干净序列,能比对到单基因簇序列的 reads 有56 708 038条,占干净序列总数的 91.43%,其中比对到唯一指定单基因簇序列的 reads 有 49 628 128条,占 80.01%。利用 RPKM 法计算单基因簇的表达量,结果表明,冬瓜 40 611 条单基因簇序列 RPKM平均值为 16.27,最大值 51 049.34(Unigene-0022103),最小值 0.011 4(不包含 0,RPKM 值为 0 的单基因簇共 300 条)。

表 1 冬瓜单基因簇在 6 大数据库中的注释结果
Table 1 Statistics of database annotations of *Benincasa hispida* unigenes

指标 Index	Nr	Swiss Prot	GO	KOG	KEGG	Pfam	总基因数 Total number of genes
基因数量 Number of genes	27 474	19 572	10 659	16 452	10 799	17 990	40 604
占比/% Percentage	67.66	48.20	26.25	40.52	26.60	44.31	100

2.3 冬瓜单基因簇的 Nr 功能注释

冬瓜 40 604 条单基因簇在 Nr 数据库中的比对结果(图 2-A)显示,其中 27 474 条单基因簇在 Nr 数据库中比对到相似序列, $0 \leq E \leq 10^{-150}$ 的数量最多,有 10 367 条,占 27 474 条单基因簇的 37.73%,其中 E 值为 0 的单基因簇有 8 752 条(占 31.85%), $10^{-150} < E \leq 10^{-100}$ 的单基因簇有 3 194 条(占 11.63%),说明比对结果可靠度较高。从 Nr 数据

2.2 冬瓜单基因簇的总体注释结果

通过 Blastx 将测序获得的 40 604 条冬瓜单基因簇与相关数据库进行比对,结果(表 1)显示,Nr 数据库注释到最多的单基因簇,共 27 474 条,占总单基因簇的 67.66%;GO 数据库注释到的单基因簇最少,共 10 659 条,占 26.25%;Swiss Prot、KOG、KEGG、Pfam 分别注释到 19 572、16 452、10 799 和 17 990 条单基因簇。

库匹配到的同源序列物种分布情况(图 2-B)来看,注释到甜瓜(*Cucumis melo*)的单基因簇数量最多,共 11 569 条(占 Nr 注释总单基因簇的 42.11%),黄瓜(*Cucumis sativus*)8 234 条(占 29.97%),土瓶草(*Cephalotus follicularis*)864 条(占 3.14%),可可(*Theobroma cacao*)665 条(占 2.42%),剩余的 22.36%分布于其他物种。

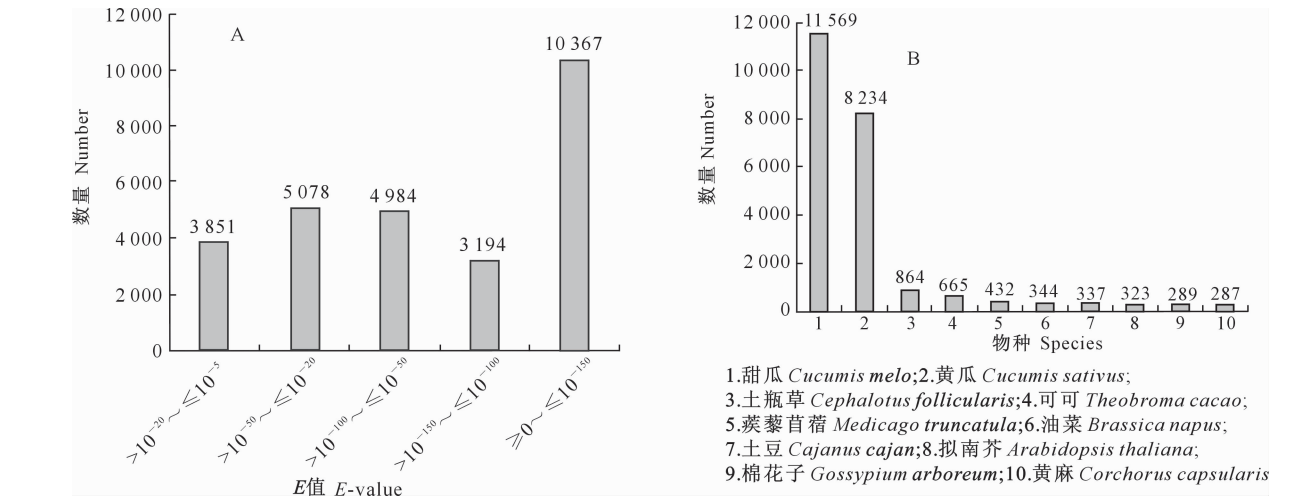


图 2 冬瓜单基因簇在 Nr 数据库中的 E 值分布(A)和物种分布(B)

Fig. 2 E-value (A) and species distribution (B) of *Benincasa hispida* unigenes in Nr database

2.4 冬瓜单基因簇的 Swiss Prot 统计分析

在 Swiss Prot 数据库中,共有 19 572 条单基因簇成功注释到,占 40 604 条单基因簇总数的 48.20%,匹配序列相似度(identity)≥80%的单基因簇有 14 456 条,匹配序列相似度在 30%~80%的单基因簇有 222 482 条,匹配序列相似度≤30%的

单基因簇有 37 282 条。

2.5 冬瓜单基因簇的 GO 功能分类

GO 功能注释是根据 Nr 注释信息得到的,是对基因功能进行分类的体系。冬瓜单基因簇的 GO 功能分类结果见表 2。

表 2 冬瓜单基因簇 GO 功能分类

Table 2 GO functional categories of *Benincasa hispida* unigenes

功能分类 Functional classification	分类名 Class name	注释数量 Number	占比/% Percentage
生物功能 Biological process	繁殖 Reproduction	385	1.53
	免疫系统进程 Immune system process	85	0.31
	代谢进程 Metabolic process	6 156	24.53
	细胞进程 Cellular process	5 711	22.76
	繁殖进程 Reproductive process	380	1.51
	生物黏附 Biological adhesion	30	0.12
	信号传导 Signaling	425	1.69
	多细胞进程 Multicellular organismal process	574	2.29
	发育进程 Developmental process	804	3.20
	生长 Growth	85	0.34
	细胞定位 Locomotion	3	0.01
	单一有机体进程 Single-organism process	4 344	17.31
	节律进程 Rhythmic process	12	0.05
	应急反应 Response to stimulus	1 485	5.92
	定位 Localization	1 462	5.82
	多个有机体进程 Multi-organism process	215	0.88
	生物调节 Biological regulation	1 785	7.11
	细胞组分分布或生物合成 Cellular component organization or biogenesis	1 141	4.55
	脱毒 Detoxification	11	0.04
分子功能 Molecular function	转录因子活性、蛋白质结合 Transcription factor activity,protein binding	11	0.10
	核苷酸结合的转录因子活性 Nucleic acid binding transcription factor activity	135	1.19
	催化活性 Catalytic activity	6 029	53.26
	信号传感器活动 Signal transducer activity	96	0.85
	结构分子活动 Structural molecule activity	95	0.84
	转运活性 Transporter activity	512	4.52
	结合活性 Binding	4 314	38.11
	电子载体活性 Electron carrier activity	12	0.10
	抗氧化活性 Antioxidant activity	24	0.21
	翻译调控活性 Translation regulator activity	1	0.10
	分子转导活性 Molecular transducer activity	54	0.48
	分子功能调节 Molecular function regulator	36	0.31
细胞组分 Cellular component	胞外区域 Extracellular region	99	0.63
	细胞 Cell	3 541	22.41
	拟核 Nucleoid	3	0.02
	膜结构 Membrane	2 095	13.26
	病毒体 Virion	33	0.21
	细胞连接 Cell junction	164	1.04
	胞外基质 Extracellular matrix	2	0.01
	膜封闭内腔 Membrane-enclosed lumen	66	0.42
	大分子复合体 Macromolecular complex	886	5.61
	细胞器 Organelle	2 655	16.80
	胞外区域部分 Extracellular region part	2	0.01
	细胞器部分 Organelle part	1 149	7.27
	病毒体部分 Virion part	33	0.21
	膜部分 Membrane part	1 528	9.67
	细胞部分 Cell part	3 541	22.40
	超分子纤维 Supramolecular fiber	2	0.01

从表 2 可以看出,冬瓜 40 604 条单基因簇中有 10 659 条单基因簇被注释到 GO 数据库内,分别匹配到生物功能、分子功能和细胞组分 3 个本体的 47 个功能组中。根据 GO 功能分类,注释到生物功能

的有 25 093 条单基因簇(占 3 个本体总数(52 211 条单基因簇)的 48.06%),注释到分子功能的有 11 319 条单基因簇(占 21.68%),注释到细胞组分的有 15 799 条单基因簇(占 30.26%)。其中生物功能

包含 19 个功能组,以代谢进程(6 156 条)、细胞进程(5 711 条)和单一有机体进程(4 344 条)占比较大,分别占 25 093 条生物功能单基因簇的 24. 53%, 22. 76%和 17. 31%,而节律进程(12 条)、脱毒(11 条)、细胞定位(3 条)注释到的单基因簇较少;分子功能包含 12 个功能组,催化活性(6 029 条)和结合活性(4 314 条)占比较大,分别占 11 319 条单基因簇的 53. 26%和38. 11%,电子载体活性(12 条)、转录因子活性(11 条)、翻译调控活性(1 条)比例较小;细胞组分包含 16 个功能组,细胞、细胞部分和细胞器占比较大,细胞、细胞部分各注释到 3 541 条单基因簇,细胞器注释到 2 655 条单基因簇,胞外基质、胞外区域部分和超分子纤维注释的单基因簇较少,均仅为 2 条。

表 3 冬瓜单基因簇 KOG 功能分类
Table 3 KOG functional categories of *Benincasa hispida* Cogn unigenes

功能类别 Functional-Categories	注释数量 Number	占比/% Percentage
RNA 加工和修饰 RNA processing and modification	1 605	5. 74
染色质结构与变化 Chromatin structure and dynamics	425	1. 52
能量生产和转换 Energy production and conversion	1 023	3. 66
细胞周期控制、细胞分裂、染色体划分 Cell cycle control, cell division, chromosome partitioning	696	2. 49
氨基酸运输和代谢 Amino acid transport and metabolism	832	2. 98
核苷酸运输和代谢 Nucleotide transport and metabolism	252	0. 90
碳水化合物的运输和代谢 Carbohydrate transport and metabolism	1 125	4. 02
辅酶运输和代谢 Coenzyme transport and metabolism	210	0. 75
脂质运输和代谢 Lipid transport and metabolism	1 028	3. 68
翻译、核糖体结构和生物合成 Translation, ribosomal structure and biogenesis	1 206	4. 31
转录 Transcription	1 469	5. 26
复制、重组和修复 Replication, recombination and repair	837	2. 99
细胞壁/膜生物发生 Cell wall/membrane/envelope biogenesis	303	1. 08
细胞活性 Cell motility	11	0. 04
翻译后修饰、蛋白质折叠、分子伴侣 Posttranslational modification, protein turnover, chaperones	3 112	11. 14
无机离子的运输和代谢 Inorganic ion transport and metabolism	567	2. 03
次生代谢产物生物合成、转运和分解代谢 Secondary metabolites biosynthesis, transport and catabolism	771	2. 76
一般功能基因 General function prediction only	5 975	21. 38
未知功能 Function unknown	1 273	4. 55
信号传导机制 Signal transduction mechanisms	2 773	9. 92
膜泡运输和胞内分泌 Intracellular trafficking, secretion, and vesicular transport	1 460	5. 22
防御机制 Defense mechanisms	187	0. 67
胞外结构 Extracellular structures	105	0. 37
核结构 Nuclear structure	91	0. 32
细胞骨架 Cytoskeleton	608	2. 18

2. 7 冬瓜单基因簇的 KEGG 代谢通路分析

将测序得到的 40 604 条冬瓜单基因簇比对到 KEGG 数据库,得到单基因簇的 Pathway 注释(表略)。结果表明,注释匹配到的 10 799 条冬瓜单基因簇,可分为 5 个大类、19 个亚类、125 条代谢途径。5 个大类中代谢注释到的单基因簇最多,共 5 868 条,占带有通路注释的 9 623 条单基因簇的

2. 6 冬瓜单基因簇的 KOG 功能分类

将冬瓜单基因簇比对到 KOG 数据库中发现,共有 16 452 条单基因簇注释到 KOG 数据库,共获得 27 944 条单基因簇,被注释到 25 种 KOG 分类中,功能区中单基因簇数量分布有所差异(表 3),其中一般功能基因注释到的单基因簇数量最多,包含 5 975 条单基因簇,占 27 944 条单基因簇的 21. 38%;其次是翻译后修饰、蛋白质折叠和分子伴侣共包含 3 112 条单基因簇,占 11. 14%;信号传导机制(2 773 条)、RNA 加工和修饰(1 605 条)、转录(1 469 条)、膜泡运输和胞内分泌(1 460 条)注释到的单基因簇也较多,最少的是细胞活性,仅有 11 条单基因簇(占 0. 04%)。

60. 98%;遗传信息处理次之,注释匹配到 2 572 条单基因簇;生物系统最少,仅 211 条。进一步分析冬瓜 125 条代谢途径发现,核糖体的单基因簇数量最多(363 条),碳代谢和植物激素信号转导居次,分别有 337 条和 311 条单基因簇。在 KEGG 数据库功能注释中预测到的 125 条代谢途径中,含有氨基酸合成、不饱和脂肪酸的生物合成、维生素 B₆ 代谢、类

黄酮生物合成等相关的代谢途径分别有 295,47,20 和 18 条单基因簇,其中部分单基因簇涉及不饱和脂肪酸、类黄酮、维生素等生理活性物质合成,这些物质可能与冬瓜的营养保健功效有关。

2.8 冬瓜单基因簇的 Pfam 统计分析

利用 Pfam 数据库对冬瓜蛋白结构域进行预测,对蛋白家族进行分析。将冬瓜 40 604 条单基因簇与 Pfam 数据库比对,其中有 17 990 条单基因簇成功注释,所有注释到的单基因簇编码蛋白序列在数据库中匹配结构的数目均为 1 个,分成 369 个类群,三角状五肽重复结构域(1 096 条)、三角状五肽蛋白家族(907 条)、富亮氨酸重复(769 条)、WD 重复域的 G-beta 蛋白(641 条)、蛋白激酶结构域(568 条)在整体中占比较大。

表 4 冬瓜转录因子家族的单基因簇数量及种类

Table 4 Number and species of unigenes in the transcription factor family of *Benincasa hispida* Cogn

家族名称 Family	数量 Count	占比/% Percentage	家族名称 Family	数量 Count	占比/% Percentage
bHLH	81	7.37	YABBY	8	0.73
ERF	74	6.73	BES1	8	0.73
C2H2	73	6.64	ARR-B	7	0.64
MYB_related	70	6.37	CO-like	7	0.64
NAC	62	5.64	NF-YC	7	0.64
MYB	60	5.46	SRS	7	0.64
bZIP	54	4.91	BBR-BPC	6	0.55
WRKY	48	4.37	Nin-like	6	0.55
C3H	46	4.19	CAMTA	6	0.55
GRAS	40	3.64	WOX	6	0.55
HD-ZIP	35	3.18	DBB	6	0.55
G2-like	34	3.09	HB-other	6	0.55
Dof	32	2.91	M-type	6	0.55
HSF	29	2.64	CPP	5	0.45
FAR1	24	2.18	LSD	4	0.36
B3	23	2.09	E2F/DP	4	0.36
TCP	23	2.09	GRF	4	0.36
Trihelix	22	2.00	GeBP	4	0.36
GATA	22	2.00	RAV	3	0.27
SBP	18	1.64	EIL	3	0.27
ARF	17	1.55	SlFa-like	2	0.18
LBD	17	1.55	HB-PHD	2	0.18
TALE	15	1.36	Whirly	2	0.18
MIKC	13	1.18	VOZ	2	0.18
NF-YA	12	1.09	NF-X1	2	0.18
ZF-HD	11	1.00	STAT	1	0.09
AP2	10	0.91	HRT-like	1	0.09
NF-YB	9	0.82			

2.10 冬瓜转录组的 SSR 位点分析

对冬瓜转录组进行测序分析,共获得 40 604 条单基因簇,其中 5 086 条单基因簇获得 5 474 个 SSR 位点(表 5)。SSR 的发生频率为 12.53%,出现频率 13.48%,平均 7.08 kb 出现一个 SSR。冬瓜转录组

2.9 冬瓜单基因簇的转录因子分析

对冬瓜的单基因簇进行转录因子分析(表 4),在 40 604 条单基因簇中,具有转录因子序列特征的有 1 099 条,分为 55 个转录因子家族。家族中所占比例最大的是 bHLH(81 条,占 1 099 条单基因簇的 7.37%),然后依次是 ERF(74 条,占 6.73%)、C2H2(73 条,占 6.64%)、MYB_related(70 条,占 6.37%)。转录因子是读取并解释 DNA 中遗传“蓝图”的蛋白质组之一,其与 DNA 结合,可以帮助启动一个增加或减少基因转录的程序,因此转录因子对于许多细胞过程是至关重要的。前人研究表明,bHLH、ERF 在植物抗逆境胁迫和生长发育、生物合成及信号转导方面均有作用^[23-26]。

SSR 重复单元次数在 4~30 次,类型丰富,其中 4~10 次重复的 SSR 有 5 127 个,占 5 474 个位点的 93.66%;11~30 次重复的 SSR 有 347 个,占 6.34%。

由表 6 可以看出,本研究中 SSR 有二、三、四、

五、六核苷酸 5 种重复类型,共含 170 种重复基序,二、三、四、五、六核苷酸重复基序数量分别为 4,10,24,42 和 90 个。二核苷酸重复基序中 AG/CT 发生频率最高,占 5 474 个位点的 26.21%,占二核苷酸重复总数(2 047 个)的 70.10%;三核苷酸重复基序中 AAG/CTT 占比较大,在重复基序中占 24.08%;而五核苷酸和六核苷酸重复基序数量较少,总体占比也较小。

表 5 冬瓜叶转录组的 SSR 类型与数量

Table 5 Number of SSR types in the transcription group of <i>Benincasa hispida</i>							
重复单元数 Number of repeat unit	二碱基重复 Di-	三碱基重复 Tri-	四碱基重复 Tetra-	五碱基重复 Penta-	六碱基重复 Hexa-	数量 Number	占比/% Percentage
4	0	0	288	152	162	602	11.00
5	0	1 274	89	44	55	1 462	26.71
6	726	643	22	5	2	1 398	25.54
7	468	333	7	3	8	819	14.96
8	284	74	4	1	4	367	6.70
9	228	38	1	1	0	268	4.90
10	124	85	1	0	1	211	3.85
11	102	40	1	0	0	143	2.61
12	19	13	0	0	0	32	0.58
13	11	17	0	0	0	28	0.51
14	9	15	0	0	0	24	0.44
≥15	76	44	0	0	0	120	2.19
							100
总计 Total	2 047	2 576	413	206	232	5 474	

表 6 冬瓜转录组的 SSR 分类重复类型频率

Table 6 Frequency of SSR categories repetition in the transcription group of <i>Benincasa hispida</i>				
重复类型 Repeat types	重复基序 Repeat motif	数量 Number	发生频率/% Frequency	占比/% Percentage
二核苷酸 Di	AC/GT	180	3.29	8.79
	AG/CT	1 435	26.21	70.10
	AT/A	431	7.87	21.06
	CG/CG	1	0.02	0.05
三核苷酸 Tri	AAC/GTT	133	2.43	5.16
	AAG/CTT	1 318	24.08	51.16
	AAT/ATT	246	4.49	9.55
	ACC/GGT	134	2.45	5.20
	ACG/CGT	62	1.13	2.41
	ACT/AGT	24	0.44	0.93
	AGC/CTG	132	2.41	5.12
	AGG/CCT	186	3.40	7.22
	ATC/ATG	179	3.27	6.95
	CCG/CGG	162	2.96	6.29
四核苷酸 Tetra		413	7.54	100
五核苷酸 Penta		206	3.76	100
六核苷酸 Hexa		232	4.24	100

3 讨 论

冬瓜营养丰富,含有多 种生理活性成分。近年来有研究表明,冬瓜的多种提取物具有抗氧化^[27]、治疗荨麻疹^[28]、降压利尿^[29]、降糖^[30]等功效。冬瓜优良的药用价值和食疗保健价值使其市场需求量旺盛。然而,目前关于冬瓜相关代谢途径的机理研究还未见报道,本研究利用 Illumina HiSeq™ 2000 技术进行转录组测序,通过拼接获得了大量冬瓜单基

因簇,为冬瓜后续相关功能基因研究奠定了基础。本研究通过高通量测序平台对冬瓜转录组进行测序及分析发现,覆盖 50%核苷酸的最大单基因簇长度为 1 591 bp,表明组装质量较好,适合进行下一步分析。组装后获得 40 611 条单基因簇,平均长 955 bp,对其中 40 604 条单基因簇进行注释,仍有 13 004条单基因簇未成功注释,占总数的 32.03%,这可能与序列片段过短、注释信息缺乏以及初次从冬瓜中获得新基因等有关^[31]。用 GO 数据库对冬

瓜单基因簇进行功能分类,匹配到 3 个本体的 47 个功能组中,其中生物功能中代谢过程注释到的单基因簇数量最多,共 6 156 条。从功能注释结果可以看出,冬瓜在基因表达水平上与代谢过程相关的表达较多,表明植物存在着不同代谢途径,且不同的代谢途径同时在植物体内与相应的生理过程联系密切。KEGG 代谢通路分析得到单基因簇的 Pathway 注释,显示成功注释了 10 799 条冬瓜的单基因簇,其中代谢途径占 5 868 条,代谢途径中次生代谢途径有 268 条。同时,KEGG 数据库注释将冬瓜单基因簇分为 125 条代谢途径。用 Pfam 数据库对冬瓜蛋白结构域进行预测、对蛋白家族进行分析,结果显示注释到的 17 990 条单基因簇可分成 369 个类群,匹配结构数目均为 1 个。

转录因子能够调控基因的表达水平及调控网络。本研究与 TF 数据库进行 hmmscan 比对,注释到 1 099 条单基因簇分为 55 个转录因子家族,占比最大的是 bHLH 家族,其次是 ERF、C2H2、MYB_related、NAC、MYB 家族。bHLH 是真核生物中存在最广泛的一大类转录因子,Cui 等^[23]发现其在花药发育调控方面具有重要作用,何洁等^[24]认为其在西瓜抵御低温、ABA 和盐胁迫中有可能发挥作用,Gonzalez 等^[32]认为 TTG1/bHLH/Myb 转录因子对拟南芥幼苗花青素生物合成途径具有调控作用。另外,Jung 等^[33]研究表明,bHLH 转录因子可能与水稻雄性不育有关。转录因子在植物的生长发育^[34]、进化^[35]、抗逆境^[36]等方面具有重要的调控作用,这为冬瓜后续在生长和抗逆性的研究上提供了有力的依据。

冬瓜资源较为丰富,利用表型观察虽然能够直观地确定作物的产量、品质以及相关的性状特征表现,但同时也受到很多条件的影响与限制,因而需要进一步结合分子标记进行亲缘关系鉴定。本研究通过转录组测序得到冬瓜 5 474 个 SSR 位点,SSR 发生频率为 12.53%,出现频率 13.48%,低于丝瓜^[37]和莲雾^[38],但高于南瓜^[39]和黄秋葵^[40],这可能与物种、单基因簇总量及长度、筛选条件等因素不同有关^[41]。冬瓜 SSR 有二、三、四、五、六核苷酸 5 种重复类型,其中二核苷酸重复类型和三核苷酸重复类型占比较大,而五核苷酸和六核苷酸重复占比较小。这与丝瓜^[37]、莲雾^[38]、南瓜^[39]和黄秋葵^[40]的结果相似,可能与 SSR 搜索标准有关。冬瓜中 AG/CT 和 AAG/CTT 为主要重复基序,AG/CT 为主要基序,这与前人对平欧杂交榛^[42]和菜薹^[43]的研究结果一

致;AAG/CTT 在双子叶植物中是常见基序,与 Morgante 等^[44]的研究结论相似。本研究结果可为后续的冬瓜研究提供参考数据。

[参考文献]

[1] Morton J F. The wax gourd, a year-round Florida vegetable with unusual keeping quality [C]//Florida State Horticultural Society Proceeding. Miami: Florida State Horticultural Society, 1971:104-109.

[2] 张帅中,梁雪.冬瓜皮药用价值及综合利用研究进展[J].现代农业科技,2016(9):286-288.

Zhang S Z, Liang X. Research progress on officinal value and comprehensive utilization of wax gourd peel [J]. Modern Agricultural Sciences and Technology, 2016(9):286-288.

[3] Huang H Y, Huang J J, Tso T K, et al. Antioxidant and angiotension-converting enzyme inhibition capacities of various parts of *Benincasa hispida* (wax gourd) [J]. Molecular Nutrition, 2004, 48(3):230-233.

[4] 旷碧峰,唐 镔,余席茂,等.冬瓜搭架栽培技术[J].上海蔬菜,2016(5):40-42.

Kuang B F, Tang E, Yu X M, et al. Technology of wax melon viaduct cultivation [J]. Shanghai Vegetables, 2016(5):40-42.

[5] 骆焕彬,易文红,卢文佳,等.特黑杂交黑皮冬瓜的特征特性及高产栽培技术[J].广东农业科学,2007(8):22-23.

Luo H B, Yi W H, Lu W J, et al. Characteristic and high yield cultivation techniques of extra-black hybrid black gourd [J]. Guangdong Agricultural Science, 2007(8):22-23.

[6] 张淑瑛.冬瓜的美容及药用价值[J].中国民间疗法,1996(4):49.

Zhang S Y. Beauty and medicinal value of wax gourd [J]. China's Naturopathy, 1996(4):49.

[7] 夏 明,阮叶萍.冬瓜提取物抗小鼠胃溃疡活性研究[J].食品科学,2005(4):243-246.

Xia M, Ruan Y P. Study on anti-gastric ulcer activity of winter melon extract in mice [J]. Food Science, 2005(4):243-246.

[8] Grover J K, Adiga G, Vats V, et al. Extracts of *Benincasa hispida* prevent development of experimental ulcers [J]. Journal of Ethnopharmacology, 2001, 78:159-164.

[9] 车双辉.冬瓜抗胃溃疡活性组分的筛选[D].杭州:浙江工商大学,2004.

Che S H. Screening of active components of winter melon against gastric ulcer [D]. Hangzhou: Zhejiang Gongshang University, 2004.

[10] 邹宇晓,徐玉娟,廖森泰,等.冬瓜的营养价值及其综合利用研究进展[J].中国果菜,2006(5):46-47.

Zou Y X, Xu Y J, Liao S T, et al. Research progress on nutrition value and comprehensive utilization of winter melon [J]. Chinese Fruit and Vegetable, 2006(5):46-47.

[11] Wu C M, Liou S E, Chang Y H, et al. Volatile compounds of the wax gourd (*Benincasa hispida* Cogn) and a wax gourd beverage [J]. Journal of Food Science, 1987, 52(1):132-134.

[12] Pandey S,Kumar S,Mishra U,et al. Genetic diversity in Indian ash gourd (*Benincasa hispida*) accessions as revealed by quantitative traits and RAPD markers [J]. HortScience, 2008,118:80-86.

[13] Jiang B,Xie D,Liu W,et al. De novo assembly and characterization of the transcriptome,and development of SSR markers in wax gourd (*Benincasa hispida*) [J]. PLoS One,2013,8(8): e71054.

[14] Jiang B,Liu W R,Xie D,et al. High-density genetic map construction and gene mapping of pericarp color in wax gourd using specific-locus amplified fragment (SLAF) sequencing [J]. BMC Genomics,2015,16:1035.

[15] Costa V,Angelini C,Feis D I,et al. Uncovering the complexity of transcriptomes with RNA-Seq [J]. Biomedicine and Biotechnology,2010(1):853-916.

[16] Morozova O,Marra M A. Applications of next-generation sequencing technologies in functional genomics [J]. Genomics, 2008,92:255-264.

[17] Wang Z,Gerstein M,Snyder M. RNA-Seq;a revolutionary tool for transcriptomics [J]. Nat Rev Genet,2009,10:57-63.

[18] Cao J,Schneeberger K,Ossowski S,et al. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations [J]. Nature Genet,2011,43:956-963.

[19] Zhan J,Thakare D,Ma C,et al. RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation [J]. Plant Cell,2015,27:513-531.

[20] Zhang X,Song Z Q,Liu T,et al. De novo assembly and comparative transcriptome analysis provide insight into lysine biosynthesis in *Toona sinensis* Roem [J]. Journal of Genomics, 2016(3):1-9.

[21] 张少平,邱珊莲,郑云云,等. 紫色黄秋葵转录组功能基因测序及分析 [J]. 核农学报,2017,31(4):643-653.

Zhang S P,Qiu S L,Zheng Y Y,et al. Functional gene sequencing and analysis of the transcription group of purple okra [J]. Nuclear Agricultural Sciences,2017,31(4):643-653.

[22] Altschul S F,Gish W,Miller W,et al. Basic local alignment search tool [J]. Molecular Biology,1990,215(3):403-410.

[23] Cui J,You C,Zhu E,et al. Feedback regulation of DYT1 by interactions with downstream bHLH factors promotes DYT1 nuclear localization and anther development [J]. Plant Cell, 2016,28(5):1078.

[24] 何 洁,顾秀容,魏春华,等. 西瓜 bHLH 转录因子家族基因的鉴定及其在非生物胁迫下的表达分析 [J]. 园艺学报, 2016,43(2):281-294.

He J,Gu X R,Wei C H,et al. Nuclear agricultural sciences, dentification and expression analysis under abiotic stresses of the bHLH transcription factor gene family in watermelon [J]. Acta Horticultural Sinica,2016,43(2):281-294.

[25] Riechmann J L,Heard J,Martin G,et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes [J]. Science,2000,290:2105-2110.

[26] 刘建光,王永强,张寒霜,等. ERF 转录因子在植物抗逆境胁迫的研究进展 [J]. 华北农学报,2013,28(S1):214-218.

Liu J G,Wang Y Q,Zhang H S,et al. Research progress of ERF transcription factors in plant biotic and abiotic stress responses [J]. Acta Agriculturae Boreali Sinica,2013,28(S1): 214-218.

[27] 张金彪,周碧青,雷永程,等. 植物源 POD 和 CAT 的筛选及其与重金属的相互作用 [J]. 福建农林大学学报(自然科学版), 2009(3):295-300.

Zhang J B,Zhou B Q,Lei Y C,et al. Screening of botanical POD and CAT and their interaction with heavy metals [J]. Journal of Fujian Agriculture and Forestry University (Natural Science Edition),2009(3):295-300.

[28] 黄建华. 冬瓜皮汤治疗急性荨麻疹 50 例 [J]. 山东中医杂志, 1995(6):252-253.

Huang J H. Fifty cases of acute urticaria were treated by donggua skin decoction [J]. Shandong Journal of Traditional Chinese Medicine,1995(6):252-253.

[29] 陈晓娟,溪 茜,黄兆富. 中药饮剂对大鼠降压和利尿的效果 [J]. 第四军医大学学报,2000(12):S237-S238.

Chen X J,Xi Q,Huang Z F. Effects of traditional Chinese medicine decoction on hypertension and diuresis in rats [J]. Journal of the Fourth Military Medical University,2000(12): S237-S238.

[30] Gu M,Fan S,Liu G,et al. Extract of wax gourd peel prevents high-fat diet-induced hyperlipidemia in C57BL/6 mice via the inhibition of the PPAR γ pathway [J]. Evidence-Based Complementary and Alternative Medicine,2013(10):3-4.

[31] 魏利斌,苗红梅,张海洋. 芝麻发育转录组分析 [J]. 中国农业科学,2012(7):1246-1256.

Wei L B,Miao H M,Zhang H Y. Transcriptomic analysis of sesame development [J]. Scientia Agricultura Sinica, 2012 (7):1246-1256.

[32] Gonzalez A,Zhao M,Leavitt J M,et al. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings [J]. Plant Journal,2008,53(5):814-827.

[33] Jung K H,Han M J,Lee Y S,et al. Rice *Undeveloped Tapetum1* is a major regulator of early tapetum development [J]. Plant Cell,2005,17(10):2705-2722.

[34] 范中奇,邝健飞,陆旺金,等. 转录因子调控果实成熟和衰老机制研究进展 [J]. 园艺学报,2015,42(9):1649-1663.

Fan Z Q,Kuang J F,Lu W J,et al. Advances in research of the mechanism of transcription factors involving in regulating fruit ripening and senescence [J]. Acta Horticultural Sinica, 2015,42(9):1649-1663.

[35] 孙丽芳,邢少辰,张 君,等. 转录因子在植物进化和抗逆中的作用 [J]. 基因组学与应用生物学,2009,28(3):569-577.

Sun L F,Xing S C,Zhang J,et al. Function of the transcription factors in plant domestication and stress resistance [J]. Genomics and Applied Biology,2009,28(3):569-577.

[36] 王 娜,张振葆,黄凤珠,等. WRKY 转录因子参与植物非生

物胁迫应答的研究进展 [J]. 核农学报, 2014, 28(10): 1819-1827.

Wang N, Zhang Z B, Huang F Z, et al. Research progress on transcriptional factors involved in plant abiotic stress response [J]. Nuclear Agricultural Sciences, 2014, 28(10): 1819-1827.

[37] 朱海生, 张前荣, 刘建汀, 等. 丝瓜转录组 SSR 位点分析及其分子标记开发 [J]. 中国细胞生物学学报, 2016, 38(9): 1118-1127.

Zhu H S, Zhang Q Q, Liu J T, et al. Analysis on SSR information in transcriptome and development of molecular markers in *Luffa cylindrica* [J]. Chinese Journal of Cell Biology, 2016, 38(9): 1118-1127.

[38] 魏秀清, 许 玲, 章希娟, 等. 莲雾转录组 SSR 信息分析及其分子标记开发 [J]. 园艺学报, 2018, 45(3): 541-551.

Wei X Q, Xu L, Zhang X J, et al. Analysis on SSR information in transcriptome and development of molecular markers in wax apple [J]. Acta Horticulturae Sinica, 2018, 45(3): 541-551.

[39] 王洋洋, 单文琪, 徐文龙, 等. 印度南瓜转录组 SSR 信息分析及其多态性研究 [J]. 园艺学报, 2016, 43(3): 578-586.

Wang Y Y, Shan W Q, Xu W L, et al. Analysis on SSR information in transcriptome and the polymorphism of *Cucurbita maxima* [J]. Acta Horticulturae Sinica, 2016, 43(3): 578-586.

[40] 李永平, 刘建汀, 陈敏氢, 等. 利用黄秋葵转录组信息挖掘 SSR 标记及用于种质分析 [J]. 园艺学报, 2018, 45(3): 579-590.

Li Y P, Liu J T, Chen M D, et al. SSR markers excavation and germplasm analysis using the transcriptome information of *Hibiscus esculentus* [J]. Acta Horticulturae Sinica, 2018, 45(3): 579-590.

[41] Zhai L, Xu L, Wang Y, et al. Novel and useful genic-SSR markers from de novo transcriptome sequencing of radish (*Raphanus sativus* L.) [J]. Molecular Breed, 2014, 33(3): 611-624.

[42] 程云清, 张丽娜, 赵永斌, 等. 平欧杂交榛转录组中 SSR 信息分析和引物筛选 [J]. 园艺学报, 2018, 45(1): 139-148.

Cheng Y Q, Zhang L N, Zhao Y B, et al. Analysis of SSR markers information and primer selection from transcriptome sequence of Hybrid Hazelnut *Corylus heterophylla* × *C. avellane* [J]. Acta Horticulturae Sinica, 2018, 45(1): 139-148.

[43] 李荣华, 王直亮, 陈静芳, 等. 菜薹转录组中 SSR 信息与可用性分析 [J]. 园艺学报, 2016, 43(9): 1816-1824.

Liu R H, Wang Z L, Chen J F, et al. Analysis of SSR information in transcriptome and their usability in Flowering Chinese Cabbage [J]. Acta Horticulturae Sinica, 2016, 43(9): 1816-1824.

[44] Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes [J]. Nature Genetics, 2002, 30(2): 194-200.

.....

(上接第 53 页)

[22] 张阿凤, 邵慧芸, 成 功, 等. 小麦生物质炭对烤烟生长及根际土壤理化性质的影响 [J]. 西北农林科技大学学报(自然科学版), 2018, 46(6): 85-93.

Zhang A F, Shao H Y, Cheng G, et al. Effects of biochar amendment on flue-cured tobacco growth and soil physical and chemical properties of rhizosphere [J]. Journal of Northwest A&F University(Natural Science Edition), 2018, 46(6): 85-93.

[23] 赵殿峰. 不同生物炭施用量对烤烟土壤理化性状及烤烟生长的影响 [D]. 陕西杨凌: 西北农林科技大学, 2014.

Zhao D F. Effects of different biochar application amount on physicochemical properties of tobacco soil and tobacco growth [D]. Yangling, Shaanxi: Northwest A&F University, 2014.

[24] 惠锦卓, 张爱平, 刘汝亮, 等. 添加生物炭对灌淤土壤养分含量和氮素淋失的影响 [J]. 中国农业气象, 2014, 35(2): 156-161.

Hui J Z, Zhang A P, Liu R L, et al. Effects of biochar on soil nutrients and nitrogen leaching in anthropogenic-alluvial soil [J]. Chinese Journal of Agrometeorology, 2014, 35(2): 156-161.

[25] 尤方芳, 赵铭钦, 陈发元, 等. 生物炭与不同肥料配施对镉胁迫下烟株生长的影响 [J]. 浙江农业学报, 2016, 28(3): 489-495.

You F F, Zhao M Q, Chen F Y, et al. Effect of combined application of biochar and fertilizer on growth of tobacco under Cd stress [J]. Zhejiang Agricultural Journal, 2016, 28(3): 489-495.

[26] Bradl H B. Adsorption of heavy metal ions on soils and soils constituents [J]. Journal of Colloid and Interface Science, 2004, 277(1): 1-18.

[27] 王丽渊, 刘国顺, 王林虹, 等. 生物质炭对烤烟干物质积累量及根际土壤理化性质的影响 [J]. 华北农学报, 2014, 29(1): 140-144.

Wang L Y, Liu G S, Wang L H, et al. Effects of biochar on dry matter accumulation of flue-cured tobacco and physical and chemical property of rhizosphere soil [J]. Agricultural Journal of North China, 2014, 29(1): 140-144.