

杨宇婷, 张 强. 甜叶菊多倍体变异机制的转录组及 SSR 分析[J]. 江苏农业科学, 2023, 51(2): 44–50.

doi:10.15889/j.issn.1002-1302.2023.02.006

# 甜叶菊多倍体变异机制的转录组及 SSR 分析

杨宇婷, 张 强

(成都中医药大学附属医院, 四川成都 610000)

**摘要:**采用 BGI PE150 测序平台, 以二倍体甜叶菊及其同源四倍体为材料进行测序。共获得 93 742 个非重复序列基因(Unigene), 平均长度为 754, N50 达到 1 278 nt。共有 83 234 个 Unigene 获得注释, 其中, 注释到基因本体(GO)、京都基因和基因组百科全书(KEGG)、NR 库分别为 50 980、28 399、65 371 个。对转录组数据差异表达进行分析, 筛选得到 5 780 个差异表达基因(DEGs), 其中, 上调差异基因数目为 1 832 个, 下调差异基因数目为 3 948 个。基因本体 GO 功能分析表明, 这些差异基因主要参与细胞生长和代谢过程等相关功能。KEGG 通路富集分析表明, 二、四倍体代谢水平差异是由基因层次决定的。在对甜叶菊植株加倍过程中的样品进行转录组测序中发现, 10 777 个简单重复序列(SSR)位点, 最多的是二核苷酸, 为 4 386 个, 占 SSR 总数的 40.70%。结果表明, 四倍体甜叶菊表现出的明显器官增大、光合作用增强与差异基因显著富集在分解代谢、信号转导、碳水化合物代谢、能量代谢和氨基酸代谢等代谢通路相关。

**关键词:**甜叶菊; 多倍体; 转录组; 简单重复序列(SSR)

**中图分类号:**R285 **文献标志码:**A **文章编号:**1002-1302(2023)02-0044-06

同源四倍体具有产量高、抗逆性强、品质优良等优点, 在药用植物育种方面备受关注<sup>[1]</sup>。多倍体植物细胞内的染色体加倍导致细胞核和细胞体积增大, 在外形上引起植物形态发生明显变化, 叶片变大变厚, 茎变粗, 植株的生物产量和药用活性成分均有不同程度的增加<sup>[2-3]</sup>。与二倍体相比, 甜叶菊四倍体植株茎秆粗壮、叶片加厚、叶面积增大、植株变矮, 糖苷含量等均发生显著变化, 在农艺性状上, 四倍体甜叶菊株系也表现出更高的品质和更强的抗逆性<sup>[4-5]</sup>。

甜叶菊[*Stevia rebaudiana* (Bertoni) Hemsl.] 是菊科甜叶菊属多年生宿根草本植物, 它所含有的甜菊糖苷的甜度是蔗糖的 250~300 倍, 但所含热量仅是蔗糖的 1/300<sup>[6]</sup>。甜叶菊四倍体株系的高品质和高产量在甜叶菊优势品种选育中应用广泛。目前, 关于四倍体甜叶菊的研究大多集中在四倍体诱导鉴定<sup>[7]</sup>、生物学特性<sup>[8]</sup>、遗传物质分子标记<sup>[9]</sup>等方面。随着生物学的发展, 下一代高通量测序技术使

得基因信息处理速度得到极大提升。RNA 测序技术已成功应用于白桦<sup>[10]</sup>、人工合成芸薹属<sup>[11]</sup>、十字花科<sup>[12]</sup>等倍性植物株系转录组测序。为探究甜叶菊同源四倍体的变异机制, 本研究以甜叶菊二倍体植株为对照, 利用 BGI PE150 测序平台和生物信息学分析方法, 分别取四倍体和二倍体叶片进行转录组测序, 获得 2 组差异表达基因(DEGs), 通过对这些差异表达基因数目进行功能富集分析, 为甜叶菊同源四倍体表型变异形成和有效成分基因调控机制相关研究提供基础资料。

## 1 材料与方法

### 1.1 试验材料

以成都中医药大学中药品种质量资源研究所种植的甜叶菊二倍体及其四倍体 1 年生植株为试验材料, 样本植株经成都中医药大学附属医院杨宇婷药师鉴定为菊科植物甜叶菊。分别随机选取二倍体和四倍体甜叶菊栽培苗各 5 株, 采样时间为 2020 年 5 月 13 日, 采集甜叶菊叶片进行转录组测序, 做 3 次生物学重复, 试验时间为 2020 年 6 月 11 日至 7 月 21 日。

### 1.2 方法

1.2.1 总 RNA 提取 样品送至成都百纳特生物科技有限公司, 采用 Trizol 法提取总 RNA, 利用

收稿日期: 2021-01-14

基金项目: 四川省科技计划(编号: 18YYJC0960)。

作者简介: 杨宇婷(1987—), 女, 四川成都人, 药师, 研究方向为药理学。E-mail: 191403281@qq.com。

通信作者: 张 强, 药师, 研究方向为药理学。E-mail: zhangqiang19912021@163.com。

Nanodrop 2000 检测 RNA 的纯度 (吸光度  $D_{260\text{ nm}}/D_{280\text{ nm}} > 2.0$ )、Agilent 2100 检测 RNA 的完整性 ( $RIN > 8.5$ ),检测合格后构建 cDNA 文库。

1.3 转录组数据分析

1.3.1 数据处理 通过去除污染的 reads、未知碱基 N 含量  $> 5\%$  的 reads 和低质量的 reads,得到高质量可用片段(clean reads)。使用 Trinity 软件对 clean reads 进行组装,之后利用 Tgicl<sup>[13]</sup> 对转录本聚类去冗余,最终得到非重复序列基因(Unigene)作为后续分析的参考序列。

1.3.2 基因表达水平分析 利用 RSEM,调用 bowtie2 的比对结果进行统计,得到每个样本比对到每个转录本上的 reads 数目,并对其进行 FPKM (Fragments Per Kilobase per Million bases)转换,来自同一个片段(fragment)的成对末端读取序列(paired-end reads)计数为 1 个 fragment,进而得到基因和转录本的表达水平。

1.3.3 差异基因表达分析 采用  $\log_2$  (Fold Change)  $\geq 2.00$ , Adj. 假定值 ( $P$  值)  $\leq 0.01$  或 0.001 的方式筛选差异基因,基于负二项分布原理

并根据 Michael 等所述方法<sup>[14]</sup> 进行差异表达基因检测。基于 GOseq 所述方法<sup>[15]</sup> 对筛选到的 DEGs 进行 GO 富集分析并进行 GO 功能统计。基于京都基因和基因组百科全书 (KEGG) 注释结果,使用 KOBAS,设置参数——fdr 为 BH (即使用 BH 校正) 进行 Pathway 富集分析。

2 结果与分析

2.1 RNA-Seq 测序原始数据分析

通过过滤去除低质量、接头污染以及未知碱基含量过高的 reads,6 个样品 (T2-1、T2-2、T2-3、T4-1、T4-2 和 T4-3) 测序得到的 clean reads 的数量分别为 10 542 231、19 757 350、12 758 156、20 225 185、21 463 543、24 421 653 个, Q20 的比例  $\geq 95.2\%$ , 低质量 (Quality  $< 20$ ) 的碱基比例较低,说明测序质量较好,可用于后续分析,结果见表 1。利用 Trinity 软件对 clean reads 进行合并组装,并得到转录本长度分布图 (图 1),其中最小长度为 143,最大长度为 1 321,组装转录本 N50 数值为 1 278,数值较大,说明组装效果较好,结果见表 2。

表 1 过滤后的 reads 质量统计

样本	过滤后的 reads 数	过滤后的碱基总数 (bp)	长度	Q20 比例 (%)	Q30 比例 (%)	GC 比例 (%)
T2-1	10 542 231	107 467 654	180	95.4	90.5	45.0
T2-2	19 757 350	123 342 674	180	96.2	92.3	44.2
T2-3	12 758 156	108 223 653	180	95.4	90.1	45.8
T4-1	20 225 185	200 112 235	180	95.2	90.8	45.6
T4-2	21 463 543	201 135 678	180	95.7	89.0	45.3
T4-3	24 421 653	198 674 345	180	96.8	92.4	47.3

注:T2 表示甜叶菊二倍体叶;T4 表示甜叶菊四倍体叶;Q20 和 Q30 的测序准确率均达到 99% 以上。

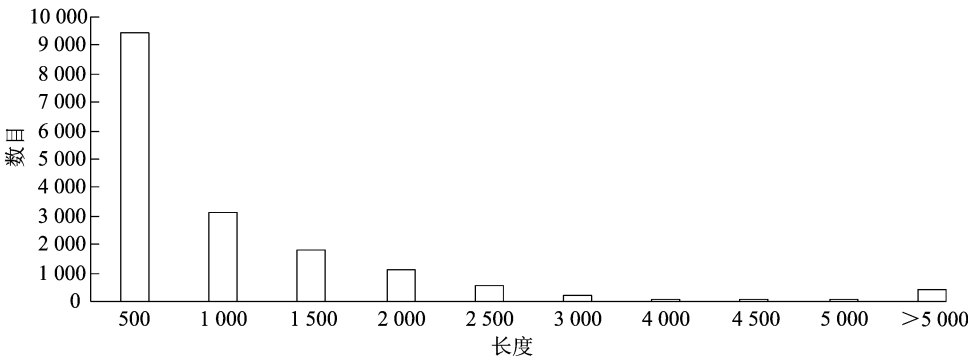


图1 转录本长度分布

2.2 转录本表达水平分析

利用 bowtie2<sup>[16]</sup>,将每个样本质控后的二代序列比对至参考转录本序列,多重比对序列的平均比

对率则为 54.35%,唯一比对上参考基因组的平均比对率为 29.01%,结果见表 3。利用 RSEM,调用 bowtie2 的比对结果进行统计,得到每个样本比对到

表 2 组装长度分布统计

长度	转录本数目
< 500 bp	11 543
500 ~ < 1 000 bp	10 236
1 000 ~ 2 000 bp	17 354
> 2 000 bp	10 357
总数目	93 742
平均长度	754
最小长度	143
最大长度	1 321
N50	1 278

注:无参转录组中,按照长度将拼接得到的转录本从大到小排序,依次累加转录本的长度,到不小于总长 50% 的拼接转录本的长度就是 N50。

表 3 reads 与组装转录本比对结果

样本	总序列数	总比对序列	唯一比对序列	多重比对序列
T2 - 1	752 134	636 287 (83.27%)	221 489 (31.18%)	554 568 (53.78%)
T2 - 2	757 350	629 486 (82.38%)	660 270 (27.80%)	529 319 (55.21%)
T2 - 3	758 156	640 490 (83.82%)	678 517 (30.15%)	596 778 (54.47%)
T4 - 1	937 294	614 968 (80.48%)	693 921 (28.35%)	507 030 (54.53%)
T4 - 2	987 690	626 583 (82.00%)	703 472 (27.24%)	524 888 (54.63%)
T4 - 3	910 777	621 387 (81.32%)	764 915 (29.31%)	546 002 (53.47%)

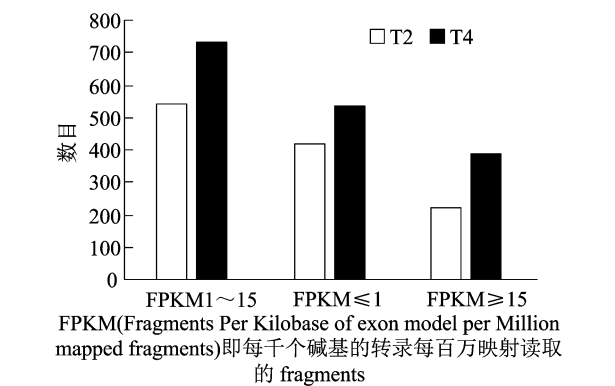


图2 样品基因表达量分布

表 4 注释结果统计

数据库	转录本数目及比例
总计	83 234 (100%)
KOG 注释	21 746 (26.13%)
KEGG 注释	28 399 (34.12%)
NR 注释	65 371 (78.54%)
GO 注释	50 980 (61.25%)
Swiss - Prot 注释	48 666 (58.47%)
未知数据库	15 248 (18.32%)

注:NR 表示 NCBI 蛋白非冗余数据库;NT 表示 NCBI 核苷酸序列数据库;Swis - Prot 表示蛋白质序列数据库;GO 为基因本体数据库;COG/KOG 为蛋白质真核同源群;KEGG 为京都基因和基因组百科全书。

每个转录本上的 reads 数目,并对其进行 FPKM 分析,结果见图 2。6 个样品中,转录本的表达丰度主要集中在 FPKM 1 ~ 15 (低水平表达)、 $FPKM \leq 1$  (极低表达水平)这 2 个区域,而  $FPKM \geq 15$  (高水平表达)的基因数目则较少。

2.3 测序基因(或转录本)功能注释

对 6 个样本的 83 234 个转录本做功能注释,通过与五大数据库进行比对,结果表明,注释到最多转录本数目的是 NR 65 371 (78.54%)数据库;其他数据库中注释到的转录本数目分别为 21 746 (26.13%)、28 399 (34.12%)、50 980 (61.25%)和 48 666 (58.47%),其中未知数据库注释的转录本数目为 15 248,占总数的 18.32%,结果见表 4。

2.3 差异表达基因统计及 GO 和 KEGG 分析

2.3.1 差异表达基因统计 四倍体甜叶菊中上调表达量高于二倍体甜叶菊,基因数目为 1 832,二倍体甜叶菊中下调表达量高于四倍体甜叶菊,基因数目为 3 948,结果见图 3。将二倍体甜叶菊与其同源四倍体中表达的差异基因做层次聚类分析,其中 X 轴代表进行聚类分析的差异比对,Y 轴代表差异基因;每行表示 1 个基因,每列表示 1 个样品。颜色代表差异倍数,颜色越红代表上调倍数越大,越蓝代表下调倍数越大。结果表明,甜叶菊植株加倍后其上调差异基因数目显著大于下调数目;从颜色深浅可以看出,上调差异基因倍数较下调倍数大。说明在甜叶菊植株加倍过程中有相关差异基因参与表达,其中上调的基因调控作用更加显著,其差异基因的层次聚类热图见图 4。

2.3.2 差异表达基因 GO 富集分析 为进一步了解加倍后甜叶菊差异基因的表达情况,对 DEGs 进行 GO 富集分析。GO 分析结果表明,这些差异表达基因分布在分子功能(molecular function)、细胞组分(cellular component)和生物过程(biological process)三大类中的 43 个类别,包括代谢过程、细胞过程、细胞组分、结合和催化活性等。其中,差异基因大部

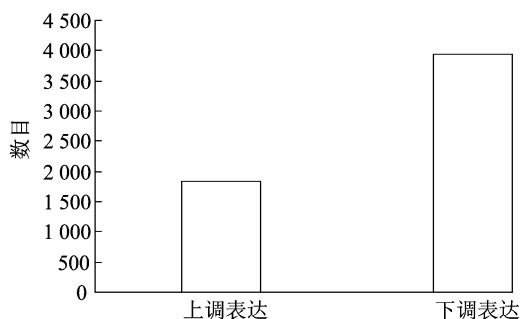


图3 差异表达基因统计数目

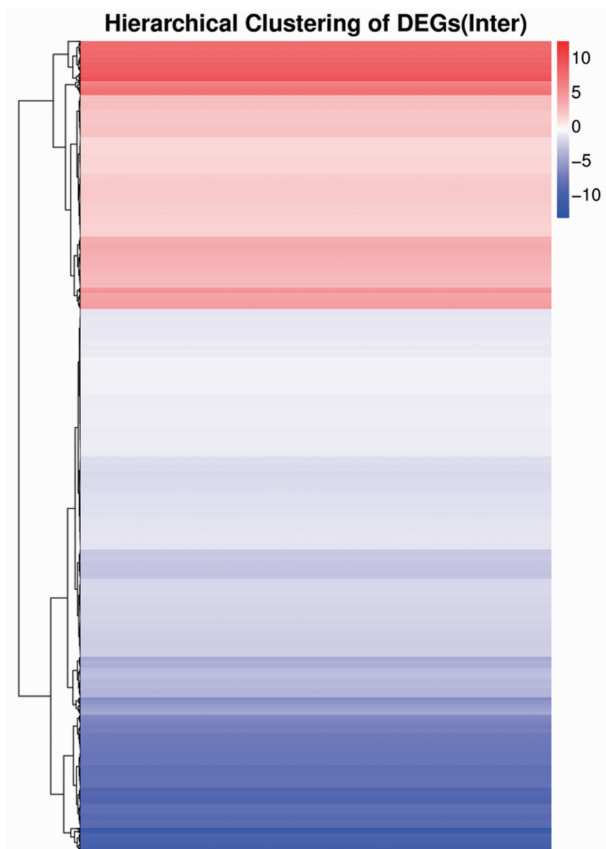


图4 差异基因表达层次聚类分析

分富集在生物过程中,而参与分子功能的差异基因相对较少。在生物学进程中,以参与代谢过程、细胞过程、单有机体过程、对刺激的反应、生物调控和生物过程调控的差异基因富集最明显,参与细胞成分的差异基因主要富集在细胞、细胞组分、细胞器、细胞膜等类别;分子功能中差异基因显著富集在结合与催化活性类别,结果见图5。

**2.3.3 差异表达基因 KEGG 分析** 通过对甜叶菊加倍后的差异表达基因进行 KEGG 代谢路径富集分析,结果显示,差异表达基因被富集到 pathway 代谢路径的 6 个大类、21 个亚类。靠前的代谢途径分

别为翻译(translation)、折叠(folding)、分类和降解(sorting and degradation)、运输和分解代谢(transport and catabolism)、碳水化合物代谢产物(carbohydrates metabolites)、氨基酸代谢(amino acid metabolism)、脂质代谢(lipid metabolism)和能量代谢(energy metabolism),结果见图6。

## 2.4 简单重复序列(SSR)信息分析

根据 *de novo* 组装结果,使用 MISA 软件对甜叶菊植株加倍过程中获得的 Unigene 进行 SSR 检测。SSR 检测结果表明,共检测出的 SSR 的位点数目为 10 777,最多的是二核苷酸 SSR,数目为 4 386,占 SSR 总数的 40.70%;三核苷酸 SSR 次之,数目为 3 412,占 31.66%,五核苷酸最少,占 0.78%,数目为 84。具体结果见表5。

## 3 讨论

多倍体具有器官增大、营养成分增多和抗逆性增强等多个优点,使得多倍体成为农学和园艺学的研究热点。目前,多倍体器官巨大性的原因是由于细胞增大而不是细胞数目增多的说法已被证实,但是细胞体积受倍性调控的分子机制仍不清楚。本研究以甜叶菊二倍体及其同源四倍体植株为试验材料,利用 BGI PE150 测序平台和生物信息学分析方法,分别取四倍体和二倍体的叶片进行转录组测序并获得 2 组差异表达基因,通过对这些 DEGs 进行功能富集分析,为甜叶菊同源四倍体表型变异形成机制和代谢途径研究提供新的线索。

对转录组数据差异表达基因进行分析,筛选得到 5 780 个差异表达基因,其中 1 832 个基因上调表达,3 948 个下调。GO 分析表明,四倍体与二倍体的差异表达基因主要集中在生物学过程(biological process)和细胞成分(cellular component)2 个功能区,对分子功能(molecular function)的调控作用不大。在细胞成分功能区发现差异基因在细胞、细胞组分和细胞器类别显著富集,表明倍性水平会影响细胞体积,在此过程中大量差异基因在运输和分解代谢、翻译、折叠、分类和降解等代谢途径显著富集促进相关蛋白质的合成,蛋白质行使调控功能从而增加细胞体积,这也是四倍体甜叶菊叶片变大变厚、营养器官增大的原因。同源多倍体药用植物育种的一个主要目的是增加次生代谢产物的含量,尤其是药用成分的增加。研究报道,同源四倍体桔梗叶片的叶绿素 a、叶绿素 b、胡萝卜素、总叶绿素的含

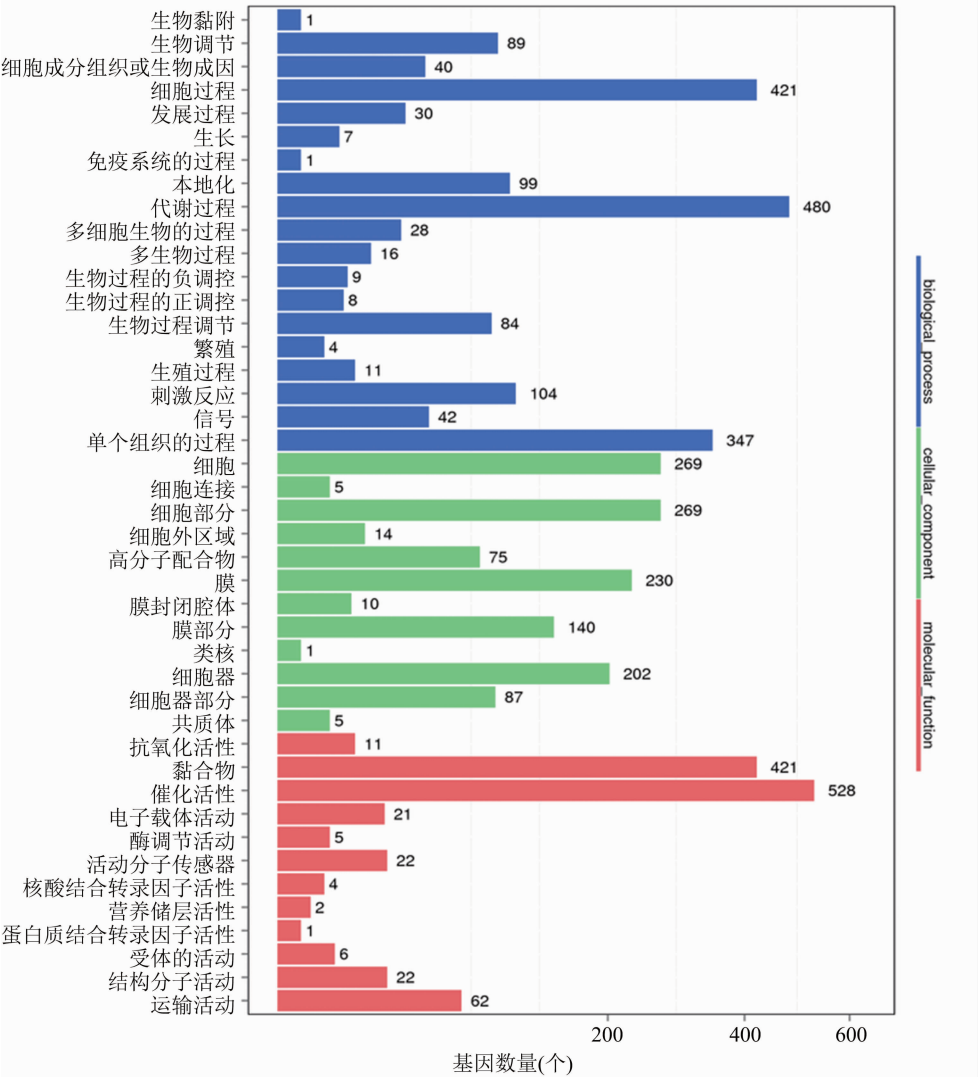


图5 差异表达基因的 GO 功能分布

量比二倍体分别增加 36.11%、69.23%、31.71%、38.85%<sup>[17]</sup>,同源四倍体青蒿的青蒿素含量是二倍体的 1.5 倍<sup>[18]</sup>。甜叶菊四倍体叶中的活性成分均高于其二倍体,据文献报道,甜叶菊四倍体的甜菊糖苷总含量(13.212%)比二倍体的甜菊糖苷总含量(9.998%)提高了 32.15%<sup>[5]</sup>。甜叶菊多倍体的转录组及 SSR 分析结果表明,甜叶菊植株面对一些生物和非生物胁迫时,会出现一系列应激反应,形成多种信号转导途径。同源四倍体甜叶菊具有更强的抗逆性,在面对寒冷刺激时,甜叶菊四倍体会调控相关差异基因上调表达来表现出更强的适应性,本研究中,一些与光合作用相关的差异基因在光合作用、光合生物的固碳作用等路径富集,显著增强了甜叶菊四倍体株系的光合效率。本研究共发现 10 777 个 SSR 位点,优势重复序列为二核苷

酸,占总 SSR 位点的 40.70%,这一研究结果与李俊仁等的结果<sup>[19-20]</sup>一致。

本研究以甜叶菊二倍体及其同源四倍体为材料,利用转录组测序技术并结合二者的生理特征,分析同源四倍体产生的基因表达变化,发现二者在外观性状、生长发育以及抗逆抗胁迫等方面存在差异,以上结果可为进一步定位关键调控基因及基因工程育种等奠定基础。

参考文献:

[1] 马艳芝,客绍英. 菰蓝四倍体株系细胞形态学和流式细胞术鉴定[J]. 中药材,2014,37(5):736-740.  
[2] 邵帅旭,陈 丽,杨 阳,等. 同源四倍体紫菜薹的诱导鉴定及其营养品质比较[J]. 西北植物学报,2018,38(7):1262-1269.  
[3] 潘红艳. 菰蓝多倍体诱导及外源褪黑素对低温胁迫下菰蓝抗氧化性的影响[D]. 西安:西北大学,2013.

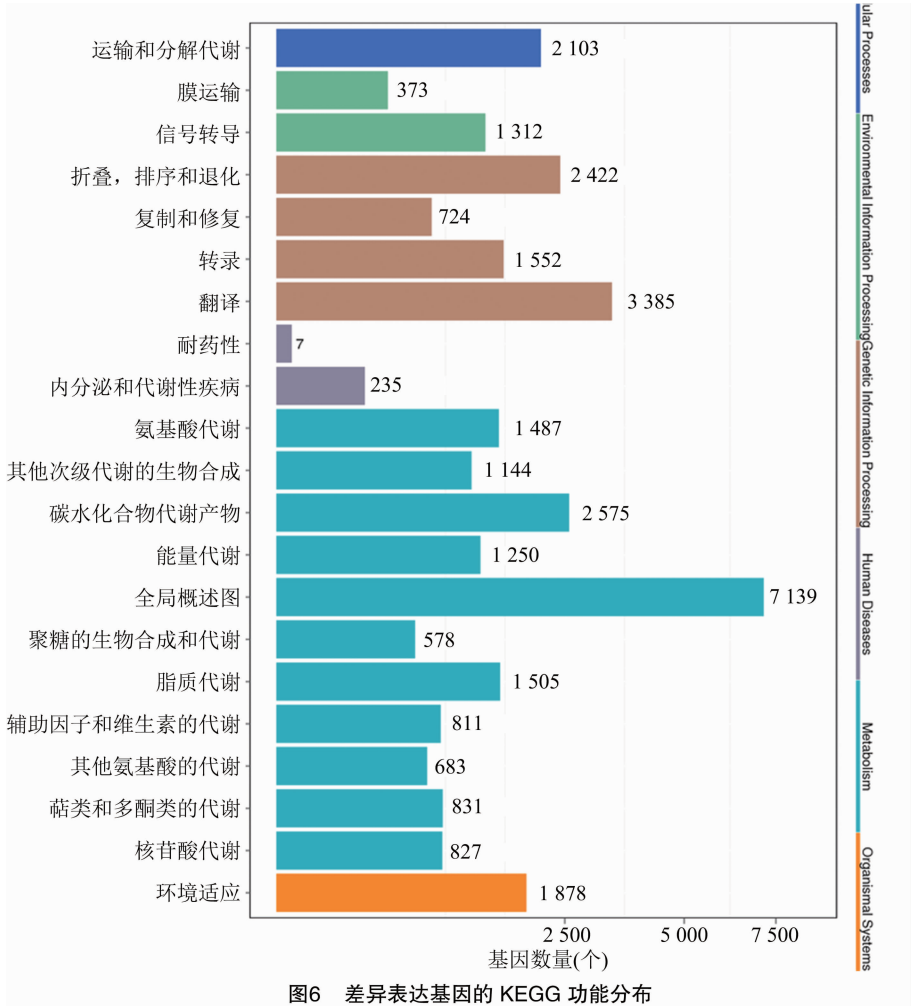


图6 差异表达基因的 KEGG 功能分布

表 5 SSR 的类型、数量及分布比率

SSR 类型	数量									比例 (%)
	5 次	6 次	7 次	8 次	9 次	10 次	11 次	>11 次	总计	
单核苷酸重复	0	0	0	0	0	0	0	2 426	2 426	22.51
二核苷酸重复	0	1 744	932	267	403	419	372	243	4 386	40.70
三核苷酸重复	1 386	1 046	473	331	49	51	25	51	3 412	31.67
四核苷酸重复	173	68	9	7	8	11	3	4	283	2.63
五核苷酸重复	65	11	3	2	0	2	0	1	84	0.78
六核苷酸重复	94	48	18	13	8	5	3	3	186	1.73
总计	1 718	2 917	1 435	620	468	488	403	2 728	10 777	100
比例	15.94%	27.07%	13.32%	5.75%	4.34%	4.53%	3.74%	25.31%	100%	

[4]陈桂平,客绍英. 二倍体及四倍体菰蓝氨基酸含量的测定分析[J]. 中药材,2011,34(4):506-508.

[5]段英姿. 菰蓝二倍体及其同源四倍体遗传差异的 ISSR 分析[J]. 西北植物学报,2012,32(8):1534-1538.

[6]李雅婷. 甜叶菊二倍体与同源四倍体生物学特性及遗传差异研究[D]. 南京:南京农业大学,2015.

[7]李雅婷,王红娟,向增旭. 甜叶菊二倍体与同源四倍体生理特征及 AFLP 分析[J]. 核农学报,2015,29(11):2103-2109.

[8]李 红,杨 岚,向增旭. 甜叶菊同源四倍体离体诱导及鉴定[J]. 西北植物学报,2012,32(8):1692-1697.

[9]程晓素,朱 平,崔广荣. 甜叶菊三倍体培育及其农艺性状分析[J]. 安徽科技学院学报,2016,30(3):25-30.

[10]杨 岚,李 红,向增旭. 二倍体和同源四倍体甜叶菊的 DNA 甲基化差异分析[J]. 核农学报,2013,27(8):1125-1130.

[11]穆怀志. 白桦四倍体的生长生殖特征及基因转录组分析[D]. 哈尔滨:东北林业大学,2013.

张艺欣,段菲,樊洁,等. 江西铅山红芽芋蔗糖合成酶基因的克隆和表达分析[J]. 江苏农业科学,2023,51(2):50-57.  
doi:10.15889/j.issn.1002-1302.2023.02.007

# 江西铅山红芽芋蔗糖合成酶基因的克隆和表达分析

张艺欣<sup>1,2,3,4</sup>, 段菲<sup>1</sup>, 樊洁<sup>1</sup>, 符卉卉<sup>1</sup>, 顾佳怡<sup>1</sup>, 兰佳欣<sup>1</sup>, 李娇<sup>1</sup>

(1. 上饶师范学院生命科学院,江西上饶 334001; 2. 上饶农业技术创新研究院,江西上饶 334001;

3. 上饶市药食同源植物资源保护与利用重点实验室,江西上饶 334001;

4. 上饶市薯芋类作物种质保存与利用重点实验室,江西上饶 334001)

**摘要:**对江西铅山红芽芋蔗糖合成酶基因进行克隆和表达分析,旨在为今后进一步研究江西铅山红芽芋蔗糖代谢机制和利用基因工程手段提高江西铅山红芽芋的品质奠定基础。从江西铅山红芽芋试管苗转录组数据库中筛选到江西铅山红芽芋蔗糖合成酶基因的核心片段,用 RT-PCR 技术克隆江西铅山红芽芋蔗糖合成酶基因,并用生物信息学方法和实时定量 PCR 方法进行序列分析和器官表达分析。结果显示,江西铅山红芽芋蔗糖合成酶基因的 cDNA 总长度为 2 256 bp, G + C 含量为 56.21%;江西铅山红芽芋蔗糖合成酶由 751 个氨基酸组成,相对分子量为 85 350.64 u,等电点为 5.83,为亲水性蛋白,二级结构由  $\alpha$ -螺旋(36.88%)、 $\beta$ -片层(16.91%)、无规则卷曲(46.21%)构成,三级结构为同源四聚体;江西铅山红芽芋蔗糖合成酶主要存在于细胞质、线粒体中,在进化水平上与芋(*Colocasia esculenta*)、眼子菜(*Potamogeton distinctus*)的亲缘关系较近,尤其是与 *C. esculenta* hypothetical protein Taro\_012658(MQI80204\_c0\_g1)在进化上具有最高的亲缘关系。实时定量 PCR 结果显示,蔗糖合成酶基因在江西铅山红芽芋中的表达存在器官特异性,在根和球茎膨大初期的表达量最高。由研究结果可知,江西铅山红芽芋蔗糖合成酶具有典型蔗糖合成酶的结构特征,氨基酸序列及核酸序列与同源物种相似度高,在进化上高度保守,对进一步揭示该酶的生物学功能具有重要意义。

**关键词:**江西铅山红芽芋;蔗糖合成酶;基因克隆;表达分析

**中图分类号:**S632.301 **文献标志码:**A **文章编号:**1002-1302(2023)02-0050-08

江西铅山红芽芋(*Colocasia esculenta* L. Schoot var. *cormosus* ‘Hongyayu’)为天南星科芋属多子芋类型草本宿根植物,为江西省铅山县“两红”(红芽

芋和河红茶)产业之一,2013 年 4 月 15 日被核准为国家地理标志农产品,其肉白芽红,营养丰富,肉质细嫩,口感细松滑,糯香可口,是药食兼优、老幼皆宜的上好滋补佳品<sup>[1-2]</sup>。当植物处于逆境时,适度积累蔗糖可保证细胞质膜的稳定性,避免胞内外相关蛋白的降解,还可作为逆境胁迫下植物代谢所需能量的来源<sup>[3]</sup>。作为蔗糖代谢过程中的关键酶,蔗糖合成酶主要负责植物组织细胞壁成分或淀粉合成的调控,直接参与植物的生长发育过程<sup>[4]</sup>。因此,克隆江西铅山红芽芋蔗糖合成酶基因并检测其

收稿日期:2021-12-07

基金项目:国家自然科学基金(编号:32060092、31860084);2021 年度江西省教育厅科学技术研究项目(编号:GJJ211729);上饶市科技局平台载体建设项目(编号:2020I001);上饶市科技局平台载体建设项目(编号:2019I017)。

作者简介:张艺欣(1991—),女,江西上饶人,硕士,讲师,主要从事植物生物技术研究工作。E-mail:474303981@qq.com。

[12] 张大为. 人工合成芸蓼属杂种及异源四倍体的转录组研究[D]. 武汉:华中农业大学,2016.

[13] 吕炜. 十字花科主要作物同源四倍体创制及转录组分析[D]. 南京:南京农业大学,2015.

[14] Pertea G, Huang X Q, Liang F, et al. TIGR Gene Indices clustering tools (TGICL): a softwaresystem for fast clustering of large EST datasets[J]. *Bioinformatics*, 2003, 19(5): 651-652.

[15] Kramarz B, Lovering R C. Gene Ontology: a resource for analysis and interpretation of Alzheimer's disease data [J]. *Exon Publications*, 2019: 23-36.

[16] Langmead B. Aligning short sequencing reads with bowtie [J]. *Current Protocols in Bioinformatics*, 2010, 32(1): 13-17.

[17] 韩盼盼. 桔梗同源四倍体与二倍体生理特性和遗传变异研究[D]. 南京:南京农业大学,2017.

[18] 马莉,高山林,黄和平,等. 青蒿同源四倍体组培优化及青蒿素含量测定[J]. *药物生物技术*, 2010, 17(4): 317-320.

[19] 李俊仁,陈秀珍,汤小婷,等. 穿心莲转录组 SSR 位点信息分析[J]. *中国中药杂志*, 2018, 43(12): 2503-2508.

[20] 张成才,张子璇,张文静,等. 利用转录组测序分析与华重楼种子休眠解除相关差异基因[J]. *中国中药杂志*, 2020, 45(8): 1893-1900.