

研究论文

Research Article

甘蔗属及其近缘属代表性种质的转录组比较分析

李思程^{1,2*} 冯梦凡^{1,2*} 杨细平^{1,2**}

1 广西大学, 亚热带农业生物资源保护与利用国家重点实验室, 广西甘蔗生物学重点实验室, 南宁, 530004; 2 广西大学农学院, 植物科学国家级实验教学示范中心, 南宁, 530004

* 同等贡献作者

** 通信作者, xipingyang@gxu.edu.cn

摘要 甘蔗近缘种斑茅 (*Tripsacum daniellii*)、甘蔗属内野生种割手密 (*Saccharum spontaneum*) 和大茎野生种 (*Saccharum robustum*) 具有茎部强壮、耐旱、耐涝、耐霜和抗病性等对现代甘蔗品种改良具有重要价值的性状。本研究通过比较甘蔗属及其近缘属四个代表性种质, 包括一个甘蔗属近缘种质 (斑茅)、两个甘蔗属野生种 (割手密、大茎野生种) 和一个甘蔗属驯化种 (热带种) 的转录组, 以探讨它们在转录层面的差异。研究发现, 甘蔗近缘种斑茅与甘蔗属野生种割手密 (BM vs Y83)、甘蔗属内的两个野生种割手密与大茎野生种 (Y83 vs N57) 以及甘蔗属内的野生种割手密和驯化种热带种 (N57 vs NJ) 之间的差异表达基因主要涉及光合作用、次级代谢和信号转导等, 这可能导致甘蔗属及其近缘属物种间在光合能力、抗生物和非生物胁迫能力以及营养利用方面存在差异。割手密在氮代谢表现出较斑茅和大茎野生种更强的潜力, 表明其在氮素利用上可能更具优势。此外, 割手密的特异表达基因涉及光感受器、昼夜节律时钟和开花途径, 这可能与开花特性相关。本研究还鉴定了甘蔗属及其近缘属、甘蔗属内野生种和驯化种之间的 4 325 个直系同源基因, 这些基因的正选择分析结果表明割手密可能在自然选择下逐渐改变基因表达模式以适应环境, 而热带种从大茎野生种驯化后的分化和适应可能与植物激素信号转导的改变相关。

关键词 甘蔗; 甘蔗近缘种; 比较转录组; RNA-seq; 斑茅; 割手密; 大茎野生种

Comparative Transcriptome Analysis of Representative Germplasm of *Saccharum* and Its Closely Related Genera

LI Sicheng^{1,2*} FENG Mengfan^{1,2*} YANG Xiping^{1,2**}

1 State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Nanning, 530004; 2 National Demonstration Center for Experimental Plant Science Education, College of Agriculture, Guangxi University, Nanning, 530004

* These authors contributed equally to this work

** Corresponding author, xipingyang@gxu.edu.cn

DOI: 10.13417/j.gab.042.001323

Abstract *Tripsacum daniellii*, a close relative of sugarcane, and *Saccharum spontaneum* and *Saccharum robustum*, wild species within sugarcane, exhibit valuable traits for improvement of modern sugarcane cultivar, such as robust stems, drought, flood, and frost tolerance, and disease resistance. In this study, we compared the transcriptomes of representative germplasm from four sugarcane and closely related species, including a close relative of sugarcane (*T. daniellii*), two wild species within the *Saccharum* genus (*S. spontaneum* and *S. robustum*), and a domesticated sugarcane (*S. officinarum*), to investigate their transcriptomic differences. The research revealed that differentially expressed genes, between the sugarcane relative *T. daniellii* and the wild species *S. spontaneum* of *Saccharum* genus (BM vs Y83), between two wild species *S. spontaneum* and *S. robustum* within the *Saccharum* genus (Y83 vs N57), and between wild species

基金项目: 本研究由广西自然科学基金项目 (2020GXNSFAA297039)、广西科技重大专项项目 (桂科 AA22117002) 和广西百人计划项目共同资助。

引用格式: 李思程, 冯梦凡, 杨细平, 2023. 甘蔗属及其近缘属代表性种质的转录组比较分析. 基因组学与应用生物学, 42 (12): 1323-1337. [Li S C, Feng M F, Yang X P, 2023. Comparative transcriptome analysis of representative germplasm of *Saccharum* and its closely related genera. Genomics and Applied Biology, 42(12): 1323-1337.]

S. robustum and domesticated species *S. officinarum* within the *Saccharum* genus (N57 vs NJ), primarily involved functions related to photosynthesis, secondary metabolism, and signal transduction, which likely led to variations in their photosynthetic capacity, resistance to biotic and abiotic stresses, and nutrient utilization. *S. spontaneum* exhibited greater potential in nitrogen metabolism compared to *T. arundinaceum* and *S. robustum*, suggesting its potential advantages in nitrogen utilization. Furthermore, species-specific expressed genes in *S. spontaneum* were associated with photoreceptors, circadian rhythm clocks, and flowering pathways, indicating its unique flowering feature. The study also identified 4 325 orthologous genes among sugarcane, its close relatives, wild species, and domesticated varieties within the *Saccharum* genus, and positive selection analysis of these genes indicated that *S. spontaneum* gradually altered its gene expression patterns under natural selection to adapt to its environment, while the differentiation and adaptation of *S. officinarum* during the domestication process from *S. robustum* were likely related to the change of plant hormone signal transduction.

Keywords Sugarcane; Sugarcane-related species; Comparative transcriptome; RNA-seq; *Tripsidium arundinaceum*; *Saccharum spontaneum*; *Saccharum robustum*

甘蔗(*Saccharum* spp.)在热带和亚热带地区的90多个国家被广泛种植,中国是继巴西和印度之后的第三大甘蔗生产国。甘蔗是生产食品、饲料、生物燃料、生物制品等的重要原料,为全球100多个国家提供80%以上的蔗糖(Paidipati et al., 2022)。长期以来,甘蔗栽培品种遗传背景狭窄、品种退化(Scortecchi et al., 2012),以及缺乏对生物和非生物胁迫的抗性(Cursi et al., 2022)。全球环境变化和需求的不断增长进一步提高了对作物改良的需求,利用野生遗传资源通过品种改良来抵御生物和非生物胁迫并提高生物产量是现代育种的重要目标。

甘蔗育种者将现有甘蔗品种与甘蔗近缘属野生种进行杂交,以丰富栽培甘蔗的基因库。这一举措有助于进一步提高甘蔗的产量,并引入一些对生产有益的重要特性,诸如增加分蘖数、提高株高、增加茎径等(Sanghera et al., 2019)。研究指出,斑茅(*Tripsidium arundinaceum*)和割手密(*Saccharum spontaneum*)具有对根腐病、线虫和烟粉虱的抗性(Croft et al., 2015);斑茅显示出较高的耐寒性和生物产量,且在渗透调节、抗氧化系统以及物质代谢方面具有显著的优势(肖芙蓉等, 2018);大茎野生种(*Saccharum robustum*)具有高纤维含量、抗氧化成分丰富以及生物产量高等特点(Alarmelu et al., 2018; Chandran et al., 2020)。这些甘蔗属及其近缘属种质资源可作为甘蔗品种改良的基因库。鉴于此,本研究的主要目的在于比较甘蔗属及其近缘属代表性种质的转录组,以深入了解它们在转录层面的差异。比较转录组的方法被广泛用于研究不同生物之间在转录水平的差异,是深入探究物种生物学特性的重要组学手段之一(Evangelistella et al., 2017; Rathinam et al., 2019; Li et al., 2020; Chen et al., 2022)。甘蔗属及其近缘属物种的比较转录组研究

不仅能深化对这些物种转录特异表达的认识,还能丰富基因组资源。

本文研究了甘蔗近缘种斑茅与甘蔗属野生种割手密(BM vs Y83)、甘蔗属内的两个野生种割手密与大茎野生种(Y83 vs N57)以及甘蔗属内的野生种割手密和驯化种热带种(N57 vs NJ)之间的转录组特征。本研究的主要目的为:(1)比较甘蔗属中近缘野生种与野生种之间、甘蔗属中野生种之间、甘蔗属中野生种与驯化种之间的基因表达差异,从转录水平研究解析物种间特异表达的基因,挖掘野生种质中可用于遗传改良的优势基因(如光合作用、生物产量、抗性基因等);(2)挖掘物种间在选择压力下的正选择基因,阐明进化或驯化过程中的正选择基因及其表达情况,鉴定分化和适应的相关功能基因。本研究结果将为未来的甘蔗基因功能研究和分子标记辅助育种提供宝贵的资源。

1 结果与分析

1.1 测序数据处理与参考序列组装评估

四个代表物种斑茅(*T. arundinaceum*)(BM)、割手密(*S. Spontaneum*)(Y83)、大茎野生种(*S. robustum*)(N57)和热带种(*S. officinarum*)(NJ)通过Illumina测序产生67 119 545个平均干净读数,Q20和(G+C)含量的平均值分别为98.18%和53.77%(表1),测序干净读数的质量满足后续分析需求。在斑茅、割手密、大茎野生种和热带种的参考转录本中,分别有294 615、471 251、355 994和255 387个唯一基因序列(unigenes),平均长度分别为1 002 bp、829 bp、829 bp和1 080 bp,(G+C)含量分别为49.12%、49.37%、50.03%和49.59%,使用BUSCO评估组装质量,其完整性分别为72.37%、69.39%、68.96%和76.89%(表2)。

表 1 转录组测序质控数据统计表
Table 1 Statistics of transcriptome sequencing quality control data

物种 Species	样本名 Sample name	干净读数 Clean reads	干净碱基 Clean bases	Q20 占比/% Q20 percent/%	(G+C) 含量/% (G+C) content/%
<i>T. arundinaceum</i>	BM_1_Le	62 997 854	9 425 968 358	98.25	52.60
<i>T. arundinaceum</i>	BM_1_RO	62 662 590	9 367 070 589	98.34	55.15
<i>T. arundinaceum</i>	BM_2_Le	57 786 612	8 647 575 094	98.45	52.51
<i>T. arundinaceum</i>	BM_2_RO	69 405 670	10 376 543 075	98.17	54.48
<i>T. arundinaceum</i>	BM_3_Le	66 469 980	9 938 618 901	97.98	52.66
<i>T. arundinaceum</i>	BM_3_RO	82 750 466	12 365 355 897	98.18	55.02
<i>S. robustum</i>	N57_1_Le	71 299 020	10 645 724 016	98.11	52.21
<i>S. robustum</i>	N57_1_RO	80 475 232	12 014 263 455	98.32	54.11
<i>S. robustum</i>	N57_2_Le	63 931 562	9 522 420 611	97.91	49.77
<i>S. robustum</i>	N57_2_RO	71 184 784	10 627 411 423	98.05	54.06
<i>S. robustum</i>	N57_3_Le	63 080 558	9 415 353 133	98.30	47.20
<i>S. robustum</i>	N57_3_RO	65 112 728	9 722 588 570	98.02	54.16
<i>S. officinarum</i>	NJ_1_Le	63 079 880	9 416 349 063	98.11	54.88
<i>S. officinarum</i>	NJ_1_RO	69 218 142	10 334 199 855	98.11	53.57
<i>S. officinarum</i>	NJ_3_Le	61 398 728	9 168 710 175	98.02	54.71
<i>S. officinarum</i>	NJ_3_RO	70 892 496	10 586 371 278	98.11	53.19
<i>S. officinarum</i>	NJ_4_Le	58 701 100	8 777 470 006	98.20	55.59
<i>S. officinarum</i>	NJ_4_RO	52 695 306	7 880 508 656	98.00	54.38
<i>S. spontaneum</i>	Y83_1_Le	55 979 770	8 375 796 973	98.31	56.21
<i>S. spontaneum</i>	Y83_1_RO	76 921 728	11 485 760 548	98.26	53.80
<i>S. spontaneum</i>	Y83_2_Le	63 084 960	9 427 919 349	98.27	55.25
<i>S. spontaneum</i>	Y83_2_RO	72 597 854	10 841 606 554	98.21	54.99
<i>S. spontaneum</i>	Y83_3_Le	63 816 138	9 541 129 810	98.33	55.03
<i>S. spontaneum</i>	Y83_3_RO	85 325 922	12 736 122 455	98.28	54.99
平均值 Average		671 19 545	10 026 701 576	98.18	53.77

表 2 参考转录本 BUSCO 组装质量评估数据统计表
Table 2 Statistical table of BUSCO assembly quality assessment data for reference transcripts

项目 Item	斑茅 <i>T. arundinaceum</i>	大茎野生种 <i>S. robustum</i>	热带种 <i>S. officinarum</i>	割手密 <i>S. spontaneum</i>
完整 BUSCOs	1 168	1 120	1 113	1 241
Complete BUSCOs	(72.37%)	(69.39%)	(68.96%)	(76.89%)
片段化 BUSCOs	300	342	341	265
Fragmented BUSCOs	(18.59%)	(21.19%)	(21.13%)	(16.42%)
缺失 BUSCOs	146	152	160	108
Missing BUSCOs	(9.05%)	(9.42%)	(9.91%)	(6.69%)
总 BUSCOs	1 614	1 614	1 614	1 614
Total BUSCOs	(100%)	(100%)	(100%)	(100%)

注：括号内比例为所属 BUSCOs 类别占总 BUSCOs 的比例。
Note: The proportion in parentheses represents the proportion of the BUSCO category to the total BUSCOs.

对处理后的数据进行转录本组装，并剔除冗余的转录本后，通过 Nr、Swiss-Prot、KOG 和 PDB 四大数据库进行注释。大多数转录本在 Nr 数据库中获得注释，在 PDB 数据库中获得注释最少。在斑茅、割手密、大茎野生种和热带种中，四个数据库的注释率分别约为 54.94%、56.02%、58.25% 和 54.94% (表 3)。

表 3 斑茅、割手密、大茎野生种和热带种转录组的注释结果

Table 3 Annotation results for *T. arundinaceum*, *S. spontaneum*, *S. robustum*, and *S. officinarum* transcriptome reference

特种 Species	Nr	Swiss-Prot	KOG	PDB	注释唯一基因序列的个数 Number of annotated unigenes	注释率/% Rate of annotations/%
斑茅 <i>T. arundinaceum</i>	161 175	95 035	98 830	58 156	161 863	54. 94
割手密 <i>S. spontaneum</i>	206 681	150 593	151 611	95 689	263 987	56. 02
大茎野生种 <i>S. robustum</i>	263 116	121 774	120 556	73 040	207 373	58. 25
热带种 <i>S. officinarum</i>	139 886	79 576	82 492	48 506	140 321	54. 94

使用 BLASTx 将转录本比对到多个物种，观察到四个物种与甘蔗杂交种 R570 的比对率极低，可能是由于 R570 的基因组组装和注释不完整所导致。

表 4 甘蔗属及其近缘属代表种质的转录组与 Nr 数据库的比较分析

Table 4 Comparative analysis of transcriptome and Nr database of representative germplasm of sugarcane and its relatives

物种 Species	比对率/% Mapping rate/%			
	斑茅 <i>T. arundinaceum</i>	割手密 <i>S. spontaneum</i>	大茎野生种 <i>S. robustum</i>	热带种 <i>S. officinarum</i>
高粱 <i>Sorghum bicolor</i>	45. 4	46. 0	46. 0	46. 0
玉米 <i>Zea mays</i>	21. 5	13. 6	13. 9	14. 1
南荻 <i>Miscanthus lutarioriparius</i>	14. 3	15. 7	16. 0	15. 8
福尼奥小米 <i>Digitaria exilis</i>	2. 0	2. 3	2. 1	2. 1
柳枝稷 <i>Panicum virgatum</i>	1. 9	2. 7	2. 5	2. 7
稷 <i>Panicum miliaceum</i>	1. 8	2. 2	2. 0	2. 0
小米 <i>Setaria italica</i>	1. 7	1. 7	1. 5	1. 6
二型花小米草 <i>Dichanthelium oligosanthes</i>	1. 3	—	—	—
甘蔗杂交品种 R570 <i>Saccharum hybrid R570</i>	—	1. 8	2. 0	2. 0
其他 Other	10. 1	14. 0	14. 1	13. 7

注：“—”表示该值未检测。

Note: “—” indicates that the value is not detected.

四个物种共有唯一基因序列共 58 258 个(图 1A)，分别通过 GO 数据库和 KEGG 数据库注释共有唯一基因序列。有 9 883 个候选基因注释到 KEGG 通路中，它们被分为 5 类，涵盖 138 个不同的 KEGG 途径。其中，“糖酵解/糖异生”、“柠檬酸循环(TCA 循环)”和“磷酸戊糖途径”是三条最具

代表性的途径(图 1B)。注释结果显示，甘蔗属及其近缘属的转录本主要涉及能量代谢和碳水化合物代谢。在三个核心 GO 注释类别中，生物过程占注释总数的 46. 62% (27 164)，分子功能和细胞组分分别占 49. 64% (28 917) 和 49. 11% (28 614) (图 1C)。

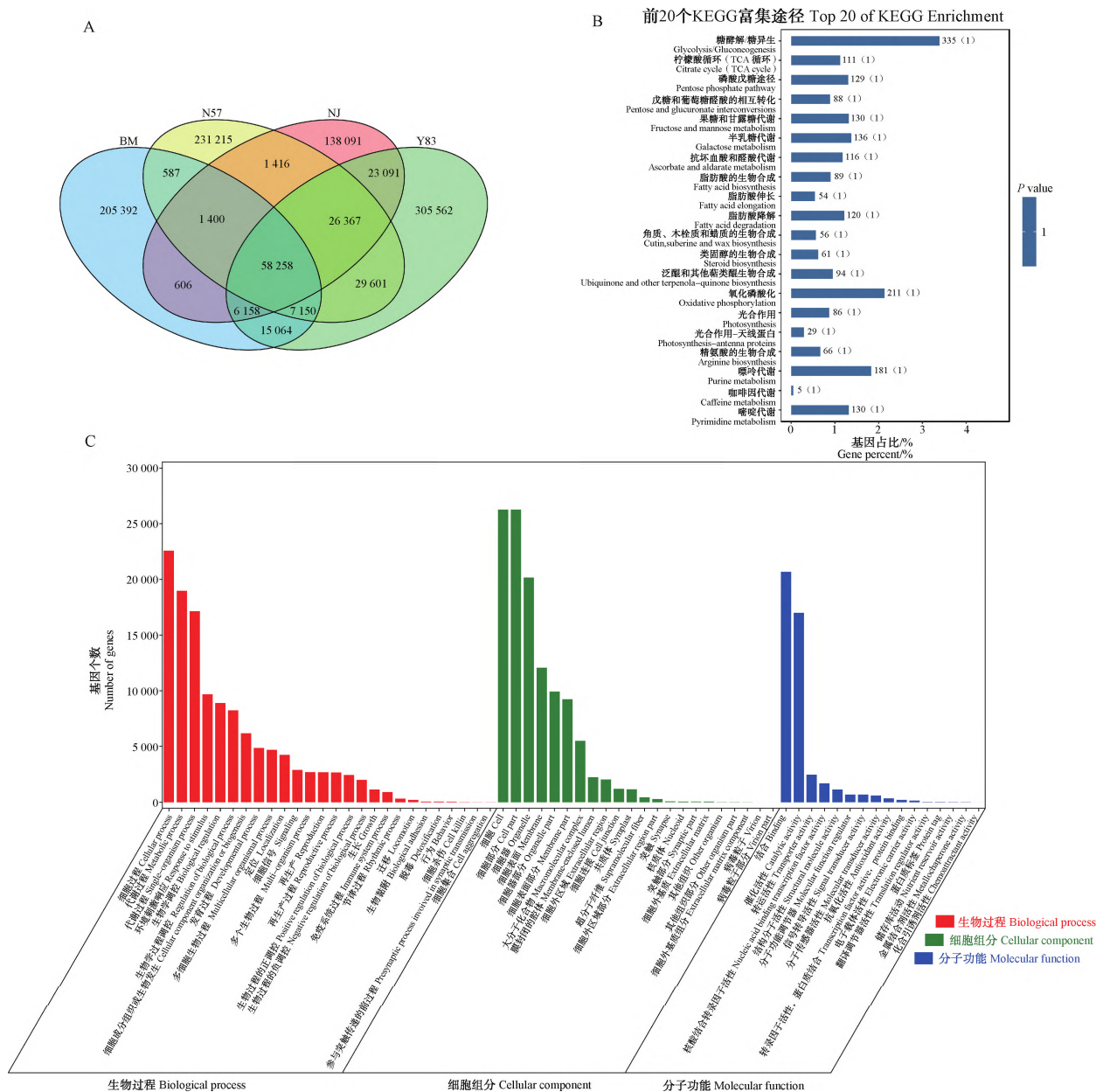


图1 斑茅、割手密、大茎野生种和热带种中转录组注释情况统计

(A) 四个物种共享基因的维恩图; (B) 四个物种共享基因的前 20 个 KEGG 通路注释;
(C) 四个物种共享基因的前 64 个 GO 分类

Figure 1 Statistical analysis of transcriptomes in *T. arundinaceum*, *S. spontaneum*,
S. robustum, and *S. officinarum*

(A) Venn diagram of shared genes in four species; (B) Top-20 KEGG pathway annotation of the four species shared genes; (C) Top-64 GO classification of the four species shared genes

1.2 组织特异性和物种特异性表达分析

为了比较四个种质的组织特异性和物种特异性的转录本表达特征,将四个种质之间共有的 58 258 个转录本用于构建每百万转录本数 (transcripts per million, TPM) 矩阵。组织特异性表达基因定义为在四分之三的物种中表达,一个组织中“TPM>1”,其他组织中“TPM=0”。物种特异性表达基因定义为在

一个物种中以阈值“TPM>1”,在其余物种中以“TPM=0”特异表达。

在组织特异性表达的基因中,有 10 个基因在根组织中呈现特异性表达,这些基因被注释到不同的生物学通路,包括“苯丙氨酸代谢”、“核糖体”以及“角质、木栓质和蜡质的生物合成”等途径。在叶组织中,检测到共有 114 个基因呈现特

异性表达。这些基因分别被注释到多个生物学通路,其中包括“光合作用-天线蛋白”、“光合作用”、“乙醛酸和二羧酸代谢”以及“光合生物中的碳固定”等途径。这些结果表明,根组织的特异表达基因与植物次生代谢和细胞壁合成有关,而叶组织特异性表达基因与其形态组成和生理功能密切相关。

在物种特异性表达基因(表5)中,斑茅具有最多的特异性表达基因,包括354个叶组织特异表达基因、232个根组织特异表达基因,以及109个同时在叶和根组织中表达的基因,它们分别被注释到

“内吞”、“生物素代谢”和“*N*-聚糖生物合成”等途径。大茎野生种和热带种的物种特异性表达基因都显著富集到“苯丙氨酸代谢”、“苯丙酸生物合成”和“苯丙酸合成”等途径中,表明这两个物种特异性表达基因类似。割手密的物种特异性表达基因主要被注释到基本代谢途径中,如“嘌呤代谢”和“果糖和甘露糖代谢”等。割手密的组织特异性表达基因中,叶和根组织中的特异表达基因显著富集到“磷酸和亚磷酸代谢”以及“烟酸和烟酰胺代谢”途径中,根组织特异表达基因富集到“磷酸和亚磷酸代谢”途径中。

表5 斑茅、割手密、大茎野生种和热带种中的组织特异性表达基因的统计

Table 5 Statistics of tissue specific expressed genes in *T. arundinaceum*, *S. spontaneum*, *S. robustum*, and *S. officinarum*

	叶组织 Leaf tissue	根组织 Root tissue	叶和根组织 Leaf tissue and root tissue
斑茅 <i>T. arundinaceum</i>	354	232	109
割手密 <i>S. spontaneum</i>	265	145	59
大茎野生种 <i>S. robustum</i>	64	152	24
热带种 <i>S. officinarum</i>	262	135	63

1.3 物种间差异表达及 KEGG 富集分析

差异表达分析采用了三个比较组,分别是甘蔗近缘属野生种与甘蔗属野生种(BM *vs* Y83)、甘蔗属内的野生种(Y83 *vs* N57),以及甘蔗属内中的野生

种和驯化种(N57 *vs* NJ)。将获得注释的四个物种的共享转录本用作背景基因集,对三个比较组中的差异表达基因(differentially expressed genes, DEGs)进行 KEGG 富集分析(P -value<0.01)(表6)。

表6 叶和根组织中差异性表达基因的 KEGG 通路富集情况

Table 6 Enrichment of KEGG pathway of differentially expressed genes in leaf and root tissues

比较组 Comparison group	叶组织中的富集通路 Leaves tissue enriched pathway	ko	P -value	根组织中的富集通路 Roots tissue enriched pathway	ko	P -value
Y83 <i>vs</i> BM	光合作用-天线蛋白 Photosynthesis-antenna proteins	ko00196	0.000 111	苯丙酸生物合成 Phenylpropanoid biosynthesis	ko00940	0.006 616
	乙醛酸和二羧酸代谢 Glyoxylate and dicarboxylate metabolism	ko00630	0.001 248			
	卟啉代谢 Porphyrin metabolism	ko00860	0.002 217			
	花生四烯酸代谢 Arachidonic acid metabolism	ko00590	0.007 94			
	亚油酸代谢 Linoleic acid metabolism	ko00591	0.002 253	苯丙酸生物合成 Phenylpropanoid biosynthesis	ko00940	0.007 999
BM <i>vs</i> Y83	磷酸戊糖途径 Pentose phosphate pathway	ko00030	0.002 959			
	硫代谢 Sulfur metabolism	ko00920	0.008 312			
	泛醌和其他萜类醌生物合成 Ubiquinone and other terpenoid-quinone biosynthesis	ko00130	0.009 855			

续表
Continuing table

比较组 Group	叶组织中的富集通路 Leaves tissue enriched pathway	ko	P-value	根组织中的富集通路 Roots tissue enriched pathway	ko	P-value
N57 vs Y83	蛋白质出口 Protein export	ko03060	0.009 579			
Y83 vs N57	乙醛酸和二羧酸代谢 Glyoxylate and dicarboxylate metabolism	ko00630	0.002 141	苯丙素类化合物生物合成 Phenylpropanoid biosynthesis	ko00940	0.000 168
	卟啉代谢 Porphyrin metabolism	ko00860	0.002 586	植物激素信号转导 Plant hormone signal transduction	ko04075	0.001 565
	昼夜节律-植物 Circadian rhythm-plant	ko04712	0.009 384	类黄酮生物合成 Flavonoid biosynthesis	ko00941	0.002 528
				芪类、二芳基庚烷类和姜醇生物合成 Stilbenoid, diarylheptanoid and gingerol biosynthesis	ko00945	0.006 283
NJ vs N57	α-亚麻酸代谢 Alpha-linolenic acid metabolism	ko00592	0.000 595	α-亚麻酸代谢 Alpha-linolenic acid metabolism	ko00592	0.000 856
	光合作用 Photosynthesis	ko00195	0.001 405	糖酵解/糖异生 Glycolysis/Gluconeogenesis	ko00010	0.001 885
	糖酵解/糖异生 Glycolysis/Gluconeogenesis	ko00010	0.002 981	碳代谢 Carbon metabolism	ko01200	0.002 086
	萜类骨架生物合成 Terpenoid backbone biosynthesis	ko00900	0.005 281	苯丙氨酸代谢 Phenylalanine metabolism	ko00360	0.005 006
N57 vs NJ	光合作用 Photosynthesis	ko00195	0.000 134	丙氨酸、天冬氨酸和谷氨酸代谢 Alanine, aspartate and glutamate metabolism	ko00250	0.008 558

注：在对比组中，基因上调表达指前面的物种的基因表达水平高于后面的物种（例如，在 Y83 与 BM 的对比组 Y83 vs BM 中，Y83 的基因表达相较于 BM 的基因表达上调）。

Note: In the comparison group, upregulation of gene expression refers to a higher level of gene expression in the preceding species than in the following species (for example, in the Y83 vs BM comparison group, Y83 gene expression is upregulated compared to BM gene expression).

在割手密和斑茅 (BM vs Y83) 的比较组内，在叶组织中，鉴定到 7 106 个割手密中上调 DEGs，鉴定到 6 175 个斑茅上调 DEGs；在根组织中，鉴定到 6 954 个割手密上调 DEGs，鉴定到有 6 036 个斑茅上调的 DEGs。在叶组织中，割手密最显著富集的通路是“光合作用-天线蛋白”“乙醛酸和二羧酸代谢”“卟啉代谢”和“花生四烯酸代谢”，而斑茅中最显著富集的通路是“亚油酸代谢”“磷酸戊糖途径”“硫代谢”和“泛醌和其他萜类醌生物合成”。在割手密和斑茅的根和叶组织中，最主要的富集通路都是“苯丙酸生物合成”。

在大茎野生种和割手密 (Y83 vs N57) 的比较组内，在叶组织中，鉴定到 4 024 个大茎野生种中上调 DEGs，鉴定到 5 696 个割手密的上调 DEGs。在根组织中，鉴定到 5 762 个大茎野生种的上调 DEGs，而鉴定到 5 303 个割手密的上调 DEGs。在叶组织中，大茎野生种最显著富集的通路是“蛋白质输出”，在割手密中最显著富集的通路是“乙醛酸和二羧酸代谢”“卟啉代谢”和“昼夜节律-植

物”。在根组织中，割手密最明显富集的通路是“苯丙素类化合物生物合成”“植物激素信号转导”“类黄酮生物合成”和“芪类、二芳基庚烷类和姜醇生物合成”，而在大茎野生种根组织中没有发现明显富集的途径。

在热带种和大茎野生种 (N57 vs NJ) 的比较组内，在叶组织中，鉴定到 5 168 个热带种中上调 DEGs，有 3 279 个大茎野生种中上调 DEGs；在根组织中有 5 017 个热带种中上调 DEGs，有 4 815 个大茎野生种中上调 DEGs。在叶组织中，热带种最显著富集的通路是“α-亚麻酸代谢”“光合作用”“糖酵解/糖异生”和“萜类骨架生物合成”，而大茎野生种的差异表达基因最显著富集于“光合作用”通路。在根组织中，大茎野生种中最显著富集的通路是“丙氨酸、天冬氨酸和谷氨酸代谢”通路。相比之下，热带种在根组织中最显著富集的通路包括“碳代谢”和“苯丙氨酸代谢”通路。此外，在根组织和叶组织中，“α-亚麻酸代谢”和“糖酵解/糖异生”途径也都表现出显著富集。这些结果提供了大茎野生种和热

带种在根组织中生物学通路的差异,突显了它们在根部代谢活动方面的差异性。

1.4 同源基因对及正选择分析

正选择基因通常反映物种对自然选择或人工选择的适应,这与它们在不同环境条件下的生存密切相关。对甘蔗近缘属的野生种与甘蔗属的野生种之间、甘蔗属的野生种之间以及甘蔗属的驯化种与野生种之间受到选择(自然选择或人工选择)的基因进行分析和比较。

去冗余后的序列经过开放阅读框(open reading frame, ORF)预测后,使用 OrthoFinder 软件在斑茅、割手密、大茎野生种和热带种四个物种之间鉴定了 4 325 个直系同源基因。如图 2 所示,在斑茅和割手

密的比较组之间,有 136 个直系同源基因对的 $Ka/Ks>1$, 这表明它们正在经历正向选择。通过 KEGG 通路富集来研究正选择基因的生物学功能,136 个直系同源基因被分配到 19 条 KEGG 途径,显著富集到的通路是“核糖体”、“D-氨基酸代谢”和“氧化磷酸化”。在割手密和大茎野生种的比较组之间,鉴定了 3 964 个直系同源基因,512 个直系同源基因的 $Ka/Ks>1$, 这些基因被分配到 64 条 KEGG 途径,最具代表性的通路是“硫代谢”。在大茎野生种和热带种的比较组之间共鉴定了 3 643 个直系同源基因,其中 714 个直系同源基因的 $Ka/Ks>1$, 这些基因被分配到 72 条 KEGG 途径,最具代表性的通路是“DNA 复制”、“植物激素信号转导”、“碱基切除修复”和“蛋白酶体”。

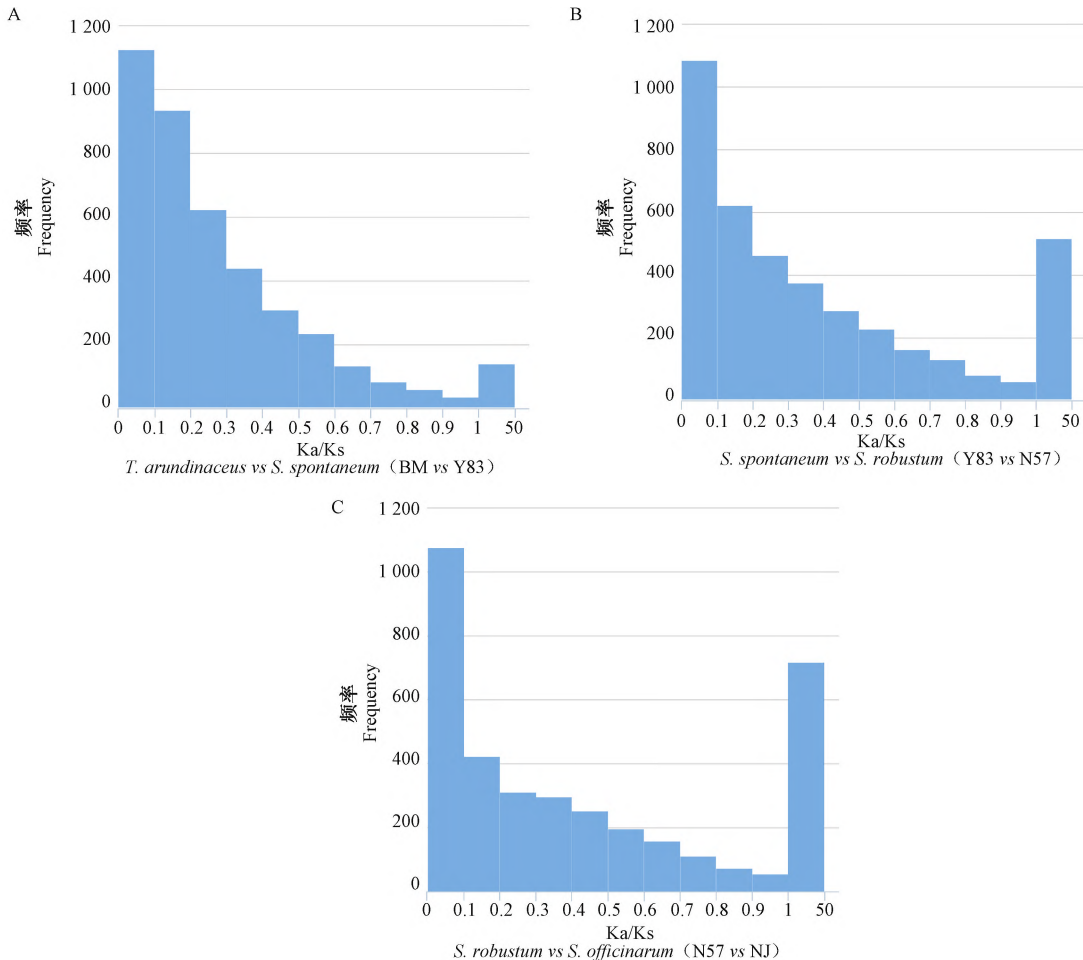


图2 斑茅、割手密、大茎野生种和热带种中同源基因的选择压力分析
(A) 斑茅与割手密比较组的频率分布图(BM vs Y83); (B) 割手密与大茎野生种比较组的频率分布图(Y83 vs N57);
(C) 大茎野生种与热带种比较组的频率分布图(N57 vs NJ)

Figure 2 Selection pressure analysis of homologous genes in *T. arundinaceum*,
S. spontaneum, *S. robustum*, and *S. officinarum*

(A) Frequency distribution diagram of *T. arundinaceum* vs *S. spontaneum* (BM vs Y83);
(B) Frequency distribution diagram of *S. spontaneum* vs *S. robustum* (Y83 vs N57);
(C) Frequency distribution diagram of *S. robustum* vs *S. officinarum* (N57 vs NJ)

进一步对三个比较组中的受到正选择的基因和组织特异性表达基因进行交集，以深入了解这些基因在不同组织中的调控和演化模式。在斑茅和割手密的比较组中没有鉴定到组织特异性表达并同时受到正选择的基因。在割手密和大茎野生种的比较组中，有 7 个直系同源基因 (*Sspon. 03G0003510-1A*、*Sspon. 04G0011670-3D*、*Sspon. 08G0026130-2D*、*Y83_*

TRINITY_GG_13372_c64_g1_i1、*Y83_Trinity_GG_25353_c59_G11_i4*、*Y83_TRINITY_GG_3580_c112_g1_i2*、*Y83_TRINITY_GG_44371_c231_g1_i2*) 在根组织中特异性表达。在大茎野生种和热带种的比较组中，3 个直系同源基因 (*Sspon. 03G0003510-1A*、*Y83_TRINITY_GG_13372_c64_g1_i1*、*Y83_Trinity_GG_50772_c43_g2_i1*) 在根组织中特异性表达(表 7)。

表 7 正选择和组织特异性表达基因交集基因的注释情况

Table 7 Annotation of the intersection of positive selection and tissue-specific expressed genes			
比较组 Comparison group	组织 Tissue	基因 ID Gene ID	基于 Nr 蛋白数据库的注释信息 Annotation information based on the Nr database
Y83 vs N57	根	<i>Sspon. 03G0003510-1A</i>	γ-干扰素诱导的溶酶体硫醇还原酶 Gamma-interferon-inducible lysosomal thiol reductase
		<i>Sspon. 04G0011670-3D</i>	致病相关 PRB 蛋白 Pathogenesis-related PRB protein
		<i>Sspon. 08G0026130-2D</i>	AT 基序核定位蛋白 AT-hook motif nuclear-localized protein
		<i>Y83_TRINITY_GG_13372_c64_g1_i1</i>	富含甘氨酸的蛋白质 Glycine-rich protein
		<i>Y83_TRINITY_GG_25353_c59_g1_i4</i>	<i>Panicum virgatum</i> rRNA 2'-O-甲基转移酶纤维蛋白样 <i>Panicum virgatum</i> rRNA 2'-O-methyltransferase fibrillar-like
		<i>Y83_TRINITY_GG_3580_c112_g1_i2</i>	毛状体双折射样蛋白 34 Protein trichome birefringence-like 34
		<i>Y83_TRINITY_GG_44371_c231_g1_i2</i>	推定脂质转移蛋白 DIR1 Putative lipid-transfer protein DIR1
N57 vs NJ	根	<i>Sspon. 03G0003510-1A</i>	γ-干扰素诱导的溶酶体硫醇还原酶 Gamma-interferon-inducible lysosomal thiol reductase
		<i>Y83_TRINITY_GG_13372_c64_g1_i1</i>	富含甘氨酸的蛋白质 Glycine-rich protein
		<i>Y83_TRINITY_GG_50772_c43_g2_i1</i>	木聚糖半乳糖基转移酶 KATAMARI 同源物 Xyloglucan galactosyltransferase KATAMA-RII homolog

2 讨论与结论

在本研究中，利用斑茅、割手密、大茎野生种和热带种的叶、根组织的转录组数据，分别探究物种间特异性和差异表达基因。在这里，使用有参加无参组装的策略来保留四个物种中所有可能的转录本，并通过将每个物种的组装结果与割手密参考基因序列及割手密的转录本数据集相关联，构建了一个甘蔗属内跨物种的全面参考转录本数据集。

组织特异性表达基因主要参与叶片光合作用和根系形态建成，这与其特有的生理功能相关；而物种特异性表达基因主要涉及植物的免疫响应、次生代谢产物合成和养分吸收利用等。此外，在甘蔗近缘属野生种与甘蔗属野生种之间(*BM vs Y83*)、甘蔗属内野生种之间(*Y83 vs N57*)以及甘蔗属野生种与驯化种之间(*N57 vs NJ*)中鉴定到的 DEGs 主要涉及光合作用、次级代谢产物合成和信号转导等功能，这在一定程度上反映了它们在光合能力、对生物和

非生物胁迫的抗性以及养分利用方面的差异。

2.1 甘蔗属及其近缘属种质的光合作用差异

光合作用对植物至关重要，为植物提供必要的能量来源。在本研究中，主要关注参与光合作用的重要 KEGG 通路，如“光合作用-天线蛋白”、“光合作用”和“光合生物中的碳固定”。本研究对三个比较组中的两两物种间进行转录组比较，以观察光合作用相关基因的差异表达情况，这有助于深入了解这些物种在光合作用方面的生物学差异。

在绿色植物的光合色素蛋白复合体中，天线蛋白充当着外围天线系统的角色，在光能的吸收中起着重要作用。在类囊体膜中，电子转移过程主要由光合作用复合体 (light-harvesting complexes, LHCs) 吸收的光能引发。植物中的 LHCs 含有光合色素，包括叶绿素 a 和叶绿素 b，以及与 LHCs 家族蛋白相关联的少量类胡萝卜素。LHC II 是主要的天线复合体，它结合大约 50% 的叶绿素，约占类囊体蛋白的一

半,由蛋白超家族的三个成员(Lhcb1、Lhcb2和Lhcb3)组成。与BM(*T. arundinaceum*)和N57(*S. robustum*)相比,Y83(*S. spontaneum*)中显著上调的基因是光系统I中的*Lhca1*、*Lhca3*、*Lhca5*和光系统II中的*Lhcb1*、*Lhcb4*、*Lhcb6*。Y83的光系统II的实际量子产量(Φ_{PSII})和净光合速率比BM、N57和NJ的高(冯梦凡等,2022),这些研究结果表明光捕获化合物基因表达量的上调可能与Y83的光合能力增加存在密切关系。

光合作用需要进一步将在光合作用复合体(例如光合作用复合体中的光系统II)中捕获的太阳能转换成化学能。光系统II(photosystem II, PS II)是与光合作用中的水分解反应和光保护密切相关的关键组件。无论是在正常生长条件下(PsbO和PsbP)(Ifuku et al., 2005; Yi et al., 2005; Yi et al., 2007),还是在低光强度生长条件下(PsbQ)(Yi et al., 2006),光系统II的PsbO、PsbP和PsbQ蛋白在高等植物PS II反应中心起着至关重要的作用(Bricker and Frankel, 2011)。在BM vs Y83和N57 vs NJ中,两个比较组都观察到编码光合中心的基因表达有差异,这表明它们光能的吸收效率的差异可能与这些基因相关。

2.2 割手密易开花原因的启示

在甘蔗育种中,种质的花期是否接近或是否能够开花都对育种进程有影响。然而,开花受到环境条件的调控,包括纬度和温度等因素(Manechini et al., 2021)。在甘蔗属物种中,割手密通常较易开花(Durai et al., 2014; Lu et al., 2015)。在Y83 vs N57和BM vs Y83比较组中差异表达基因富集在影响光周期的KEGG通路中,其中专门编码光敏色素用于感测红光/远红光光感受器的植物光敏色素B(phytochrome B, *PHYB*)和植物光敏色素A(phytochrome A, *PHYA*)、在割手密中表达上调(Lu et al., 2015);而隐花色素1(cryptochrome 1, *PHY1*),作为蓝光光感受器在植物中产生多种发育和昼夜节律反应(Brautigam et al., 2004),也在割手密中表达上调。节律时钟中央振荡器中保守时钟成分,如编码PRR(pseudo-response regulator)蛋白家族的基因、节律基因*LHY*(late elongated hypocotyl)、时钟基因*TOC1*(timing of CAB expression 1)和时钟基因*ZTL*(zeite-locus protein),这些基因的表达上调可能与光周期的刺激相关。通过物种间的比较发现,与开花途径相关的典型基因,如成花素(flowering locus T,

FT)、节律钟输出基因(*GIGANTEA*, *GI*)和*CONSTANS*基因(*CO*)在物种间比较发现其上调,这类类似于先前关于甘蔗光周期的研究结果(Manechini et al., 2021),这些基因在开花前的成熟叶片中的信号传导阶段开始时变得更为活跃。*FKF1*(flavin-binding, kelch repeat, F-BOX 1)对拟南芥的日长识别至关重要,有研究发现其在割手密中表达上调(Imaizumi et al., 2003)。拟南芥中的早花基因3(early flowering 3, *ELF3*)通过影响时钟相关基因,如节律基因*LHY*(late elongated hypocotyl, *LHY*)、节律基因*CCA1*(circadian clock associated 1)和*GI*的转录来调节循环节律(Murakami et al., 2007; Matsubara et al., 2012)。在Y83 vs N57的比较组中,鉴定到割手密中的*ELF3*基因差异表达。相对于甘蔗属中的其他物种,割手密易开花的现象与光感受器、昼夜节律时钟的中心振荡器等开花途径的差异表达基因密切相关。

2.3 野生种质为甘蔗育种提供了基因库

植物中通过碳固定产生的碳骨架和由植物脂氧合酶生成的氧化脂质参与多种生理过程,如生长发育、分蘖等,最终影响植物生物量的增加(Yan et al., 2021)。在BM vs Y83的比较组中,差异基因富集的KEGG通路是“碳代谢”、“亚油酸代谢”和“光合生物中的碳固定”。碳固定是指将二氧化碳(CO_2)固定并合成为有机化合物的过程(Bar-Even et al., 2010),这些有机化合物不仅被用于储存能量,而且也作为植物生物体内其他重要分子的关键组成部分。通过增加碳固定的过程,可以改善植物的生长,并促进生物体内的生物量积累。这被认为是一种可行的方法,相关研究表明这种策略有助于提高植物的生态效益和生产效率(Ducat and Silver, 2012; Betti et al., 2016; Bar-Even, 2018)。在本研究中, C_4 二羧酸途径中的丙酮酸磷酸二激酶(phosphoenolpyruvate phosphate dikinase, *ppdk*)和磷酸烯醇式丙酮酸羧化酶(phosphoenolpyruvate carboxylase, *ppc*)与亚油酸代谢途径中负责氧化多不饱和脂肪酸的脂氧合酶(lipoxygenase, *LOX*)基因的在斑茅和割手密之间差异表达,可能是斑茅和割手密之间生物量存在差异的原因之一。

研究表明可以通过将现有甘蔗品种与割手密进行杂交来实现甘蔗根系的改良(Sakaiguchi et al., 2007; Takaragawa et al., 2022)。在割手密与斑茅(BM vs Y83)和割手密与大茎野生种(Y83 vs N57)的比较组中,通过分析根组织中的DEGs,发现这些基

因主要被注释到氮代谢通路,并且与氮源的吸收和同化密切相关。这表明割手密在氮素利用方面相对斑茅和大茎野生种具有优势。这一研究结果为利用割手密的优良性状(如氮素高效利用)来改良甘蔗品种提供了重要线索。

2.4 甘蔗对非生物胁迫反应的转录组学

植物采用一系列的免疫机制来抵御由病原体引起的感染(Jones and Dangl, 2006),其中包括利用抗性蛋白识别特异性病原体效应物,从而启动效应物触发的免疫反应。在病原体感染后,高水平的水杨酸(salicylic acid, SA)会激活病原相关基因的表达,其中一些基因编码具有抗微生物活性的蛋白质(van Loon et al., 2006)。割手密的差异表达基因还在植物激素信号传导途径和色氨酸代谢途径中富集,其中一些相关基因的表达也呈现出不同程度的上调,例如生长素响应因子(auxin/indole-3-acetic acid, AUX/IAA)等。苯丙氨酸代谢相关基因在割手密的根组织中表达上调,特别是调节蛋白 NPR1(nonexpresser of pathogenesis-related genes 1, NPR1)、转录因子 TGA(transcription factor TGA, TGA)、病原相关蛋白 1(pathogenesis-related protein 1, PR1)(Kumar et al., 2022)等。PR1 在多种植物中参与抗病反应(Jones and Dangl, 2006),包括甘蔗赤条病原体引起的病害中,已经显示出该基因对提高抗病能力起到了作用(Brigida et al., 2016),同时 PR1 基因在甘蔗的耐寒性中发挥了一定的作用(Fouladvand et al., 2022)。总之,研究结果表明这些差异基因可能在割手密抗病、根系生长甚至抗寒方面发挥作用。

研究表明,甘蔗可以通过调节碳水化合物分解、氧化还原、黄酮类和其他次生代谢物合成等途径的基因表达来应对非生物和生物胁迫(Pereira-Santana et al., 2017)。虽然本研究的材料没有经过环境胁迫处理,但仍观察到胁迫相关基因在比较组中的差异表达,如“泛醌和其他萜类醌生物合成”、“苯丙酸生物合成”和“类黄酮生物合成”等次级代谢物合成途径的相关基因。黄酮类化合物是一类重要的植物次生代谢产物,包括强抗氧化剂(Pourcel et al., 2007),可增强植物对高光强和紫外线辐射的耐受性,并吸引昆虫传粉(Agati and Tattini, 2010)。萜类化合物等次生代谢产物也具有广泛的生理和生态功能,如氧化应激、保护植物细胞免受活性氧(ROS)的损害、调节生长和发育、抵抗环境压力,并参与害虫的防御反应等(Dudareva et al., 2013)。苯丙氨

酸是一大类次级代谢物合成的前体,包括异黄酮类、木质素、木质纤维素、黄酮类和链烷类(Cuong et al., 2019)。前人研究表明,苯丙氨酸代谢、苯丙烷类物质和类黄酮物质合成显示出有利于甘蔗对利夫森氏细菌亚种(*Leifsonia xyli* subsp. *xyli*)(Fu et al., 2019)和黏虫(*Mythimna separata*)(Wang et al., 2021)的防御响应。

2.5 正选择压力分析

为深入了解不同种质在适应性进化方面的差异,本文对斑茅、割手密、大茎野生种和热带种的正选择基因进行了比较分析,研究结果有助于更全面地理解它们在自然环境中的演化策略和适应性特征。从甘蔗近缘属野生种与甘蔗属野生种、甘蔗属野生种以及甘蔗属的驯化种与野生种这三个比较组中得到明显不同的正选择基因数量和注释结果。正选择基因在近缘物种之间数量的不同反映了它们在物种演化过程中面临不同的环境压力和生态适应要求。通过研究这些正选择基因的分布和数量变化,能够更深入地了解每个物种在演化中的独特适应性路径,从而揭示它们在不同环境中的生态适应性演化。

与甘蔗属野生种割手密相比,在甘蔗属近缘物种斑茅中鉴定到核糖体蛋白大亚基和核糖体蛋白小亚基的编码基因、C₃型光合作用向 C₄光合转变过程中的基因和“氧化磷酸化”途径中 NADH-泛醌氧化还原酶(NADH-ubiquinone oxidoreductase)亚单位的基因受到强烈正选择。在割手密和大茎野生种的比较组中,经历正选择的基因主要富集在“硫代谢”通路中。在这一通路中,硫酸盐还原通过植物激素的诱导协调了多种生理过程和分子机制,进一步诱导了对盐胁迫的耐受性。这表明这些基因在植物对环境压力的适应中发挥着关键的作用(Fatma et al., 2013)。在大茎野生种和热带种中发现的正选择基因主要富集在“植物激素信号转导”通路中,其中 ABA 反应元件结合因子(abscisic acid response element binding factor, ABF)参与非生物胁迫反应(Amir Hossain et al., 2010),其可提高对植物环境胁迫的耐受性(Nakashima and Yamaguchi-Shinozaki, 2013)。乙烯不敏感蛋白基因(ethylene insensitive 2, EIN2)作为乙烯信号通路、ABA 信号通路等胁迫信号通路的交叉连接节点(Wang et al., 2007),同样鉴定其受到强烈的正选择信号,这表明 EIN2 在植物应对乙烯响应、干旱胁迫、盐碱胁迫等多种信号通路的交互中可能发挥着关键的调控作用,提示其在植

物进化过程中的重要性。

在自然选择压力下，割手密、斑茅和大茎野生种的正选择基因的数量和类型有所不同，尤其体现在氧化磷酸化和硫代谢方面。在热带种中，正选择基因主要类型与环境适应性信号方面有关(尤其是植物激素信号转导)，这与大茎野生种的正选择基因也存在差异。这些研究结果有助于从转录水平上了解甘蔗属及其近缘属物种在驯化过程中的差异和各自的进化特征。

3 材料与方法

3.1 实验材料

供试的4个种质分别为斑茅(*T. arundinaceum*) (BM)、割手密(*S. spontaneum*) (Y83)、大茎野生种(*S. robustum*) (N57)和热带种(*S. officinarum*) (NJ)。所有实验材料均种植于广西大学试验田(108°17'E, 22°50'N)，于苗期收集叶、根的组织样品，其中每个样品的叶和根都取3个生物学重复，详细信息如表8所示。

表8 本研究使用的样品信息表
Table 8 Table of information about the samples used in this study

物种 Species	样本名 Sample name	组织名称 Tissue name	重复样本 Sample duplicates
<i>T. arundinaceum</i>	BM_1_Le	叶	1
<i>T. arundinaceum</i>	BM_2_Le	叶	2
<i>T. arundinaceum</i>	BM_3_Le	叶	3
<i>T. arundinaceum</i>	BM_1_RO	根	1
<i>T. arundinaceum</i>	BM_2_RO	根	2
<i>T. arundinaceum</i>	BM_3_RO	根	3
<i>S. robustum</i>	N57_1_Le	叶	1
<i>S. robustum</i>	N57_2_Le	叶	2
<i>S. robustum</i>	N57_3_Le	叶	3
<i>S. robustum</i>	N57_1_RO	根	1
<i>S. robustum</i>	N57_2_RO	根	2
<i>S. robustum</i>	N57_3_RO	根	3
<i>S. officinarum</i>	NJ_1_Le	叶	1
<i>S. officinarum</i>	NJ_3_Le	叶	2
<i>S. officinarum</i>	NJ_4_Le	叶	3
<i>S. officinarum</i>	NJ_1_RO	根	1
<i>S. officinarum</i>	NJ_3_RO	根	2
<i>S. officinarum</i>	NJ_4_RO	根	3
<i>S. spontaneum</i>	Y83_1_Le	叶	1
<i>S. spontaneum</i>	Y83_2_Le	叶	2
<i>S. spontaneum</i>	Y83_3_Le	叶	3
<i>S. spontaneum</i>	Y83_1_RO	根	1
<i>S. spontaneum</i>	Y83_2_RO	根	2
<i>S. spontaneum</i>	Y83_3_RO	根	3

3.2 参考转录组的组装、构建和注释

使用 fastp 软件(v0.23.1) (Chen et al., 2018) 对 RNA-seq 测序产生的原始数据进行质控。该软件能够自动识别并去除测序数据中的接头序列，删除包含“N”的读数以及质量较低的读数，最终得到经过处理的高质量、无污染的测序读数用于下一步分析。使用 hisat2 软件(v2.2.1) (Kim et al., 2015) 将干净读数比对到割手密参考基因组 AP85-441 (2n = 4x = 32) (<http://sugarcane.zhangjisenlab.cn/sgd/html/download.html>)。比对上参考基因组的读数使用 Trinity 软件(v2.1.1) (Haas et al., 2013) 进行基因组引导的有参转录本组装，而未比对上的读数经过无参从头组装获得的转录本作为补充，得到每个甘蔗种质的转录本数据集。以割手密参考基因序列及 *S. spontaneum* (Y83) 的转录本数据集作为比对中心，与其他种质的转录本数据集建立对应关系，这一过程使用 GMAP 软件(v2021-08-25) (Wu and Watanabe, 2005) (参数: -min-Identity 0.9) 及 CD-HIT-EST 软件(v4.8.1) (Li and Godzik, 2006) 删除冗余后的序列 (参数: -c 0.8)，最终得到这四个种质的代表性转录本数据集，即参考转录本。BUSCO 用于转录本组装完整性评估(使用数据库: mbryophyta_odb10); (G+C) 含量、Contig N50、平均长度等指标用于组装质量评估。功能注释由多个公共数据库参与注释，参考转录本通过 BLASTx 分别比对到 Nr (ProteinBLAST: <http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Protein>)、PDB (PDB: <https://www.rcsb.org/>)、KOG (KOG: <http://ftp.ncbi.nih.gov/pub/COG/KOG/kyva>) 和 Swiss-Prot (<http://www.gpmaw.com/html/swiss-prot.html>) 数据库，阈值为 $E\text{-value} < 10^{-5}$ 。

3.3 共有基因注释及特异性表达基因富集分析

将经过质控处理的 RNA-seq 干净读数使用 Bowtie2 软件(v2.2.5) (Langmead and Salzberg, 2012) 与各个物种的参考转录本进行比对。然后，从比对结果中提取唯一比对的读数，并利用 Salmon 软件(v1.6.0) (Patro et al., 2017) 进行定量分析。

3.4 差异表达分析及 KEGG 富集分析

使用 edgeR 软件(Robinson et al., 2010) 对转录本的表达量(以 TPM 表示)进行了 TMM (trimmed mean of M-values) 标准化(Blake et al., 2020)。在标准化后，通过设定条件，即 $|\log_2 \text{Fold Change}| \geq 2$ 和 $\text{FDR} < 0.01$ ，筛选出具有显著差异表达的基因。

随后,对这些显著差异表达基因进行了 KEGG 富集分析,以进一步了解它们在生物学通路中的功能和调控。

3.5 同源分析及正选择分析

使用 TransDecoder 软件 (<http://transdecoder.github.io>) (5.5.0) 预测转录本的 ORF, 并将其翻译成蛋白序列, 然后使用 OrthoFinder 软件 (2.3.8) (Emms and Kelly, 2019) 进行直系同源基因筛选。OrthoFinder 软件基于 BLAST 的全对全双向 (all-vs-all bidirectional) 比对, 使用 MSA 软件 (Bodenhofer et al., 2015) 进行多重序列比对, MAFFT (Katoh and Standley, 2013) 进行多序列联配。对单拷贝基因使用 ParaAT 软件 (Zhang et al., 2012) 生成下一步计算 Ka/Ks 的文件。下一步使用 MUSCLE 软件 (Edgar, 2004) 进行蛋白序列比对, 根据蛋白比对序列回译成密码子对应的核酸序列比对结果, 通过 KaKs-Calculator 软件 (Wang et al., 2010) 计算 Ka/Ks 值。

3.6 数据获取

原始转录组数据已上传至中国生物信息中心 (Genome Sequence Archive) (索引号: PRJCA020103)。

作者贡献

冯梦凡是本研究的执行人; 冯梦凡和李思程完成数据分析和论文初稿的写作; 冯梦凡和李思程完成试验结果分析; 杨细平是项目的构思者及负责人, 指导实验设计、数据分析和论文写作与修改。全体作者都阅读并同意最终的文本。

参考文献

冯梦凡, 汤震, 黄有总, 等, 2022. 甘蔗净光合速率及其影响因素的初步分析. 分子植物育种, (2023-09-07) [2022-07-28]. <https://kns.cnki.net/kcms/detail/46.1068.S.20220727.1715.010.html>. [FENG M F, TANG Z, HUANG Y Z, et al., 2022. Preliminary analysis of net photosynthetic rate and its affecting factors of sugarcane. Molecular Plant Breeding, (2023-09-07) [2022-07-28]. <https://kns.cnki.net/kcms/detail/46.1068.S.20220727.1715.010.html>.]

肖芙蓉, 曹哲群, 徐荣, 等, 2018. 斑茅野生种抗旱性鉴定. 西南农业学报, 31(8): 1611-1616. [XIAO F R, CAO Z Q, XU R, et al., 2018. Identification of drought resistance of several wild species of *Erianthus arundinaceum*. Southwest China Journal of Agricultural Sciences, 31(8): 1611-1616.]

AGATI G, TATTINI M, 2010. Multiple functional roles of flavonoids in photoprotection. New Phytol., 186(4): 786-793.

ALARMELU S, PAZHANY A S, JAYABOSE C, 2018. Genetic improvement and development of genetic stocks in sugarcane through backcross breeding. J. Sugarcane Res., 8(1): 36-42.

BAR-EVEN A, 2018. Daring metabolic designs for enhanced plant carbon fixation. Plant Sci., 273: 71-83.

BAR-EVEN A, NOOR E, LEWIS N E, et al., 2010. Design and analysis of synthetic carbon fixation pathways. Proc. Natl. Acad. Sci. USA, 107(19): 8889-8894.

BETTI M, BAUWE H, BUSCH F A, et al., 2016. Manipulating photorespiration to increase plant productivity: recent advances and perspectives for crop improvement. J. Exp. Bot., 67(10): 2977-2988.

BLAKE L E, ROUX J, HERNANDO-HERRAEZ I, et al., 2020. A comparison of gene expression and DNA methylation patterns across tissues and species. Genome Res., 30(2): 250-262.

BODENHOFER U, BONATESTA E, HOREJŠ-KAINRATH C, et al., 2015. msa: an R package for multiple sequence alignment. Bioinformatics, 31(24): 3997-3999.

BRAUTIGAM C A, SMITH B S, MA Z Q, et al., 2004. Structure of the photolyase-like domain of cryptochrome 1 from *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA, 101(33): 12142-12147.

BRICKER T M, FRANKEL L K, 2011. Auxiliary functions of the PsbO, PsbP and PsbQ proteins of higher plant Photosystem II: a critical analysis. J. Photochem. Photobiol. B, 104(1-2): 165-178.

BRIGIDA A B S, ROJAS C A, GRATIVOL C, et al., 2016. Sugarcane transcriptome analysis in response to infection caused by *Acidovorax avenae* subsp. *avenae*. PLoS ONE, 11(12): e0166473.

CHANDRAN K, NISHA M, GIREESAN P P, 2020. Characterization of progenies from polycrosses of *S. robustum* clones *f. sanguineum*. Sugar Tech, 22(3): 379-388.

CHEN N, ZHANG H, ZANG E, et al., 2022. Adaptation insights from comparative transcriptome analysis of two *Opisthopappus* species in the Taihang mountains. BMC Genomics, 23(1): 466.

CHEN S F, ZHOU Y Q, CHEN Y R, et al., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17): i884-i890.

CROFT B, BHUIYAN S, MAGAREY R, et al., 2015. New sources of resistance to major diseases from wild relatives of sugarcane. Proc Aust Soc Sugar Cane Technol., 37: 218-226.

- CUONG D M, HA T W, PARK C H, et al., 2019. Effects of LED lights on expression of genes involved in phenylpropanoid biosynthesis and accumulation of phenylpropanoids in wheat sprout. *Agronomy*, 9(6): 307.
- CURSI D E, HOFFMANN H P, BARBOSA G V S, et al., 2022. History and current status of sugarcane breeding, germplasm development and molecular genetics in Brazil. *Sugar Tech*, 24(1): 112-133.
- DUCAT D C, SILVER P A, 2012. Improving carbon fixation pathways. *Curr. Opin. Chem. Biol.*, 16(3-4): 337-344.
- DUDAREVA N, KLEMPIEN A, MUHLEMANN J K, et al., 2013. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol.*, 198(1): 16-32.
- DURAI A A, GOVINDARAJ P, PAZHANY A S, 2014. Flowering behaviour of sugarcane genotypes from different agro climatic zones of India. *Sugar Tech*, 16(2): 157-163.
- EDGAR R C, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32(5): 1792-1797.
- EMMS D M, KELLY S, 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.*, 20(1): 238.
- EVANGELISTELLA C, VALENTINI A, LUDOVISI R, et al., 2017. De novo assembly, functional annotation, and analysis of the giant reed (*Arundo donax* L.) leaf transcriptome provide tools for the development of a biofuel feedstock. *Bio-technol. Biofuels*, 10(1): 138.
- FATMA M, KHAN M I R, MASOOD A, et al., 2013. Coordinate changes in assimilatory sulfate reduction are correlated to salt tolerance: involvement of phytohormones. *Annu. Res. Rev. Biol.*, 3: 267-295.
- FOULADVAND M, EBRAHIMI A, RAHAEI M, et al., 2022. Identification of pathways and genes with differential expression in sugarcane after cold stress, through RNA-seq analysis. *Env. Stresses Crop Sci.*, 15(2): 541-550.
- FU Y H, WEI J J, PAN Y B, et al., 2019. Comparative analysis reveals changes in transcriptomes of sugarcane upon infection by *Leifsonia xyli* subsp. *xyli*. *J. Phytopathol.*, 167(11-12): 633-644.
- HAAS B J, PAPANICOLAOU A, YASSOUR M, et al., 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.*, 8(8): 1494-1512.
- HOSSAIN M A, LEE Y, CHO J I, et al., 2010. The bZIP transcription factor OsABF₁ is an ABA responsive element binding factor that enhances abiotic stress signaling in rice. *Plant Mol. Biol.*, 72(4-5): 557-566.
- IFUFU K, YAMAMOTO Y, ONO T A, et al., 2005. PsbP protein, but not PsbQ protein, is essential for the regulation and stabilization of photosystem II in higher plants. *Plant Physiol.*, 139(3): 1175-1184.
- IMAZUMI T, TRAN H G, SWARTZ T E, et al., 2003. FKF₁ is essential for photoperiodic-specific light signalling in *Arabidopsis*. *Nature*, 426(6964): 302-306.
- JONES J D G, DANGL J L, 2006. The plant immune system. *Nature*, 444(7117): 323-329.
- KATOH K, STANDLEY D M, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, 30(4): 772-780.
- KIM D, LANGMEAD B, SALZBERG S L, 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, 12(4): 357-360.
- KUMAR S, ZAVALIEV R, WU Q L, et al., 2022. Structural basis of NPR1 in activating plant immunity. *Nature*, 605(7910): 561-566.
- LANGMEAD B, SALZBERG S L, 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, 9(4): 357-359.
- LI J J, YE C L, CHANG C F, 2020. Comparative transcriptomics analysis revealing flower trichome development during flower development in two *Lonicera japonica* Thunb. cultivars using RNA-seq. *BMC Plant Biol.*, 20(1): 341.
- LI W Z, GODZIK A, 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13): 1658-1659.
- LU X D, ZHOU C M, XU P B, et al., 2015. Red-light-dependent interaction of phyB with SPA1 promotes COP1-SPA1 dissociation and photomorphogenic development in *Arabidopsis*. *Mol. Plant*, 8(3): 467-478.
- MANECHINI J R V, DA SILVA SANTOS P H, ROMANEL E, et al., 2021. Transcriptomic analysis of changes in gene expression during flowering induction in sugarcane under controlled photoperiodic conditions. *Front. Plant Sci.*, 12: 635784.
- MATSUBARA K, OGISO-TANAKA E, HORI K, et al., 2012. Natural variation in *Hd17*, a homolog of *Arabidopsis* *ELF3* that is involved in rice photoperiodic flowering. *Plant Cell Physiol.*, 53(4): 709-716.
- MURAKAMI M, TAGO Y, YAMASHINO T, et al., 2007. Comparative overviews of clock-associated genes of *Arabidopsis thaliana* and *Oryza sativa*. *Plant Cell Physiol.*, 48(1): 110-121.
- NAKASHIMA K, YAMAGUCHI-SHINOZAKI K, 2013. ABA signaling in stress-response and seed development. *Plant Cell Rep.*, 32(7): 959-970.
- PAIDIPATI K K, BANIK A, SHAH B, et al., 2022. Forecasting of sugarcane productivity estimation in India: a comparative study with advanced non-parametric regression models. *J.*

- Algebr. Stat., 13(2): 760-778.
- PATRO R, DUGGAL G, LOVE M I, et al., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods, 14(4): 417-419.
- PEREIRA-SANTANA A, ALVARADO-ROBLEDO E J, ZAMORA-BRISEÑO J A, et al., 2017. Transcriptional profiling of sugarcane leaves and roots under progressive osmotic stress reveals a regulated coordination of gene expression in a spatiotemporal manner. PLoS ONE, 12(12): e0189271.
- POURCEL L, ROUTABOUL J M, CHEYNIER V, et al., 2007. Flavonoid oxidation in plants: from biochemical properties to physiological functions. Trends Plant Sci., 12(1): 29-36.
- RATHINAM M, MISHRA P, VASUDEVAN M, et al., 2019. Comparative transcriptome analysis of pigeonpea, *Cajanus cajan* (L.) and one of its wild relatives *Cajanus platycarpus* (Benth.) Maesen. PLoS ONE, 14(7): e0218731.
- ROBINSON M D, MCCARTHY D J, SMYTH G K, 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1): 139-140.
- SAKAIGAICHI T, TERAJIMA Y, SUGIMOTO A, et al., 2007. Comparison of root distribution and root growth direction in two sugarcane hybrids with contrasting tolerance to water stress. Proc. Int. Soc. Sugar Cane Technol., 26: 754-758.
- SANGHERA G S, MALHOTRA P, SINGH H, et al., 2019. Climate change impact in sugarcane agriculture and mitigation strategies//Malik C P, Trivedi P C. Harnessing plant biotechnology and physiology to stimulate agricultural growth. New Delhi, India: OM Publication: 99-115.
- SCORTECCI K C, CRESTE S, CALSA T, et al., 2012. Challenges, opportunities and recent advances in sugarcane breeding. Plant Breeding, 1: 267-296.
- TAKARAGAWA H, OKAMOTO K, TERAJIMA Y, et al., 2022. Evaluation of root distribution and nitrate leaching in sugarcane, *Erianthus*, and their intergeneric hybrid at new planting. Plant Prod. Sci., 25(3): 298-310.
- VAN LOON L C, REP M, PIETERSE C M J, 2006. Significance of inducible defense-related proteins in infected plants. Annu. Rev. Phytopathol., 44: 135-162.
- WANG D P, ZHANG Y B, ZHANG Z, et al., 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genom. Proteom. Bioinform., 8(1): 77-80.
- WANG Y N, LIU C, LI K X, et al., 2007. *Arabidopsis* EIN2 modulates stress response through abscisic acid response pathway. Plant Mol. Biol., 64(6): 633-644.
- WANG Y R, ZHANG J S, WANG R, et al., 2021. Unveiling sugarcane defense response to *Mythimna separata* herbivory by a combination of transcriptome and metabolic analyses. Pest Manag. Sci., 77(10): 4799-4809.
- WU T D, WATANABE C K, 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics, 21(9): 1859-1875.
- YAN H F, ZHOU H W, LUO H M, et al., 2021. Characterization of full-length transcriptome in *Saccharum officinarum* and molecular insights into tiller development. BMC Plant Biol., 21(1): 228.
- YI X P, HARGETT S R, FRANKEL L K, et al., 2006. The PsbQ protein is required in *Arabidopsis* for photosystem II assembly/stability and photoautotrophy under low light conditions. J. Biol. Chem., 281(36): 26260-26267.
- YI X P, HARGETT S R, LIU H J, et al., 2007. The PsbP protein is required for photosystem II complex assembly/stability and photoautotrophy in *Arabidopsis thaliana*. J. Biol. Chem., 282(34): 24833-24841.
- YI X P, MCCHARGUE M, LABORDE S, et al., 2005. The manganese-stabilizing protein is required for photosystem II assembly/stability and photoautotrophy in higher plants. J. Biol. Chem., 280(16): 16170-16174.
- ZHANG Z, XIAO J F, WU J Y, et al., 2012. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. Biochem. Biophys. Res. Commun., 419(4): 779-781.

(责任编辑 张积森)
(责任执行主编 陈玲玲)