植物遗传资源学报 2025，26（6）：1080-1090
Journal of Plant Genetic Resources DOI：10.13430/j.cnki.jpgr.20241029003

两个甘薯野生近缘种的比较基因组分析

肖世卓[1]，许　攀[2]，王　珧[1]，戴习彬[1]，周志林[1]，曹清河[1]
（[1]江苏徐淮地区徐州农业科学研究所/农业农村部甘薯生物学与遗传育种重点实验室，徐州 221131；[2]兰州大学草地农业科技学院/草种创新与草地农业生态系统全国重点实验室/农业农村部草牧业创新重点实验室，兰州 730000）

摘要：甘薯是重要的农作物，其野生近缘种数量庞大，蕴含着丰富的基因变异。本研究利用公开发表的、基因组组装质量较高的两个甘薯野生种进行了比较基因组分析。发掘了覆盖全基因组的1798184个SNP，同时还发掘了2076个易位和84个倒位，倒位中包含2106个基因，富集分析显示这些基因主要富集于次级代谢产物的生物合成等通路。其中34段倒位的断点位于基因内部，可能影响到基因的结构。两个基因组之间存在10226个插入型和11411个缺失型的结构变异，对这些结构变异进行了注释，并对可能影响基因功能或表达的结构变异进行了基因富集分析，结果发现这些可能受影响的基因主要富集于次级代谢产物通路，DNA修复和复制等相关功能的基因也出现了富集。通过统计两个野生种和其他近缘种共线性基因对的同义替换率（Ks）和四倍简并位点颠换率（4DTv），说明番薯属植物的共同祖先在发生了全基因组三倍化事件之后才分化成不同的种。以上研究将为甘薯野生近缘种优异变异的发掘和物种分化研究提供有力支撑。

关键词：甘薯；野生近缘种；比较基因组学

Comparative Genomic Analysis of Two Sweetpotato Wild Relatives

XIAO Shizhuo[1]，XU Pan[2]，WANG Yao[1]，DAI Xibin[1]，ZHOU Zhilin[1]，CAO Qinghe[1]
（[1]*Xuzhou Institute of Agricultural Sciences in Jiangsu Xuhuai District/Key Laboratory of Biology and Genetic Breeding of Sweetpotato，Ministry of Agriculture and Rural Affairs，Xuzhou 221131*；[2]*College of Pastoral Agriculture Science and Technology，Lanzhou University/State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems/Key Laboratory of Grassland Livestock Industry Innovation，Ministry of Agriculture and Rural Affairs，Lanzhou 730000*）

Abstract：Sweetpotato is a globally important crop with numerous wild relatives harboring extensive genetic diversity. In this study，a comparative genomic analysis was conducted using two publicly available high-quality genomes of wild relatives. Our analysis identified 1798184 genome-wide single nucleotide polymorphisms（SNPs），along with 2076 translocations and 84 inversions encompassing 2106 genes. Enrichment analysis showed that these genes are mainly enriched in pathways such as the biosynthesis of secondary metabolites. Among them，the breakpoints of 34 inversions are located inside the genes and may affect the structure of these genes. There are 10226 insertion-type and 11411 deletion-type structural variations between the two genomes. We annotated these structural variations and conducted gene enrichment analysis on those structural variations that may affect gene function or expression. We found that these potentially affected genes were mainly enriched in pathways of secondary metabolites，while genes related to functions such as DNA repair and replication were enriched. Through deploying statistical results of the synonymous substitution rate（Ks）and fourfold degenerate synonymous site（4DTv）of collinear gene pairs of two wild species and other related species，it was indicated that the common ancestor of the genus *Ipomoea* differentiated into different species after a whole-genome

收稿日期：2024-10-29　　**网络出版日期：**2025-01-17
URL：https://doi.org/10.13430/j.cnki.jpgr.20241029003
第一作者研究方向为甘薯种质资源保存、鉴定以及优异基因的挖掘，E-mail：xiaoshizhuo@jaas.ac.cn
通信作者：曹清河，研究方向为甘薯种质资源，E-mail：caoqinghe@jaas.ac.cn
基金项目：江苏省自然科学基金（BK20221213）；国家甘薯产业技术体系（CARS-10-GW01）
Foundation projects：Natural Science Foundation of Jiangsu Province（BK20221213）；China Agriculture Research System（CARS-10-GW01）

triplication event. Collectively, the findings will provide insights for the excavation of excellent variations in wild relatives of sweetpotato as well as for analyzing species differentiation

Key words：sweetpotato；wild relatives；comparative genomics

作物的野生近缘种蕴含着丰富且有价值的基因资源，不仅可以弥补栽培种长期驯化过程中丢失的重要性状，还可以为栽培作物的起源和驯化提供重要的遗传信息[1]。甘薯［*Ipomoea batatas*（L.）Lam.］是重要的粮食、饲料以及能源作物，其高产、广适的特性为中国乃至世界的粮食安全做出了重要贡献。甘薯在植物分类学上属于旋花科（Convolvulaceae）番薯属（*Ipomoea*）甘薯组（*batatas*），番薯属包含了约800种植物，是旋花科中最大的属，同时也是植物界较大的属之一，为植物分类学的研究提供了丰富的资源[2]。其中甘薯所在的甘薯组包含了约15个物种，这些种被认为是与甘薯亲缘关系最近的野生种，大部分为二倍体，也包含了四倍体和六倍体[2]。这些物种可能参与到了甘薯的起源或进化，其中三浅裂野牵牛（*I.trifida*）和毛果甘薯（*I. cordatotriloba*）就属于甘薯组中两个重要的野生近缘种。

三浅裂野牵牛被认为是甘薯的二倍体祖先种之一，是甘薯驯化研究中最受关注的物种，包含二倍体和六倍体。该种与甘薯杂交亲和，可以作为亲本与甘薯杂交，拓宽甘薯栽培种遗传背景。早在1975年，日本学者就利用六倍体的三浅裂野牵牛与栽培种进行杂交，最终获得了高淀粉、高产、抗茎线虫病的品种南丰[3]。因其在甘薯起源中扮演了重要角色，二倍体的三浅裂野牵牛是最早开展全基因组测序和组装的甘薯野生种。2015年，日本学者分别组装了三浅裂野牵牛自交系Mx23Hm和高度杂合系0431-1的基因组草图[4]。2018年，由美国和中国等国家合作，共同绘制了首个染色体级别的三浅裂野牵牛基因组[5]。2019年，Li等[6]绘制了一个具有膨大根的三浅裂野牵牛基因组。但是上述三浅裂野牵牛基因组的组装都是以二代测序技术为主，质量仍有待提高。2022年，中日韩三国学者，利用PacBio测序技术，重新对自交系Mx23Hm进行了测序、组装和注释[7]，获得了总长为502.2 Mb的高质量基因组，包含34386个基因。毛果甘薯基因组的纯合度较高，自交亲和，结实率高。2024年，由本团队主导，绘制了毛果甘薯的端粒到端粒（T2T，telomere-to-telomere）基因组，是旋花科番薯属中第一个真正意义上的T2T基因组。该基因组全长454.8 Mb，由无间隙的15条染色体组成，包含40238个蛋白编码基因[8]。

植物比较基因组学是指通过比较不同植物物种的完整基因组序列，研究它们之间的基因组结构、功能和演化关系的科学。比较基因组不仅可以用来揭示植物之间的遗传相似性，还为理解植物的进化历史、适应机制以及功能基因的预测提供了重要工具。到2024年为止，已经公布了超过1000个植物基因组，涵盖大约800个物种[9]。借助测序技术和算法，比较基因组能够在全基因组范围内识别SNP以及复杂的结构变异（通常大于50 bp），包括拷贝数变异、存在/缺失变异、倒位和易位等，用来进行物种进化的分析和解析复杂性状的遗传基础，从而服务于育种实践。He等[10]利用全球377份栽培稻和野生稻的三代测序数据，构建了迄今为止最大规模的水稻群体水平倒位变异图谱，发现与逆境响应相关的基因在倒位内部或附近发生了明显富集，揭示了倒位变异对水稻抗逆性状的重要影响。Liu等[11]通过对26份大豆种质进行高质量的基因组组装，并与已公开的大豆基因组进行比较基因组分析，揭示了结构变异在调控关键农艺性状方面扮演的重要角色：例如在野生大豆与栽培大豆之间，查尔酮合成酶基因簇的结构变异是导致种皮颜色从黑色向黄色转变的主要驯化因素；*SoyZH13_14G179600*基因的结构变异也造成了该基因在不同种质资源中表达模式的差异。

甘薯基因组较为复杂，高质量组装较为困难，比较基因组研究较少，主要在野生二倍体种中开展。通过比较基因组分析，Wu等[5]发现贮藏蛋白基因在三浅裂野牵牛、三裂叶薯（*I. triloba*）以及马铃薯（*Solanum tuberosum* L.）中发生了明显扩张，暗示贮藏蛋白在根茎类作物进化中较为重要。但是上述比较基因组研究以二代测序为基础，无法对结构变异等大片段变异进行发掘，对基因组间的变异认识受限。本研究通过比较甘薯组中组装质量较高的两个二倍体野生种基因组，来发掘甘薯组基因组上的变异，尤其是大的结构变异，并对变异进行系统分析，为甘薯组植物的进化和分化，以及野生近缘种优异变异的育种利用奠定基础。

1 材料与方法

1.1 试验数据的获取

本研究中用到的基因组数据均来自公开数据库,具体下载地址如下所示:三浅裂野牵牛(https://plantgarden.jp/en/list/t35884/genome/t35884.G002)、毛果甘薯(https://ngdc.cncb.ac.cn/gwh/Assembly/85948/show)、马鞍藤(*I. pes-caprae*)(https://ngdc.cncb.ac.cn/gwh/Assembly/82945/show)以及马铃薯(http://www.bioinformaticslab.cn/pubs/dm8/)。

1.2 基因组结构变异的发掘

本研究使用两种方法来发掘基因组的结构变异。首先是基因组之间的比对,即使用Smartie-sv[12]将毛果甘薯基因组比对至三浅裂野牵牛基因组,获得长度不小于50 bp的结构变异。同时,使用NGMLR 0.2.7[13]将毛果甘薯的HiFi reads比对至三浅裂野牵牛基因组,然后使用Sniffles[14]来获得结构变异。将两种方法产生的结构变异进行整合,重复检测到的结构变异,即插入或者缺失类型相同、检出位置有重叠区域的结构变异被认为是高置信度的结构变异。使用ANNOVAR[15]对这些结构变异进行注释。使用SyRI[16]来鉴定倒位和重复序列,最后使用SyRI将比对结果可视化。

1.3 Ks和4DTv的分析

将三浅裂野牵牛、毛果甘薯、马鞍藤以及马铃薯基因组的氨基酸序列进行比对,使用MCScanX 101[17]鉴定共线性基因,参数设为默认。使用PAML[18]来评估共线性基因间的同义替换率(Ks)。使用KaKs_Calculator[19]分析保守同源基因中的4D位点及四倍简并位点颠换率(4DTv),筛选得到位点数大于20个的结果,分析各物种比较的4DTv值,预测物种之间的分化事件,分析各物种的Ks值,预测各自全基因组加倍事件。

1.4 KEGG和GO分析

采用InterProScan v5.52-86.0[20]对基因进行本体论(GO)注释。运用EggNOG-mapper v2.1.4[21]进行基因表达通路(KEGG)和蛋白结构域(Pfam)注释。借助topGO[22]软件包和REVIGO[23]进行GO富集分析,使用CirGo[24]和ggplot2[25]软件包将结果可视化。通过clusterProfiler[26]软件包进行KEGG富集分析。

2 结果与分析

2.1 基因组比较和SNP发掘

三浅裂野牵牛和毛果甘薯的基因组长度分别为502237654 bp和454814757 bp,contig N50长度分别为3.73 Mb和30.93 Mb,BUSCO评估基因组的完整度分别为98.5%和97.6%,预测的基因数量分别为34386个和40238个,三浅裂野牵牛的测序方法主要为PacBio和Hi-C技术,毛果甘薯除了上述两种测序技术之外,还使用了Nanopore技术(表1)。

表1 三浅裂野牵牛和毛果甘薯的基因组统计

Table 1 Statistics of *I. trifida* and *I. cordatotriloba* genome parameters

参数 Parameter	三浅裂野牵牛 *I. trifida*	毛果甘薯 *I. cordatotriloba*
种质 Germplasm	Mx23Hm	*xiaoshu*
基因组长度(bp) Length of genome	502237654	454814757
Contig N50(Mb)	3.73	30.93
测序方法 Sequencing method	PacBio,Hi-C	PacBio,Nanopore,Hi-C
BUSCO(%)	98.5	97.6
基因数量 Gene number	34386	40238

通过比较三浅裂野牵牛和毛果甘薯的基因组发现,三浅裂野牵牛和毛果甘薯两个基因组之间存在1798184个SNP位点。相比于三浅裂野牵牛,毛果甘薯的重复序列多出112个拷贝,总长220663 bp,减少152个拷贝,总长559140 bp。两个基因组存在50802个高度分化的区域,这段区域在三浅裂野牵牛中的长度是228414336 bp,在毛果甘薯中为254217620 bp(表2)。

表2 三浅裂野牵牛和毛果甘薯基因组的序列变异

Table 2 The sequence variations between *I. trifida* and *I. cordatotriloba* genomes

变异类型 Variation type	数目 Count	长度(bp) Length
SNPs	1798184	1798184/1798184
拷贝数增加 Copy-gains	112	−/220633
拷贝数减少 Copy-losses	152	559140/−
高度分化 Highly diverged	50802	228414336/254217620
串联重复 Tandem repeats	14	15681/25346

/之前为三浅裂野牵牛基因组中的长度,/之后为毛果甘薯基因组中的长度;−:无数据

The number before the / symbol is the length in the *I. trifida* genome, and the number after the / symbol is the length in the *I. cordatotriloba* genome;−:No data

2.2　基因组共线性分析

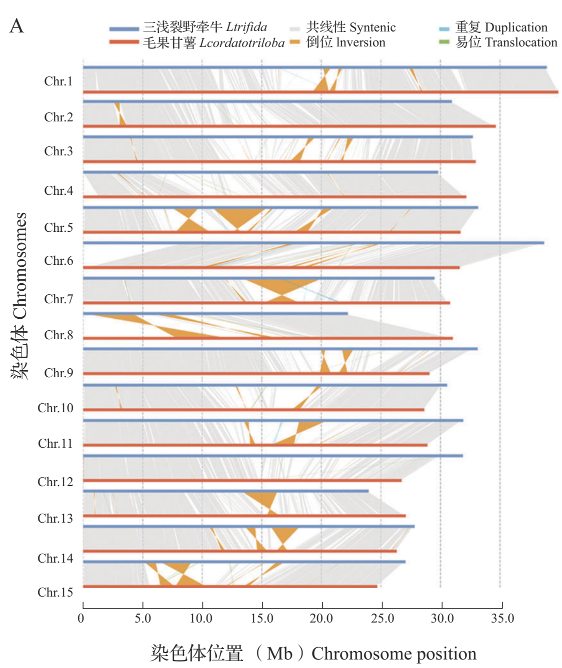
染色体倒位和易位会改变基因顺序，影响基因表达，同时也会增加物种基因资源的遗传多样性。为了探究三浅裂野牵牛和毛果甘薯之间可能存在的遗传多样性，对两个基因组进行了共线性分析。结果显示两个基因组之间存在1582个共线性区域，这段区域在三浅裂野牵牛基因组中的长度为271201177 bp，占基因组全长的54.0%；在毛果甘薯基因组中的长度则为304956210 bp，占基因组全长的67.0%。两个基因组之间存在84个倒位，倒位在三浅裂野牵牛基因组中的长度为41239691 bp，占比为8.2%；在毛果甘薯基因组中的长度为32262135 bp，占比为7.1%。两个基因组之间还存在2076个易位，易位的长度在三浅裂野牵牛基因组中为13473827 bp，占比为2.7%，在毛果甘薯基因组中的长度则为13973042 bp，占比为3.1%（表3，图1A）。

倒位可能会对基因的表达产生影响。通过对倒位中基因的研究发现，84段倒位中共包含2106个基因，对这2106个基因分别进行KEGG和GO富集分析。KEGG分析显示基因富集较多的通路是二苯乙烯类、二芳基庚酸类和姜酚（Stilbenoid, diarylheptanoid and gingerol biosynthesis, map00945）以及类黄酮（Flavonoid biosynthesis, map00941）等次级代谢产物的生物合成途径（图1B）。除此之外，错配修复（Mismatch repair, map03430）、碱基切除修复（Base excision repair, map03410）和DNA复制（DNA replication, map03030）等通路的基因也出现富集（图1B）。GO富集分析同样显示，端粒维持（Telomere maintenance, GO：0000723）、DNA解螺旋活性（DNA helicase activity, GO：0003678）、双链DNA绑定（Double-stranded DNA binding, GO：0003690）、DNA介导的DNA聚合酶活性（DNA-directed DNA polymerase activity, GO：0003887）以及DNA修复（DNA repair, GO：0006281）相关的基因出现显著富集（图1C）。这些基因跟倒位的形成和固定可能存在某种联系。

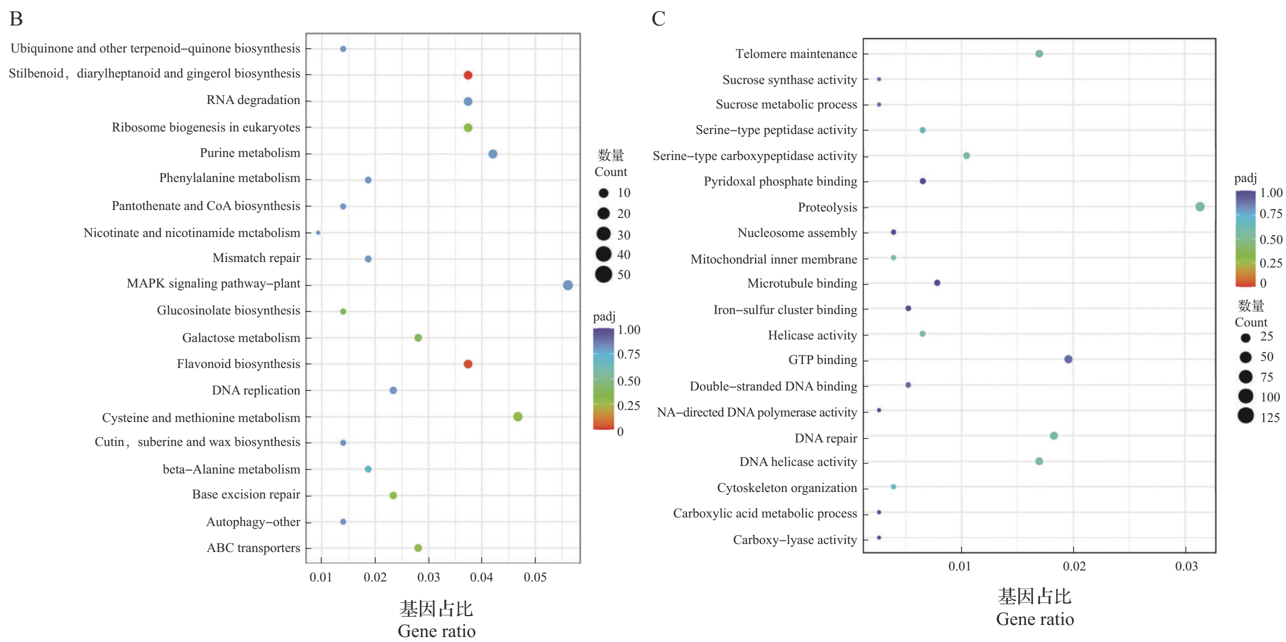
表3　三浅裂野牵牛和毛果甘薯基因组的共线性分析

Table 3　Syntenic analysis between *I. trifida* and *I. cordatotriloba* genomes

变异类型 Variation type	三浅裂野牵牛 *I. trifida*			毛果甘薯 *I. cordatotriloba*		
	数量 Count	长度（bp） Length	占比（%） Proportion	数量 Count	长度（bp） Length	占比（%） Proportion
共线性区域 Syntenic regions	1582	271201177	54.0	1582	304956210	67.0
倒位 Inversions	84	41239691	8.2	84	32262135	7.1
易位 Translocations	2076	13473827	2.7	2076	13973042	3.1
重复 Duplications	918	4724592	0.9	2106	7285309	1.6
未匹配序列 Not aligned sequences	4131	135011642	26.9	5913	98197220	21.6



（图1）



A:三浅裂野牵牛和毛果甘薯基因组之间的共线性;B:倒位中基因的KEGG富集分析;C:倒位中基因的GO富集分析

A: The synteny between *I. trifida* and *I.cordatotriloba* genomes; B: KEGG enrichment analysis of genes in inversion;

C: GO enrichment analysis of genes in inversion

图1 三浅裂野牵牛和毛果甘薯基因组之间的共线性以及倒位中基因的富集分析

Fig. 1 The synteny between *I. trifida* and *I.cordatotriloba* genomes and gene enrichment analysis in inversions

倒位会产生染色体断点,断点如果发生在基因内部会引起基因结构的变化从而影响其功能。经过统计发现,共有34段倒位的断点发生于三浅裂野牵牛或毛果甘薯中基因内部,其中1段倒位的两端断点均位于基因内部,该段倒位位于三浅裂野牵牛14号染色体24546584~24707043 bp,倒位两端基因分别为 *Itr_chr14CG16870* 和 *Itr_chr14CG16990*(表4)。34段倒位共影响到35个基因,这些基因行使的功能包含转录因子、功能蛋白和酶等,参与到植物生长发育的不同阶段,无明显规律,呈现出随机特点(表4)。

表4 位于倒位断点处的基因统计

Table 4 Listed genes residing at the inversion breakpoints

编号 No.	基因组 Genomes	染色体 Chromosomes	倒位起止位置(bp) The starting and ending position of the inversions	断点所在基因 The genes covering the breakpoints	基因起始位置(bp) The starting and ending position of the genes	基因功能注释 Functional annotation of genes
1	三浅裂野牵牛	Itr_chr01	11703012~11703394	*Itr_chr01CG07580*	11702887~11704047	–
2	三浅裂野牵牛	Itr_chr03	19922914~20345663	*Itr_chr03CG11900*	20340885~20345769	Poly(A) RNA polymerase GLD2
3	三浅裂野牵牛	Itr_chr03	21722914~22704229	*Itr_chr03CG12960*	21714866~21724114	–
4	三浅裂野牵牛	Itr_chr04	2752356~2873657	*Itr_chr04CG04770*	2870724~2887446	–
5	三浅裂野牵牛	Itr_chr04	5793962~5795348	*Itr_chr04CG08650*	5784645~5795829	–
6	三浅裂野牵牛	Itr_chr06	3927167~4015859	*Itr_chr06CG06040*	4013606~4016405	Proton antiporter activity
7	三浅裂野牵牛	Itr_chr06	21569168~21569727	*Itr_chr06CG16690*	21567864~21570685	Light-harvesting complex II chlorophyll a/b binding protein 1
8	三浅裂野牵牛	Itr_chr06	22213921~22728379	*Itr_chr06CG17140*	22712790~22738139	–
9	三浅裂野牵牛	Itr_chr07	26827512~26828003	*Itr_chr07CG19600*	26821927~26828700	Phosphorelay sensor kinase activity

表 4（续）

编号 No.	基因组 Genomes	染色体 Chromosomes	倒位起止位置(bp) The starting and ending position of the inversions	断点所在基因 The genes covering the breakpoints	基因起始位置(bp) The starting and ending position of the genes	基因功能注释 Functional annotation of genes
10	三浅裂野牵牛	Itr_chr08	27788218~27801091	*Itr_chr08CG18580*	27785342~27790952	Lysosomal Pro-X carboxypeptidase
11	三浅裂野牵牛	Itr_chr08	27816710~27817090	*Itr_chr08CG18590*	27814863~27820819	Lysosomal Pro-X carboxypeptidase
12	三浅裂野牵牛	Itr_chr09	911603~963422	*Itr_chr09CG01660*	961189~969086	Sterol 24-C-methyltransferase
13	三浅裂野牵牛	Itr_chr09	13287680~16314621	*Itr_chr09CG15430*	16314335~16315624	–
14	三浅裂野牵牛	Itr_chr10	15750888~18193749	*Itr_chr10CG14040*	15742310~15753395	Serine-type endopeptidase activity
15	三浅裂野牵牛	Itr_chr13	11841296~11924724	*Itr_chr13CG12110*	11922947~11937966	AP-2 complex subunit mu-1
16	三浅裂野牵牛	Itr_chr14	15978288~16031423	*Itr_chr14CG11650*	16021901~16031479	–
17	三浅裂野牵牛	Itr_chr14	23223138~23349472	*Itr_chr14CG15830*	23221295~23223298	Zinc ion binding
18	三浅裂野牵牛	Itr_chr14	24546584~24707043	*Itr_chr14CG16870*	24543680~24547627	Pectinesterase activity
				Itr_chr14CG16990	24706321~24709541	–
19	三浅裂野牵牛	Itr_chr14	24861187~25244280	*Itr_chr14CG17470*	25242071~25245346	–
20	三浅裂野牵牛	Itr_chr15	106470~4084032	*Itr_chr15CG05070*	4080322~4084326	Recognition of pollen
21	三浅裂野牵牛	Itr_chr15	5682723~7141479	*Itr_chr15CG06760*	5682643~5684909	Pectinesterase
22	毛果甘薯	Chr11	24496445~24496831	*Icor_Chr11.01919*	24494734~24497336	Lysosomal Pro-X carboxypeptidase-like isoform X2
23	毛果甘薯	Chr12	12598615~12598997	*Icor_Chr12.01502*	12594644~12598979	B3 domain-containing protein At1g05920-like
24	毛果甘薯	Chr13	969575~1024868	*Icor_Chr13.00176*	968757~971843	Cycloartenol-C-24-methyltransferase 1-like isoform X1
25	毛果甘薯	Chr13	15430378~16541871	*Icor_Chr13.01604*	15429529~15439711	Katanin p80 WD40 repeat-containing subunit B1 homolog isoform X2
26	毛果甘薯	Chr14	25785754~25821651	*Icor_Chr14.02284*	25777482~25787636	Vetispiradiene synthase 3-like
27	毛果甘薯	Chr15	11927803~12008190	*Icor_Chr15.00998*	11927513~11928117	–
28	毛果甘薯	Chr2	3042641~3683581	*Icor_Chr02.00501*	3682866~3684458	Cysteine-rich repeat secretory protein 60
29	毛果甘薯	Chr3	4464551~4571946	*Icor_Chr03.00679*	4463814~4466804	Cation/H(+) antiporter 15-like
30	毛果甘薯	Chr3	14014929~14036263	*Icor_Chr03.01538*	14008217~14016412	G-type lectin S-receptor-like serine/threonine-protein kinase
31	毛果甘薯	Chr5	16136141~16136714	*Icor_Chr05.01181*	16129497~16153598	Alcohol dehydrogenase-like 2 isoform X2

表4(续)

编号 No.	基因组 Genomes	染色体 Chromosomes	倒位起止位置(bp) The starting and ending position of the inversions	断点所在基因 The genes covering the breakpoints	基因起始位置(bp) The starting and ending position of the genes	基因功能注释 Functional annotation of genes
32	毛果甘薯	Chr5	17695783~18808792	*Icor_Chr05.01360*	18805788~18810010	WAT1-related protein At1g43650-like
33	毛果甘薯	Chr6	2894889~2930955	*Icor_Chr06.00381*	2930162~2947252	Heavy metal-associated isoprenylated plant protein 39-like isoform X2
34	毛果甘薯	Chr7	13546317~13782991	*Icor_Chr07.01385*	13780765~13783068	Uncharacterized protein LOC109158392

2.3 插入缺失型结构变异的鉴定和分析

基因组上插入和缺失型的结构变异是引起表型变异的重要因素。本研究分别使用毛果甘薯三代测序得到的长序列以及组装完成的T2T基因组与三浅裂野牵牛基因组进行比较,得到全基因组的结构变异。通过基因组之间比较发现,相比于三浅裂野牵牛,毛果甘薯基因组分别存在45154个插入型和34638个缺失型的结构变异。通过HiFi reads与基因组比较,发现相比于三浅裂野牵牛,毛果甘薯基因组分别存在60284个插入型和83558个缺失型的结构变异。由于结构变异检出存在较高的假阳性,为了降低假阳性率,将两种方法共同检测到的结构变异进行整合,保留重复检测到的结构变异,最终获得了10226个插入和11411个缺失(表5)。虽然这一过程可能会丢失一些真实存在的结构变异,但是保留的结构变异被认为是具有高可信度的。通过序列注释发现,保留的结构变异中插入型结构变异有1494个位于基因上游, 1431个位于基因下游,4162个位于基因间,2581个位于内含子,173个位于外显子,380个位于基因上游或者下游,5个位于剪接位点。缺失型的结构变异中,有1648个位于基因上游,1450个位于基因下游,5663个位于基因间,1998个位于内含子,348个位于外显子,295个位于基因上游或者下游,9个位于剪接位点(表6)。

表5 毛果甘薯基因组的插入和缺失型结构变异统计

Table 5 Statistics of structural variations at *I. cordatotriloba* genomes

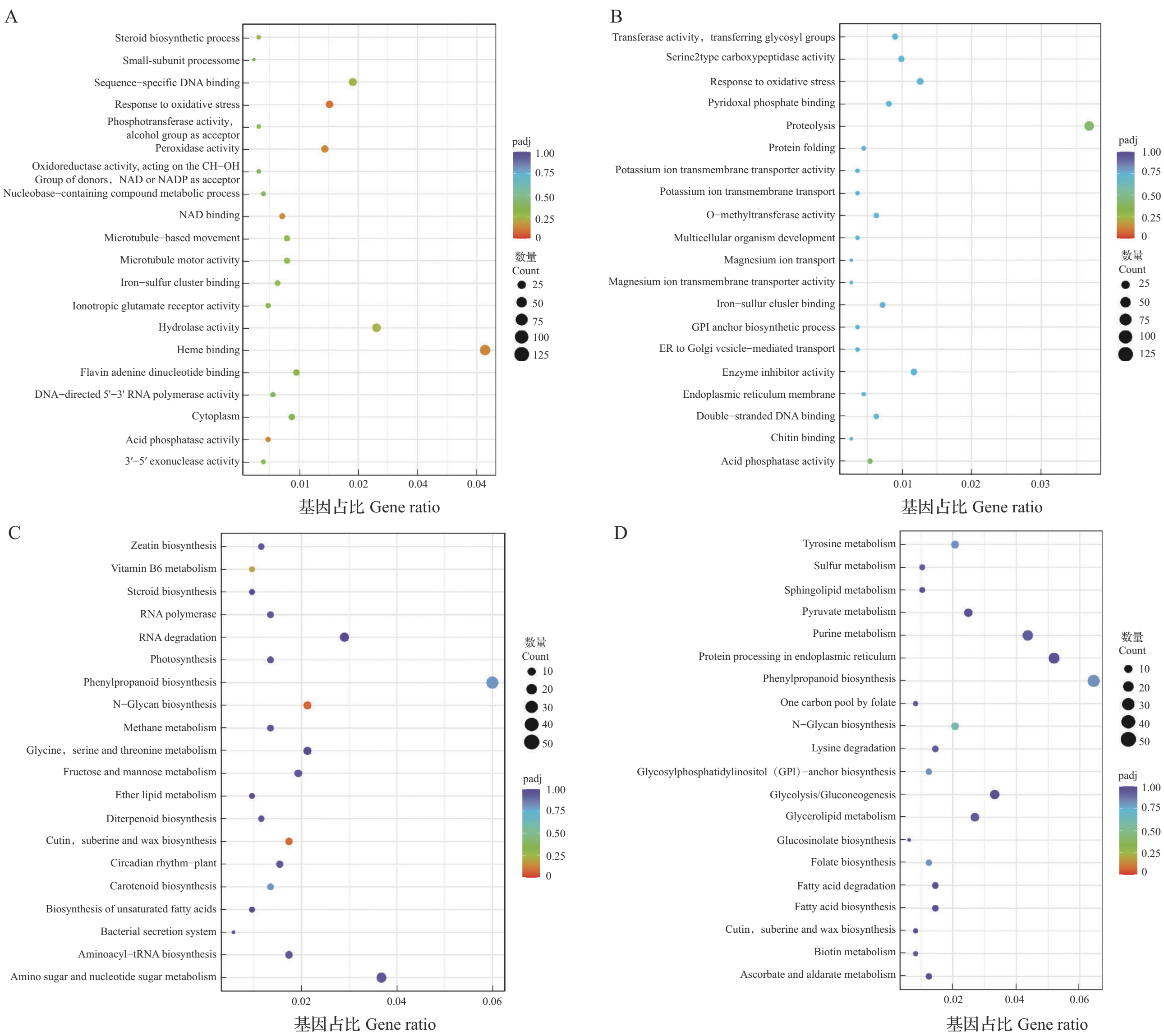
比对方法 Align methods	插入 Insertions	缺失 Deletions
基因组比对基因组 Genome to genome	45154	34638
HiFi reads比对基因组 HiFi reads to genome	60284	83558
整合 Merge	10226	11411

表6 毛果甘薯结构变异的位置注释

Table 6 Annotation of positions of structural variations at *I.cordatotriloba* genomes

注释 Annotation	插入 Insertions	缺失 Deletions
上游 Upstream	1494	1648
下游 Downstream	1431	1450
基因间 Intergenic	4162	5663
内含子 Intronic	2581	1998
外显子 Exonic	173	348
上游或者下游 Upstream or downstream	380	295
剪接位点 Splicing	5	9

位于基因上游和外显子的结构变异可能会引起基因表达或者功能的变化,进而造成表型变异,本研究对可能受结构变异影响的基因(下文简称结构变异基因)进行了富集分析。GO富集分析显示,插入型结构变异基因主要富集于氧化胁迫应答(Response to oxidative stress,GO:0006979)、过氧化物酶活性(Peroxidase activity,GO:0004601)、血红素结合(Heme binding,GO:0020037)等胁迫响应通路(图2A);缺失型结构变异基因主要富集于酸性磷酸酶活性(Acid phosphatase activity,GO:0003993)、蛋白质水解作用(Proteolysis,GO:0006508)和双链DNA 结合(Double-stranded DNA binding, GO:0003690)等通路(图2B)。KEGG富集分析结果显示,插入型结构变异基因主要富集于角质,亚硫酸和蜡的生物合成(Cutin, suberine and wax biosynthesis, map00073)、N糖的生物合成(N-Glycan biosynthesis, map00510)和维生素B6的代谢(Vitamin B6 metabolism,map00750)等途径(图2C);缺失型结构变异基因主要富集于N糖的生物合成(N-Glycan biosynthesis,map00510)、苯丙烷类物质的生物合成(Phenylpropanoid biosynthesis,map00940)以及叶酸生物合成(Folate biosynthesis,map00790)(图2D)。富集分析结果说明,两个甘薯野生近缘种之间的结构变异,主要用来改变次级产物的变化,在保证了基本生存功能基因保守性的同时,增加了两个物种对不同环境的适应性。



A:插入型结构变异基因的GO富集分析;B:缺失型结构变异基因的GO富集分析;C:插入型结构变异基因的KEGG富集分析;D:缺失型结构变异基因的KEGG富集分析

A: GO enrichment analysis of insertion structural variant genes; B: GO enrichment analysis of deletion structural variant genes; C: KEGG enrichment analysis of insertion structural variant genes; D: KEGG enrichment analysis of deletion structural variant genes

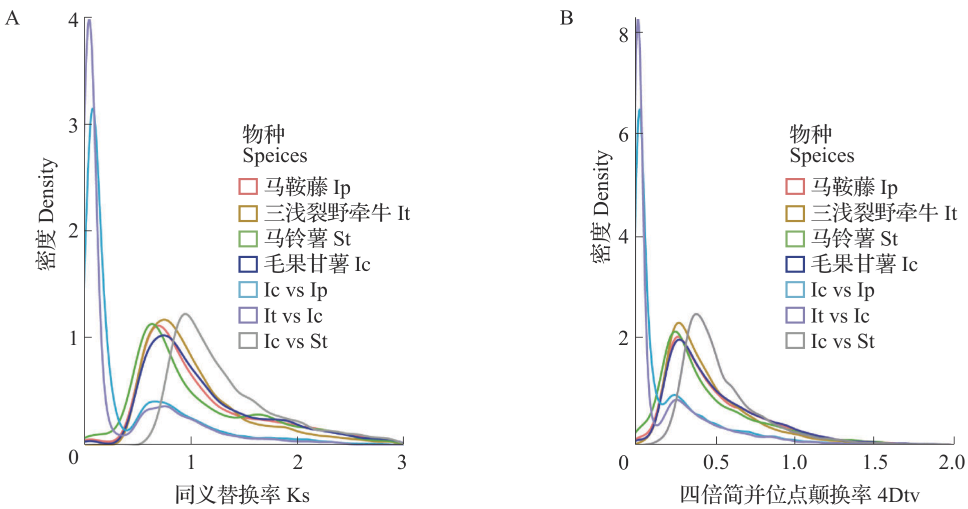
图2　三浅裂野牵牛和毛果甘薯基因组间结构变异基因的富集分析

Fig. 2　Enrichment analysis of structural variant genes between *I.trifida* and *I.cordatotriloba* genomes

2.4　两个近缘野生种的进化和分化

为了探讨毛果甘薯和三浅裂野牵牛在番薯属中的进化历程,本研究分析了三浅裂野牵牛、毛果甘薯以及番薯属近缘种马鞍藤之间共线性基因对的Ks值,以马铃薯基因组[27]为参照。结果显示三浅裂野牵牛的峰对应的Ks值(Ks^{lt})和毛果甘薯的峰对应的Ks值(Ks^{lc})最为接近,其次是马鞍藤的峰对应的Ks值(Ks^{lp}),三者的峰都在Ks = 0.75左右(图3A),说明番薯属这三个种在约53.18百万年前发生过加倍事件。而番薯属的加倍事件主要为三倍化[28]。进一步研究不同物种之间的Ks值分布,结果显示三浅裂野牵牛和毛果甘薯相比主峰值对应

的Ks值($Ks^{It\ vs\ Ic}$)略小于毛果甘薯和马鞍藤相比主峰值对应的Ks值($Ks^{Ic\ vs\ Ip}$),证实了三浅裂野牵牛和毛果甘薯的分化时间略晚于毛果甘薯和马鞍藤的分化时间。Ks^{It}、Ks^{Ic}和Ks^{Ip}均大于$Ks^{It\ vs\ Ic}$和$Ks^{Ic\ vs\ Ip}$,而小于$Ks^{Ic\ vs\ St}$,这说明番薯属的全基因组三倍化事件发生于属内不同种的分化之前,而在番薯属与马铃薯分化之后。除此之外,三浅裂野牵牛和毛果甘薯以及毛果甘薯和马鞍藤之间比较得到的Ks分布各存在一个较低的峰,相应的Ks值与3个物种各自的峰值Ks接近;而各自较低峰对应的Ks值又大于主峰对应的Ks值,对应了全基因组三倍化事件,也进一步证实了番薯属的全基因组三倍化事件发生于属内不同种的分化之前。而马铃薯对应的峰Ks^{St}与番薯属较为接近,说明番薯属的祖先与马铃薯的祖先在相近时期分别独立发生了全基因组加倍事件。上述这些物种共线性碱基对之间的四倍简并位点颠换率(4DTv)统计得到了相同的结果(图3B)。



A:三浅裂野牵牛、毛果甘薯、马鞍藤和马铃薯以及它们之间的Ks值分布;B:三浅裂野牵牛、毛果甘薯、马鞍藤和马铃薯以及它们之间的4DTv值分布

A: Distributions of Ks value of *I.trifida*, *I.cordatotriloba*, *I.pes-caprae*, *S. tuberosum* and that between them; B: Distributions of 4DTv value of *I. trifida*, *I.cordatotriloba*, *I.pes-caprae*, *S. tuberosum* and that between them

图3 三浅裂野牵牛、毛果甘薯、马鞍藤和马铃薯以及它们之间的Ks和4DTv值分布

Fig. 3 Distributions of Ks and 4DTv value of *I.trifida*, *I.cordatotriloba*, *I.pes-caprae*, *S. tuberosum* and that between them

3 讨论

中国甘薯遗传基础较为狭窄,大部分推广品种具有南瑞苕或者胜利百号的血缘[29]。甘薯野生种质资源丰富,挖掘野生近缘种中的优异变异应用于甘薯育种,有利于拓宽甘薯的遗传背景。同时,甘薯的起源尚不清晰,了解甘薯野生近缘种的分化可为甘薯的起源进化提供参考。本研究以甘薯野生近缘种中基因组组装质量较高的两个种为研究对象,进行了比较基因组分析。挖掘了覆盖全基因组的SNP,可以此为基础,进一步筛选开发分子标记,快速简便地对甘薯野生种进行分型和鉴定。

染色体倒位和易位是一种重要的遗传变异,可以迅速改变染色体的结构和基因组合,创造出新的基因连锁关系,并通过长时间的进化,在群体中维持一定的频率。新的基因组合可能赋予个体新的适应性特征,有助于物种适应环境变化。同时,倒位还可能会影响基因的表达。三浅裂野牵牛和毛果甘薯之间存在2076个易位和84个倒位。人类染色体倒位和易位最主要的遗传效应是育性受损。天然三浅裂野牵牛自交育性较差,且基因组是杂合的,这些倒位和易位可能会存在于三浅裂野牵牛亚基因组之间,从而在减数分裂时期产生染色体异常的配子,最终导致自交亲和性较差。栽培甘薯的自交亲和性差可能也与此相关,但需要更高质量的单倍型基因组进行深入研究。一些大型的倒位会抑制染色体的局部重组,产生生殖隔离进而推动物种分化[30]。三浅裂野牵牛和毛果甘薯的物种分化可能与本研究中发掘的84段倒位存在联系,但是倒位产生的原因仍然不清晰。三浅裂野牵牛和毛果甘薯基因组倒位断点处的基因有35个,这些基因功能注释多样,并未发现与倒位的发生存在明显相关性。说明倒位的发生可能并不是由断点处基因起作用,可能与染色体的空间结构存在联系。

基于测序技术和算法的影响，以往对基因组变异的认识主要集中于SNP，基因组变异与表型变异之间的关联也以SNP为主。但是随着技术的进步，越来越多的证据证明结构变异也是基因组变异的重要部分。利用短读长挖掘的结构变异假阳性较高，甚至达到了89%[31]。本研究基于长读长测序挖掘插入和缺失型的结构变异，并结合两种不同比对方法进行筛选，最终保留的结构变异不足四分之一，大大提高了结构变异的可靠性。结构变异在物种分化和遗传多样性中扮演重要角色，一些影响物种生长发育的结构变异会导致物种本身无法生存，这些变异在进化中不会被保留，而相对轻微的变异，在自然选择和进化过程中可能会被保留下来，为物种的进化和适应提供了原材料。本研究通过比较三浅裂野牵牛和毛果甘薯基因组，发现结构变异基因主要富集于次级代谢通路，倒位中的基因也主要富集于次级代谢通路，次级代谢通路基因的改变赋予了甘薯野生种相应表型的变异以及对不同环境的适应性，丰富了甘薯野生近缘种的遗传多样性。

全基因组加倍是物种进化和分化的重要驱动力，由于缺乏高质量参考基因组，从全基因组加倍角度对番薯属植物进化历程的研究还较少。甘薯及其野生近缘种在进化过程中经历了全基因组三倍化事件，通过分析甘薯野生种与其它物种（马铃薯）共线性基因对的Ks和4DTv，证实在漫长的进化历程中，番薯属植物先是与马铃薯所在的属分道扬镳，之后各自经历了全基因组三倍化事件。番薯属的全基因组三倍化与马铃薯的全基因组加倍发生的时间相近，约在53.18百万年以前。被子植物大规模全基因组加倍事件主要发生在三个地质历史时期，其中第二次大规模的加倍化事件发生在白垩纪-古近纪灭绝时期，距今约65百万年以前[32]。番薯属祖先和马铃薯祖先的全基因组加倍可能就发生在这个时期。全基因组加倍使其增强了对环境的适应性，在剧烈的环境变化中幸存了下来。随后番薯属植物分化成不同的种。三浅裂野牵牛和毛果甘薯（或者它们的祖先）先与马鞍藤分化，之后三浅裂野牵牛和毛果甘薯发生了分化。这与前人对番薯属野生种的研究结果相吻合[5, 33]。

参考文献

［1］ 乔卫华，张宏斌，郑晓明，陈宝雄，陈彦清，李垚奎，程云连，张丽芳，方沩，孙玉芳，杨庆文．我国作物野生近缘植物保护工作近20年的成就与展望．植物遗传资源学报，2020，21(6)：1329-1336

Qiao W H，Zhang H B，Zheng X M，Chen B X，Chen Y Q，Li Y K，Cheng Y L，Zhang L F，Fang W，Sun Y F，Yang Q W. Achievements of the conservation of wild relatives of crops in the past 20 years and the prospects in China. Journal of Plant Genetic Resources，2020，21(6)：1329-1336

［2］ Muñoz-Rodríguez P，Wood J R I，Wells T，Carruthers T，Sumadijaya A，Scotland R W. The challenges of classifying big genera such as *Ipomoea*. Taxon，2023，72(6)：1201-1215

［3］ 刘旭．作物种质资源学教程．北京：高等教育出版社，2024：248

Liu X. Crop germplasmics. Beijing：Higher Education Press，2024：248

［4］ Hirakawa H，Okada Y，Tabuchi H，Shirasawa K，Watanabe A，Tsuruoka H，Minami C，Nakayama S，Sasamoto S，Kohara M，Kishida Y，Fujishiro T，Kato M，Nanri K，Komaki A，Yoshinaga M，Takahata Y，Tanaka M，Tabata S，Isobe S N. Survey of genome sequences in a wild sweet potato，*Ipomoea trifida*（H. B. K.）G. Don. DNA Research，2015，22(2)：171-179

［5］ Wu S，Lau K H，Cao Q，Hamilton J P，Sun H，Zhou C，Eserman L，Gemenet D C，Olukolu B A，Wang H，Crisovan E，Godden G T，Jiao C，Wang X，Kitavi M，Manrique-Carpintero N，Vaillancourt B，Wiegert-Rininger K，Yang X，Bao K，Schaff J，Kreuze J，Gruneberg W，Khan A，Ghislain M，Ma D，Jiang J，Mwanga R O M，Leebens-Mack J，Coin L J M，Yencho G C，Buell C R，Fei Z. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. Nature Communications，2018，9(1)：4580

［6］ Li M，Yang S，Xu W，Pu Z，Feng J，Wang Z，Zhang C，Peng M，Du C，Lin F，Wei C，Qiao S，Zou H，Zhang L，Li Y，Yang H，Liao A，Song W，Zhang Z，Li J，Wang K，Zhang Y，Lin H，Zhang J，Tan W. The wild sweetpotato（*Ipomoea trifida*）genome provides insights into storage root development. BMC Plant Biology，2019，19(1)：119

［7］ Yoon U-H，Cao Q，Shirasawa K，Zhai H，Lee T-H，Tanaka M，Hirakawa H，Hahn J-H，Wang X，Kim H S，Tabuchi H，Zhang A，Kim T-H，Nagasaki H，Xiao S，Okada Y，Jeong J C，Nagano S，Shin Y，Lee H-U，Park S-U，Lee S J，Lee K，Yang J-W，Ahn B O，Ma D，Takahata Y，Kwak S-S，Liu Q，Isobe S. Haploid-resolved and chromosome-scale genome assembly in hexa-autoploid sweetpotato（*Ipomoea batatas*（L.）Lam）.（2022-12-25）［2024-4-01］. https://doi. org/10.1101/2022.12.25.521700

［8］ Xiao S，Wang Y，Zhou Z，Zhao L，Zhao L，Gao B，Dai X，Xu P，Cao Q. *Xiaoshu*，a simple genetic model system for sweetpotato（*Ipomoea batatas*（L.）Lam.). Plant Biotechnology Journal，2025，23(2)：674-676

［9］ Shi J，Tian Z，Lai J，Huang X. Plant pan-genomics and its applications. Molecular Plant，2023，16(1)：168-186

[10] He W，He H，Yuan Q，Zhang H，Li X，Wang T，Yang Y，Yang L，Yang Y，Liu X，Wei H，Zhang H，Zhang B，Guo M，Leng Y，Shi C，Lyu Y，Chen W，Wang X，Zhang Z，Yu B，Zhang B，Xu Q，Qian H，Zhou Y，Wang S，Qian Q，Shang L. Widespread inversions shape the genetic and phenotypic diversity in rice. Science Bulletin，2024，69(5)：593-596

[11] Liu Y，Du H，Li P，Shen Y，Peng H，Liu S，Zhou G A，Zhang H，Liu Z，Shi M，Huang X，Li Y，Zhang M，Wang Z，Zhu B，Han B，Liang C，Tian Z. Pan-genome of wild and cultivated soybeans. Cell，2020，182(1)：162-176 e113

[12] Kronenberg Z N，Fiddes I T，Gordon D，Murali S，Cantsilieris S，Meyerson O S，Underwood J G，Nelson B J，Chaisson M J P，Dougherty M L，Munson K M，Hastie A R，Diekhans M，Hormozdiari F，Lorusso N，Hoekzema K，Qiu R，Clark K，Raja A，Welch A E，Sorensen M，Baker C，Fulton R S，Armstrong J，Graves-Lindsay T A，Denli A M，Hoppe E R，Hsieh P，Hill C M，Pang A W C，Lee J，Lam E T，Dutcher S K，Gage F H，Warren W C，Shendure J，Haussler D，Schneider V A，Cao H，Ventura M，Wilson R K，Paten B，Pollen A，Eichler E E. High-resolution comparative analysis of great ape genomes. Science，2018，360(6393)：eaar6343

[13] Sedlazeck F J，Rescheneder P，Smolka M，Fang H，Nattestad M，von Haeseler A，Schatz M C. Accurate detection of complex structural variations using single-molecule sequencing. Nature Methods，2018，15(6)：461-468

[14] Smolka M，Paulin L F，Grochowski C M，Horner D W，Mahmoud M，Behera S，Kalef-Ezra E，Gandhi M，Hong K，Pehlivan D，Scholz S W，Carvalho C M B，Proukakis C，Sedlazeck F J. Detection of mosaic and population-level structural variants with Sniffles2. Nature Biotechnology，2024，42(10)：1571-1580

[15] Wang K，Li M，Hakonarson H. ANNOVAR：Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research，2010，38(16)：e164-e164

[16] Goel M，Sun H，Jiao W B，Schneeberger K. SyRI：Finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biology，2019，20(1)：277

[17] Wang Y，Tang H，DeBarry J D，Tan X，Li J，Wang X，Lee T-h，Jin H，Marler B，Guo H，Kissinger J C，Paterson A H. MCScanX：A toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Research，2012，40(7)：e49

[18] Yang Z. PAML 4：Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution，2007，24(8)：1586-1591

[19] Zhang Z. KaKs_Calculator 3.0：Calculating selective pressure on coding and non-coding sequences. Genomics，Proteomics & Bioinformatics，2022，20(3)：536-540

[20] Jones P，Binns D，Chang H-Y，Fraser M，Li W，McAnulla C，McWilliam H，Maslen J，Mitchell A，Nuka G，Pesseat S，Quinn A F，Sangrador-Vegas A，Scheremetjew M，Yong S-Y，Lopez R，Hunter S. InterProScan 5：Genome-scale protein function classification. Bioinformatics，2014，30(9)：1236-1240

[21] Cantalapiedra C P，Hernández-Plaza A，Letunic I，Bork P，Huerta-Cepas J. eggNOG-mapper v2：Functional annotation，orthology assignments，and domain prediction at the metagenomic scale. Molecular Biology and Evolution，2021，38(12)：5825-5829

[22] Alexa A，Rahnenführer J，Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics，2006，22(13)：1600-1607

[23] Supek F，Bošnjak M，Škunca N，Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE，2011，6(7)：e21800

[24] Kuznetsova I，Lugmayr A，Siira S J，Rackham O，Filipovska A. CirGO：An alternative circular way of visualising gene ontology terms. BMC Bioinformatics，2019，20(1)：84

[25] Ginestet C. ggplot2：Elegant graphics for data analysis. Journal of the Royal Statistical Society Series A：Statistics in Society，2011，174(1)：245-246

[26] Yu G，Wang L G，Han Y，He Q Y. clusterProfiler：An R package for comparing biological themes among gene clusters. Omics，2012，16(5)：284-287

[27] Yang X，Zhang L，Guo X，Xu J，Zhang K，Yang Y，Yang Y，Jian Y，Dong D，Huang S，Cheng F，Li G. The gap-free potato genome assembly reveals large tandem gene clusters of agronomical importance in highly repeated genomic regions. Molecular Plant，2023，16(2)：314-317

[28] Yang J，Moeinzadeh M H，Kuhl H，Helmuth J，Xiao P，Haas S，Liu G，Zheng J，Sun Z，Fan W，Deng G，Wang H，Hu F，Zhao S，Fernie A R，Boerno S，Timmermann B，Zhang P，Vingron M. Haplotype-resolved sweet potato genome traces back its hexaploidization history. Nature Plants，2017，3(9)：696-703

[29] Liu D，Ning Z，Hong Z，Yu X，Qin J，Wang L，He S，Liu Q. AFLP fingerprinting and genetic diversity of main sweetpotato varieties in China. Journal of Integrative Agriculture，2012，11(9)：1424-1433

[30] Hu H，Scheben A，Wang J，Li F，Li C，Edwards D，Zhao J. Unravelling inversions：Technological advances，challenges，and potential impact on crop breeding. Plant Biotechnology Journal，2024，22(3)：544-554

[31] Yuan Y，Bayer P E，Batley J，Edwards D. Current status of structural variation studies in plants. Plant Biotechnology Journal，2021，19(11)：2153-2163

[32] Fawcett J A，Maere S，Van de Peer Y. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proceedings of the National Academy of Sciences，2009，106(14)：5737-5742

[33] Jiang F，Wang S，Wang H，Wang A，Xu D，Liu H，Yang B，Yuan L，Lei L，Chen R，Li W，Fan W. A chromosome-level reference genome of a Convolvulaceae species *Ipomoea cairica*. G3 (Bethesda)，2022，12(9)：jkac187