

玉米C型不育系、保持系和恢复系 线粒体基因组的比较分析

张启明¹, 石海春^{1,2}, 余学杰^{1,2}, 杨 殊¹, 赵长云², 许秀兰², 柯永培^{1,2}

(1. 四川农业大学农学院, 成都 611130; 2. 四川正红生物技术有限公司, 成都 610213)

摘 要: 以不育系 K932S 和 K169S、保持系 K169 和恢复系 K932R 为供试材料, 通过二代测序技术对线粒体基因组序列进行比较分析, 挖掘玉米 CMS-C 败育相关基因(*ORFs*)。结果表明, 不育胞质 K932S、K169S 和 K932R 线粒体基因组大小分别为 739 676、739 723、739 765 bp, 正常胞质 K169 为 569 617 bp。注释到蛋白编码基因 33 个, tRNA 基因 15~16 个和 rRNA 基因 3~4 个。大于 100 bp 重复序列在不育胞质和正常胞质基因组中分别为 19、10 个, 小于 100 bp 重复序列差异不大。同源片段在 3 个不育胞质基因组中排列顺序一致, 共线性达 99% 以上。正常胞质中 3 个片段反转, 排列顺序也发生了重大改变, 相似度为 95%。4 个基因组中发现 *ORFs* 基因 145~151 个, 其中, K169S 与 K169 间差异基因(*ORFs*)高达 13 个, 与 K932S 或 K932R 间差异仅 1 个, K169S 特有 *ORFs* 基因 8 个。筛选出 orf429 (*atp6-c*)、orf104-s、orf410、orf349 和 orf246 共 5 个 *ORFs*, 可能与 CMS-C 败育相关。

关键词: 玉米; 细胞质雄性不育; 线粒体基因组; 差异基因(*ORFs*)

中图分类号: S513.035.3

文献标识码: A

Comparative Analyses of Mitochondrial Genomes of C-type Sterile Line, Maintainer Line and Restorer Line in Maize

ZHANG Qi-ming¹, SHI Hai-chun^{1,2}, YU Xue-jie^{1,2}, YANG Shu¹, ZHAO Chang-yun², XU Xiu-lan², KE Yong-pei^{1,2}

(1. College of Agronomy, Sichuan Agricultural University, Chengdu 611130;

2. Sichuan Zhenghong Bio Co., Ltd., Chengdu 610213, China)

Abstract: In order to dig up genes(*ORFs*) related to CMS-C sterility in maize, the sequences of the mitochondrial genome of four materials, sterile line K932S and K169S, maintainer line K169 and restorer line K932R, were comparatively analyzed by using next-generation sequencing technology. The results showed that the mitochondrial genome length of sterile cytoplasm K932S, K169S and K932R were 739 676 bp, 739 723 bp and 739 765 bp, respectively; and that of normal cytoplasm K169 was 569 617 bp; 33 protein coding genes, 15-16 tRNA genes and 3-4 rRNA genes were annotated. There were 19 and 10 repetitive sequences longer than 100 bp in the sterile cytoplasmic and normal cytoplasmic genomes, respectively, and repetitive sequences shorter than 100 bp presented little differences; Among the three sterile cytoplasmic genomes, the homologous fragments are arranged in the same order, with a collinearity of 99%, however, three reversed fragments were found in the normal cytoplasm, and the arrangement order also changed, with a similarity of 95%. Among the four genomes, 145-151 *ORFs* were found, there were 13 differential genes(*ORFs*) between K169S and K169, but only 1 of that was extracted between K932S and K932R, 8 unique *ORFs* was discovered in K169S mitochondrial genome. Five *ORFs* including orf429(*atp6-c*), orf104-s, orf410, orf349 and orf246 were screened, which may be related to CMS-C abortion. Those results may provide some genomic information for dissecting the mechanism of CMS-C abortion in maize.

Key words: Maize; Cytoplasmic male sterility; Mitochondrial genome; Differential genes(*ORFs*)

录用日期: 2020-10-12

基金项目: 四川省科技支撑计划项目(2016NYZ0006)

作者简介: 张启明(1994-), 甘肃天水人, 硕士, 研究专业为作物遗传育种。E-mail: 945935557@qq.com

柯永培为本文通讯作者。E-mail: keyp169@163.com

玉米是最早利用细胞质雄性不育(cytoplasmic male sterility, CMS)生产杂交种子的作物之一^[1]。二十世纪70年代初,国外玉米小斑病T小种爆发,给玉米生产造成了严重损失。研究者更加注重CMS败育及恢复机理研究,以期为CMS材料的合理应用提供理论指导。Levings^[2]证明,玉米CMS与线粒体基因(Mitochondrial gene, mtDNA)有关,认为mtDNA是细胞质雄性不育因子的载体。随后,在多作物上都鉴定出与CMS育性相关的mtDNA,如玉米 *T-urf13*、水稻 *orf79*、油菜 *orf224*、萝卜 *orf125* 和小麦 *orf256* 等^[3]。前人研究认为,植物CMS主要是由于mtDNA的突变和重组引起的^[4,5],与重排产生的特异基因嵌合基因密切相关。

玉米CMS-T育性相关基因 *T-urf13* 的发现,被誉为植物CMS分子生物学的研究的典范^[6,7],该基因编码1个13 kDa多肽,第39位氨基酸会结合T毒素,使得线粒体内膜通道孔径加大,导致线粒体功能异常,该基因的表达能够导致花粉降解^[8]。在CMS-S中,存在1个R区,该区域与 *atp9* 的侧翼区及编码区高度同源,其中两个嵌合的 *orf355-orf77* 是导致CMS-S败育的原因^[9-11]。关于CMS-C,目前认为 *atp6-c* 可能与育性有关^[12],但尚缺乏直接证据,也未发现可能与CMS-C育性相关的候选基因或者ORFs,其败育机理仍然未知。本研究以不育系K932S与K169S、保持系K169和近等基因恢复系K932R为供试材料,其中,不育细胞质来源于K932S,属于CMS-C材料,从二分体时期开始败育^[13]。利用二代测序技术组装线粒体基因组,通过线粒体差异基因的比较分析,从mtDNA一级结构上筛选与玉米CMS-C育性相关的ORFs,为进一步在分子水平上阐明其败育机理提供依据。

1 材料与方法

1.1 供试材料

供试材料为CMS-C不育系K932S与K169S、保持系K169和恢复系K932R。其中,K932S、K169S和K932R为相同不育胞质、不同核背景材料,K932R为K932S的近等基因系;K169为正常可育胞质,与K169S核背景相同,由四川正红生物技术有限责任公司提供。

1.2 实验方法

1.2.1 玉米线粒体基因组提取

用植物线粒体DNA萃取试剂盒(GENMED公司)提取线粒体DNA。

1.2.2 玉米线粒体基因组测序

采用Illumina Hiseq高通量测序平台,2×150 bp双末端测序策略对文库进行测序。由上海天昊生物技术有限公司完成。

1.2.3 玉米线粒体基因组组装

用软件metaSPAdes对过滤后的clean reads进行拼接和gap补平,并比对到参考基因组上进行优化,获得完整的基因组序列。

1.2.4 玉米线粒体基因组注释

不育胞质与可育胞质分别以B37核背景下的CMS-C(GenBank登录号:DQ645536)和NB(GenBank登录号:AY506529)线粒体基因组为参考基因组,用mitoMaker脚本对组装结果进行注释,通过在线软件Mitofy对注释结果加以校正,用在线软件tRNAscan-SE(<http://lowelab.ucsc.edu/tRNAscan-SE>)对tRNA基因进行注释,注释得到gene、CDS、tRNA、rRNA 4种特征。

1.2.5 玉米线粒体基因组的比较分析

利用NCBI上Blastn suite-2 sequence(<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)搜索线粒体基因组重复序列,用软件SSRHunter搜索SSR位点,利用Mauve软件确定各线粒体基因组间具有共线性关系的序列信息。

1.2.6 玉米线粒体差异基因分析

用NCBI的在线软件ORF Finder(<http://www.ncbi.nlm.nih.gov/projects/gorf/>)预测线粒体基因组的所有ORFs(≥300 bp)。用DNAMAN和MEGA5.0对所有的ORFs以及注释到的蛋白编码基因进行序列比较,并用在线软件TMHMM Server v.2.0(<http://www.cbs.dtu.dk/services/TMHMM/>)预测相应的蛋白跨膜结构域。根据ORFs嵌合结构、与功能基因的位置和蛋白跨膜结构域,尤其是与ATP合成酶以及电子呼吸链编码功能基因的相关性等特点,筛选CMS-C败育相关基因。

2 结果与分析

2.1 玉米线粒体基因组测序结果

2.1.1 测序数据统计

通过Illumina Hiseq高通量测序平台,采用2×150 bp双末端策略,对4个材料的文库进行线粒体基因组测序。根据原始数据过滤后结果, Q_{30} 值均为100%, Q_{30} 值均在94%以上,Clean reads所占比例均在98%左右(表1)。说明测序数据整体质量较好,可用于后续线粒体基因组的组装。

2.1.2 线粒体基因组组装

对测序片段进行拼接,K169片段数最少,为8

个,片段长度为 9 909 ~ 158 211 bp;K932R 最多,为 32 个,片段长度为 1 728 ~ 80 313 bp; K932S 和 K169S 介于二者之间,说明 2 个不育系及保持系 DNA 片段较为完整,恢复系 DNA 片段较为破碎。对

存在 gap 的序列补齐后,获得 4 个材料的线粒体基因组序列,其基因组大小在 3 个胞不育质材料间差异不大,总长变幅为 739 676 ~ 739 765 bp;正常胞质 K169 仅 569 617 bp,GC 含量差异不大(表 2)。

表 1 4 个供试材料过滤后测序数据统计
Table 1 Sequence data statistics of 4 test materials after filtration %

材 料 Material	质控后数据 Clean read	质控后碱基 Clean base	比例 Ratio	R1		R2	
				Q20	Q30	Q20	Q30
K932S	20624266	3085925737	98.40	100	98.56	100	94.91
K169S	23280960	3475320234	98.28	100	98.68	100	95.35
K169	38931048	5817575865	97.94	100	98.61	100	95.52
K932R	20451880	3059097822	98.02	100	98.66	100	94.17

注:R1、R2 表示双末端测序;Q₂₀表示序列中质量大于 20 的碱基比例;Q₃₀表示序列中质量大于 30 的碱基比例。
Note: R1 and R2 represent double-terminal sequencing;Q₂₀ represents the proportion of bases with a mass greater than 20 in the sequence.Q₃₀: ratio of bases with a mass greater than 30 in the sequence.

表 2 序列拼接统计分析
Table 2 Statistical analysis of sequence splicing

项 目 Item	K932S	K169S	K169	K932R
基因总长(bp)	739 676	739 723	569 617	739 765
GC 含量(%)	43.97	43.96	43.93	43.97
拼接片段数	15	17	8	32
最大片段长度(bp)	170 685	166 160	158 211	80 313
最小片段长度(bp)	5 135	4 476	9 909	1 728

2.1.3 线粒体基因组注释

在 4 个线粒体基因组中,均注释到蛋白编码基因 33 个,不育胞质中注释到 tRNA 15 个和 rRNA 3 个,正常胞质分别为 16 个和 4 个(表 3),其中,*nad1*、*cob*、*atp1* 和 *rps7* 在 4 个基因组均为双拷贝,但正常胞质中 *rps7* 基因有两个拷贝不同,1 个来源于叶绿体基

因组。*cox2*、*cox3*、*atp9*、*rps2B*、*rps4* 和 *rps13* 在不育胞质中为双拷贝,正常胞质中为单拷贝。不育胞质中 *trnM* 和 *trnS* 比正常胞质多 1 个拷贝,而 *trnR*、*trnV* 和 *trnL* 比正常可育胞质少 1 个拷贝。另外,正常胞质中还存在两个叶绿体来源的非编码 RNA 基因 *rrn16* 和 *trnI*(图 1)。

表 3 4 个线粒体基因组注释基因数量
Table 3 Annotation gene number of 4 mitochondrial genomes

基因数量 Number of genes	K932S	K169S	K169	K932R
蛋白编码基因	33	33	33	33
转运 RNA	15	15	16	15
核糖体 RNA	3	3	4	3

2.2 玉米线粒体基因组比较分析

2.2.1 重复序列分析

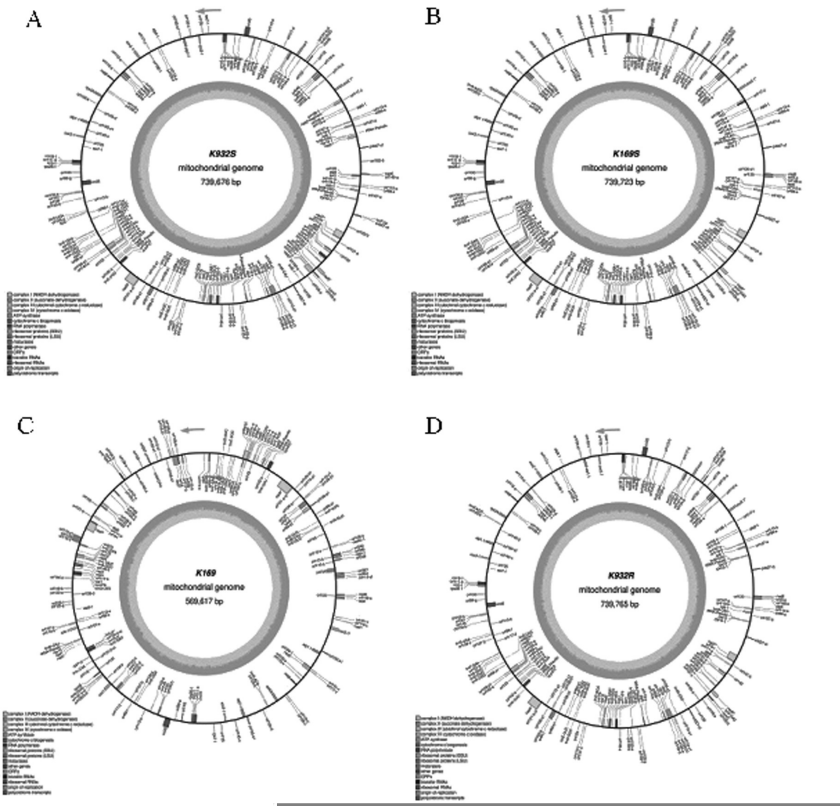
高等植物线粒体基因组中存在大量的重复序列,主要包括随机重复序列和串联重复序列,这些重

复序列并非冗余 DNA,在影响进化、变异和遗传等生命进程中都具有一定的作用。由图 2 可见,在 4 个线粒体基因组中存在着大量的重复序列,其中,30 ~ 40 bp 的重复序列最多,其次是 20 ~ 30、40 ~ 50、

100 ~ 500 bp 重复序列。虽然短片段重复序列活性不高,不能引起线粒体基因组结构和大小的变化,但也能介导重组的发生。小于 100 bp 短片段重复序列数量在 4 个线粒体基因组中差别不大,但大于 100 bp 的重复序列则存在较大的差异。

100 ~ 1 000 bp 重复序列比较,在 K932S、K169S 和 K932R 中差异不大,在 K93R 中仅少了 1 个四拷贝

重复序列,共有拷贝数高达 11 个。在 K169 中,仅双拷贝数相同,三拷贝重复序列数少 4 个,也不存在四拷贝重复序列,与其他 3 个材料共有数仅 3 个(表 4)。由此可见,该不育胞质在不同核背景下 100 ~ 1 000 bp 重复序列差异小,而在在正常胞质中显著减少,这也是造成两种类型胞质线粒体基因组大小差异的重要原因之一。



注: A、B、C 和 D 图分别表示 K932S、K169S、K169 和 K932R 线粒体基因组。
Note: A, B, C, and D in the figure represent K932S, K169S, K169, and K932R mitochondrial genomes, respectively.

图 1 4 个玉米线粒体基因组图谱
Fig.1 Maps of four mitochondrial genomes in maize

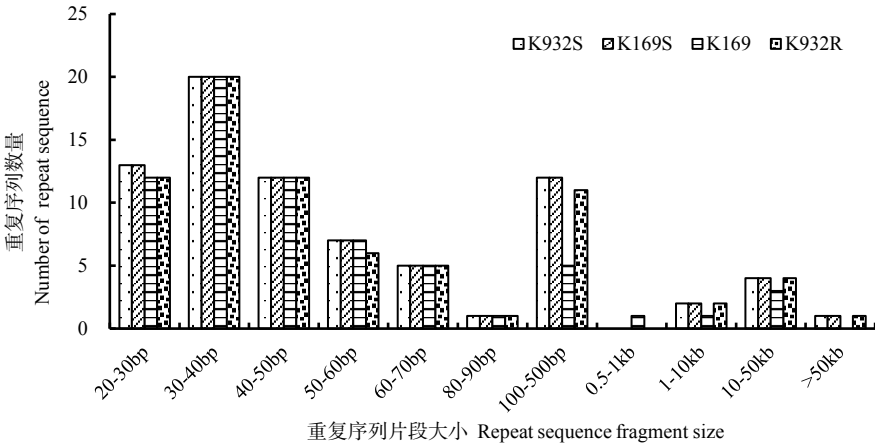


图 2 4 个线粒体基因组中重复序列的数量
Fig.2 The number of repetitive sequences in the four mitochondrial genomes

表 4 4 个线粒体基因组中 100 ~ 1 000 bp 重复序列相似性
Table 4 Similarity of 100–1 000 bp repetitive sequences in four mitochondrial genomes

基因组 Genome	总 数 Sum	双拷贝 Dual copy	三拷贝 Three copy	四拷贝 Four copy	K932S 共有 Shared in K932S	K169S 共有 Shared in K169S	K169 共有 Shared in K169	K932R 共有 Shared in K932R
K932S	12	4	6	2		11	3	11
K169S	12	4	6	2			3	11
K169	6	4	2					
K932R	11	4	6	1	11	11	3	

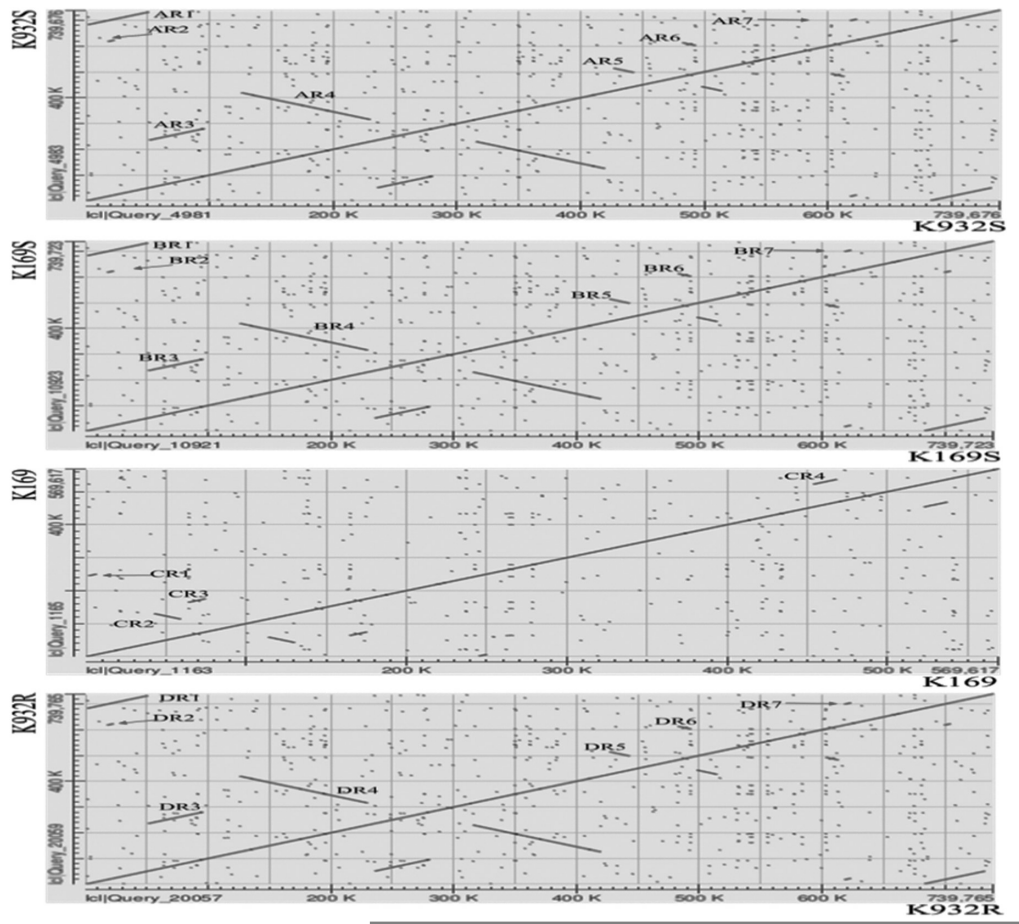


图 3 4 个线粒体基因组中大于 1 kb 重复序列分布图
Fig.3 Distribution of repetitive sequences greater than 1 kb in four mitochondrial genomes

表 5 4 个线粒体基因组 SSR 统计分析
Table 5 SSR statistical analysis of four mitochondrial genomes

重复单元类型 Repeating element type	K932S	K169S	K169	K932R
两碱基重复	230	233	175	228
三碱基重复	41	41	27	37
四碱基重复	2	2	2	2
五碱基重复	38	38	21	38
六碱基重复	16	16	8	16
合 计	327	330	233	321

大于1 kb重复序列比较,在4个玉米线粒体基因组各包含4~7个,其中,保持系K169中4个,其余3个材料中均有7个(图3)。这些大于1 kb的长片段重复序列总长为88 952~238 682 bp,占整个基因组的8.46%~32.27%。说明这些大片段重复序列的存在,是导致不育胞质和可育胞质线粒体基因组大小变异的主要原因。

用软件 SSRHunter 搜索4个线粒体基因组的SSR位点。表5所示,4个材料线粒体基因组中的SSR总数为233~330个,其中,四碱基重复单元在4个材料间无差异,均为2个;其余重复单元保持系K169明显少于其他3个材料,而且恢复系K932R中

二、三碱基重复单元少于2个不育系,说明核背景的改变对SSR位点数有一定影响,预示核质基因组间存在着频繁的序列迁徙现象。

2.2.2 序列共线性分析

图4可以看出,4个基因组极为相似,存在高度同源有片段9个。在K932S、K169S和K932R中,这些共线性片段的排列顺序一致,共线性高达99%。但与K169相比,其中3个片段发生反转,尤其是序列的排列顺序发生了重大改变,不育系中还存在个别不匹配的片段,相似度为95%,说明玉米线粒体基因组在组成上极为相似,但不育胞质与正常胞质间线粒体基因组在结构上存在很大差异。

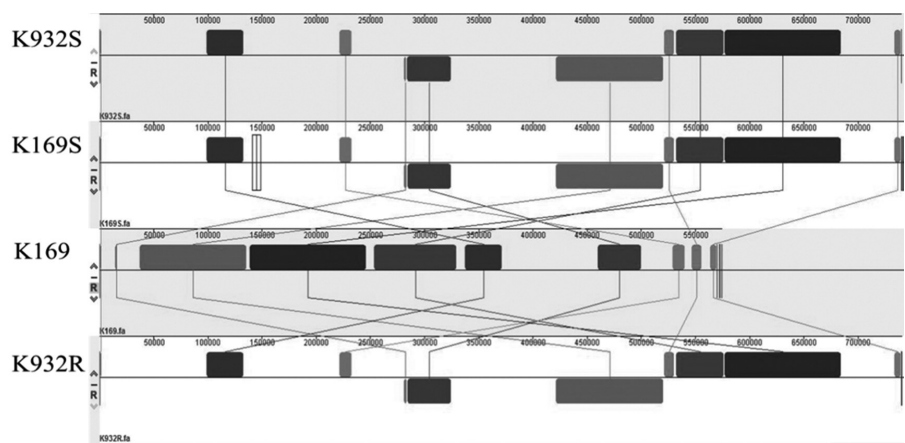


图4 4个线粒体基因组的共线性图

Fig.4 Collinearity diagram of four mitochondrial genomes

2.3 玉米线粒体差异基因分析

2.3.1 线粒体基因组 ORFs 预测

预测线粒体基因组的所有 ORFs(≥ 300 bp)发现,4个线粒体基因组的 ORFs 为145~151个。其中,

正常胞质K169中单拷贝140个,双拷贝11个,其余3个不育胞质材料间的拷贝数差异不大(图5)。这种差异是由于不育胞质比正常胞质存在更大的长片段重复序列的原因所致。

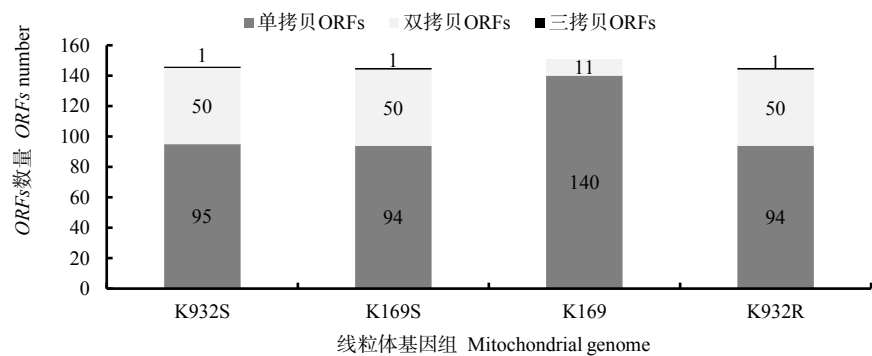


图5 4个线粒体基因组中的 ORFs 数量

Fig.5 Number of ORFs in four mitochondrial genomes

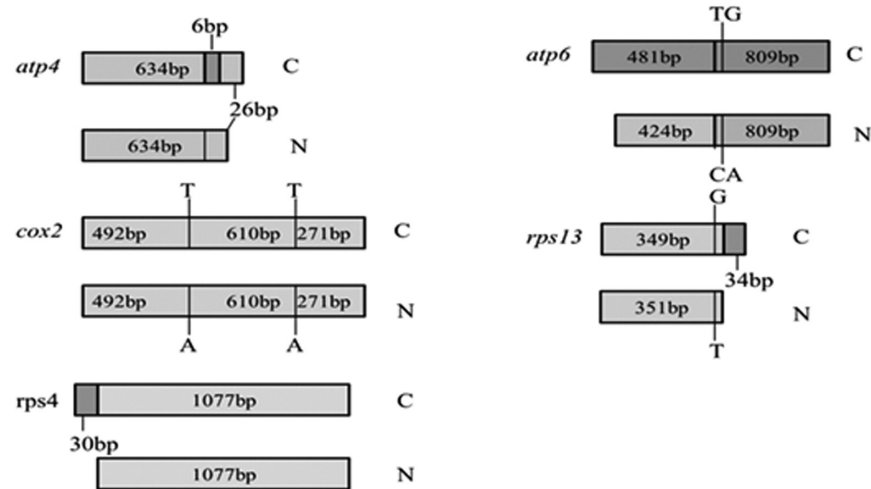
2.3.2 已知功能差异基因分析

通过不育胞质和可育胞质蛋白编码基因序列比对,发现 *atp4*、*atp6*、*cox2*、*rps4* 和 *rps13* 这5个线粒体

基因在不同胞质间存在序列差异(图6),仅 *atp6* 和 *cox2* 蛋白具有跨膜结构域(图7)。其中, *atp4* 在5'端下游634 bp处有6个碱基(AGAGAT)的插入,导致基

因的编码产物由1个苏氨酸(ACC)变异为赖氨酸(AAG)—精氨酸(AGA)—丝氨酸(TCC);*atp6*在5'端嵌入了481 bp的未知来源的序列,而3'端809bp高度保守,仅在537 bp处有两个碱基(TG-CA)的突变,这种改变导致比正常*atp6*多了1个蛋白跨膜结构域;*cox2*在493、1 103 bp处各有1个碱基(T-A)的同义突

变;*rps13*在349 bp处1个碱基T-G的突变导致基因延长了34 bp,且在不育胞质中具有两个拷贝;*rps4*在不育胞质中除1 077 bp拷贝外,还含有1个1 107 bp的拷贝,这个拷贝与第1个拷贝相比仅是在5'端多了30 bp,其余序列完全一致。



注:图中C表示K932S、K169S、K932R中的基因;N表示在K169中的基因;相同颜色表示序列相同。

Note: C in the figure represents the genes in K932S, K169S and K932R. N represents the gene in K169; the same color means the same sequence.

图6 已知功能线粒体差异基因示意图

Fig.6 Schematic diagram of mitochondrial differential genes with known functions

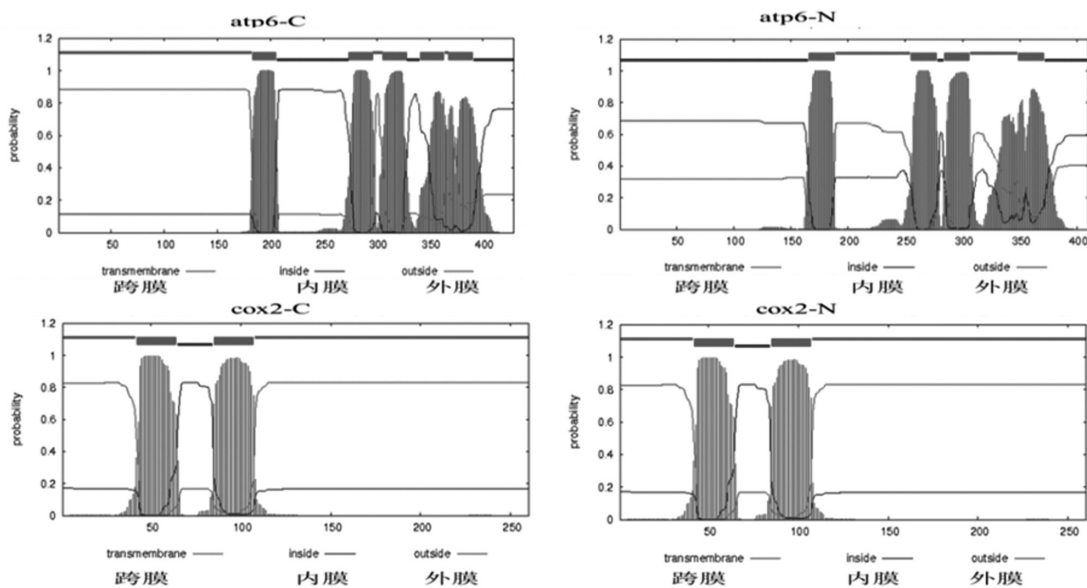


图7 已知功能线粒体差异基因蛋白跨膜结构域预测

Fig.7 Prediction of protein transmembrane domain of mitochondrial differential genes with known functions

2.3.3 差异ORFs分析

对同质异核不育系K932S与K169S线粒体基因组比对分析,发现K169S中orf653由于在K932S线粒体基因组中有44 bp的碱基插入,形成了orf419和orf238两个ORFs,但都不存在蛋白跨膜结构域。不

育系K932S与恢复系K932R相比,除具有与以上相同的差异外,还存在1个ORF差异。orf104-s与orf106-b相比,在205 bp处丢失了6个碱基(ACAA-GA),orf104-s和orf106-b都位于*atp4*和*rpl16*之间,距*rpl16*仅有350 bp,可能与*rpl16*共转录,orf104-s

和orf106-b均具有1个蛋白跨膜结构域(图8、图9)。

同核异质不育系K169S与保持系K169线粒体基因组比对分析,除已知功能差异基因 $atp6$ 等5个外,K932S与K932R差异基因 $orf104-s$ 与 $orf106-b$,在K169S和K169中存在同样差异,其余7个差异 $ORFs$,其中双拷贝5个。另外,K169S中特有 $ORFs$ 8个,K169中特有 $ORFs$ 14个,其中双拷贝5个(表6)。

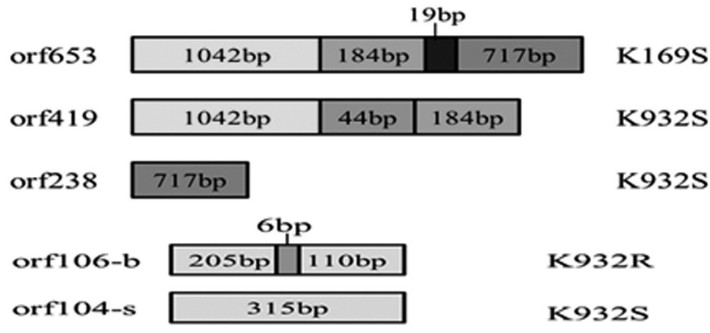


图8 K932S与K169S或K932R线粒体差异ORFs示意图

Fig.8 Schematic diagram of mitochondrial differential ORFs between K932S and K169S or K932R

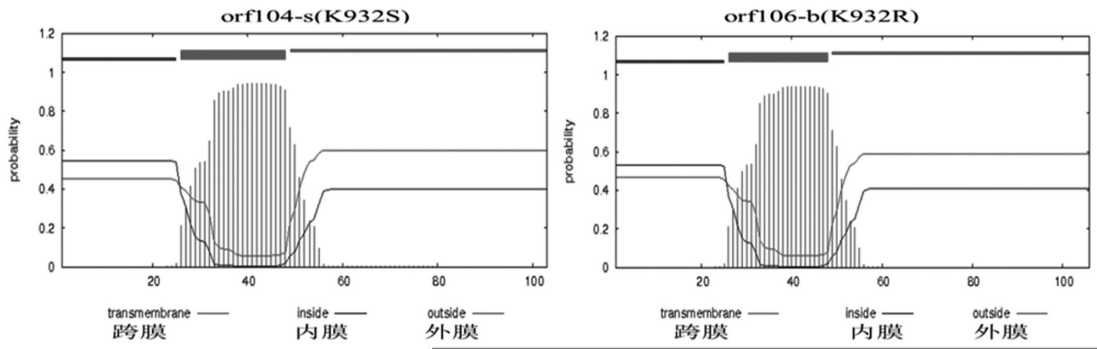


图9 K932S与K932R差异ORF蛋白跨膜结构域预测

Fig.9 Prediction of protein transmembrane domain of mitochondrial differential ORF between K932S and K932R

表6 K169S与K169差异ORFs及特有ORFs

Table 6 Differential ORFs and unique ORFs in K169S and K169

差异ORFs Differential ORFs		特有ORFs Unique ORFs		
K169S	K169	K169S	K169	
orf111(2+)	orf100-c	orf100-s1(2+)	orf173	orf1159
orf173	orf115(2+)	orf100-s2	orf134	orf115
orf149(2+)	orf145	orf105-s(2+)	orf274	orf156
orf163	orf159	orf106-s(2+)	orf115	orf270
orf246(2+)	orf248	orf112	orf734	orf204
orf410	orf396	orf120	orf444	orf107
orf349(2+)	orf522	orf264(2+)	orf387	orf131
-	-	orf307(2+)	-	-

注:括号内数字表示拷贝数。
Note: The numbers in brackets represent copies.

对差异ORFs分析发现,orf173在5'端嵌合了orf115-a的282 bp序列;orf149在322 bp处有48 bp未知来源的序列;orf410在473 bp处嵌合了112 bp的 $atp6$ 序列;orf349在419 bp处嵌合了129 bp的 $atp6$ 序列,在668 bp处嵌合了79 bp的orf186的序列,orf186本身也有34 bp未知序列的嵌入,而这34 bp也在嵌入orf349的79 bp中;orf163在5'端的10 bp来源于orf911,13 bp处到396 bp处来源于orf127,

419 bp处到492 bp处来源于 *rps3*, 而其他 ORFs 均无嵌合结构(图 10)。进一步预测这些 ORFs 的蛋白跨

膜结构域, 发现只有 orf246、orf410 和 orf349 存在蛋白跨膜结构域(图 11)。

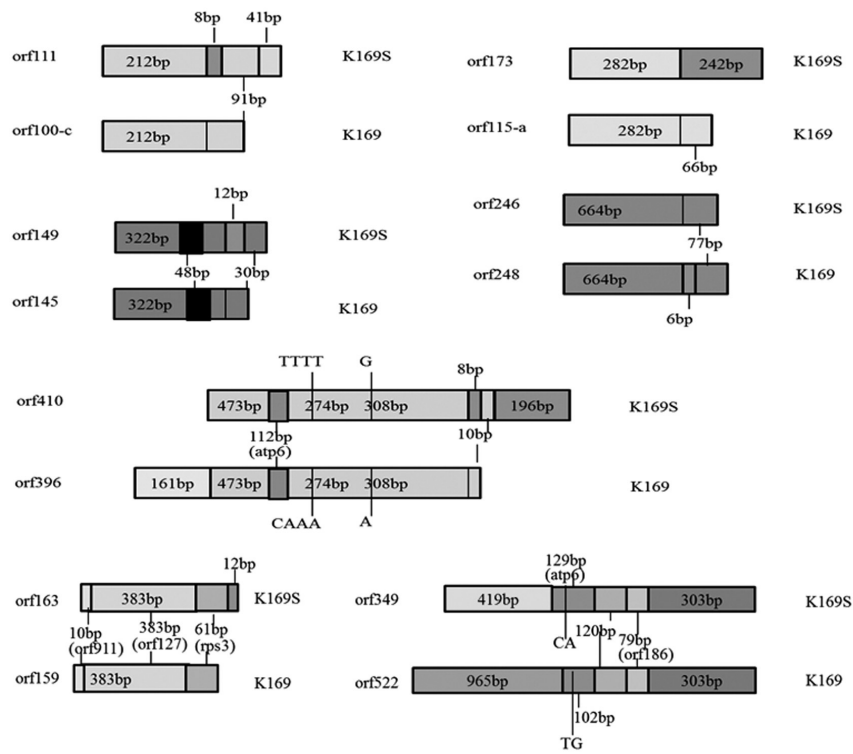


图 10 K169S与K169线粒体差异ORFs示意图

Fig.10 Schematic diagram of mitochondrial differential ORFs between K169S and K169

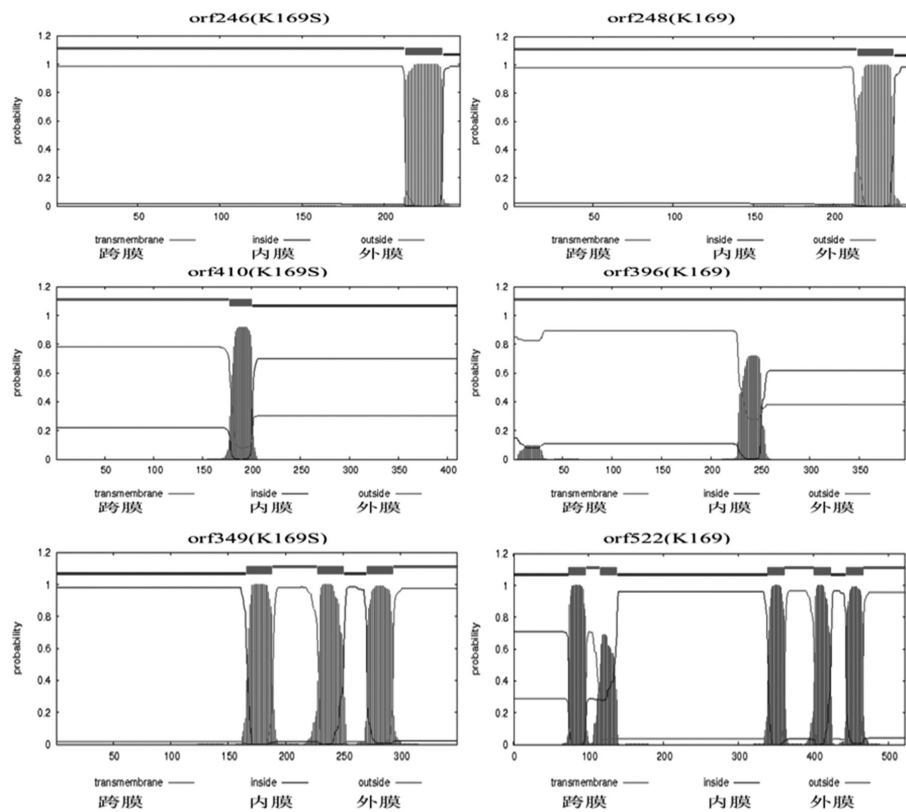


图 11 K169S与K169线粒体差异ORFs跨膜蛋白结构域预测

Fig.11 Prediction of protein transmembrane domain of mitochondrial differential ORFs between K169S and K169

表7 K169S特有ORFs同源序列
Table 7 Homologous sequences of unique ORFs in K169S

开放阅读框 ORFs	同源序列 Homologous sequences	同源序列基因组 Homologous sequence genomes	蛋白跨膜结构域 Protein transmembrane domains
orf264	632bp from CMS-S, 203bp from teosinte mtDNA	CMS-S, teosinte mtDNA	0
orf100-s1	92bp from orf100-c	NB, CMS-S	0
orf100-s2	146bp from orf100-a	NB, CMS-T, CMS-S	1
orf105-s	238bp from teosinte mtDNA	NA, CMS-S	0
orf106-s	308bp from NA	teosinte mtDNA, CMS-S	0
orf112	339bp from NA	NA, CMS-T	1
orf120	142bp from NB mtDNA orf127	NB	0
orf307	812bp from CMS-S	NA, (628 bp A-C) teosinte mtDNA, CMS-S	0

进一步对K169S线粒体基因组中的特异ORFs,在NCBI-nr数据库中进行同源序列搜索,发现这些ORFs大多在玉米其他类型的细胞质中存在同源序列,其中,只有orf264和orf307具有嵌合结构,orf100-s2和orf112具有蛋白跨膜结构域(表7)。

综上所述,同核异质K169S与K169相比,具有差异和不育系中特异基因(*ORFs*)共有21个,其中,具有蛋白跨膜结构域的*ORFs*仅7个。orf173、orf149、orf410、orf163、orf349、orf264、orf307和orf106-s 8个*ORFs*具有嵌合结构;而orf104-s位于*atp4*和*rpl16*之间,且距*rpl16*仅有350 bp,可能与*rpl16*共转录。

3 结论与讨论

3.1 玉米CMS-C同质异核线粒体基因组比较

现有研究发现,植物线粒体基因组处于一种动态变化的状态,有来自于细胞核和叶绿体基因序列的水平转移或迁移^[14]。张东旭^[15]等发现,相同胞质(Milo)来源的高粱CMS线粒体基因组具有异质性,验证了CMS基因的表达受核质相互作用的影响以及核质协同进化的理论。张艳^[16]研究发现,两组Ogu型甘蓝CMS同质异核材料酶切图谱具有差异。NCBI上公布的水稻线粒体有20多种,基因组大小为400~600 kb,造成差异较大的原因可能是由于不同品种中重复序列的不同^[17]。B37核背景下玉米CMS-C线粒体基因组大小为739 719 bp(DQ645536)。邱涛^[18]测序结果,4个不同核背景玉米CMS-C不育系的线粒体基因组大小为739 543~739 660 bp。本研究表明,在不同核背景下基因组大小为739 676~739 765 bp。

本研究发现,K932S、K169S和K932R线粒体基因组存在排列顺序一致的高度同源片段,共线性在

99%以上;K932S中的SSR位点数比K169S少3个,不存在其他差异;K932S比K932R多6个,且多了1个100~1 000 bp的重复序列。差异ORFs分析发现,在K932S中由于44 bp的插入,导致了K169S中的orf653变成了两个较短的ORFs(orf419和orf238);另一个差异ORF在K932S和K169S中为orf104-s,而在K932R中有6个碱基的插入变成orf106-b。另外,基因组的组装以B37核背景下CMS-C为参考序列,可能造成同质异核玉米CMS-C差异。目前已经明确了植物核基因组、线粒体基因组以及叶绿体基因组存在频繁的迁移^[19],核背景的改变必然导致这种迁移的改变。由此可见,不同核背景下CMS-C线粒体基因组差异很小,主要是非编码区以及部分小片段重复序列差异。

3.2 玉米CMS-C同核异质线粒体基因组比较

NCBI上公布的在B37核背景下,CMS-C胞质和正常胞质线粒体基因组大小分别为739 719、569 630 bp。本研究在K169核背景下测序结果分别为739 723、569 617 bp。Allen等^[20]研究认为,0.5~120 kb的DNA片段是造成玉米3类CMS以及正常胞质线粒体基因组大小差异的主要原因,玉米也是最初报道由于大重复序列重组产生可以产生异质性的作物^[21]。本研究发现,两个线粒体基因组小于100 bp的重复序列相差不大,100~1 000 bp和大于1 kb的重复序列在不育系中有12个和7个,但在保持系中只有6个和4个,SSR位点数不育系也远多于保持系。两个基因组比较,其中3个同源片段发生反转,片段排列顺序也发生了重大改变,相似度为95%。由此说明,CMS-C不育胞质和正常胞质线粒体基因组在组成上极为相似,但结构上存在很大差异,大于100 bp的大片段重复序列差异是造成基因组大小差异的主要原因。

高等植物中,同核异质材料线粒体基因组间的差异基因,主要体现在ORFs类型基因组上。高粱不育系Tx623A和保持系Tx623B分别有ORFs 147个和145个,两者共有139个。衡双平^[22]通过芥菜型油菜hau CMS同核异质材料线粒体基因组比较,鉴定出CMS相关基因orf288。曹琼文^[23]通过胡萝卜同核异质材料P2S及P2M的线粒体基因组比较,初选出CMS相关候选基因6个。本研究通过玉米CMS-C同核异质线粒体基因组间所有差异基因(ORFs, ≥ 300 bp)比较,发现K169S与K 169间差异基因(ORFs)高达13个;K169S特有ORFs 8个,K169特有ORFs 14个。结合基因嵌合结构、与功能基因的位置及蛋白跨膜结构域等综合比较分析,筛选出orf429 (*atp6-c*)、orf104-s、orf410、orf349和orf246共5个ORFs,可能与CMS-C败育相关,为进一步阐明玉米CMS-C败育机理提供了基因组信息。

参考文献:

- [1] Forde B G, Leaver C J. Nuclear and cytoplasmic genes controlling synthesis of variant mitochondrial polypeptides in male-sterile maize [J]. *Proc. Natl Acad Sci., USA*, 1980,77(1): 418-422.
- [2] Levings C R, Pring D R. Restriction endonuclease analysis of mitochondrial DNA from normal and Texas cytoplasmic male-sterile maize [J]. *Science*, 1976, 193(4248): 158-160.
- [3] Chen L, Liu Y. Male sterility and fertility restoration in crops [J]. *Annual Review of Plant Biology*, 2014, 65(1): 579-606.
- [4] Schnable P S, Wise R P. The molecular basis of cytoplasmic male sterility and fertility restoration [J]. *Trends in Plant Science*, 1998, 3 (5): 175-180.
- [5] Bentolila S, Alfonso A A, Hanson M R. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants [J]. *Proc. Natl Acad Sci., USA*, 2002, 99(16): 10887-10892.
- [6] Forde B G, Robin J C O, Leaver C J. Variation in mitochondrial translation products associated with Male-Sterile cytoplasm in maize [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1978, 75(8): 3841-3845.
- [7] Banting G S, Glerum D M. Mutational analysis of the *Saccharomyces cerevisiae* cytochrome c oxidase assembly protein Cox11p [J]. *Eukaryot Cell*, 2006, 5(3): 568-578.
- [8] Dewey R E, Timothy D H, Levings C S. A mitochondrial protein associated with cytoplasmic male sterility in the T cytoplasm of maize [J]. *Proc Natl Acad Sci., U S A*, 1987, 84(15): 5374-5378.
- [9] 秦雅婷,赵衍鑫,岳兵,等. CCACA 碱基序列是玉米 CMS-S 线粒体 orf355-orf77 转录本剪切识别位点 [J]. *中国生物化学与分子生物学报*, 2013, 29(6): 555-562.
Qin Y T, Zhao Y X, Yue B, et al. Identification of sequence CCACA as a recognition motif of mitochondrial transcripts Orf355- orf77 cleavage sites in CMS-S maize [J]. *Chinese Journal of Biochemistry and Molecular Biology*, 2013, 29(6): 555-562. (in Chinese)
- [10] Zabala G, Gabay- Laughnan S, Laughnan J R. The nuclear gene *R/3* affects the expression of the mitochondrial chimeric sequence R implicated in S-type male sterility in maize [J]. *Genetics*, 1997, 147(2): 847-860.
- [11] Paillard M S R R. Nucleotide sequence of the S-1 mitochondrial DNA from the S cytoplasm of maize [J]. *EMBO J.*, 1985, 5(4): 125-1128.
- [12] Dewey R E, Timothy D H, Levings R C S. Chimeric mitochondrial genes expressed in the C male-sterile cytoplasm of maize [J]. *Current genetics*, 1991, 20(6): 475-482.
- [13] Zhou G C, Shi H C, Yu X J, et al. Genetic characterisation and cytological identification of a male sterile mutant in maize (*Zea mays* L.) [J]. *Cereal Research Communications*, 2018, 46(2): 344-354.
- [14] Stupar R M, Lilly J W, Town C D, et al. Complex mtDNA constitutes an approximate 620 kb insertion on arabidopsis thaliana chromosome 2: Implication of potential sequencing errors caused by Large-Unit repeats [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2001, 98(9): 5099-5103.
- [15] 张东旭,李润植. 高粱 CMS 材料线粒体基因组 DNA 指纹研究 [J]. *上海交通大学学报(农业科学版)*, 2002(3): 173-177.
Zhang D X, Li R Z. DNA finger prints of mitochondrial genomes in sorghum bicolor CMS lines [J]. *Journal of Shanghai Jiaotong University(Agricultural Science)*, 2002(3): 173-177. (in Chinese)
- [16] 张艳. 甘蓝细胞质雄性不育性对核背景响应的研究 [D]. 西南大学, 2010.
- [17] 王平,丛玲,王春语,等. 高粱 A1 型细胞质雄性不育系与保持系线粒体基因组分析比较 [J]. *生物技术通报*, 2019, 35(5): 42-47.
Wang P, Cong L, Wang C Y, et al. Comparison of mitochondrial genome between A1 cytoplasmic male sterile line and maintainer line of sorghum bicolor [J]. *Biotechnology Bulletin*, 2019, 35(5): 42-47. (in Chinese)
- [18] 邱涛. 玉米同核异质及同质异核雄性不育系线粒体基因组比较分析 [D]. 四川农业大学, 2017.
- [19] 刘英,刘广达,苏小娟,等. 高等植物水平基因转移研究进展 [J]. *植物生理学报*, 2015, 51(7): 985-995.
Liu Y, Liu G D, Su X J, et al. Research progress of horizontal gene transfer between higher plants [J]. *Plant Physiology Journal*, 2015, 51(7): 985-995. (in Chinese)
- [20] Allen J O, Fauron C M, Minx P, et al. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize [J]. *Genetics*, 2007, 177(2): 1173-1192.
- [21] Small I D, Isaac P G, Leaver C J. Stoichiometric differences in DNA molecules containing the *atpA* gene suggest mechanisms for the generation of mitochondrial genome diversity in maize [J]. *Embo J*, 1987, 6(4): 865-869.
- [22] 衡双平. 油菜hau CMS线粒体基因组和不育基因的研究 [D]. 华中农业大学, 2015.
- [23] 曹琼文. 胡萝卜瓣化型细胞质雄性不育系及其保持系的线粒体基因组分析 [D]. 中国农业科学院, 2019.

(责任编辑:朴红梅)