

瑟伯氏棉着丝粒重复序列鉴定

陈晔^{1,2}, 刘玉玲², 韦洋洋², 杨太有¹, 彭仁海^{1,2}

(1.河南师范大学 生命科学学院,河南 新乡 453007;2.安阳工学院 生物与食品工程学院,河南 安阳 455000)

摘要:棉花着丝粒染色质主要由重复序列组成,包括高度重复的着丝粒反转录转座子和卫星重复序列等,但其着丝粒重复序列进化机制尚不清楚。为了深入了解着丝粒重复序列的演化,利用染色质免疫共沉淀(ChIP)和中期染色体荧光原位杂交(FISH)技术对瑟伯氏棉(*Gossypium thurberi*, D₁)着丝粒序列进行了分析。结果表明,相应的 ChIP-DNA 序列属于着丝粒功能区域,发现了 10 个较为富集的着丝粒重复序列。中期染色体 FISH 定位发现,CL179 和 CL234 在四倍体陆地棉中的 A 和 D 染色体组中都有杂交信号,在二倍体 A 组亚洲棉中没有检测到信号,由此可推断在形成四倍体后一些着丝粒特异的重复序列进行了扩增。重复序列 CL95 在瑟伯氏棉以及陆地棉部分染色体的着丝粒区域出现杂交信号,而在二倍体 D 组雷蒙德氏棉和 A 组亚洲棉的着丝粒区域并没有出现信号,可认为该序列是瑟伯氏棉特异的且在四倍体形成后被保留了下来。通过对瑟伯氏棉着丝粒序列及演化的分析能够为揭示棉属植物着丝粒重复序列及染色体组的演化关系提供重要的理论依据。

关键词:瑟伯氏棉;着丝粒重复序列;染色质免疫沉淀;荧光原位杂交

中图分类号:Q756

文献标志码:A

着丝粒是染色体的重要组成部分,在细胞有丝分裂和减数分裂过程中通过与纺锤体微管相互作用来介导染色体分离,维持着生物体的稳定^[1-2]。在大多数真核生物中着丝粒是由高度重复的卫星序列和反转录转座子组成,而且着丝粒的功能是保守的,着丝粒 DNA 序列在物种间有很大的差异,甚至在同一类物种中也存在较大差异^[3-6]。着丝粒特异性组蛋白 H3(CENH3),在哺乳动物中是 CENP-A^[7],是组蛋白 H3 的突变体,也是功能着丝粒染色质的典型特征。研究表明,CENH3 在着丝粒中起着重要作用,而且只存在活性着丝粒中^[8-10]。因此,CENH3 已被用作活性着丝粒的可靠标记。在所研究的物种中,尽管着丝粒区域的卫星 DNA 序列极其多样化,但与 CENH3 结合的 DNA 序列大致是一致的,其单体大小一般为 100~200 bp^[11]。因此,物种间着丝粒卫星的共同特征可能暗示了着丝粒的功能相关性。近年来,已对多种植物的着丝粒进行了深入的研究,水稻^[6]、拟南芥^[12]、玉米^[13]、甘蔗^[14]等植物通过进行染色质免疫共沉淀实验,结果表明所得到的 DNA 序列是着丝粒的功能序列,且着丝粒是由卫星 DNA 序列和反转录转座子组成的。

棉花属锦葵科(*Malvaceae*)棉属(*Gossypium*),是世界上重要的天然纤维作物,包括 A,B,C,D,E,F,G,K8 个二倍体基因组和一个四倍体 AD 基因组^[15]。在四倍体 AD 基因组的 D 亚组供体种来源方面不同研究者存在争议,但更多倾向于多系起源,其中二倍体 D 基因组的三裂棉、瑟伯氏棉和雷蒙德氏棉均可能是陆地棉的祖先种^[15-16]。随着棉花基因组测序的完成,棉花基因组结构得以逐步解析,重复序列含量高达 60%~70%^[17-18]。其中,棉花着丝粒重复序列来源于反转录转座子相关序列并在四倍体演化过程中保存下来并大量扩增^[19-21]。基于此,本研究运用染色质免疫共沉淀(ChIP)和中期染色体荧光原位杂交(FISH)技术对瑟伯氏棉进行着丝粒重复序列的初步鉴定,并进行基因组间比较分析,以期能为棉花着丝粒重复序列演化机制、不同四倍体棉种 D 亚组供体种来源提供支持。

收稿日期:2019-03-18;**修回日期:**2019-10-18.

基金项目:国家重点研发计划(2018YFD0100300);河南省科技创新杰出人才(184200510009);河南省科技攻关(182102110048).

作者简介:陈晔(1993-),女,河南永城人,河南师范大学硕士研究生,研究方向为分子细胞生物学,E-mail:1264149789@qq.com.

通信作者:杨太有,教授,E-mail:13803731625@163.com;彭仁海,教授,博士,E-mail:aydxprh@163.com.

1 材料与方法

1.1 实验材料

本实验材料为二倍体 D 基因组瑟伯氏棉 *G. thurberi* (DD₁)、雷蒙德氏棉 *G. raimondii* (DD₅)、A 基因组亚洲棉 *G. arboretum* (AA₂)、四倍体陆地棉 *G. hirsutum* (AADD₁)、黄褐棉 *G. mustelinum* (AADD₄)，其中瑟伯氏棉的叶片用于 ChIP 实验，其他棉种的根尖用于 FISH 检测。

1.2 染色质免疫共沉淀

染色质免疫共沉淀(ChIP)实验是参照 Nagaki 等^[12]的实验方法并加以优化.主要步骤如下.

将研磨好的棉花叶片粉末,快速加入等体积的 M1(0.1 M Potassium Phosphate 2 mL, 1 M NaCl 2 mL, β -巯基乙醇 16 μ L, Hexylene glycoL 2.37 mL, ddH₂O 定容 20 mL)抽提缓冲液,并用药匙搅匀,埋于冰中 10 min,在摇床上轻轻摇动.用 2 层纱布把样品过滤到一个提前放在冰上的新的离心管中,水平转子 1 200 g, 4 $^{\circ}$ C, 离心 10 min, 去除上清液.加入 10 mL M2(0.1 M Potassium Phosphate 3 mL, 1 M NaCl 3 mL, 1 M MgCl₂ 300 μ L, β -巯基乙醇 23.43 μ L, Hexylene glycoL 3.54 mL, Triton-100 100 μ L, ddH₂O 定容 30 mL), 混匀,水平转子 1 100 g, 4 $^{\circ}$ C, 离心 10 min, 去除上清液,重复操作 3 次,干净即可.加入 8 mL MNB(50%(m/V) Sucrose 2 mL, 1 M Tris-HCl 500 μ L, 1 M MgCl₂ 40 μ L, 1 M CaCl₂ 10 μ L, ddH₂O 定容 10 mL)缓冲液混匀,水平转子 1 100 g, 4 $^{\circ}$ C, 离心 10 min, 去除上清液.用去尖的枪头加入 1.8 mL 的 MNB 缓冲液,轻轻抽吸混匀,避免气泡产生,将样品液体均匀分到 3 个提前预冷的 1.5 mL 的离心管中.利用 MNase 对细胞核进行酶切,37 $^{\circ}$ C 水浴酶切 10 min,快速加入 0.5 M EDTA 终止酶切反应后放置冰上 2 min, 13 000 r \cdot min⁻¹, 4 $^{\circ}$ C 离心 10 min.用去尖枪头将上清液转移到 1.5 mL 的离心管中,取各样品分别进行 DNA 抽提,2%(m/V)琼脂糖胶电泳,100 V, 35 min, 观察拍照.选择最佳酶切浓度样品离心进一步去除杂质,加入 Incubation buffer (1 M NaCl 300 μ L, 1 M Tris-HCl 40 μ L, 0.5 M EDTA 20 μ L, 0.1 M PMSF 4 μ L, Complete Mini 20 μ L, ddH₂O 定容 2 mL)温浴缓冲液,轻轻混匀,加入 3 μ L 的 CENH3 抗体(抗体制备成 1 μ g/ μ L),混匀后用封口膜封住,4 $^{\circ}$ C 摇晃过夜.用 rPAS, Incubation buffer 温浴缓冲液洗涤珠子,颠倒混匀,用封口膜封住,4 $^{\circ}$ C 摇晃 3~4 h.用不同浓度的盐溶液洗去与 beads 结合的非特异性染色质,加 400 μ L 洗提缓冲液(42 $^{\circ}$ C 预热),在 65 $^{\circ}$ C 水浴锅水浴 15 min,每 5 min 颠倒混匀一次,于 13 000 r \cdot min⁻¹, 室温离心 1 min, 静置 1 min, 吸取上清液,重复此次操作,收集合并两次上清,然后进行 ChIP-DNA 的纯化.

1.3 ChIP-DNA 文库的构建

利用 End-itTM 试剂盒对 ChIP-DNA 进行末端修复, AMPure XP beads 磁珠吸附,酒精洗涤, Elution Buffer 洗涤,加 A-base, 37 $^{\circ}$ C 温浴 30 min, 然后加 Adapter 连接, 室温 15 min, AMPure XP beads 磁珠吸附,酒精洗涤, Elution Buffer 洗涤,产物按 98 $^{\circ}$ C 45 S; 5 个循环的 98 $^{\circ}$ C 15 S, 63 $^{\circ}$ C 30 S, 72 $^{\circ}$ C 30 S; 72 $^{\circ}$ C 1 min 的程序进行第一次扩增,用 2%(m/V)的琼脂糖凝胶电泳并用胶回收试剂盒进行产物回收,然后按 98 $^{\circ}$ C 45 S; 13 个循环的 98 $^{\circ}$ C 15 S, 63 $^{\circ}$ C 30 S, 72 $^{\circ}$ C 30 S; 72 $^{\circ}$ C 1 min 的程序进行第二次扩增并用胶回收试剂盒进行产物回收,所得的产物即为 ChIP-DNA 文库,放 -80 $^{\circ}$ C 保存.

1.4 ChIP-Seq 数据分析

建好的文库使用 HiSeq 2500 平台测序.经测序后会产生大量的数据,首先使用 FastUniq 和 Trimmomatic 软件处理 ChIP 和 Input 的序列读取,产生 151 bp 双端测序的 ChIP 及 Input 数据^[22-23].根据相似性的聚类方法,随机选取 5 000 000 条 reads,用 repeatexplorer 软件进行重复序列的预测,生成 Clusters 重复序列,将 CenH3 和 Input 数据比对到 Clusters 重复序列,以确定富集的着丝粒重复序列^[24-25].

1.5 荧光原位杂交

1.5.1 中期染色体的制备

将棉花种子埋于经煮沸消毒的潮湿沙土中,光照培养箱中培养,去主根,以收获大量侧根.取根尖,清水冲洗,25 ppm 放线菌酮室温处理 90 min,以聚集中期分裂象,卡诺固定液($V_{\text{无水乙醇}} : V_{\text{冰乙酸}} = 3 : 1$)固定,4 $^{\circ}$ C 存放备用.取根尖,清水浸泡、冲洗 15 min \times 2,取根尖端部白色生长点部分放入 4%(m/V)的纤维素酶(纤维

素酶 R-10)与 2%(m/V)果胶酶的 60 μL 混合酶解液,37 $^{\circ}\text{C}$ 水浴处理 40 min.吸取根尖 1~2 个,放到干净的载玻片上,滤纸吸干残留水分,加入 10 μL 45%(m/V)的乙酸,用尖头镊子夹碎根尖,盖上盖片,压片,镜检,选择中期分裂象数量与染色体分散度合适的制片,-80 $^{\circ}\text{C}$ 保存备用.

1.5.2 探针标记

8 μL 的 DNA 中加入 2 μL DIG-Nick Translation Mix,标记程序为 15 $^{\circ}\text{C}$,90 min,65 $^{\circ}\text{C}$,10 min 终止反应.标记的探针用琼脂糖凝胶检测,片段大小在 100~200 bp 之间,之后探针放-20 $^{\circ}\text{C}$ 保存备用.

1.5.3 FISH 检测

从-80 $^{\circ}\text{C}$ 取出染色体制片,用刀片迅速从盖片一角撬开,65 $^{\circ}\text{C}$ 烘烤.在染色体制片上加 40 μL 4%的多聚甲醛到载片上,37 $^{\circ}\text{C}$ 或室温下固定 10 min;1 \times PBS 洗 1 min,2 \times SSC 洗 5 min,甩干载片.加 50 μL 70%甲酰胺于玻片上,避光,盖上盖片在 85 $^{\circ}\text{C}$ 杂交炉中变性 4 min.迅速甩掉盖玻片,立即浸入 70%酒精,脱色摇床上 5 min,然后取出置于 100%酒精在脱色摇床上 5 min,取出晾干.配置杂交液(含 ChIP-DNA 探针)置于 100 $^{\circ}\text{C}$ 中变性 10 min,取出置于冰上 5~10 min.加 20 μL 杂交液到片子上,盖上盖玻片于室温放置 5~10 min,再放入 85 $^{\circ}\text{C}$ 杂交炉共变性 2 min,放入保温盒中 37 $^{\circ}\text{C}$ 过夜.将片子取出,用 1 \times PBS 清洗 3 次,每次 5 min.1 \times TNT 清洗 1 \times 5 min,空气干燥玻片.取 1 μL anti-Dig-Rod(抗体)加于 30 μL TNB 中混合均匀,加到玻片上,盖上盖片,37 $^{\circ}\text{C}$ 1 h.取出玻片,1 \times PBS 洗 3 次,每次 5 min,暗室完全晾干.用 DAPI 染色(5 μL 即可),盖上盖片,使用安装有 CCD 摄像头的 ZEISS Axioskop2 mot pLus 的荧光显微镜观察荧光信号拍照,并使用 ZEISS-ISIS 成像系统软件对图像进行采集和加工,使用 Photoshop 7.0 软件对图片处理.

2 结果和分析

2.1 MNase 酶切浓度的检测

选择合适的酶切浓度是进行后续实验的前提,为了选择最佳的酶切浓度,设置了不同梯度的浓度.本次实验设置了 3 个梯度 1.5 U、2.2 U、3 U,分别酶切后取 100 μL 作为样品进行 DNA 的抽提,通过琼脂糖凝胶电泳检测.结果表明,酶浓度在 3 U 时细胞核被充分酶切,主条带在 100~200 bp 之间,且次要条带不明显(图 1),因此,选择浓度为 3U 的样品进行下一步的实验.

2.2 ChIP-DNA 的 FISH 分析

使用 CENH3 抗体进行的染色质免疫沉淀得到的 ChIP-DNA,为了验证该 ChIP-DNA 是否位于着丝粒区域,将 ChIP-DNA 标记探针,以瑟伯氏棉的中期染色体作为靶染色体进行荧光原位杂交.根据杂交的检测结果,在染色体的着丝粒区域产生明显的荧光信号,蓝色信号是染色体,红色是 ChIP-DNA 的荧光信号,而在染色体其他位置没有明显的信号且背景清晰(图 2).因此说明该 ChIP-DNA 位于瑟伯氏棉的着丝粒区,是和 CENH3 特异性结合的 DNA 序列.

另外,为了验证 ChIP-DNA 是否在其他棉种上有信号,以陆地棉、亚洲棉、雷蒙德氏棉以及黄褐棉的根尖为靶染色体分别作了杂交.结果表明,红色标记的 ChIP-DNA 在雷蒙德氏棉的全部染色体着丝粒区,陆地棉和黄褐棉一半的染色体着丝粒区域都产生了荧光信号,而在亚洲棉上没有相应的荧光信号(图 3).

2.3 着丝粒重复序列的分离鉴定与进化分析

将构建的文库利用测序平台进行测序,为了从棉花基因组中分离得到重复序列,所产生的数据使用 repeatexplorer 软件进行序列的聚类分析,生成 clusters 重复序列,将 CENH3 和 Input 数据比对到 clusters 重复序列,以确定富集的着丝粒重复序列.将通过信息学分析得到的序列进行 PCR 扩增并标记为探针在瑟伯氏棉上进行荧光杂交分析(图 4 和表 1).结果在瑟伯氏棉的着丝粒区域看到了荧光信号,且有的探针是产生相对较弱的 FISH 信号,甚至观察不到信号,说明它们在基因组中所占的比例较低.

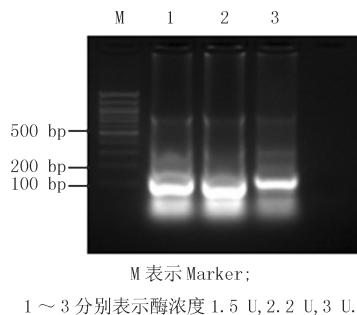
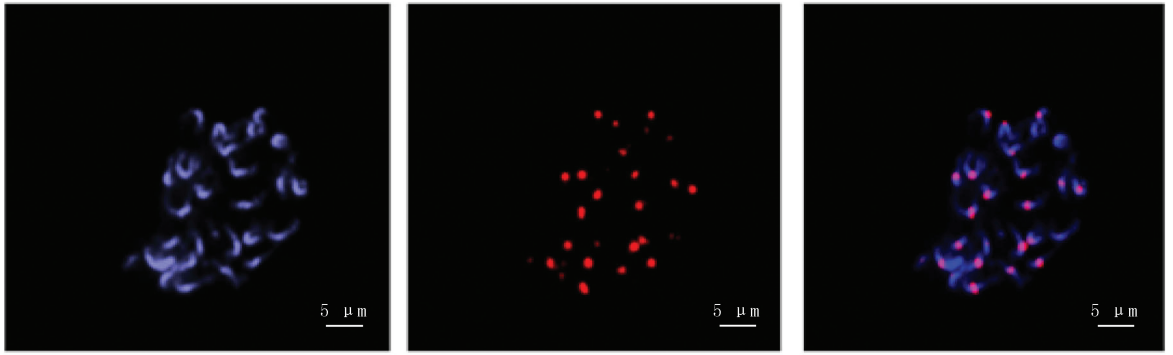


图 1 MNase 酶切浓度检测

Fig1 MNase digestion concentration detection

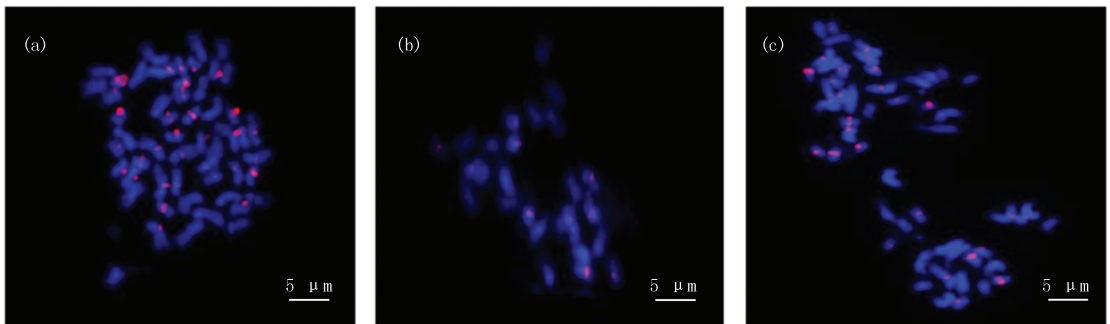


(a) 瑟伯氏棉中期染色体

(b) ChIP-DNA 荧光信号

(c) a 和 b 整合

图 2 瑟伯氏棉中期染色体 ChIP-DNA FISH 检测

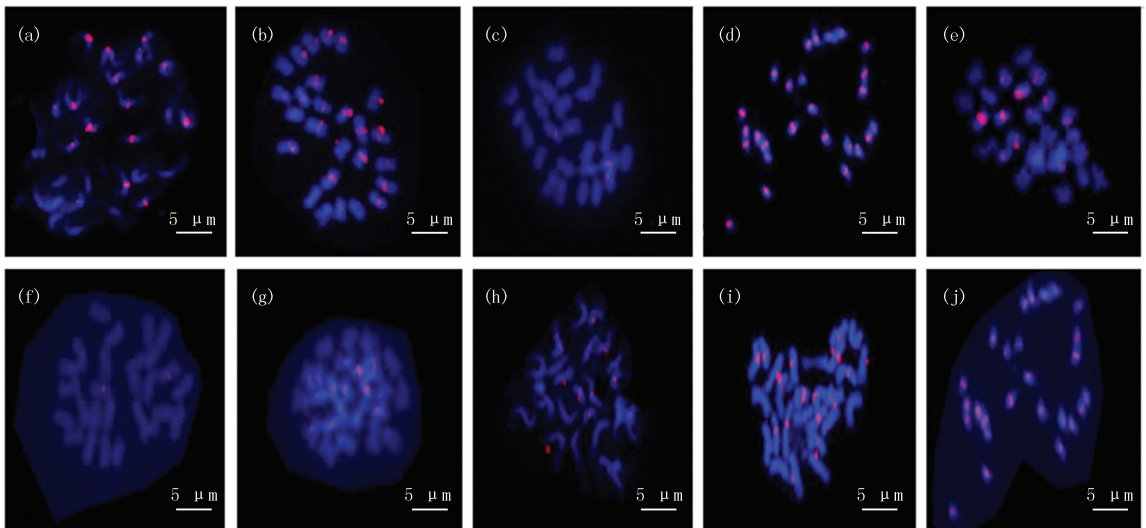
Fig. 2 Chip-DNA FISH detection of mid chromosome of *G. thurberi*

(a) 陆地棉有丝分裂中期染色体为靶 DNA 的杂交图; (b) 雷蒙德氏棉有丝分裂中期染色体为靶 DNA 的杂交图;

(c) 黄褐棉有丝分裂中期染色体为靶 DNA 的杂交图.

图 3 ChIP-DNA 在不同棉种的 FISH 验证

Fig. 3 FISH validation of Chip-DNA in different cotton species



(a) ~ (e) CL179、CL95、CL185、CL23、CL186 5 个重复序列分别在瑟伯氏棉上的杂交图;

(f) ~ (j) CL209、CL42、CL49、CL149、CL234 5 个重复序列分别在瑟伯氏棉上的杂交图.

图 4 不同重复序列在以瑟伯氏棉中期染色体为靶 DNA 的荧光原位杂交图

Fig. 4 FISH detection of different repeats on the mid chromosome of *G. thurberi*

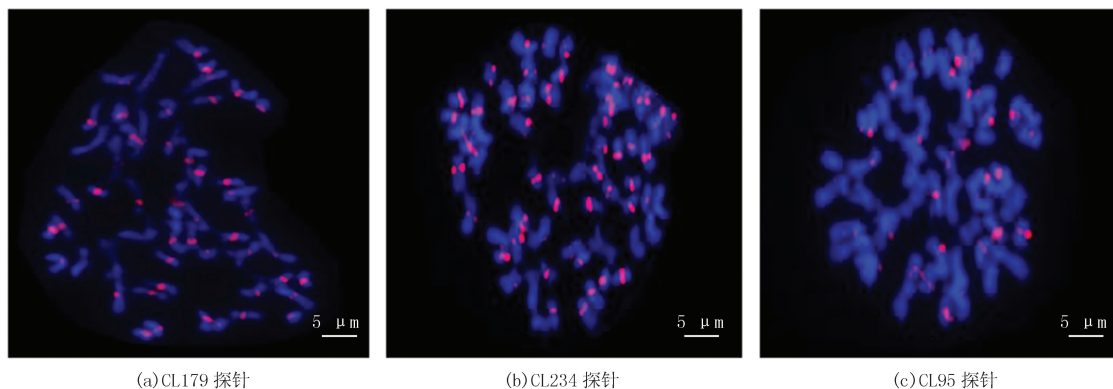
为了表征着丝粒重复序列的演化,将所得着丝粒重复序列与 Han^[19]等文章中的序列进行 blast 序列比对,从中挑选了 3 个特殊的序列,来进一步分析.重复序列 CL179 和 CL234 分别标记探针并在陆地棉上进行 FISH 分析(图 5(a)和图 5(b)).结果显示,这两个重复序列在陆地棉的 AD 染色体上的着丝粒区域都产生

了荧光信号,同时这两个序列在亚洲棉的中期染色体上未检测到信号.由此说明在异源四倍体形成过程中,D亚组着丝粒序列的拷贝数增加并传播到了A基因组的着丝粒区.另外发现,CL95这个重复序列在瑟伯氏棉以及陆地棉部分染色体的着丝粒区域出现杂交信号,而在雷蒙德氏棉和亚洲棉的着丝粒区域并没有出现荧光信号(图4b和图5c).由此现象可推论出该重复序列是瑟伯氏棉特有的且在四倍体演化过程中被保存下来.

表 1 瑟伯氏棉着丝粒序列鉴定

Tab.1 Centromere sequence identification of *G. thurberi*

重复序列	长度/bp	引物序列
CL179	667	F:CTGTCCATTCGGCTATCACA R:AATCAGAATGGAAAGGCTATCG
CL95	1 229	F:TGAAAAAGAGCAAAAAGAGATTGA R:TGGAGATAAGTCAACCACATTGA
CL185	321	F:CGGTTCCGGTATTGCTAGTT R:CATGAAAGTACTAGGAGCGTTGG
CL23	414	F:AAATGGAAATTCGGCCAAG R:TCAAGATTCACCTTCTTGTTTTCAA
CL186	2 863	F:GCGTGAGTGTTAACATTTTCTTTC R:CTTCGGATTTGGCTAATGCT
CL209	553	F:TTCAATGAAGAACCGAAAATTGA R:ATGTGCATGCAGCTCATCT
CL42	1 563	F:TGTGGCCAATGAAAGTGTTCC R:CGGAGCTTCAATTCCAAATC
CL49	1 319	F:AGGGCATGCAAATGACTTTC R:TTCAGCAAAAAGTCCTCTCTCT
CL149	2 145	F:GACCAGAAAAATTTGTGGCTAGTAA R:TCATTCAAACCTCAAAGAGTGGA
CL234	1 160	F:TATTTGGTTGGGCTTCCATC R:TTGTGAGCGAATTTGTGAGG



(a)CL179 探针

(b)CL234 探针

(c)CL95 探针

图 5 不同的重复序列为探针在陆地棉中期染色体上 FISH 验证

Fig.5 FISH validation of different repeat sequences on the mid chromosome of *G. hirsutum*

3 讨 论

ChIP 是近年来在全基因组水平上研究 DNA 和蛋白质相互作用的重要手段,CENH3 在真核生物的功

能着丝粒中,通过 CENH3 抗体进行 ChIP 实验,并分析可得到着丝粒功能 DNA 序列,通过该方法已对多种植物进行了鉴定.而酶切浓度的选择是 ChIP 实验的关键环节.本研究结合已有的方法以及棉花的特征,对瑟伯氏棉进行 ChIP,设置三个酶切梯度,以便于确定最佳酶切浓度,以便后续实验.而基因组和样品量是影响酶切的重要因素,因此不同的材料要根据其自身的要求,来设置合适的酶浓度.

通过 FISH 进一步检测分析所得的 ChIP-DNA 序列是瑟伯氏棉着丝粒功能区域.对 ChIP-DNA 序列在其他几个棉种上进行 FISH 检测,在 D 基因组以及四倍体 D 亚组上的着丝粒区都产生信号,而在 A 亚组以及亚洲棉上没有相应信号,可能表明着丝粒序列在进化的过程中 A 和 D 基因组的同源性很低.全基因组多倍体是植物进化过程中的一个突出的过程,多倍体常常伴随多种基因组的变化,包括反转录转座子等^[26]. Han 等表明在陆地棉 A 和 D 亚组都可见信号的着丝粒重复序列来源于反转录转座子^[19].本次实验也发现类似现象.另外发现,在陆地棉中的着丝粒荧光信号要比在瑟伯氏棉着丝粒信号要亮.由此可能得出:这两序列也来自反转录转座子,并在异源多倍化后进行了大量扩增.CL95 在瑟伯氏棉和陆地棉上均有信号出现,而在雷蒙德和亚洲棉上并没有观察到信号,由此可以推论:该重复序列是瑟伯氏棉特异的且在四倍体演化过程中被保存下来,也为瑟伯氏棉作为四倍体供体种之一提供证据.

棉花的演化机制是否和雷蒙德氏棉有着相似还需要更进一步的研究.本研究对棉花以及其他植物着丝粒相关研究有一定的借鉴意义,同时也为异源四倍体棉及其他植物多倍体的演化模式提供理论基础.

参 考 文 献

- [1] Jiang J M, James A B, Wayne A P, et al. A molecular view of plant centromeres[J]. Trends in Plant Science, 2003, 8(12): 570-575.
- [2] Tatsuo F, William C. Earnshaw. The centromere: chromatin foundation for the Kinetochore Machinery[J]. Developmental Cell, 2014, 30(5): 496-508.
- [3] 夏薇, 李静, 周丽霞, 等. 高等植物着丝粒序列的结构与特征研究[J]. 安徽农业科学, 2014, 42(10): 2859-2862.
- [4] Sun X P, Hiep D L, Janice M, Wahlstrom, et al. Sequence analysis of a functional drosophila centromere[J]. Genome, 2003, 13: 182-194.
- [5] Kamm A, Galassol I, Schmidt T, et al. Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationships between *Arabidopsis* species[J]. Plant Molecular Biology, 1995, 27: 853-862.
- [6] Cheng Z K, Fenggao D, Tim L, et al. Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon[J]. The Plant Cell, 2002, 14: 1691-1704.
- [7] William C E, Naomi R. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma[J]. Chromosoma, 1985, 91: 313-321.
- [8] Kiyotaka N, Kaori T, Munenori W, et al. Centromere targeting of alien CENH3s in Arabidopsis and tobacco cells[J]. Chromosome Research, 2010, 18: 203-211.
- [9] Minh B, Emiliós K D, Christian H, et al. Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo[J]. Cell, 2012, 150: 317-326.
- [10] Hitoshi K, Naoki H, Hiroaki T. Current progress on structural studies of nucleosomes containing histone H3 variants[J]. Elsevier, 2013, 23: 109-115.
- [11] Daniël P M, Keith R B, Hugh A Y, et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution[J]. Genome Biology, 2013, 14: R10.
- [12] Kiyotaka N, Paul B T, Cathy X Z, et al. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres[J]. Genetics, 2003, 163: 1221-1225.
- [13] Zhao H N, Zhu X B, Wan K, et al. Gene expression and chromatin modifications associated with maize centromeres[J]. Genetics, 2015, 6(1): 183-192.
- [14] Zhang W P, Zuo S, Li Z J, et al. Isolation and characterization of centromeric repetitive DNA sequences in *Saccharum spontaneum*[J]. Scientific Reports, 2017, 7: 41659.
- [15] Jonathan F W, Richard C C. Polyploidy and the evolutionary history of cotton[J]. Academic Press, 2003, 78: 139-180.
- [16] 王坤波. 棉属 21 个种基于原位杂交的核型分析[D]. 北京: 中国农业科学院, 2009.
- [17] Li F G, Fan G Y, Lu C R, et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution[J]. Nat. Biotechnol, 2015, 33: 524-530.
- [18] Wang K B, Wang Z W, Li F G, et al. The draft genome of a diploid cotton *Gossypium raimondii*[J]. Nat. Genet, 2012, 44(10): 1098-1103.
- [19] Han J L, Rick E M, Shan W B, et al. Rapid proliferation and nucleolar organizer targeting centromeric retrotransposons in cotton. The Plant Journal, 2016, 88: 992-1005.

- [20] Luo S, Jennifer M, Bradley A, et al. The cotton centromere contains a Ty3-gypsy-like LTR retroelement[J]. *PLoS One*, 2012, 7(4): e35261.
- [21] Zhang W P, Cao Y J, Wang K, et al. Identification of centromeric regions on the linkage map of cotton using centromere-related repeats. *Genomics*, 2014, 104(6 Pt B): 587-593.
- [22] Xu H B, Luo X, Jun Q, et al. FastUniq: A fast de novo duplicates removal tool for paired short reads[J]. *PLoS One*, 2012, 7(12): e52249.
- [23] Anthony M B, Marc L, Bjoern U. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. *Bioinformatics*, 2014, 30: 2114-2120.
- [24] Gong Z Y, Wu Y F, Andrea K, et al. Repeatless and repeatbased centromeres in potato: Implications for centromere evolution[J]. *Plant Cell*, 2012, 24(9): 3559-3574.
- [25] Giovana A T, Gong Z Y, Marina L, et al. Organization and evolution of subtelomeric satellite repeats in the Potato genome[J]. *G3: Genes-Genomes-Genetics*, 2011, 1: 85-92.
- [26] Wendel J F. The wondrous cycles of polyploidy in plants[J]. *Am J Bot*, 2015, 102(11): 1753-1756.

Identification of centromere repeat sequence in *Gossypium thurberi*

Chen Ye^{1,2}, Liu Yuling², Wei Yangyang², Yang Taiyou¹, Peng Renhai^{1,2}

(1. College of Life Sciences, Henan Normal University, Xinxiang 453007, China;

2. College of Biology and Food Engineering, Anyang Institute of Technology, Anyang 455000, China)

Abstract: The centromere chromatin of cotton is mainly composed of repetitive sequences, including highly repetitive centromere reverse transcriptional transposons and satellite repeats, etc. However, the evolutionary mechanism of the centromere repeats is still unclear. To further understand the evolution of centromere repeats, chromatin immunoprecipitation (ChIP) and metaphase chromosome fluorescence in situ hybridization (FISH) techniques were used to analyze the centromere sequences of *Gossypium thurberi* (D1). The results showed that the corresponding ChIP-DNA sequences belong to the functional region of centromere, and 10 relative abundant centromere repeats were found. FISH localization of metaphase chromosomes found that in the A and D sub-genomes chromosomes of tetraploid upland cotton CL179 and CL234 appeared FISH signals, but in *G. arboreum* of diploid A genome no signals were detected. Therefore, it could be inferred that some specific repeat sequences of centromeres were amplified after the formation of tetraploid. Repeated sequence CL95 showed hybrid signals in centromere regions of *G. thurberi* and part of upland cotton chromosomes, but no signals in centromere regions of *G. raimondii* *G. arboreum*, suggesting that the sequence was specific to *G. thurberi* and retained after tetraploid formation. The preliminary identification of *G. thurberi* centromere sequence and its evolutionary process provides a basis for further study of the structure and function of cotton centromere, which is helpful to reveal the evolutionary mechanism of the repeated centromere sequence.

Keywords: *Gossypium thurberi*; centromere repeat sequence; chromatin immunoprecipitation; fluorescence in situ hybridization

[责任编辑 王凤产 杨浦]