Genomics 104 (2014) 587–593

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Identification of centromeric regions on the linkage map of cotton using centromere-related repeats

Wenpan Zhang [1], Yujie Cao [1], Kai Wang, Ting Zhao, Jiedan Chen, Mengqiao Pan, Qiong Wang, Shouli Feng, Wangzhen Guo, Baoliang Zhou *, Tianzhen Zhang *

National Key Laboratory of Crop Genetics and Germplasm Enhancement, Cotton Research Institute, Nanjing Agricultural University, Nanjing 210095, China

ARTICLE INFO

Article history:
Received 12 June 2014
Accepted 7 September 2014
Available online 16 September 2014

Keywords:
Cotton
Centromere
Retrotransposon
Tandem repeat

ABSTRACT

Centromere usually contains high-copy-number retrotransposons and satellite repeats, which are difficult to map, clone and sequence. Currently, very little is known about the centromere in cotton. Here, we sequenced a bacterial artificial chromosome (BAC) mapping to the centromeric region and predicted four long-terminal-repeat (LTR) retrotransposons. They were located in the heterochromatic centromeric regions of all 52 pachytene chromosomes in *Gossypium hirsutum*. Fiber-FISH mapping revealed that these retrotransposons span an area of at least 1.8 Mb in the centromeric region. Comparative analysis showed that these retrotransposons generated similar, strong fluorescent signals in the D progenitor *Gossypium raimondii* but not in the A progenitor *Gossypium herbaceum*, suggesting that the centromere sequence of tetraploid cotton might be derived from the D progenitor. Centromeric regions were anchored on 13 chromosomes of D-genome sequence. Characterization of these centromere-related repeats and regions will enhance cotton centromere mapping, sequencing and evolutionary studies.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

In most eukaryotes, the centromere is an essential chromosomal region required for faithful chromosome segregation. Centromeres in higher eukaryotes typically consist of megabase arrays of satellite repeats and retroelements, both of which are highly diverse. Budding yeast (*Saccharomyces cerevisiae*) has the simplest centromere sequence, consisting of ~125 bp of DNA lacking repetitive elements [1]. Conversely, the centromeres of the fission yeast (*Schizosaccharomyces pombe*) span several hundred kbs and comprise essential core sequences unique to each chromosome flanked by repetitive sequences [2,3]. Highly repeated satellites have been identified in the centromeres of animals and plants [4–9]. However, most centromeric satellite repeats are species-specific or specific to closely related species [6,10].

In contrast to satellites, which are highly diverse, centromeric retroelements are relatively conserved in the plants investigated to date. An intriguing example is the centromeric retrotransposon (CR) family identified in grass. This *Ty3/gypsy*-like retroelement was first found in *Brachypodium sylvaticum* and was then confirmed within the centromeres of all grass species studied and even in the dicotyledonous *Beta* species [11–13]. However, centromeric satellite repeats are most likely derived from retrotransposons [14,15]. The amplification and insertion of retrotransposons may have given rise to the current repeat-based centromere [14].

Molecular linkage maps based on DNA markers provide essential tools for genetic research, such as QTL mapping, marker-assisted breeding and map-based cloning [16]. These maps also provide the backbone for sequence assembly during the whole genome sequencing of complex genomes such as those of wheat and cotton. The precise positioning of centromere determines the linkage relationship and relative positions between markers in or near the centromere region. Although the centromere is one of the most conspicuous markers on the cytological map, the identification of centromeres on genetic maps has been difficult to achieve [17–20]. To date, only a few centromeres have been localized to their genetic maps in plants [21–23]. Due to the high content of repetitive sequences and the low recombination rate in centromere regions, specific markers or genetic stocks have often been required to map centromeres on linkage maps [21–23]. For example, to locate the positions of rice centromeres, secondary and telotrisomics were used to assign restriction fragment length polymorphism markers to specific chromosome arms. Then the centromeres of all 12 linkage maps were mapped using dosage analysis [21]. In *Arabidopsis* and maize, the meiotic mutant line *quartet1* and specific markers derived from the centromere-related transposon were applied, respectively, to map the genetic locations of centromeres [22,23].

Cotton is the leading fiber crop worldwide. However, due to its large genome size (885–2500 Mb) and high repetitive element content (>50%), studies of the cotton genome and centromeres have lagged behind those of other important agronomic crops. In addition,

* Corresponding authors.
E-mail addresses: baoliangzhou@njau.edu.cn (B. Zhou), cotton@njau.edu.cn (T. Zhang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ygeno.2014.09.002
0888-7543/© 2014 Elsevier Inc. All rights reserved.

there is currently no available information about the genetic localization of the cotton centromere. In this study, we sequenced a centromere-related BAC that hybridized to the centromeric region of cotton [24]. Using sequence analysis, we identified four intact LTR retrotransposons. We elucidated their organization and construction using cytogenetic mapping. Based on whole-genome sequence analysis, we identified the centromeric regions and their corresponding SSR markers on the linkage map.

2. Results and discussion

2.1. Identification of centromere retrotransposons in cotton

In a previous study, we found that BAC 97G20 hybridized to the centromeric regions of all 52 chromosomes of *Gossypium hirsutum* (Supplementary Fig. S1A) [24]. Further analyses showed that this BAC probe generated signals in the centromeric regions of all 13 chromosomes of the D subgenome progenitor *Gossypium raimondii* (Supplementary Fig. S1B). We wanted to determine whether there were conserved centromere-related sequences in this BAC. To identify centromere-related sequences, we sequenced the BAC clone. By searching the NCBI databases (BLASTn) and the cotton genome database (CottonDB — http://www.cottondb.org/cgi-bin/ace/search/cottondb; CottonGen — http://www.cottongen.org/), we found several regions that were similar to the centromeric retrotransposons of *Pisum sativum* (GU136552, 68% in 2.2 kb) and rice (DQ458292, 69% in 1.5 kb). Two cotton centromeric retrotransposons, CRG1 and CRG2 (JQ009328 and JQ009329, respectively), showed high similarity (77–99%) to this BAC (see details below). Additionally, no tandem repeat was found in this 93.9-kb fragment.

Because retrotransposons and satellites are the dominant elements in the plant centromere [25], we next focused on identifying retroelements in this BAC. Four retrotransposons, named GhCR1–4 (KF517432–KF517435), were predicted by LTR_FINDER [26], spanning 35.2 kb in this BAC (Table 1). Interestingly, highly similar segments (70–99%) were found when we aligned these GhCRs with other parts of this BAC. In total, approximately 36.3-kb of the sequence showed high sequence identity with GhCRs, indicating that these GhCRs never underwent high frequency transposition events in this region. Sequence analysis by BLASTx showed that GhCR2 and GhCR3 were intact *Ty3/gypsy* retrotransposons that contained well-defined domains of integrase, reverse transcriptase and RNase H. Both of these sequences have highly identical LTRs (99.4% in GhCR2, 98.5% in GhCR3), indicating that they are young retroelements. Searching the cotton expressed sequence tag (EST) database, we found ESTs with high similarity to the LTRs, suggesting that they may be active in cotton. For GhCR1 and GhCR4, no highly similar retroelement was found when we queried NCBI by BLASTx. We then examined these two retrotransposons by GENSCAN [27] and FGENE SH (http://linux1.softberry.com); one predicted gene was found in each retrotransposon. Querying by BLASTp showed that the predicted genes were weakly similar to the plant *gag/pol* domain (identity < 55%), indicating that these LTR retrotransposons underwent rapid degeneration.

Two *Ty3/gypsy*-like retroelements (CRG1 and CRG2) localized in cotton centromeric regions were recently identified [28]; we found that two of our newly identified GhCRs showed high sequence similarity to these two CRGs. Specifically, ~1.4-kb segments shared between CRG1 and GhCR1 showed a sequence identity of 93.6%, while ~2.9-kb segments shared between CRG2 and GhCR3 showed a sequence identity of 94.4%. As expected, all of these highly similar regions are located within the *gag/pol* domain.

2.2. GhCRs show different enrichment levels between the A- and D-genomes in cotton

We then assayed the localizations of these retrotransposons by FISH. Primers were designed according to the LTR regions (Supplementary Table S1), and PCR products were labeled and used as probes in the FISH analyses. The FISH results showed that all four of these retrotransposons were specific to the centromere region in all of the chromosomes of *G. hirsutum*, and the signals in the A genome were much weaker than those in the D genome (Fig. 1 and Supplementary Fig. S2). Bright centromeric signals were also observed in the centromeric regions of the D subgenome progenitor *G. raimondii* (Fig. 1 and Supplementary Fig. S2). However, GhCR2 and GhCR4 only generated weak signals in several centromeric regions of the A subgenome progenitor *Gossypium herbaceum* (Fig. 1). By contrast, GhCR1 and GhCR3 did not generate any signals in *G. herbaceum*. This result was confirmed by searching the current cotton database by BLASTn; no sequence showed high similarity (sequence identity > 80%) with GhCR1 or GhCR3.

Similarly, CRGs were previously found in the tetraploid and D genome diploids but not in A genome diploids [28]. Due to their high sequence similarity, we believe that GhCR1 and GhCR3 belong to the same class as the CRGs. This type of retrotransposon may have originated from the D genome species and invaded the A subgenome after tetraploid cotton formation [29]. By contrast, GhCR2 and GhCR4 might originate before the diploid divergence but they did not proliferate during evolution.

Centromeric chromatin in plants is thought to be highly methylated (DNA and histone) [30]. However, this type of chromatin can exhibit both heterochromatin and euchromatin features on chromosomes [31]. Compared with metaphase chromosomes, pachytene chromosomes present higher resolution (approximately 500 kb) [32]. Importantly, in pachytene chromosomes, heterochromatin and euchromatin can be distinguished by differential DAPI-staining; the heterochromatin region is highly stained but the euchromatin region is difficultly stained [31]. We therefore hybridized the GhCRs onto pachytene chromosomes to examine the chromatin features of these retrotransposons. Like rice, the centromeric regions of cotton pachytene chromosomes exhibited multiple highly stained heterochromatin domains (Fig. 2 and Supplementary Fig. S3). All of these GhCRs colocalized with the bright heterochromatin regions, which was also reported for rice [31].

We also hybridized the CRGs probe onto pachytene chromosomes to determine whether they would localize to the heterochromatin regions. Like the GhCRs, the CRGs also hybridized to the heterochromatic regions (Supplementary Fig. S3).

2.3. GhCRs span less than 1.8 Mb in cotton

To address the organization and construction of GhCRs, we conducted fiber-FISH using the GhCRs as probes. Fiber signals of various sizes

Table 1
LTR retrotransposons in BAC 97G20.

	Strand	Location	Size (bp)	5′-LTR		3′-LTR		LTR region similarity
				Location	Size (bp)	Location	Size (bp)	
GhCR1	+	2469–8126	5658	2469–3916	1448	6674–8126	1453	0.986
GhCR2	+	25,391–33,534	8144	25,391–26,891	1501	32,032–33,534	1503	0.994
GhCR3	+	50,180–66,254	16,075	50,180–50,518	339	65,916–66,254	339	0.985
GhCR4	−	73,857–79,229	5373	73,857–74,729	873	78,356–79,229	874	0.992

Download English Version:

https://daneshyari.com/en/article/2820686

Download Persian Version:

https://daneshyari.com/article/2820686

Daneshyari.com