

RESEARCH

Open Access



# Genomic and cytogenetic analyses reveal satellite repeat signature in allotetraploid okra (*Abelmoschus esculentus*)

Jiarui Liu<sup>1</sup>, Xinyi Lin<sup>1</sup>, Xiaojie Wang<sup>2</sup>, Liqing Feng<sup>3</sup>, Shixin Zhu<sup>1</sup>, Runmeng Tian<sup>1</sup>, Jingping Fang<sup>3</sup>, Aifen Tao<sup>1</sup>, Pingping Fang<sup>1</sup>, Jianmin Qi<sup>1</sup>, Liwu Zhang<sup>1</sup>, Yongji Huang<sup>4\*</sup> and Jiantang Xu<sup>1\*</sup>

## Abstract

**Background** Satellite repeats are one of the most rapidly evolving components in eukaryotic genomes and play vital roles in genome regulation, genome evolution, and speciation. As a consequence, the composition, abundance and chromosome distribution of satellite repeats often exhibit variability across various species, genome, and even individual chromosomes. However, we know little about the satellite repeat evolution in allopolyploid genomes.

**Results** In this study, we investigated the satellite repeat signature in five okra (*Abelmoschus esculentus*) accessions using genomic and cytogenetic methods. In each of the five accessions, we identified eight satellite repeats, which exhibited a significant level of intraspecific conservation. Through fluorescence in situ hybridization (FISH) experiments, we observed that the satellite repeats generated multiple signals and exhibited variations in copy number across chromosomes. Intriguingly, we found that five satellite repeats were interspersed with centromeric retrotransposons, signifying their involvement in centromeric satellite repeat identity. We confirmed subgenome-biased amplification patterns of these satellite repeats through existing genome assemblies or dual-color FISH, indicating their distinct dynamic evolution in the allotetraploid okra subgenome. Moreover, we observed the presence of multiple chromosomes harboring the 35 S rDNA loci, alongside another chromosomal pair carrying the 5 S rDNA loci in okra using FISH assay. Remarkably, the intensity of 35 S rDNA hybridization signals varied among chromosomes, with the signals predominantly localized within regions of relatively weak DAPI staining, associated with GC-rich heterochromatin regions. Finally, we observed a similar localization pattern between 35 S rDNA and three satellite repeats with high GC content and confirmed their origin in the intergenic spacer region of the 35 S rDNA.

**Conclusions** Our findings uncover a unique satellite repeat signature in the allotetraploid okra, contributing to our understanding of the composition, abundance, and chromosomal distribution of satellite repeats in allopolyploid genomes, further enriching our understanding of their evolutionary dynamics in complex allopolyploid genomes.

\*Correspondence:

Yongji Huang  
scolandry@foxmail.com  
Jiantang Xu  
xujiantang@fafu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Keywords** Satellite repeat, rDNA, Intergenic spacer region, Allotetraploid, *Abelmoschus esculentus*, Fluorescence in situ hybridization

## Background

Okra (*Abelmoschus esculentus*), a member of the family *Malvaceae*, is widely cultivated in tropical and subtropical regions worldwide [1]. Due to its rich nutritional elements and significant dietary fiber content, okra has been highly favored as a horticultural crop [2, 3]. Additionally, it is rich in magnesium, iron, polyphenols, and folate, which contribute to its medicinal properties, including antioxidant, anticancer and antidiabetic activities [4, 5]. The formation of *A. esculentus* might have occurred through hybridization between *A. tuberculatus*, an indigenous species from Northern India, and *A. ficulneus*, a species from East Africa. This hybridization event, followed by chromosome doubling, led to the formation of an allopolyploid *Abelmoschus* hybrid [6]. Recent finding from genome assembly of an okra variety ‘Wufu’ has provided evidence supporting its identity as an allotetraploid [7].

Satellite repeats, initially regarded as junk DNA due to their non-coding nature, have increasingly been found to have functional significance [8, 9]. Accumulating evidences suggest that they play crucial roles in heterochromatin domain formation, accurate chromosome segregation, and chromatin-mediated regulation of gene expression [10–12]. Satellite repeats are predominantly abundant in heterochromatin regions of eukaryotic chromosomes, including centromeres, pericentromeres and subtelomeres [11]. They are arranged in long arrays of highly similar head-to-tail tandemly repeated units (monomers). Intriguingly, satellite repeats are one of the most rapidly evolving parts of the genome, resulting in extreme diversity in monomer size, genomic abundance, nucleotide sequences, and chromosomal distribution, even among closely related species [8]. Another significant region of tandem repeat sequences within nuclear DNA is the ribosomal DNA (rDNA). It comprises two distinct gene classes, the major cluster 35 S rDNA and the minor cluster 5 S rDNA. The 35 S rDNA includes genes for 18 S, 5.8 S, and 28 S rRNA, along with two internal transcribed spacers (ITS1 and ITS2) and intergenic spacers (IGS). On the other hand, the 5 S rDNA gene consists of a gene transcription region and a non-transcribed spacer (NTS) [13]. Although rDNA is among the most evolutionarily conserved DNA sequences, the intergenic spacer sequences are less subject to functional constraints and evolve at higher rates [14]. Consequently, these intergenic spacer sequences are commonly employed for phylogenetic inference at low taxonomic levels due to their high substitution rates and low intragenomic sequence heterogeneity.

Allopolyploid species serve as excellent models for investigating the evolutionary trajectory of satellite repeats [15, 16]. Diverse satellite repeat variants may arise from a common ancestral sequence and undergo rapid and dynamic evolution, including the nucleotide sequence variation and genomic abundance change, particularly allopolyploidization, which combines different genomes of distinct progenitor species [17]. For instance, within allotetraploid *Chenopodium quinoa*, several satellite repeat subfamilies have swiftly adapted to become subgenome-specific and have displayed varying genomic abundance patterns [15]. Although the current okra reference genome was released in 2023, satellite repeats have yet to be conclusively identified [7]. Given the existing knowledge gap regarding satellite repeat evolution in the okra genome, the allotetraploid okra presents a potentially attractive for gaining insights into the impact of allotetraploidization on the evolution of satellite repeats in the okra genome.

Over a decade ago, satellite repeats were frequently underrepresented or even absent in genome assemblies due to their highly repetitive nature and identical or nearly identical sequences [18]. However, recent advancements in sequencing technologies, such as PacBio and Nanopore, have enabled the generation of longer reads capable of assembling large satellite arrays, ultimately allowing for the completion of telomere-to-telomere (T2T) genome assemblies [19, 20]. Despite these breakthroughs, T2T genome assembly continues to pose challenges for many species, particularly those with complex genomes. Consequently, to investigate such elusive genomic regions in the absence of T2T genome assemblies, it is crucial to employ appropriate approaches to shed light on the “dark matter” of the genome. A combination of high-throughput next-generation sequencing (NGS) and fluorescence in situ hybridization (FISH) techniques has emerged as a promising method for in-depth analysis of all components of a genome, including the chromosomal locations of satellite repeat families of interest [21]. These techniques have already proven successful in various plant species, such as *Arabidopsis thaliana*, *Brachypodium distachyon*, sugarcane, rice, potato and switchgrass [10, 22–25].

In this study, we conducted a comprehensive analysis to identify and characterize satellite repeats in five okra accessions. Eight satellite repeats were identified in each accession and FISH analysis revealed that these sequences exhibited multiple signals and variations in copy numbers across chromosomes. Interestingly, we discovered that five satellite repeat sequences were

interspersed with centromeric retrotransposons. Additionally, we confirmed the biased amplification pattern within the subgenome and the chromosomal location pattern of both 35 S rDNA and 5 S rDNA. Finally, we confirmed the origin of three satellite repeat sequences with high GC content in the intergenic spacer regions of rDNA. These findings contribute to our understanding of satellite repeats and their evolutionary dynamics in the allopolyploid genome.

## Results

### Genome-wide identification revealed eight satellite repeats in five okra accessions

To identify satellite repeats in five okra accessions, we firstly sequenced their genomes using the Illumina HiSeq X platform, generating paired-end sequencing data ranging from 3.12 to 3.38 Gb in size. Subsequently, we conducted a graph-based sequence similarity clustering analysis using RepeatExplorer 2 on randomly selected 2 M pair-end reads for each accession. Based on the star or ring cluster shape and tandem arrangement, we selected a total of 40 potential satellite repeat clusters in the studies okra accessions for further analysis (Table 1).

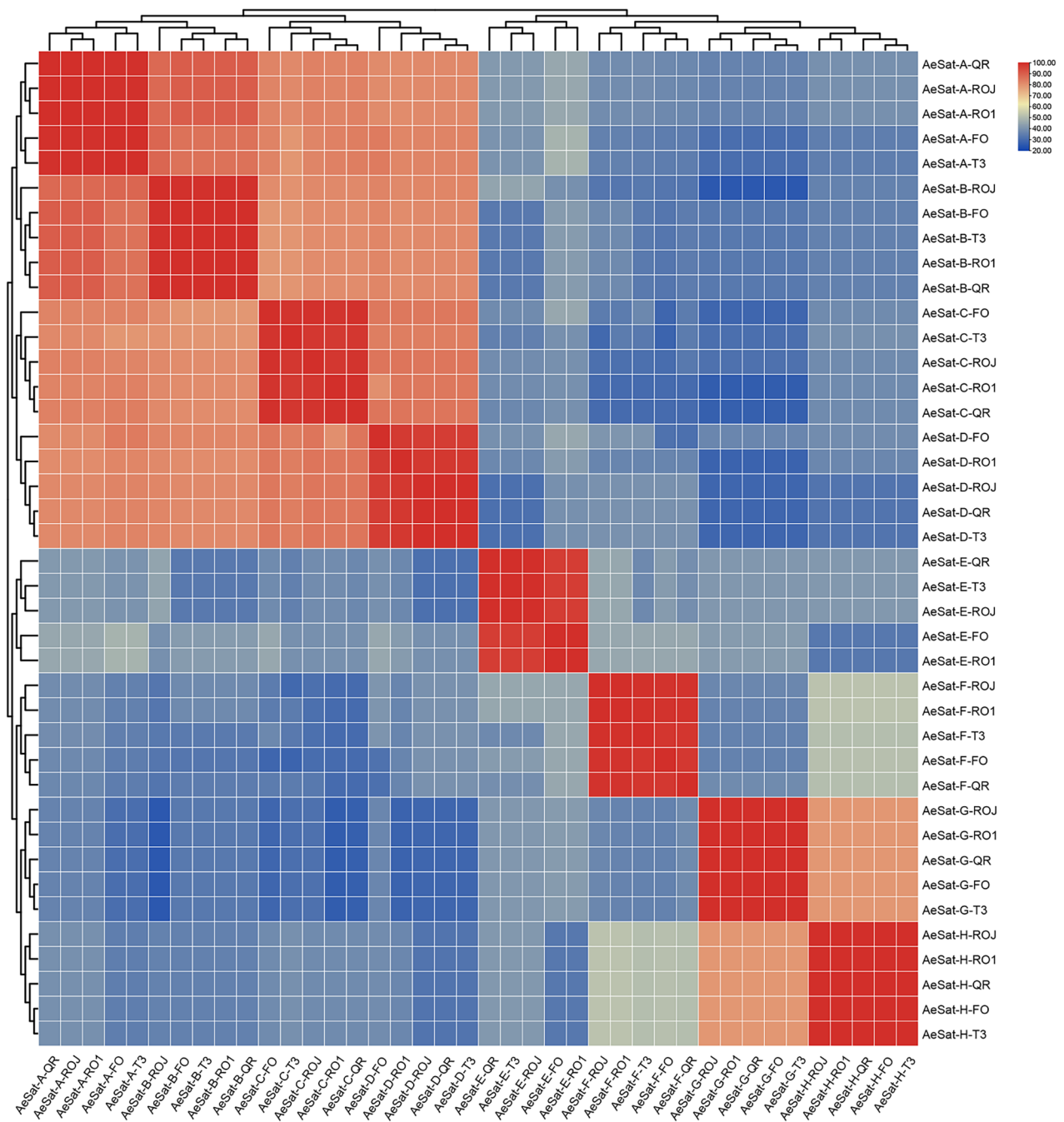
These sequences were classified into eight distinct groups based on their sequence similarities, namely AeSat-A, AeSat-B, AeSat-C, AeSat-D, AeSat-E, AeSat-F, AeSat-G and AeSat-H (Fig. 1 and S1). Within each group, we observed a high level of sequence similarity among the sequences from the five okra accessions, ranging from 97.0 to 100% (Table 1), indicating a significant level of intraspecific conservation. Interestingly, the groups AeSat-A, AeSat-B, AeSat-C and AeSat-D exhibited similar sequence characteristics (Table 1), such as higher sequence similarity (79.1–90.6%), sequence lengths approximating the size of a mononucleosome (169–170 bp), and low GC content (26–32%). These shared sequence features suggest that they have evolved from a common ancestral sequence. However, the genomic abundance of these four groups differed (Table 1). Specifically, AeSat-B had the highest genomic proportion, ranging from 1.3 to 1.5%, followed by AeSat-A (1.10–1.20%) and AeSat-D (0.27–0.31%) as the second and third most abundant, while AeSat-C had the lowest genomic abundance (0.14–0.17%). The monomer in the AeSat-E group had 99 bp in length, GC content ranging from 48 to 52%, and genomic abundance between 0.81% and 1.20% (Table 1). Additionally, the AeSat-F group had the longest monomer length (202 bp), moderate GC content (54%), and genomic abundance ranging from 0.21 to 0.30%. The AeSat-G and AeSat-H groups were GC-rich, with monomer lengths of 136 bp and 43 bp, and genomic abundance ranging from 0.53 to 0.88% for both groups (Table 1).

### Cytological assays confirmed the chromosome distribution pattern of eight satellite repeats across five accessions

To ascertain the chromosome distribution of de novo identified satellite repeats, we hybridized each candidate satellite repeat probe from the eight groups to the metaphase chromosomes of the five okra accessions. FISH analyses showed that all the satellite repeat probes produced multiple signals in all of the studies accessions, confirming their identity as satellite repeats (Figs. 2 and

**Table 1** Satellite repeat characteristics in five okra accessions

Group	Satellite repeat	Monomer length (bp)	GC content (%)	Genome proportion (%)
A	AeSat-A-FO	169	31	1.20
	AeSat-A-T3	169	31	1.20
	AeSat-A-ROJ	169	31	1.10
	AeSat-A-RO1	169	31	1.20
	AeSat-A-QR	169	31	1.20
B	AeSat-B-FO	170	28	1.50
	AeSat-B-T3	170	28	1.40
	AeSat-B-ROJ	170	29	1.30
	AeSat-B-RO1	170	28	1.40
	AeSat-B-QR	170	28	1.40
C	AeSat-C-FO	170	26	0.17
	AeSat-C-T3	170	28	0.14
	AeSat-C-ROJ	170	27	0.14
	AeSat-C-RO1	170	26	0.16
	AeSat-C-QR	170	27	0.17
D	AeSat-D-FO	170	29	0.26
	AeSat-D-T3	170	28	0.30
	AeSat-D-ROJ	170	28	0.27
	AeSat-D-RO1	170	28	0.29
	AeSat-D-QR	170	28	0.31
E	AeSat-E-FO	99	48	0.92
	AeSat-E-T3	99	52	0.81
	AeSat-E-ROJ	99	52	1.20
	AeSat-E-RO1	99	48	0.90
	AeSat-E-QR	99	52	0.91
F	AeSat-F-FO	202	54	0.21
	AeSat-F-T3	202	54	0.26
	AeSat-F-ROJ	202	54	0.30
	AeSat-F-RO1	202	54	0.21
	AeSat-F-QR	202	54	0.27
G	AeSat-G-FO	136	68	0.53
	AeSat-G-T3	136	68	0.56
	AeSat-G-ROJ	136	68	0.88
	AeSat-G-RO1	136	68	0.61
H	AeSat-H-FO	43	67	0.53
	AeSat-H-T3	43	67	0.56
	AeSat-H-ROJ	43	67	0.88
	AeSat-H-RO1	43	67	0.61
	AeSat-H-QR	43	67	0.69



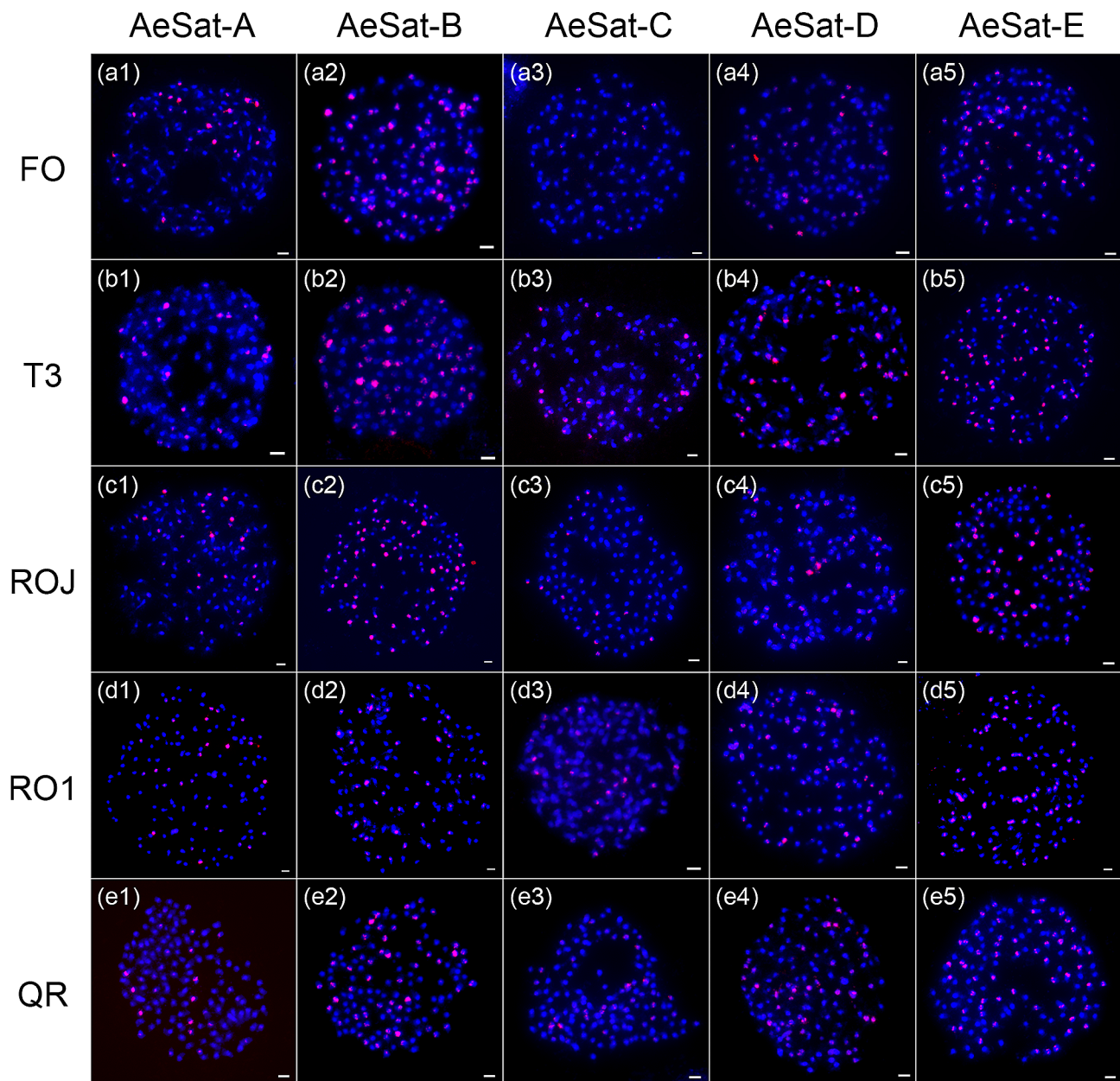
**Fig. 1** A heatmap depicting the sequence similarity of satellite repeats in five okra accessions. Satellite repeats that group together had the highest similarity indexes between them. Color scale: warmer (red) color indicated more similar, while cooler (blue) color indicated more different satellite repeats in the pairs compared

3). Furthermore, we consistently observed clear variations in signal intensities between the chromosomes for these probes (Figs. 2 and 3), which reflect their copy number variations across different chromosomes.

Regarding the signal number, AeSat-B and AeSat-E were detected in 64 chromosomes in all of the studied accessions. For AeSat-A, 29 signal loci were observed in the four accessions, except for the QR accession,

which had 20 signal loci. Polymorphism in the number of AeSat-D signal loci was observed among these accessions, with 27 loci in FO, 53 loci in T3, 42 loci in ROJ, 60 loci in RO1, and 100 loci in QR. The number of signals in AeSat-C varied among the four okra accessions, with 27, 53, 42, and 60 signals for FO, T3, ROJ, and RO1, respectively. However, the number of signals in QR (100 loci) was more than that in the other four accessions (Fig. 2).





**Fig. 2** FISH mapping of satellite repeats AeSat-A, AeSat-B, AeSat-C, AeSat-D and AeSat-E in five okra accessions. DAPI-stained mitotic chromosomes are shown in blue, and the signals of five satellite repeat probes are shown in red. FISH mapping of repeats AeSat-A, AeSat-B, AeSat-C, AeSat-D and AeSat-E in FO (a1-a5), T3 (b1-b5), ROJ (c1-c5), RO1 (d1-d5) and QR (e1-e5). The signal loci of these satellite repeats exhibit variability in five okra accessions. Scale bars: 1  $\mu$ m

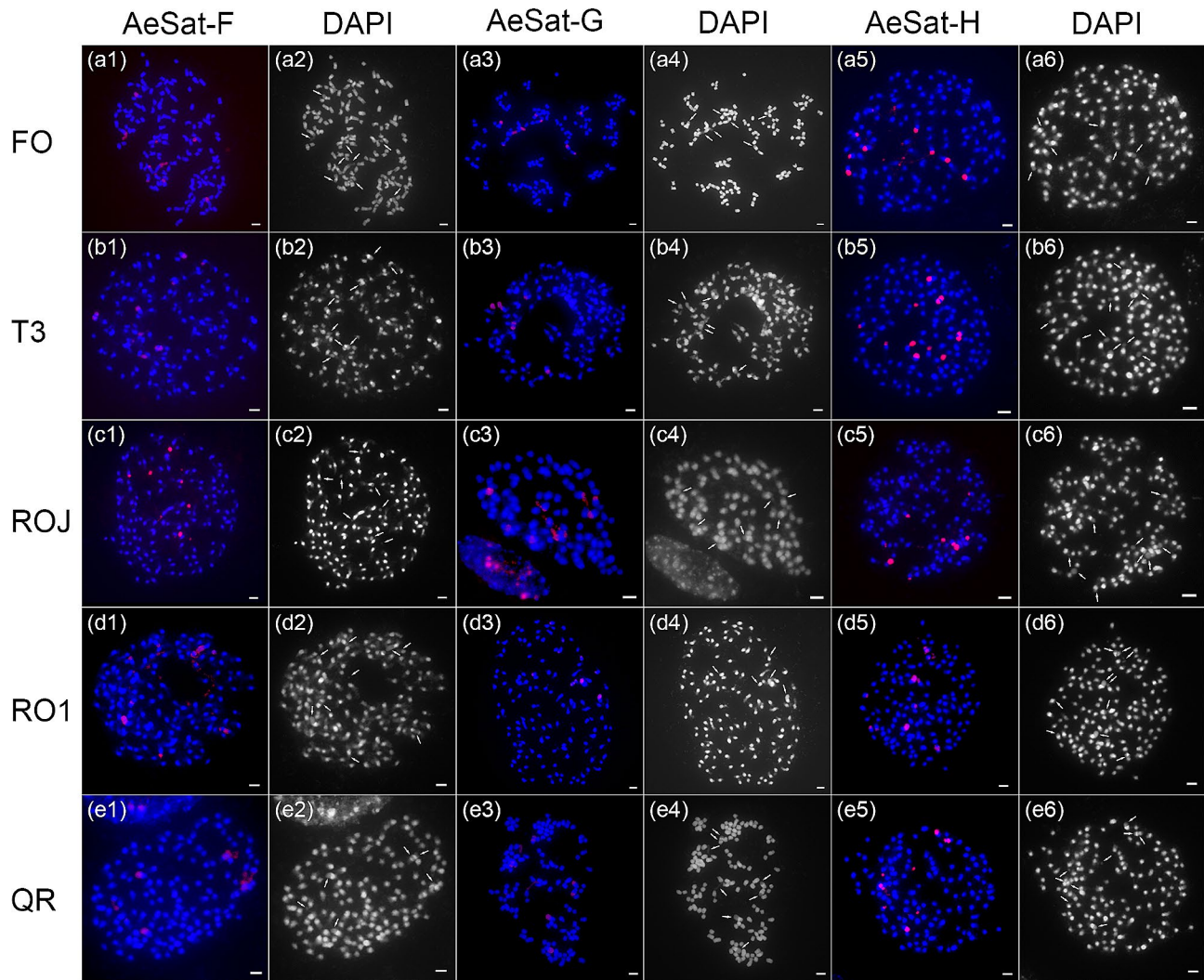
Therefore, despite the sequence conservation observed among various accessions, the signal loci of these satellite repeats display polymorphism.

The satellite repeats, AeSat-E, AeSat-G and AeSat-H, exhibit a similar pattern of multiple signal localization (Fig. 3). Notably, these signals are predominantly observed in regions with relatively weak DAPI staining. DAPI is a fluorescent dye known for its strong affinity to AT-rich regions in DNA, while it exhibits weaker binding to GC-rich heterochromatic regions [26]. This

observation is consistent with the high GC content characteristic of these three satellite repeats.

#### Five centromere-associated satellite repeats exhibit distribution bias within the two subgenomes

Recently, the genome assembly of an allotetraploid okra variety ‘Wufu’ has been publicly released, allowing us to analyze the precise distribution locations of these satellite repeats on chromosomes. Upon alignment with the assembled okra genome, we observed that five distinct

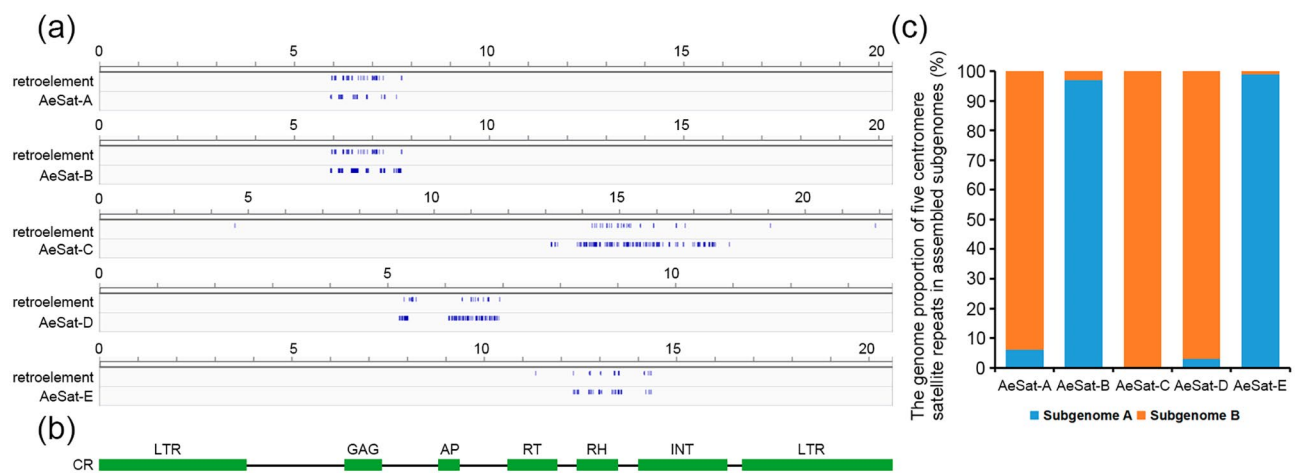


**Fig. 3** FISH mapping of satellite repeats AeSat-F, AeSat-G and AeSat-H in five okra accessions. DAPI-stained mitotic chromosomes are shown in blue and grey. The signals of three satellite repeats are shown in red. FISH mapping of repeats AeSat-F, AeSat-G and AeSat-H in FO (a1–a6), T3 (b1–b6), ROJ (c1–c6), RO1 (d1–d6) and QR (e1–e6). These signals were primarily detected in areas exhibiting low DAPI fluorescence. Scale bars: 1  $\mu$ m

satellite repeats (namely AeSat-A to AeSat-E) were interspersed with a retroelement sequence (Fig. 4a). To assign the retroelement to a specific retroelement subfamily, the retroelement sequence was then annotated to REXdb protein database using BLAST based on the similarity to conserved protein domains database. A query of this retroelement in the protein database disclosed that it bore all the genes required for transposition (Fig. 4b), indicating that it is an autonomous retrotransposon. We unequivocally assigned this okra retrotransposon to the CR-type chromoviruses (Fig. 4b), often marked by a preferential accumulation pattern within plant centromeres, belonging to the Ty3/gypsy chromovirus lineage [19, 20, 23–29]. Hence, this observation indicates their classification as centromeric satellite repeats. Additionally, we have also observed that AeSat-B and AeSat-E predominantly colonize the A subgenome, whereas AeSat-A,

AeSat-C and AeSat-D display a distinct preference for distribution within the B subgenome (Fig. 4c).

To determine the chromosomal distribution patterns of these satellite repeats in five okra accessions, we performed dual-color FISH analysis on their metaphase chromosome spreads. Dual-probe FISH using AeSat-B and AeSat-E revealed the overlapped signals from these two satellite repeats in RO1 (Fig. 5). Similar chromosomal distribution patterns were observed in the other four okra accessions (Figures S2–S5). In contrast, the signals of AeSat-B did not co-localize with those of AeSat-A and AeSat-C except on one pair of chromosomes of all five analyzed accessions (Fig. 5 and S2–S5). Furthermore, AeSat-B signals only co-localized with AeSat-D signals on a few chromosomes (Fig. 5 and S2–S5). These results are consistent with the analysis results of the okra variety ‘Wufu’ genome assembly, suggesting that these sequences



**Fig. 4** Characterization of genomic distributions of five satellite repeats in an allotetraploid okra variety 'Wufu'. **(a)** Diagrams illustrate the distributions of a retroelement sequence and satellite repeats AeSat-A, AeSat-B, AeSat-C, AeSat-D and AeSat-E in assembled genomes. **(b)** The annotation of the retroelement protein domains in REXdb protein database. **(c)** The genome proportion of AeSat-A, AeSat-B, AeSat-C, AeSat-D and AeSat-E in assembled subgenomes

are highly conserved in the chromosome distribution patterns of different okra accessions (Fig. 4b).

#### 35 S rDNA mainly resides in AT-rich heterochromatin regions of both subgenomes, while 5 S rDNA predominantly localizes in subgenome A

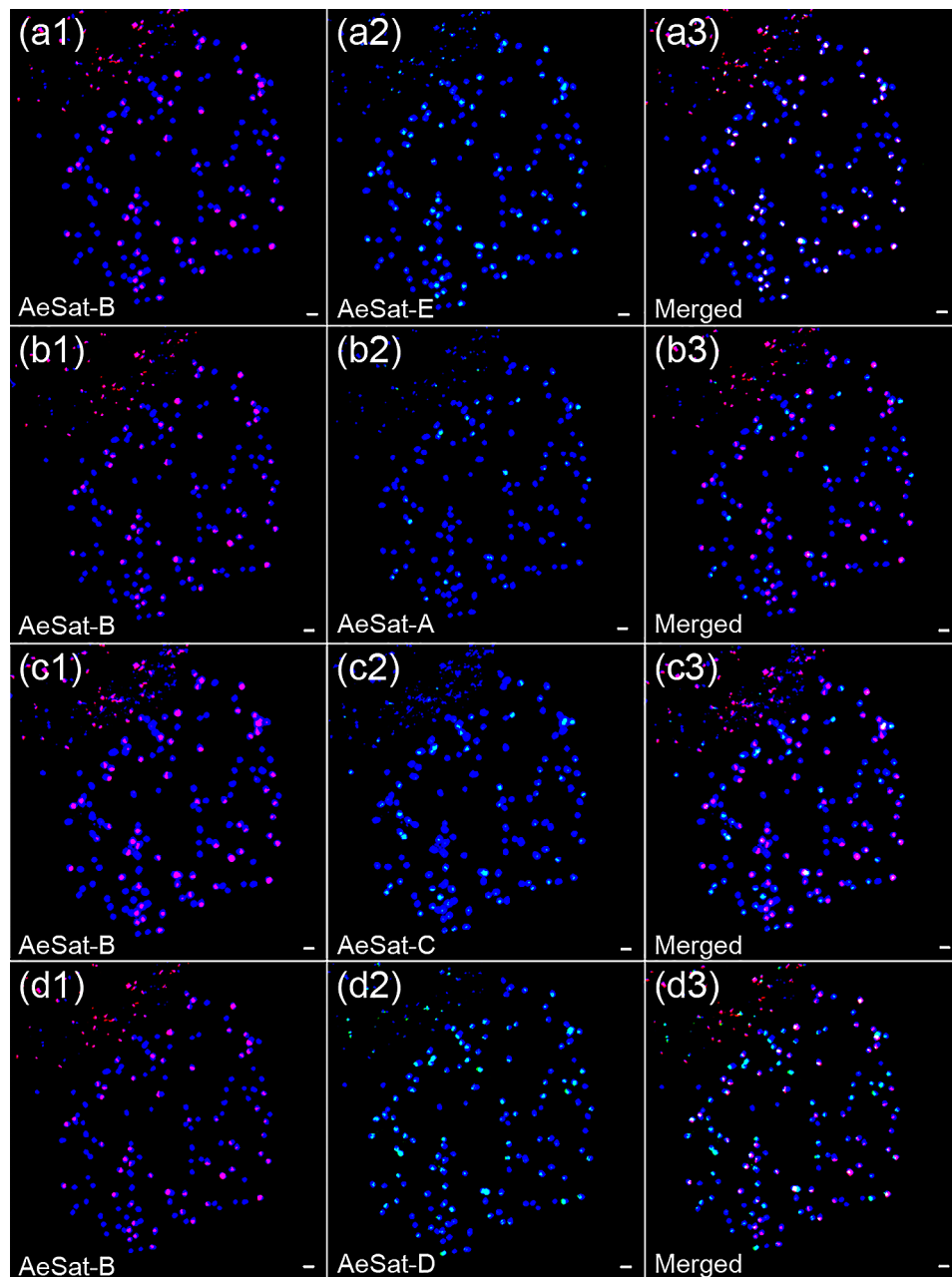
To determine the sequence conservation of the intergenic regions of 5 S rDNA and 35 S rDNA (herein ITS1+5.8 S+ITS2 sequence of 35 S rDNA and the 5 S rDNA+NTS sequence) in five okra accessions, we used DNAMAN8 software to perform multiple sequence alignments on the sequences of the five okra accessions. We found that the coding sequences of 5 S rDNA were highly conserved, including the A-, IE- and C-box regulatory elements (Figure S5). However, a base deletion was found in the NTS sequence of 5 S rDNA in the ROJ germplasm (Figure S6). In the ITS1+5.8 S+ITS2 sequence of 35 S rDNA, we found a variation of C and T bases at position +2 bp of ITS2 (Figure S7). These results indicate that the sequences of these five accessions are highly conserved, with only a few base variations.

By comparing the genomic proportions of 5 S rDNA and 35 S rDNA in NGS data, third-generation sequencing (TGS) data and assembled genomes, we found that all the data showed similar genomic proportions for 5 S rDNA, ranging from 0.13 to 0.49% (Figure S8). In addition, the NGS data showed similar genomic proportions for 35 S rDNA, ranging from 1.98 to 2.87% (Figure S8). However, the genomic proportion of assembled genomes for 35 S rDNA from the TGS and assembled genomes was much lower compared to the NGS data, suggesting 35 S rDNA is under-represented in the recently published okra genome assembly (Figure S8). To check the localization of 35 S and 5 S rDNA loci, we performed FISH analysis

on metaphase chromosomes in five okra accessions using 35 S rDNA and 5 S rDNA probes. In all of the examined accessions, we detected multiple chromosomes containing the 35 S rDNA loci, as well as another chromosomal pair harboring the 5 S rDNA loci (Fig. 6). Interestingly, the intensity of 35 S rDNA hybridization signals varied across the chromosomes, and the signals were mainly located within regions of relatively weak DAPI staining (Fig. 6).

Unlike the 35S rDNA, the 5S rDNA loci seem to be less prone to homogenization in certain allopolyploids, graph clustering of Illumina reads corresponding to 5S rDNA could produce a simple circle in diploid species and more complex structures in allopolyploid genomes [30]. In the graphs, each vertex represents a sequence read and nodes connecting vertices depict sequence similarity between the reads. The loops were composed of vertices depicted in grey representing variable NTS regions in allopolyploid genomes [30]. Using this approach, we observed two loop structures interconnected by a junction region (annotated as the coding sequence of 5S rDNA) in five okra accessions, indicating the presence of their respective copies of 5S rDNA in the okra subgenome (Figure S9). Additionally, we observed varying numbers of grey vertices in the two loops, implying differences in their copy numbers within two subgenomes (Figure S9a-e). After assessing the copy number of 5S rDNA in both subgenomes of the okra variety 'Wufu' genome assembly, we observed that the copy number of 5 S rDNA in subgenome A (26,257) greatly exceeds that in subgenome B (546) (Figure S9f). Hence, in conjunction with the FISH results, we confirmed that 5 S rDNA predominantly localizes in subgenome A.



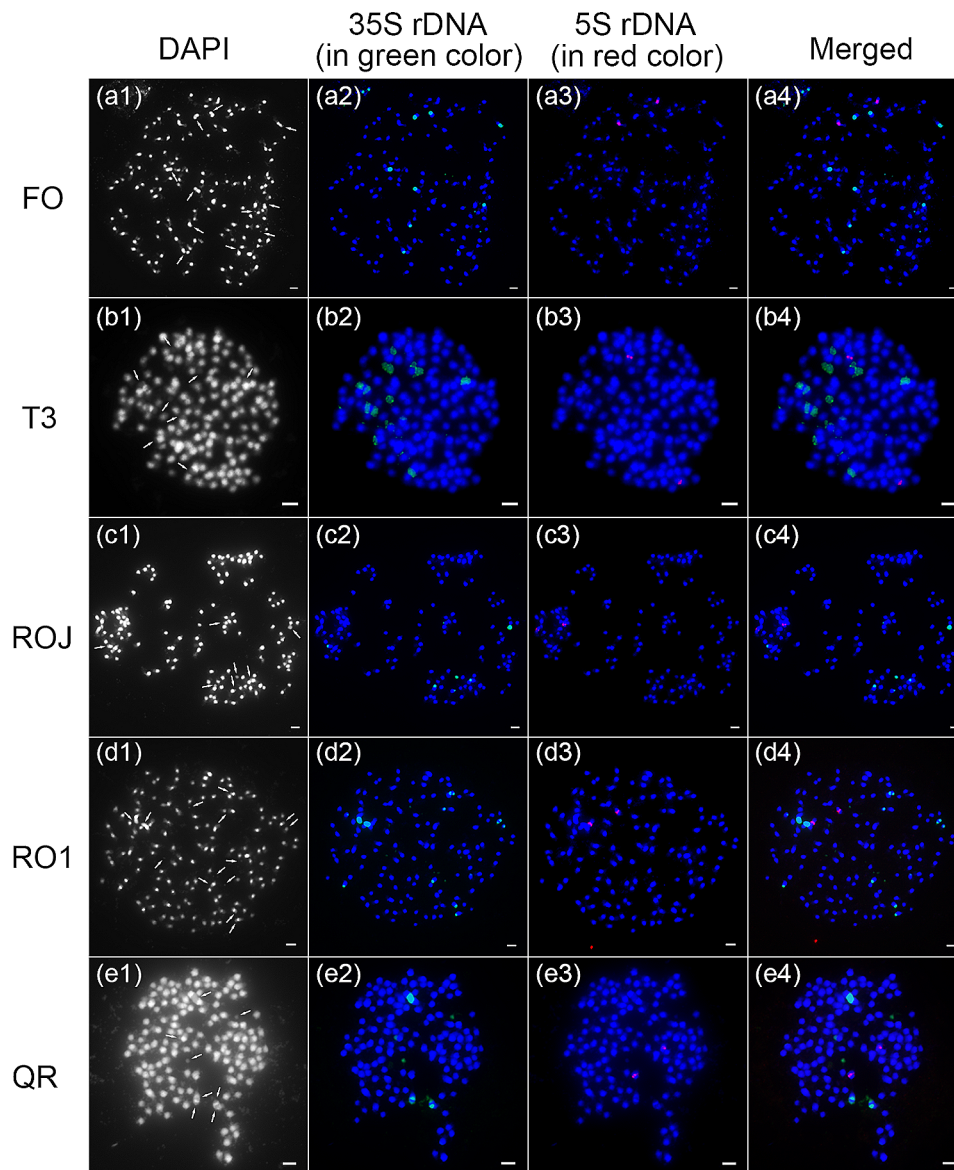


**Fig. 5** FISH localization of five centromeric satellite repeats on mitotic metaphase chromosomes of RO1 accession. **(a)** Dual-color FISH of AeSat-B (red) and AeSat-E (green), showing colocalization of these two satellite repeats on the chromosomes of RO1; **(b)** Dual-color FISH of AeSat-B (red) and AeSat-A (green), showing non-colocalization of these two satellite repeats on the chromosomes of RO1; **(c)** Dual-color FISH of AeSat-B (red) and AeSat-C (green), showing non-colocalization of these two satellite repeats except on one pair of chromosomes of RO1; **(d)** Dual-color FISH of AeSat-B (red) and AeSat-D (green), showing colocalization of these two satellite repeats on a few chromosomes of RO1. Thus, five centromeric satellite repeats exhibit distribution bias within the two subgenomes in RO1 accession. Scale bars: 1  $\mu$ m

### Three satellite repeats with high GC content co-localized with 35 S rDNA

Based on the aforementioned FISH mapping experiments, we observed a consistent localization pattern for 35 S rDNA, AeSat-F, AeSat-G and AeSat-H. These sequences exhibited a preference for localizing within regions characterized by relatively weak DAPI staining. To validate the co-localization of these sequences,

we conducted co-localization analyses using the 35 S rDNA probe in conjunction with the AeSat-E, AeSat-G, and AeSat-H satellite repeat probes on the same metaphase chromosome cells. FISH results demonstrated the co-localization of the satellite repeats from these three probes with 35 S rDNA at the same chromosomal locus (Fig. 7 and S10-11). Subsequently, we conducted sequence alignments of these three satellite repeats with



**Fig. 6** FISH localization of 5 S rDNA and 35 S rDNA families on mitotic metaphase chromosomes of five okra accessions. DAPI-stained mitotic chromosomes are shown in blue and grey. The signals of 35 S rDNA and 5 S rDNA are shown in green and red, respectively. The arrowheads point to the regions of relatively weak DAPI staining. (a1–a4) FISH mapping of 35 S rDNA and 5 S rDNA on the chromosomes of FO. (b1–b4) FISH mapping of 35 S rDNA and 5 S rDNA on the chromosomes of T3. (c1–c4) FISH mapping of 35 S rDNA and 5 S rDNA on the chromosomes of ROJ. (d1–d4) FISH mapping of 35 S rDNA and 5 S rDNA on the chromosomes of RO1. (e1–e4) FISH mapping of 35 S rDNA and 5 S rDNA on the chromosomes of QR. In the five okra accessions, we identified numerous chromosomes containing the 35 S rDNA loci, along with an additional pair of chromosomes containing the 5 S rDNA loci. Scale bars: 1  $\mu$ m

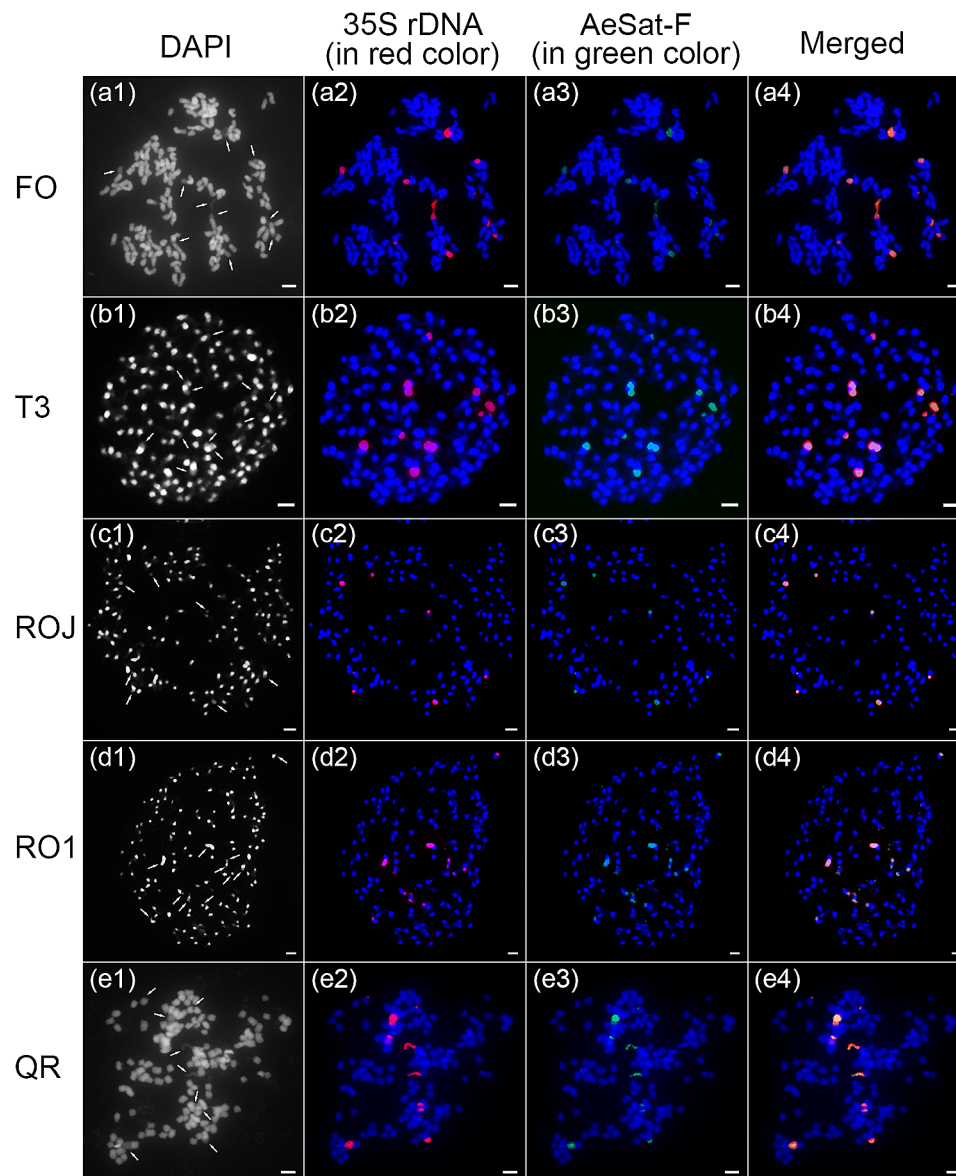
the assembled genome sequence of okra. Remarkably, all three sequences were found to be situated within the IGS sequence of the 35 S rDNA (Fig. 8), and they all exhibited high sequence similarity with the IGS sequences of two okra subgenomes (average sequence similarity ranging from 86.8 to 97.8%). Specifically, AeSat-F and AeSat-H showed higher average sequence similarity with those of subgenome A, at 97.8% and 96.0% respectively, while AeSat-G exhibited higher average sequence similarity with those of subgenome B, at 91.1% (Table 2). Therefore, our findings confirmed that AeSat-F, AeSat-G and

AeSat-H were satellite repeats derived from the IGS region of the 35 S rDNA.

### Discussion

In higher plants, centromeric sequences are generally composed of satellite repeats and centromeric retrotransposons (CRs), as demonstrated in sugarcane, rice, maize, and *B. distachyon* [19, 20, 23–29]. In our study, we discovered that five satellite repeats were consistently interspersed with the CRs lineage in okra (Fig. 4), indicating their association with centromeric satellite repeat

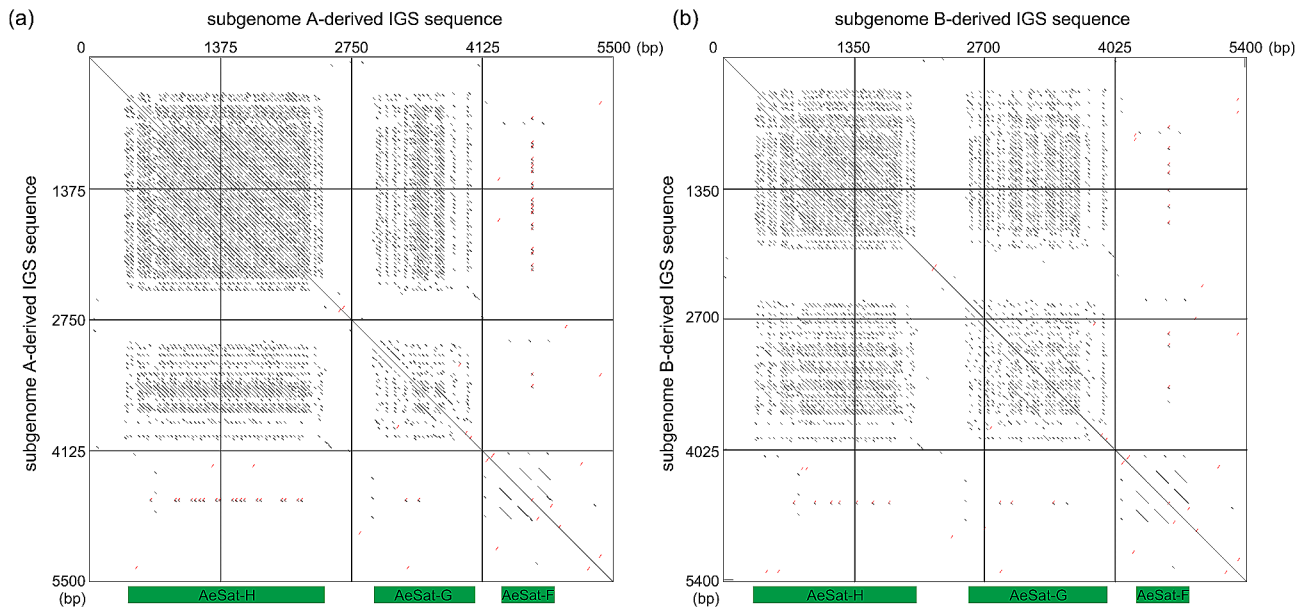




**Fig. 7** FISH mapping of 35 S rDNA and AeSat-F on the metaphase spreads of the five okra accessions. DAPI-stained mitotic chromosomes are shown in blue and grey. The signals of 35 S rDNA and AeSat-F are shown in red and green, respectively. The arrowheads point to the regions of relatively weak DAPI staining. (a1-a4) FISH mapping of 35 S rDNA and AeSat-F in FO. (b1-b4) FISH mapping of 35 S rDNA and AeSat-F in T3. (c1-c4) FISH mapping of 35 S rDNA and AeSat-F in ROJ. (d1-d4) FISH mapping of 35 S rDNA and AeSat-F in RO1. (e1-e4) FISH mapping of 35 S rDNA and AeSat-F in QR. AeSat-F exhibited a propensity for localizing in regions with relatively low DAPI staining, and it co-localized with the 35 S rDNA at the same chromosomal location. Scale bars: 1  $\mu$ m

identity in okra. Additionally, the abundant centromeric satellite repeats in higher plants form the most dominant tandem repeat families in the genomes. Similar characteristic has also been observed in okra, where five centromeric satellite repeats account for approximately 2.5% of the genome (Table 1). Intriguingly, in many higher plants, centromeres are typically characterized by the prevalence of a specific family of satellite repeats with considerable sequence diversity [31]. For instance, the centromeres in the model plant *A. thaliana* harbors several megabases of satellite repeat variants with a 79–89% sequence identity

[32]. Similarly, we found that four out of these five centromeric satellite repeats in okra (AeSat-A, AeSat-B, AeSat-C and AeSat-D) exhibit a high degree of similarity (79.4–90.6%, Fig. 1 and S1). Moreover, the centromeric satellite repeat AeSat-E exhibits a low sequence similarity of 31.3–48.2% with these four centromeric satellite repeats, indicating that the centromeric satellite repeat sequences originate from two distinct families in okra (Fig. 1 and S1). Satellite repeats undergo rapid changes during adaptation to new environments, which contribute to swift genome evolution and speciation



**Fig. 8** Dot-plot analysis of the IGS sequence of the 35 S rDNA from the assembled genome sequence of okra. Dot-plot representation illustrates the pairwise similarities between sub-repeats in the IGS sequence of the 35 S rDNA, indicating that three satellite repeats, namely AeSat-F, AeSat-G and AeSat-H, were derived from sub-repeats within the IGS sequence of subgenome A (a) and sB (b)

**Table 2** The average sequence similarity of the two subgenomic IGS sequences with AeSat-F, AeSat-G, and AeSat-H

	subgenome A-derived IGS	subgenome B-derived IGS
AeSat-F	97.8%	95.4%
AeSat-G	86.8%	91.1%
AeSat-H	96.0%	95.4%

[15–17]. In okra, there is a bias towards the predominance of centromeric repeats in a specific subgenome (Fig. 4). These findings provide novel insights into the sequence variation of centromeric repeats following polyploidization and asymmetrical evolution of centromeric repeats in allotetraploid okra subgenomes.

In most eukaryotes, centromeric satellite repeats are typically 150 to 180 bp in length, a length sufficient for wrapping around a single nucleosome [22, 33]. For example, CENP-A nucleosomes in *Homo sapiens* and CENH3 nucleosomes in *Oryza sativa* are highly phased on ~181 bp alpha satellite repeat and ~178 bp AthCEN178 satellite repeat, respectively [34, 35]. However, in some organisms like potato and chicken, centromeres can have multiple satellite arrays with varying unit sizes, including both repeat-containing and repeat-free sequences [24, 36]. In our study, our findings revealed four centromeric satellite repeats (AeSat-A, AeSat-B, AeSat-C and AeSat-D) have a near mono-nucleosomal length of 169–170 bp (Table 1). The remarkable consistency in length of the okra satellite repeats suggests the presence of a stringent length requirement. These observations suggest that these centromeric repeats may have evolved into a mature, centromere-adaptive structure in okra. However,

one centromeric satellite repeat, AeSat-E, has a significantly different mono-nucleosomal length with 99 bp (Table 1). Similar to other plant species, satellite repeats with atypical monomeric lengths are likely unsuitable for centromeres [24].

Compared to the highly conserved coding sequences of 35 S rDNA and 5 S rDNA, the intergenic spacer sequences (ITS1 and ITS2 from 35 S rDNA, NTS from 5 S rDNA) exhibit more variability. Therefore, these intergenic spacer sequences serve as useful genetic markers for understanding interspecific and even intraspecific relationships [37, 38]. Traditionally, rDNA is thought to be evolved in a strict concerted evolution pattern, with little to no variation within species, but sequence diversity between species or higher taxa [39]. In the case of okra, there is limited intraspecific variability among five accessions, with only minor variations detected in the ITS2 and NTS region sequences (Figures S6 and S7). A cytogenetic study across 2,949 land plant species demonstrated that most species possess more 35 S rDNA loci than 5 S rDNA loci (47.79% of the investigated species). Conversely, for 19.05% of the species, the opposite pattern is observed; while the remaining 33.16% of species possess an equal number of loci [40]. Furthermore, there is a prevalent trend of an increase in the number of rDNA loci with ploidy level [40]. Nevertheless, in some species like *Artemisia* and *Nicotiana*, there is a potential decrease in the number of rDNA loci per monoploid genome following polyploidization [41–43]. This finding provides evidence for the occurrence of diploidization processes in polyploid plants over millions of years. In

our study, we detected multiple 35 S rDNA loci (exceeding the expected ploidy level) and only two 5 S rDNA loci (less than the expected ploidy level) in five okra accessions using the FISH assay (Fig. 6), indicating a highly dynamic change in the number of rDNA loci during okra allopolyploidization. 35 S rDNA is fragile sites associated with epigenetic alterations and DNA damage under stress conditions, which could be one of the factors contributing to the significant variation in 35 S rDNA loci following polyploidization [44]. Furthermore, our examination of the clustering results of 5 S rDNA revealed a double ring configuration, implying the loss of 5 S rDNA from one okra subgenome after allopolyploidization (Figure S9). In chromosomes, the occurrence patterns of 5 S rDNA can be categorized into two types: the S-type arrangement, featuring long tandem repeats of regularly spaced units, and the L-type arrangement, characterized by solitary insertions are present within the intergenic spacer of the 35 S rDNA. The S-type arrangement is predominant in angiosperms, while the L-type arrangement is more characteristic of plant lineages that diverged early in angiosperm evolution and certain taxa within gymnosperms [40]. Our FISH assay revealed that 5 S rDNA are separated from 35 S rDNA in okra, thus suggesting its S-type arrangement, similar to most angiosperms.

In addition, we also found that three satellite DNAs in okra originated from the distinct IGS sub-repeats of 35 S rDNA (Fig. 8). Remarkably, several other plants, including potatoes, tomatoes, tobacco and pepper, were found to have highly amplified satellite DNAs that share sequence homology with the IGS sub-repeats of 35 S rDNA [42, 45–47]. However, in *Capsicum frutescens*, these satellite DNAs are even distributed in dispersed patterns across multiple chromosomes [42], whereas this pattern has not been observed in okra (Fig. 7). Furthermore, in both the 35 S rDNA and IGS regions of okra, we observed that they were predominantly located in DAPI-negative staining regions, suggesting the presence of GC-rich heterochromatin and the scarcity of AT-rich DNA in these regions (Figs. 3, 6 and 7, S10 and S11). A similar phenomenon has been documented in other species, such as *Rhodeus amarus* [48].

TGS technologies greatly improve the accuracy of genome assemblies, as they are capable of accurately reading and capturing complete DNA fragments of 10 kilobases or longer. This allows for the coverage of satellite repeats [18]. However, the strategies employed by these technologies are applicable to shorter arrays of less complex repeats that traverse several kilobases. In the case of okra, we identified arrays of the centromeric satellite repeats (AeSat-A, AeSat-B, AeSat-C, AeSat-D and AeSat-E) and 5 S rDNA through FISH analysis (Figs. 2 and 6), which were accurately assembled on the okra genome assembly. Nevertheless, accurately assembling

genomes is still a challenge due to the presence of complex repetitive regions, such as arrays that span hundreds to thousands of kilobases and are comprised of units that are several kilobases in length, including 35 S rDNA. For instance, the recently published near T2T genome of *A. thaliana* still contains inaccuracies specifically related to highly repeated DNA sequences including the short-arm 35 S rDNA clusters and adjacent telomeres on chromosomes 2 and 4 [20]. In our study, we also noticed that they are under-represented in the recently published okra genome assembly (Figure S8), possibly due to the aforementioned challenges. Additionally, the presence of three complex satellite sequences on the IGS, along with their highly GC-rich content, further exacerbates the difficulty in its genome assembly.

## Conclusion

In this study, we examined the satellite repeat patterns in five okra accessions. We used both genomic and cytogenetic methods to identify and analyze these satellite repeats. We discovered that each of the five accessions had eight distinct satellite repeats, which showed a significant degree of similarity among these accessions. By conducting FISH experiments, we observed that these satellite repeats were present in multiple loci and had variations in their distribution across the chromosomes. Interestingly, we found that five of these satellite repeats were interspersed with retrotransposons located near the centromeres, indicating their potential involvement in defining the identity of satellite repeats in the centromeric region. Through the analysis of existing genome assemblies and dual-color FISH, we confirmed that the amplification patterns of these satellite repeats were biased towards specific subgenomes, suggesting distinct evolutionary dynamics within the allopolyploid subgenome. Additionally, using FISH assay, we observed the presence of multiple chromosomes containing the 35 S rDNA loci, as well as another pair of chromosomes carrying the 5 S rDNA loci in okra. Notably, the intensity of the 35 S rDNA hybridization signals varied across the chromosomes, with all signals localized within regions of relatively weak DAPI staining, which are associated with GC-rich heterochromatin. Lastly, we discovered a similar localization pattern between the 35 S rDNA and three satellite repeats possessing a high GC content, confirming their origin in the intergenic spacer region of the 35 S rDNA. Overall, our study provides new insights into the unique satellite repeat patterns in okra, shedding light on their composition, distribution, and evolutionary dynamics in complex allopolyploid genomes.

## Methods

### Plant material and genomic DNA isolation

Five okra accessions, including Fruit Okra (FO), Titanic No.3 (T3), Red Okra JPN (ROJ), Red Okra No.1 (RO1) and QK-RP-2-1 (QR), were used in this study. The phenotypic traits, such as fruit color, plant height, fruit number, and fruit shape, of these five okra accessions from diverse geographical regions manifest substantial variations. In our okra hybrid breeding program, the successful production of the F<sub>1</sub> generation through the crossbreeding of five phenotypically distinct accessions indicated their potential closely related genetic relationships. Consequently, we aim to conduct satellite sequence analysis to investigate the evolutionary patterns of satellite sequences in these five allopolyploid okra genomes. These accessions were cultivated in a glasshouse located at Minjiang University in Fujian, China. To obtain DNA samples, fresh leaves were collected from each of the five accessions and subjected to DNA extraction using the CTAB method.

### Genome-wide identification of repetitive DNA

Identification of repetitive DNA was achieved using a graph-based sequence similarity clustering analysis from randomly selected 2 M pair-end reads for each accession as described by Novak et al. [21]. Initially, the genomic DNA from the five accessions was subjected to low coverage sequencing using the Illumina HiSeq platform. Quality assessment and filtering of low-quality reads were performed using FastQC and Trimmomatic v0.36. To conduct graph-based clustering analysis, a set of two million randomly selected reads was analyzed using the Galaxy RepeatExplorer2 web platform (<https://repeatexplorer-elixir.cerit-sc.cz/galaxy/>). Reads with graph nodes showing similar distribution patterns were categorized into the same repeat sequence families. For subsequent analyses, the representative sequences from each satellite sequence family were extracted based on the presence of tandem sub-repeats within their read or assembled contig sequences.

### Sequence analysis and alignment of 5 S rDNA and 35 S rDNA

Using the TAREAN tool in the Galaxy RepeatExplorer2 web platform (<https://repeatexplorer-elixir.cerit-sc.cz/galaxy/>), repeat sequence clusters of 5 S rDNA (5 S rDNA+NTS) and 35 S rDNA (ITS1+5.8 S+ITS2) were extracted from the NGS data of five okra varieties. From these two clusters, representative contigs were selected for sequence analysis and alignment. Finally, sequence analysis and alignment of 5 S rDNA (5 S rDNA+NTS) and 35 S rDNA (ITS1+5.8 S+ITS2) from the five okra varieties were performed using the default parameters of the DNAMAN 8 software.

### PCR amplification and probe preparation

PCR amplification was conducted using fusion PCR for the full-length satellite repeats (Table S2). PCR amplification was performed in a 20- $\mu$ L volume containing 100 ng genomic DNA, 200  $\mu$ M dNTPs, 1 $\times$  ExTaq Buffer, 250 nM primers, and 1 U ExTaq DNA polymerase (TaKaRa Bio, Kusatsu, Shiga, Japan). The PCR conditions were as follows: an initial denaturation at 95  $^{\circ}$ C for 3 min, followed by 35 amplification cycles of denaturation at 95  $^{\circ}$ C for 30 s, annealing at 55  $^{\circ}$ C for 30 s, extension at 72  $^{\circ}$ C for 30 s, and a final extension at 72  $^{\circ}$ C for 10 min. Following PCR amplification, the products were analyzed on a 1% (w/v) agarose gel electrophoresis. Moreover, the resulting PCR products were labeled using nick translation with either biotin dUTP or digoxigenin dUTP (Roche, <http://www.roche.com>).

### Chromosome preparation and FISH

Chromosome preparation and FISH were conducted following the previously described methods [25]. Root tips were pretreated with an 8-hydroxyquinoline solution at room temperature for 2 h before fixation in 3:1 ethanol:glacial acetic acid for 24 h. Subsequently, the root tips were treated with an enzyme mixture (1% pectolyase Y23, 2% pectolyase, 2% cellulase RS, and 4% Onozuka R-10 cellulase) at 37  $^{\circ}$ C for 3 h. The resulting cell suspension was dropped onto slides. Denatured probes were mixed with a hybridization mixture (50% formamide and 20% dextran sulfate in 2 $\times$  SSC) and applied to denatured chromosome on slides. A cover slip was added to cover the mixture, and the slides were incubated overnight in a hybridization chamber at 37  $^{\circ}$ C. Following incubation, the slides were washed with 2 $\times$  SSC three times for 5 min each and 1 $\times$  PBS for 5 min at room temperature. The labeled probes were detection using either streptavidin-conjugated Alexa Fluor 488 (Invitrogen, <http://www.invitrogen.com>) or rhodamine-conjugated anti-digoxigenin antibodies (Roche, <https://www.sigmaldrich.cn/CN/zh/product/roche>). Finally, the slides were counterstained with 4',6-diamidino-2-phenylindole (DAPI), and images were captured using an Olympus BX63 fluorescence microscope equipped with an Olympus DP80 CCD camera (Olympus, <https://www.getolympus.com>). The captured images were processed using cellsens dimension 1.9. The contrast of the images was adjusted using photoshop CC (Adobe, <https://www.adobe.com>).

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-024-04739-9>.

**Table S1.** The origin of the five okra accessions. **Table S2.** The sequences of the full-length satellite repeats. **Figure S1.** Multiple sequence alignment of eight satellite repeats in five okra accessions. **Figure S2.** FISH localization of five centromere satellite repeats on mitotic metaphase



chromosomes of FO accession. (a) Dual-color FISH of AeSat-B (red) and AeSat-E (green), showing colocalization of these two satellite repeats on the chromosomes of FO; (b) Dual-color FISH of AeSat-B (red) and AeSat-A (green), showing non-colocalization of these two satellite repeats on the chromosomes of FO; (c) Dual-color FISH of AeSat-B (red) and AeSat-C (green), showing non-colocalization of these two satellite repeats except on one pair of chromosomes of FO; (d) Dual-color FISH of AeSat-B (red) and AeSat-D (green), showing colocalization of these two satellite repeats on a few chromosomes of FO. Thus, five centromeric satellite repeats exhibit distribution bias within the two subgenomes in FO accession. Scale bars: 1  $\mu\text{m}$ . **Figure S3.** FISH localization of five centromere satellite repeats on mitotic metaphase chromosomes of T3 accession. (a) Dual-color FISH of AeSat-B (red) and AeSat-E (green), showing colocalization of these two satellite repeats on the chromosomes of T3; (b) Dual-color FISH of AeSat-B (red) and AeSat-A (green), showing non-colocalization of these two satellite repeats on the chromosomes of T3; (c) Dual-color FISH of AeSat-B (red) and AeSat-C (green), showing non-colocalization of these two satellite repeats except on one pair of chromosomes of T3; (d) Dual-color FISH of AeSat-B (red) and AeSat-D (green), showing colocalization of these two satellite repeats on a few chromosomes of T3. Thus, five centromeric satellite repeats exhibit distribution bias within the two subgenomes in T3 accession. Scale bars: 1  $\mu\text{m}$ . **Figure S4.** FISH localization of five centromere satellite repeats on mitotic metaphase chromosomes of ROJ accession. (a) Dual-color FISH of AeSat-B (red) and AeSat-E (green), showing colocalization of these two satellite repeats on the chromosomes of ROJ; (b) Dual-color FISH of AeSat-B (red) and AeSat-A (green), showing non-colocalization of these two satellite repeats on the chromosomes of ROJ; (c) Dual-color FISH of AeSat-B (red) and AeSat-C (green), showing non-colocalization of these two satellite repeats except on one pair of chromosomes of ROJ; (d) Dual-color FISH of AeSat-B (red) and AeSat-D (green), showing colocalization of these two satellite repeats on a few chromosomes of ROJ. Thus, five centromeric satellite repeats exhibit distribution bias within the two subgenomes in ROJ accession. Scale bars: 1  $\mu\text{m}$ . **Figure S5.** FISH localization of five centromere satellite repeats on mitotic metaphase chromosomes of QR accession. (a) Dual-color FISH of AeSat-B (red) and AeSat-E (green), showing colocalization of these two satellite repeats on the chromosomes of QR; (b) Dual-color FISH of AeSat-B (red) and AeSat-A (green), showing non-colocalization of these two satellite repeats on the chromosomes of QR; (c) Dual-color FISH of AeSat-B (red) and AeSat-C (green), showing non-colocalization of these two satellite repeats except on one pair of chromosomes of QR; (d) Dual-color FISH of AeSat-B (red) and AeSat-D (green), showing colocalization of these two satellite repeats on a few chromosomes of QR. Thus, five centromeric satellite repeats exhibit distribution bias within the two subgenomes in QR accession. Scale bars: 1  $\mu\text{m}$ . **Figure S6.** Multiple sequence alignment of 5S rDNA and NTS sequence from five okra accessions. Regions of the 5S rDNA and NTS sequence are shaded in the grey and blue box, respectively. The A-, IE- and C-Boxes of pol III promoters are marked in green. **Figure S7.** Multiple sequence alignment of ITS1+5.8S+ITS2 sequence from five okra accessions. Regions of the 5.8S rDNA as well as ITS1 and ITS2 sequence are shaded in the grey and blue box, respectively. **Figure S8.** The genomic proportions of 35S rDNA and 5S rDNA in NGS, TGS and genome assembly. **Figure S9.** The 5S rDNA sequence reads organized in graph structures from the RepeatExplorer2 graphical output. (a-e) Single reads are represented by vertices (nodes) and their sequence overlaps by edges. The 5S rDNA coding sequences and intergenic spacers are highlighted in green and grey vertices, respectively. Two loop structures were interconnected by a junction region (annotated as the coding sequence of 5S rDNA) in five okra accessions. Arrows indicate two intergenic spacers of 5S rDNA. (f) The copy number of 5S rDNA in both subgenomes of the okra variety 'Wufu' genome assembly. **Figure S10.** FISH mapping of 35S rDNA and AeSat-G on the metaphase spreads of the five okra accessions. DAPI-stained mitotic chromosomes are shown in blue and grey. The signals of 35S rDNA and AeSat-G are shown in red and green, respectively. (a1-a4) FISH mapping of 35S rDNA and AeSat-G in FO. (b1-b4) FISH mapping of 35S rDNA and AeSat-G in T3. (c1-c4) FISH mapping of 35S rDNA and AeSat-G in ROJ. (d1-d4) FISH mapping of 35S rDNA and AeSat-G in RO1. (e1-e4) FISH mapping of 35S rDNA and AeSat-G in QR. AeSat-G exhibited a propensity for localizing in regions with relatively low DAPI staining, and it co-localized with the 35S rDNA at the same chromosomal location. Scale bars: 1  $\mu\text{m}$ . **Figure S11.** FISH mapping of 35S rDNA and AeSat-H on the metaphase spreads of the five okra accessions. DAPI-stained mitotic chromosomes are shown in blue and grey. The signals of 35S rDNA and AeSat-H are shown in red and green, respectively. (a1-a4) FISH mapping of 35S rDNA and AeSat-H in FO. (b1-b4) FISH mapping of 35S rDNA and AeSat-H in T3. (c1-c4) FISH mapping of 35S rDNA and AeSat-H in ROJ. (d1-d4) FISH mapping of 35S rDNA and AeSat-H in RO1. (e1-e4) FISH mapping of 35S rDNA and AeSat-H in QR. AeSat-H exhibited a propensity for localizing in regions with relatively low DAPI staining, and it co-localized with the 35S rDNA at the same chromosomal location. Scale bars: 1  $\mu\text{m}$ .

chromosomes are shown in blue and grey. The signals of 35S rDNA and AeSat-H are shown in red and green, respectively. (a1-a4) FISH mapping of 35S rDNA and AeSat-H in FO. (b1-b4) FISH mapping of 35S rDNA and AeSat-H in T3. (c1-c4) FISH mapping of 35S rDNA and AeSat-H in ROJ. (d1-d4) FISH mapping of 35S rDNA and AeSat-H in RO1. (e1-e4) FISH mapping of 35S rDNA and AeSat-H in QR. AeSat-H exhibited a propensity for localizing in regions with relatively low DAPI staining, and it co-localized with the 35S rDNA at the same chromosomal location. Scale bars: 1  $\mu\text{m}$ .

#### Acknowledgements

None.

#### Author contributions

Y.H. and J.X. conceived the research. J.L., X.L., L.F., X.W., S.Z. and R.T. performed experiments. Y.H., J.L., X.W. and L.F. analyzed the data. A.T., P.F., J.Q. and L.Z. collected the plant materials, helped in the experiment and made suggestions. Y.H., J.L., J.F. and J.X. wrote the manuscript. All authors have read and approved the manuscript.

#### Funding

This work was supported by funds from the China Agriculture Research System of MOF and MARA (CARS-16), Weng Hongwu Original Research Fund of Peking University and Weng Hongwu Academic Innovation Research Fund of Peking University.

#### Data availability

The datasets necessary for supporting the results of this article are included in this manuscript and its additional files. The datasets generated and/or analyzed during the current study are available in the National Center for Biotechnology Information (NCBI) repository with accession number BioProject PRJNA1005429.

#### Declarations

#### Ethics approval and consent to participate

All methods were performed in accordance with the relevant guidelines and regulations of institutional, national, and international guidelines and legislation.

#### Consent for publication

Not applicable.

#### Competing interests

We declare that there is no competing interest among the authors.

#### Author details

<sup>1</sup>Scientific Observing and Experimental Station of Southeastern kenaf & jute, Ministry of Agriculture and Rural Affairs of the People's Republic of China, Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, Fujian Provincial Key Laboratory of Crop Breeding by Design, National Engineering Research Center for Sugarcane, College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, China

<sup>2</sup>College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China

<sup>3</sup>College of Life Science, Fujian Normal University, Fuzhou 350117, China

<sup>4</sup>Ministerial and Provincial Joint Innovation Centre for Safety Production of Cross-Strait Crops, College of Geography and Oceanography, Minjiang University, Fuzhou 350108, China

Received: 20 September 2023 / Accepted: 10 January 2024

Published online: 25 January 2024

#### References

- Elkhalifa AEO, Alshammari E, Adnan M, Alcantara JC, Awadelkareem AM, Eltoum NE, Mehmood K, Panda BP, Ashraf SA. Okra (*Abelmoschus esculentus*) as a potential dietary medicine with nutraceutical importance for sustainable



- health applications. *Molecules*. 2021;26(3):696. <https://doi.org/10.3390/molecules26030696>.
2. Fan S, Zhang Y, Sun Q, Yu L, Li M, Zheng B, Wu X, Yang B, Li Y, Huang C. Extract of okra lowers blood glucose and serum lipids in high-fat diet-induced obese C57BL/6 mice. *J Nutr Biochem*. 2014;25(7):702–9. <https://doi.org/10.1016/j.jnutbio.2014.02.010>.
  3. Graziano S, Orsolini L, Rotolo MC, Tittarelli R, Schifano F, Pichini S. Herbal highs: review on psychoactive effects and neuropharmacology. *Curr Neuropharmacol*. 2017;15(5):750–61. <https://doi.org/10.2174/1570159x14666161031144427>.
  4. Liao Z, Zhang J, Liu B, Yan T, Xu F, Xiao F, Wu B, Bi K, Jia Y. Polysaccharide from okra (*Abelmoschus esculentus* (L.) Moench) improves antioxidant capacity via PI3K/AKT pathways and nrf2 translocation in a type 2 diabetes model. *Molecules*. 2019;24(10). <https://doi.org/10.3390/molecules24101906>.
  5. Yan T, Liu B, Wang N, Liao Z, Wu B, He B, Jia Y. The flavonoids of okra insulates against oxidative stress, neuroinflammation and restores BDNF levels in Aβ(1)–(42) induced mouse model of Alzheimer's disease. *Exp Gerontol*. 2021;147:111263. <https://doi.org/10.1016/j.exger.2021.111263>.
  6. Joshi AB, Hardas MW. Allopolyploid nature of okra, *Abelmoschus esculentus* (L.) Moench. *Nature*. 1956;178(4543):1190–0. <https://doi.org/10.1038/1781190a0>.
  7. Wang R, Li W, He Q, Zhang H, Wang M, Zheng X, Liu Z, Wang Y, Du C, Du H, et al. The genome of okra (*Abelmoschus esculentus*) provides insights into its genome evolution and high nutrient content. *Hortic Res*. 2023;10(8):1–15. <https://doi.org/10.1093/hr/uhad120>.
  8. Garrido-Ramos MA. Satellite DNA: an evolving topic. *Genes*. 2017;8(9). <https://doi.org/10.3390/genes8090230>.
  9. Shatskikh AS, Kotov AA, Adashev VE, Bazylev SS, Olenina LV. Functional significance of satellite DNAs: insights from *Drosophila*. *Front Cell Dev Biol*. 2020;8:312. <https://doi.org/10.3389/fcell.2020.00312>.
  10. Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang J. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. *Genetics*. 2003;163(3):1221–5. <https://doi.org/10.1093/genetics/163.3.1221>.
  11. Garrido-Ramos MA. Satellite DNA in plants: more than just rubbish. *Cytogenet Genome Res*. 2015;146(2):153–70. <https://doi.org/10.1159/000437008>.
  12. Gong Z, Yu H, Huang J, Yi C, Gu M. Unstable transmission of rice chromosomes without functional centromeric repeats in asexual propagation. *Chromosome Res*. 2009;17(7):863–72. <https://doi.org/10.1007/s10577-009-9073-7>.
  13. Wicke S, Costa A, Muñoz J, Quandt D. Restless 5S: the re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. *Mol Phylogenet Evol*. 2011;61(2):321–32. <https://doi.org/10.1016/j.jympev.2011.06.023>.
  14. Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics*. 2007;175(2):477–85. <https://doi.org/10.1534/genetics.107.071399>.
  15. Heitkam T, Weber B, Walter I, Liedtke S, Ost C, Schmidt T. Satellite DNA landscapes after allotetraploidization of quinoa (*Chenopodium quinoa*) reveal unique A and B subgenomes. *Plant J*. 2020;103(1):32–52. <https://doi.org/10.1111/tpj.14705>.
  16. Schmidt T, Heitkam T, Liedtke S, Schubert V, Menzel G. Adding color to a century-old enigma: multi-color chromosome identification unravels the autotriploid nature of saffron (*Crocus sativus*) as a hybrid of wild *Crocus cartwrightianus* cytotypes. *New Phytol*. 2019;222(4):1965–80. <https://doi.org/10.1111/nph.15715>.
  17. Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A. Fall and rise of satellite repeats in allopolyploids of *Nicotiana glauca* over c. 5 million years. *New Phytol*. 2010;186(1):148–60. <https://doi.org/10.1111/j.1469-8137.2009.03101.x>.
  18. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet*. 2020;21(10):597–614. <https://doi.org/10.1038/s41576-020-0236-x>.
  19. Włodzimierz P, Rabanal FA, Burns R, Naish M, Primitis E, Scott A, Mandáková T, Gorringer N, Tock AJ, Hollander D, et al. Cycles of satellite and transposon evolution in *Arabidopsis* centromeres. *Nature*. 2023;618(7965):557–65. <https://doi.org/10.1038/s41586-023-0606-z>.
  20. Naish M, Alonge M, Włodzimierz P, Tock AJ, Abramson BW, Schmücker A, Mandáková T, Jamge B, Lambing C, Kuo P, et al. The genetic and epigenetic landscape of the *Arabidopsis* centromeres. *Science*. 2021;374(6569):eabi7489. <https://doi.org/10.1126/science.abi7489>.
  21. Novák P, Neumann P, Macas J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat Protoc*. 2020;15(11):3745–76. <https://doi.org/10.1038/s41596-020-0400-y>.
  22. Yang X, Zhao H, Zhang T, Zeng Z, Zhang P, Zhu B, Han Y, Braz GT, Casler MD, Schmutz J, et al. Amplification and adaptation of centromeric repeats in polyploid switchgrass species. *New Phytol*. 2018;218(4):1645–57. <https://doi.org/10.1111/nph.15098>.
  23. Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J. Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell*. 2002;14(8):1691–704. <https://doi.org/10.1105/tpc.003079>.
  24. Zhang H, Kobližková A, Wang K, Gong Z, Oliveira L, Torres GA, Wu Y, Zhang W, Novák P, Buell CR, et al. Boom-bust turnovers of megabase-sized centromeric DNA in *Solanum* species: Rapid evolution of DNA sequences associated with centromeres. *Plant Cell*. 2014;26(4):1436–47. <https://doi.org/10.1105/tpc.114.123877>.
  25. Huang Y, Ding W, Zhang M, Han J, Jing Y, Yao W, Hasterok R, Wang Z, Wang K. The formation and evolution of centromeric satellite repeats in *Saccharum* species. *Plant J*. 2021;106(3):616–29. <https://doi.org/10.1111/tpj.15186>.
  26. Barros e Silva AE, Guerra M. The meaning of DAPI bands observed after C-banding and FISH procedures. *Biotech Histochem*. 2010;85(2):115–25. <https://doi.org/10.1080/10520290903149596>.
  27. Li Y, Zuo S, Zhang Z, Li Z, Han J, Chu Z, Hasterok R, Wang K. Centromeric DNA characterization in the model grass *Brachypodium distachyon* provides insights on the evolution of the genus. *Plant J*. 2018;93(6):1088–101. <https://doi.org/10.1111/tpj.13832>.
  28. Wolfgruber TK, Sharma A, Schneider KL, Albert PS, Koo DH, Shi J, Gao Z, Han F, Lee H, Xu R, et al. Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. *PLoS Genet*. 2009;5(11):e1000743. <https://doi.org/10.1371/journal.pgen.1000743>.
  29. Lee HR, Zhang W, Langdon T, Jin W, Yan H, Cheng Z, Jiang J. Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. *Proc Natl Acad Sci U S A*. 2005;102(33):11793–8. <https://doi.org/10.1073/pnas.0503863102>.
  30. García S, Wendel JF, Borowska-Zuchowska A, Ainouche M, Kuderova A, Kovarik A. The utility of graph clustering of 5S ribosomal DNA homoeologs in plant allopolyploids, homoploid hybrids, and cryptic introgressants. *Front Plant Sci*. 2020;11:41. <https://doi.org/10.3389/fpls.2020.00041>.
  31. Ávila Robledillo L, Neumann P, Kobližková A, Novák P, Vrbová I, Macas J. Extraordinary sequence diversity and promiscuity of centromeric satellites in the Legume tribe *Fabeae*. *Mol Biol Evol*. 2020;37(8):2341–56. <https://doi.org/10.1093/molbev/msaa090>.
  32. Maheshwari S, Ishii T, Brown CT, Houben A, Comai L. Centromere location in *Arabidopsis* is unaltered by extreme divergence in CENH3 protein sequence. *Genome Res*. 2017;27(3):471–8. <https://doi.org/10.1101/gr.214619.116>.
  33. Jiang J, Birchler JA, Parrott WA, Dawe RK. A molecular view of plant centromeres. *Trends Plant Sci*. 2003;8(12):570–5. <https://doi.org/10.1016/j.tplants.2003.10.011>.
  34. Hasson D, Panchenko T, Salimian KJ, Salman MU, Sekulic N, Alonso A, Warburton PE, Black BE. The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nat Struct Mol Biol*. 2013;20(6):687–95. <https://doi.org/10.1038/nsmb.2562>.
  35. Zhang T, Talbert PB, Zhang W, Wu Y, Yang Z, Henikoff JG, Henikoff S, Jiang J. The CentO satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres. *Proc Natl Acad Sci U S A*. 2013;110(50):E4875–4883. <https://doi.org/10.1073/pnas.1319548110>.
  36. Shang WH, Hori T, Toyoda A, Kato J, Popenдорф K, Sakakibara Y, Fujiyama A, Fukagawa T. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res*. 2010;20(9):1219–28. <https://doi.org/10.1101/gr.106245.110>.
  37. Hu X, Yu F, Huang Y, Sun L, Li X, Yang S, Chen K, Huang F, Zeng K, Zhang M, et al. Characterization analysis of the 35S rDNA intergenic spacers in *Erianthus Arundinaceus*. *Gene*. 2019;694:63–70. <https://doi.org/10.1016/j.gene.2019.01.026>.
  38. Huang Y, Yu F, Li X, Luo L, Wu J, Yang Y, Deng Z, Chen R, Zhang M. Comparative genetic analysis of the 45S rDNA intergenic spacers from three *Saccharum* species. *PLoS ONE*. 2017;12(8):e0183447. <https://doi.org/10.1371/journal.pone.0183447>.
  39. Gong L, Luo H, Shi W, Yang M. Intra-individual variation and transcribed pseudogenes in the ribosomal ITS1–5.8S–ITS2 rDNA of *Paraplusia japonica* (Pleuronectiformes: Cynoglossidae). *Biochem Biophys Res Commun*. 2019;513(3):726–31. <https://doi.org/10.1016/j.bbrc.2019.04.064>.

40. Garcia S, Kovařík A, Leitch AR, Garnatje T. Cytogenetic features of rRNA genes across land plants: analysis of the plant rDNA database. *Plant J*. 2017;89(5):1020–30. <https://doi.org/10.1111/tpj.13442>.
41. Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (*Solanaceae*). *Ann Bot*. 2008;101(6):805–14. <https://doi.org/10.1093/aob/mcm326>.
42. Jo SH, Park HM, Kim SM, Kim HH, Hur CG, Choi D. Unraveling the sequence dynamics of the formation of genus-specific satellite DNAs in the family *Solanaceae*. *Heredity*. 2011;106(5):876–85. <https://doi.org/10.1038/hdy.2010.131>.
43. Renny-Byfield S, Kovařík A, Chester M, Nichols RA, Macas J, Novák P, Leitch AR. Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of *Nicotiana tabacum*. *PLoS ONE*. 2012;7(5):e36963. <https://doi.org/10.1371/journal.pone.0036963>.
44. Huang M, Li H, Zhang L, Gao F, Wang P, Hu Y, Yan S, Zhao L, Zhang Q, Tan J, et al. Plant 45S rDNA clusters are fragile sites and their instability is associated with epigenetic alterations. *PLoS ONE*. 2012;7(4):e35139. <https://doi.org/10.1371/journal.pone.0035139>.
45. Jo SH, Koo DH, Kim JF, Hur CG, Lee S, Yang TJ, Kwon SY, Choi D. Evolution of ribosomal DNA-derived satellite repeat in tomato genome. *BMC Plant Biol*. 2009;9:42. <https://doi.org/10.1186/1471-2229-9-42>.
46. Lim KY, Skalicka K, Koukalova B, Volkov RA, Matyasek R, Hemleben V, Leitch AR, Kovarik A. Dynamic changes in the distribution of a satellite homologous to intergenic 26-18S rDNA spacer in the evolution of *Nicotiana*. *Genetics*. 2004;166(4):1935–46. <https://doi.org/10.1534/genetics.166.4.1935>.
47. Stupar RM, Song J, Tek AL, Cheng Z, Dong F, Jiang J. Highly condensed potato pericentromeric heterochromatin contains rDNA-related tandem repeats. *Genetics*. 2002;162(3):1435–44. <https://doi.org/10.1093/genetics/162.3.1435>.
48. Kirtiklis L, Ocalewicz K, Wiechowska M, Boroń A, Hliwa P. Molecular cytogenetic study of the European bitterling *Rhodeus amarus* (Teleostei: Cyprinidae: Acheilognathinae). *Genetica*. 2014;142(2):141–8. <https://doi.org/10.1007/s10709-014-9761-x>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.