Cloning and preliminary verification of telomere-associated sequences in upland cotton

Yuling Liu¹, Zhen Liu¹, Yangyang Wei¹, Yanjun Wang¹, Jiaran Shuang¹, Renhai Peng¹

Anyang Institute of Technology, Anyang, Henan, 455000, China

Corresponding author: Renhai Peng (aydxprh@163.com)

Abstract
Telomeres are structures enriched in repetitive sequences at the end of chromosomes. In this study, using the telomere primer AA(CCCTAAA)³ CCC for the single primer PCR, two DNA sequences were obtained from Gossypium hirsutum (Linnaeus, 1753) accession (acc.) TM-1. Sequence analysis showed that the two obtained sequences were all rich in A/T base, which was consistent with the characteristic of the telomere-associated sequence (TAS). They were designated as GhTAS1 and GhTAS2 respectively. GhTAS1 is 489 bp long, with 57.6% of A/T, and GhTAS2 is 539 bp long, with 63.9% of A/T. Fluorescence in situ hybridization results showed that both of the cloned TASs were located at the ends of the partial chromosomes of G. hirsutum, with the strong signals, which further confirmed that GhTAS1 and GhTAS2 were telomere-associated sequences including highly tandemly repetitive sequences. Results of blast against the assembled genome of G. hirsutum showed that GhTAS sequences may be missed on some assembled chromosomes. The results provide important evidence for the evaluation of the integrity of assembled chromosome end sequences, and will also contribute to the further perfection of the draft genomes of cotton.

Keywords
G. hirsutum, telomere-associated sequence, cloning, FISH

Introduction
Telomeres are DNA-protein complexes at the ends of chromosomes (Blackburn 1991). Telomere structures are highly conserved, and vary surprisingly little between organisms (Richards and Ausubel 1988, Ganal et al. 1991, Fajkus et al. 2005, Watson and...
Riha 2010). In humans, telomere repeated sequences are composed of conserved a minisatellite sequence unit 5’-TTAGGG-3’ (Moyzis et al. 1988), whereas in Tetrahy- mena (Furgason, 1940) each chromosome end has a conserved 5’-TTGGGG-3’ tel- omere repeat unit (Blackburn and Gall 1978). The first plant telomere DNA sequence, 5’-TTTAGGG-3’ tandem repeat, was isolated from Arabidopsis thaliana (Linnaeus, 1753) (Richards and Ausubel 1988). Subsequent studies have demonstrated that the Arabidopsis-type telomeres presented in most plants (Fajkus et al. 2005, Ling et al. 2012, Schrumpfová et al. 2019). At the same time, other studies have shown that some plants lacked typical telomere tandem repeat 5’-TTTAGGG-3’, which sheds more light on telomere function and how telomeres responded to genetic change (Adams et al. 2001, Sýkorová et al. 2003a, Peška et al. 2015).

Telomere tandem repeats located at the end of chromosomes represent only a part of the end of chromosomes. Telomere-associated sequences (TASs) located directly proximal to telomere tandem repeats (Li et al. 2009) play an important role in telomere maintenance and chromosome stability through epigenetic modification or recombination (Cross et al. 1990, Zhong et al. 1998, Sýkorová et al. 2003b, Tran et al. 2015). In addition, TAS is also a good marker at the end of the genetic linkage map. Three TASs cloned from rice showed high polymorphism at the ends of chromosome arms of different rice varieties based on the results of genetic mapping (Ashikawa et al. 1994). Despite functional importance, the nucleotide sequences in the subtelomere region have not been fully resolved in many sequenced genomes (Lese et al. 1999, Mefford and Trask 2002, Mizuno et al. 2006). So, more work is needed to reveal the structure and function of the subtelomeres.

At present, there is relatively little research on cotton telomere. Combining FISH using the Arabidopsis-type telomere sequence amplified from Arabidopsis genomic DNA and BAL-31 digestion, Ling et al. (2012) published the first study on cotton telomeres, which proved the Arabidopsis-type telomere sequence existed in the cotton genome. G. hirsutum is the most important cultivated cotton species. So far, different versions of the genome sequence have been released (Li et al. 2015, Zhang et al. 2015, Wang et al. 2019, Hu et al. 2019), however, high content of repetitive sequences affects the quality of genome assembly (Sýkorová et al. 2013, Liu et al. 2016). TAS occupies a large proportion in subtelomere tandem repeats regions. Therefore, in order to improve the quality of genome assembly, nucleotide sequences in the subtelomere region need to be further analyzed.

Material and methods

Plant materials

The plant material was G. hirsutum acc. TM-1 (AADD), which was planted in the experimental field of Anyang Institute of Technology, Henan, China. Genomic DNA was isolated from fresh young leaves using the modified CTAB method (Song
et al. 1998). Root tip material used for *G. hirsutum* chromosome preparation were harvested from the about 6-day seedlings planted in an incubator and pretreated by 25 ppm cycloheximide at 20 °C for 80 min, then fixed in methanol-acetic acid (3:1) and stored at 4 °C for 24 h. Squashes of root tips were prepared according to Liu et al. (2017).

**Primers**

The eight single primers of the plant telomere repeat were selected from NCBI database (https://www.ncbi.nlm.nih.gov) according to the previous studies for single primer PCR (Burr et al. 1992, Gong et al. 1998, Weiss-Schneeweiss et al. 2004, Liu et al. 2005). The primers sequence information is shown in Table 1.

**Cloning and sequencing of telomere-associated sequences**

The selected single primers of the plant telomere repeat sequence (Table 1) were amplified by single primer PCR using the genomic DNA of *G. hirsutum* as template, to find the suitable conditions for obtaining promising products and candidates for subtelomeric regions. The amplification procedure was as 95 °C for 3 min, followed by 35 cycles of 95 °C for 15 s, 55 °C/60 °C for 15 s, 72 °C for 30 s, and a final extension at 72 °C for 5 min. The amplification products were detected by 1% agarose gel electrophoresis, and the appropriate single primer and annealing temperature were selected based on the above result. Then, PCR amplification was performed using the selected single primer in a 50 μl reaction volume containing 25 μl of 2 × Phanta Max Buffer, 1 μl of Phanta Max Super-Fidelity DNA Polymerase (Vazyme), 0.8 μmol/L of the telomeric single primer, and 10 ng of genomic DNA. The objective band from PCR was recovered by gel extraction kit (SanPrep Column DNA Gel Extraction kit, Sangon Biotech) and was cloned into *Trans*1-T1 competent cells by the *pEasy*-Blunt Simple Cloning Vector (TransGen Biotech) according to the manufacturer’s instructions. The positive clones were selected for sequencing by Shanghai Sangon.

<table>
<thead>
<tr>
<th>Name</th>
<th>Taxonomic name</th>
<th>Reference</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR1</td>
<td><em>Oryza sativa</em> (Linnaeus, 1753)</td>
<td>Gong et al. 1998</td>
<td>(TTTAGGG)</td>
</tr>
<tr>
<td>TR2</td>
<td><em>Zea mays</em> (Linnaeus, 1753)</td>
<td>Burr et al. 1992</td>
<td>(TTTAGGG)</td>
</tr>
<tr>
<td>TR3</td>
<td><em>Othocallis siberica</em> (Linnaeus, 1753)</td>
<td>Weiss-Schneeweiss et al. 2004</td>
<td>(TTTAGGG)</td>
</tr>
<tr>
<td>TR4</td>
<td><em>Ginkgo biloba</em> (Linnaeus, 1771)</td>
<td>Liu et al. 2005</td>
<td>(CCCTAAA)</td>
</tr>
<tr>
<td>TR5</td>
<td><em>Brassica campestris</em> (Linnaeus, 1753)</td>
<td>Kapila et al. 1996</td>
<td>(CCCTAAA),CCC</td>
</tr>
<tr>
<td>TR6</td>
<td><em>Othocallis siberica</em></td>
<td>Weiss-Schneeweiss et al. 2004</td>
<td>AA (CCCTAAA),CCC</td>
</tr>
<tr>
<td>TR7</td>
<td><em>Zea mays</em></td>
<td>Burr et al. 1992</td>
<td>(CCCTAAA)</td>
</tr>
<tr>
<td>TR8</td>
<td><em>Othocallis siberica</em></td>
<td>Weiss-Schneeweiss et al. 2004</td>
<td>(CCCTAAA)</td>
</tr>
</tbody>
</table>
Software and websites for sequences analysis

DNAMAN software was used for extraction and alignment of cloned sequences. Repetitive sequence analysis was performed using the online program CENSOR (https://www.girinst.org/censor/index.php). BLAST algorithm blastn (https://www.cottongen.org/blast) was used to identify GhTAS from *G. hirsutum* genome database (*Gossypium hirsutum* ZJU v2.1, a1) (Hu et al. 2019). All the above analyses were performed according to the default parameters.

FISH validation

The TAS plasmid DNA was extracted using the TIANTprep Mini Plasmid Kit according to the instructions. Then, TAS plasmid DNA was labeled with DIG-Nick Translation Mix (Roche). The 45S rDNA probes derived from *Arabidopsis thaliana* (Gan et al. 2013) were labeled with biotin-Nick Translation Mix (Roche) according to the instructions of the manufacturer. Chromosome preparation and FISH were performed according to the previous methods (Liu et al. 2017).

Results

Optimization of the single primer PCR

According to the melting temperature (Tm) value distribution of the eight candidate single primers (55 °C–62 °C), two annealing temperatures were selected, namely 55 °C and 60 °C. The results of PCR amplification showed that an obvious band of roughly 500 bp was amplified using the single primer TR6 under the two annealing temperatures, especially, the band amplified under annealing temperature of 60 °C showed better specificity and higher brightness (Fig. 1B-6). So, the primer TR6 (AA (CCCTAAA)₃CCC) was chosen for the following PCR amplification.

Cloning of TAS

A single band with a size of roughly 500 bp was amplified using the single primer TR6 under the annealing temperature of 60 °C with Phanta Max Super-Fidelity DNA Polymerase (Fig. 2A-2). After transformation, eight positive clones were obtained after a positive test from transformed clones (Fig. 2B). Then, the eight positive clones were sequenced.
Figure 1. Amplification results of candidate single primers. M Marker A and B the annealing temperature is 55 and 60 °C respectively 1–8 primers TR1–TR8.

Figure 2. Results of cloning and positive test. A PCR amplification results M Marker 1 Common Taq enzyme 2 High-fidelity enzyme. B Positive test of bacterial colony PCR M Marker 1–12 the candidate clones.

Sequences analysis

Sequence component analysis

Sequence analysis of the eight positive clones revealed that all clones had the same forward and inverted telomere primer sequence at the two ends. Sequence alignment showed that there were two different internal sequences in eight sequences. So, the two different cloned DNA sequences with different size of 488 bp and 538 bp were selected and named as GhTAS1 and GhTAS2 (Fig. 3). Their sequences had been uploaded to GenBank (accession No. MT078976 and MT078977).
The two sequences were rich in A/T bases, that is, 57.6% and 63.9% respectively. Repeat masking analysis indicated that the tandem repeats content were 31.35% in GhTAS1 and 42.38% in GhTAS2, which mainly consisted of satellite DNA and transposable elements. The above results are consistent with the typical characteristics of telomere-associated sequences (Li et al. 2009).

Homology analysis of GhTASs

Sequence alignment results of DNAMAN shown that GhTAS1 and GhTAS2 had low homology, with the sequence similarity of 38.90%, which may be due to their different chromosomal sources.

After comprehensive comparison of the obtained TASs of *G. hirsutum* and the TASs of *Arabidopsis thaliana*, *Glycine max* (Linnaeus, 1753), *Oryza sativa* (Linnaeus, 1753), *Zea mays* (Linnaeus, 1753), *Larix gmelinii* (Ruprecht, 1920) listed on NCBI, it was found that their similarity was low, ranging from 25% to 50% (Table 2). All these indicated that the cloned telomere-associated sequences had obvious species specificity.

BLAST of GhTAS1 and GhTAS2 against *G. hirsutum* genome

GhTAS1 and GhTAS2 were found using blastn with the latest *G. hirsutum* genome sequence (*Gossypium hirsutum* ZJU v2.1, a1) in Cottongen (https://www.cottongen.org/).
Table 2. Similarity of telomere-associated sequences between *G. hirsutum* and other plants.

<table>
<thead>
<tr>
<th>Species</th>
<th>NCBI accession No.</th>
<th>TASs of <em>G. hirsutum</em></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GhTAS1</td>
<td>GhTAS2</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>AC074298.1</td>
<td>39.60%</td>
<td>36.71%</td>
</tr>
<tr>
<td></td>
<td>AM177016.1</td>
<td>14.08%</td>
<td>12.94%</td>
</tr>
<tr>
<td></td>
<td>AM177019.1</td>
<td>13.52%</td>
<td>13.93%</td>
</tr>
<tr>
<td></td>
<td>AM177060.1</td>
<td>10.88%</td>
<td>10.15%</td>
</tr>
<tr>
<td><em>Glycine max</em></td>
<td>AF041999.1</td>
<td>20.24%</td>
<td>16.79%</td>
</tr>
<tr>
<td><em>Oryza sativa</em></td>
<td>U12056.1</td>
<td>28.71%</td>
<td>25.27%</td>
</tr>
<tr>
<td><em>Zea mays</em></td>
<td>S46927.1</td>
<td>48.70%</td>
<td>41.93%</td>
</tr>
<tr>
<td><em>Larix gmelinii</em></td>
<td>EF474441.1</td>
<td>31.40%</td>
<td>30.57%</td>
</tr>
</tbody>
</table>

Table 3. Partial blast results of GhTAS1 and GhTAS2 in the *G. hirsutum* genome.

<table>
<thead>
<tr>
<th>Sequence name</th>
<th>Genomic location</th>
<th>Query matches</th>
<th>Hit matches</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GhTAS1 D06</td>
<td>1–488</td>
<td>65407147–65406660</td>
<td>98.98%</td>
<td></td>
</tr>
<tr>
<td>GhTAS1 D03</td>
<td>184–281</td>
<td>26586–26683</td>
<td>78.57%</td>
<td></td>
</tr>
<tr>
<td>GhTAS1 A01</td>
<td>171–237</td>
<td>118151185–118151119</td>
<td>82.09%</td>
<td></td>
</tr>
<tr>
<td>GhTAS1 D02</td>
<td>138–219</td>
<td>69751633–69751551</td>
<td>79.52%</td>
<td></td>
</tr>
<tr>
<td>Scaffold515-obj</td>
<td>184–281</td>
<td>9914–9817</td>
<td>75.51%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 A06</td>
<td>14–537</td>
<td>126445179–126444656</td>
<td>99.62%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 D11</td>
<td>14–537</td>
<td>71336660–71336138</td>
<td>98.47%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 A13</td>
<td>14–535</td>
<td>47688–48202</td>
<td>94.08%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 A02</td>
<td>25–512</td>
<td>40084–39589</td>
<td>88.15%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 D02</td>
<td>25–512</td>
<td>69751559–69752073</td>
<td>86.68%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 A12</td>
<td>25–512</td>
<td>30186–29672</td>
<td>86.15%</td>
<td></td>
</tr>
<tr>
<td>Scaffold546-obj</td>
<td>46–455</td>
<td>8556–8146</td>
<td>89.07%</td>
<td></td>
</tr>
<tr>
<td>Scaffold515-obj</td>
<td>25–271</td>
<td>31264–31514</td>
<td>89.33%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 A09</td>
<td>25–315</td>
<td>83200103–83200398</td>
<td>86.96%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 A11</td>
<td>25–271</td>
<td>121355653–121355904</td>
<td>88.54%</td>
<td></td>
</tr>
<tr>
<td>scaffold407_obj_A03</td>
<td>59–271</td>
<td>36503–36719</td>
<td>92.24%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 A07</td>
<td>25–271</td>
<td>96580716–96580969</td>
<td>88.24%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 D10</td>
<td>278–455</td>
<td>66830830–66831007</td>
<td>93.26%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 A10</td>
<td>25–271</td>
<td>115081227–115081476</td>
<td>87.75%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 A05</td>
<td>285–455</td>
<td>39831434–39831267</td>
<td>93.60%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 A01</td>
<td>278–455</td>
<td>118169784–118169962</td>
<td>91.06%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 D08</td>
<td>278–512</td>
<td>69075939–69076196</td>
<td>84.11%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 D03</td>
<td>278–455</td>
<td>23313–23139</td>
<td>91.01%</td>
<td></td>
</tr>
<tr>
<td>GhTAS2 D09</td>
<td>278–442</td>
<td>51987281–51987445</td>
<td>91.52%</td>
<td></td>
</tr>
</tbody>
</table>

Results showed that GhTAS1 was mapped onto five chromosomes and one scaffold of *G. hirsutum*, and GhTAS2 was mapped onto all 26 chromosomes and 14 scaffolds of *G. hirsutum* with different E-value. The partial blast results with lower E-value were listed in Table 3. GhTAS1 was localized at one end of the chromosome D06, with a higher similarity of 98.98%, and was localized at the single end of chromosomes D03, A01, D02 and D01, as well as Scaffold515, with lower similarity (Fig. 4A). GhTAS2 showed
Figure 4. Localization patterns of GhTAS1 and GhTAS2 on *G. hirsutum* partial chromosomes. A Gh-TAS1 B GhTAS2.
higher chromosomes coverage than GhTAS1. Among the all blast results, GhTAS2 was localized at both ends of chromosomes D11, A13, A02 and D02 and at the single end of chromosomes A06, A12 and two scaffolds with higher similarity (Fig. 4B). At the same time, unlike GhTAS1, the GhTAS2 sequence is also mapped to other chromosomal regions in addition to the ends of chromosomes (Fig. 4B1–6).

**Chromosome localization of GhTAS1 and GhTAS2 based on FISH**

To examine the chromosome physical location of GhTAS1 and GhTAS2, we carried out FISH on *G. hirsutum* metaphase chromosomes using a digoxin-labeled GhTAS probe and a biotin-labeled 45S rDNA probe. The results showed that GhTAS1 had signals at the end of nearly half of the chromosomes of *G. hirsutum*, and most of them were distributed at the single end. The signal intensity on different chromosomes was also different (Fig. 5A-2, A-4). GhTAS2 has signals on both ends of most chromosomes of *G. hirsutum* (Fig. 5B-2, B-4). Three pairs of 45S rDNA signals were detected on the chromosomes of *G. hirsutum* (Fig. 5A-3 and 5B-3 arrows). Two pairs of GhTAS1 signals were collinear with 45S rDNA (Fig. 5A-2 arrows). Three pairs of GhTAS2 signals were collinear with 45S rDNA (Fig. 5B-2 arrows). In addition, the chromosomes carrying GhTAS2 FISH signals were much more than those with GhTAS1 FISH signals (Fig. 5A-2, B-2), which was similar to the blast results (Fig. 4, Table 3).
Discussion

In this study, the telomere primer AA(CCCTAAA)CCC was used as a single primer to obtain the TAS sequences of *G. hirsutum* by single primer PCR. The homology of the two TASs is relatively low and with the similarity of 38.90%. Chromosome FISH localization of the two sequences also showed obvious differences in chromosome distribution and signal strength (Fig. 5A, B), which may be due to the differences of chromosome specificity and sequence copy number of the two TASs. In the early study of *Chironomus palidivittatus* (Edwards, 1929) TAS, it was found that there were considerable differences in TAS between species, within species, and even in telomere of the same species (Cohn and Edstrom 1992). Gong et al. cloned six TASs in rice and found high polymorphism of these sequences through RFLP analysis (Gong et al. 1998). From then on, this phenomenon has been found in related studies of other species (Li et al. 2009). Therefore, TASs show great specificity, unlike the more conservative telomere repeated sequences (TR).

Since telomere and adjacent subtelomere regions could not be covered by PAC and BAC clones, sequencing efforts were unable to reveal the structure of these regions. In addition, the discovery of interstitial telomeric sequences (ITSs) makes telomeric minisatellites have double-faced character, which causes more problems in producing genomic assemblies (Richards et al. 1991, Sýkorová et al. 2003). Therefore, nucleotide sequences in the subtelomere regions have not been fully resolved in many genomes that have been sequenced (Mefford and Trask 2002, Mizuno et al. 2006), which greatly affects the quality of genome assembly. FISH localization can reflect the true position of DNA fragments in chromosomes (Jiang and Gill 2006). FISH combined with genomic BLAST can intuitively judge the genomic assembly quality of DNA sequences. Chromosomal locations of 45S rDNA in *G. hirsutum* had been revealed using double-probe FISH, that is, chromosomes A09, D07 and D09 (Gan et al. 2013). In this study, according to the genome BLAST and chromosome FISH localization results of GhTAS and 45S rDNA, it was found that TASs at the end of some chromosomes were not assembled in the genome sequence map. Obviously, results of blastn showed that GhTAS1 was only mapped onto chromosomes D06, D03, A01, D02 and D01 (Table 3, Fig. 4A), but FISH showed more chromosomes carried GhTAS1 signals, including two of the three chromosomes with 45S rDNA A09, D07 or D09, which had not appeared on the blastn results. That is, GhTAS1 sequences may be missed on these assembled chromosomes. The results provide important evidence for the evaluation of the integrity of assembled chromosome end sequences.

Conclusions

We cloned two telomere-associated sequences from *G. hirsutum* acc. TM-1 using the single-primer PCR, and made analysis about the sequence characteristics of two TASs. The two TASs sequences were enriched in A/T, and were flanked by the forward and in-
verted primer sequences at each end respectively. By comparative analysis based on the results of blastn and FISH localization of the two TASs, we found that TASs at the end of some chromosomes were not assembled in the genome sequence map. Our study not only contributes to the analysis of telomere structure of cotton, but also provides intuitive evidence for the evaluation of the integrity of the assembled *G. hirsutum* genome.

**Acknowledgements**

The research was sponsored by National Key R&D Program of China (No. 2018YFD0100300), Innovation Scientists and Technicians Troop Construction Projects of Henan Province (20IRTSTHN021), Science and Technology Development Program of Anyang City (2018-66-133), Science and Technology Development Project of Henan Province (182102410041).

**References**


Telomere related sequences of upland cotton


Karyotype and putative chromosomal inversion suggested by integration of cytogenetic and molecular data of the fungus-farming ant Mycetomoellerius iheringi Emery, 1888

Ricardo Micolino¹², Maykon Passos Cristiano², Danon Clemes Cardoso¹²

¹ Programa de Pós-Graduação em Genética, Departamento de Genética, Universidade Federal do Paraná (UFPR), Centro Politécnico, Jardim das Américas, 81531-990, Curitiba, PR, Brazil ² Departamento de Biodiversidade, Evolução e Meio Ambiente, Universidade Federal de Ouro Preto (UFOP), Ouro Preto, MG, Brazil

Corresponding author: Danon Clemes Cardoso (danon@ufop.edu.br)

Academic editor: V. Gokhman | Received 3 January 2020 | Accepted 28 February 2020 | Published 7 May 2020


Abstract

Comparative cytogenetic analyses are being increasingly used to collect information on species evolution, for example, diversification of closely related lineages and identification of morphologically indistinguishable species or lineages. Here, we have described the karyotype of the fungus-farming ant Mycetomoellerius iheringi Emery, 1888 and investigated its evolutionary relationships on the basis of molecular and cytogenetic data. The M. iheringi karyotype consists of 2n = 20 chromosomes (2K = 18M + 2SM). We also demonstrated that this species has the classical insect TTAGG telomere organization. Phylogenetic reconstruction showed that M. iheringi is phylogenetically closer to M. cirratus Mayhé-Nunes & Brandão, 2005 and M. kempfi Fowler, 1982. We compared M. iheringi with other congeneric species such as M. holmgreni Wheeler, 1925 and inferred that M. iheringi probably underwent a major pericentric inversion in one of its largest chromosomes, making it submetacentric. We discussed our results in the light of the phylogenetic relationships and chromosomal evolution.

Keywords

chromosomal evolution, FISH, fungus growing, karyomorphometry, TTAGG, Trachymyrmex

Copyright Ricardo Micolino et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Introduction

Fungus-farming ants (Formicidae: Myrmicinae: Attini) are exclusive to the New World and occur mainly in the Neotropical region, with some species found in the Nearctic region (Weber 1966; Rabeling et al. 2007). The most recently diverged species include the well-known leafcutter ants (genera \textit{Atta} Fabricius, 1804 and \textit{Acromyrmex} Mayr, 1865) as well as the genera \textit{Xerolitor} Sosa-Calvo et al., 2018, \textit{Sericomymex} Mayr, 1865 and \textit{Trachymyrmex} Forel, 1893. Previous phylogenetic analyses have shown that the genus \textit{Trachymyrmex} is paraphyletic (e.g., Schultz and Brady 2008; Sosa-Calvo et al. 2018; Micolino et al. 2019a). However, this taxonomic complication was recently resolved by multilocus phylogenetic analyses with a comprehensive number of species (Solomon et al. 2019). Thus, a new systematic arrangement of three clades was proposed as follows: \textit{Mycetomoellerius} Solomon et al. 2019 (former \textit{Iheringi} group), \textit{Pantrachymyrmex} Solomon et al., 2019 (former \textit{Intermedius} group), and \textit{Trachymyrmex} (based on the type species \textit{Trachymyrmex septentrionalis} McCook, 1881). Nevertheless, \textit{Trachymyrmex} sensu stricto, largely containing North American species, is still most prominently studied (e.g., Rabeling et al. 2007; Seal et al. 2015; Sánchez-Peña et al. 2017).

Cytogenetics encompasses the study of chromosomes that may have direct implications on species evolution, such as the identification of cryptic species and diversification of closely related lineages (White 1978; King 1993). In general, ants exhibit one of the largest chromosomal variability among organisms (reviewed by Lorite and Palomeque 2010), leading to the hypothesis that chromosomal rearrangements, i.e., Robertsonian fissions and fusions (known major rearrangements that can change the chromosomal number within lineages), actively contributed to the diversification of ants (Imai et al. 1988, 2001; Cardoso et al. 2018a). Despite the large number of species in the three genera formerly included into “\textit{Trachymyrmex}” (about 60 species, see above), there is limited cytogenetic information on this ant group. To date, only seven species have been karyotyped, three of which have not been identified to the species level (see Table 1). On the basis of the available data, the described chromosomal numbers appear to be stable within the three genera, ranging from $2n = 12$ to $2n = 22$ and predominantly comprising metacentric chromosomes (reviewed by Cardoso et al. 2018a).

Table 1. Former “\textit{Trachymyrmex}” species with their described karyotypes. 2n: diploid chromosome number; (n): haploid chromosome number; 2K: karyotype formula; Locality: sampling site; M: metacentric chromosomes; SM: submetacentric chromosomes.

<table>
<thead>
<tr>
<th>Species</th>
<th>2n (n)</th>
<th>2K</th>
<th>Locality</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Mycetomoellerius fuscus}$^*$</td>
<td>18 (9)</td>
<td>16M + 2SM</td>
<td>Minas Gerais State, Brazil</td>
<td>Barros et al. (2013a)</td>
</tr>
<tr>
<td>\textit{Mycetomoellerius holmgreni}</td>
<td>20 (10)</td>
<td>20M</td>
<td>Minas Gerais State, Brazil</td>
<td>Barros et al. (2018)</td>
</tr>
<tr>
<td>\textit{Mycetomoellerius iheringi}</td>
<td>20 (10)</td>
<td>18M + 2SM</td>
<td>Santa Catarina State, Brazil</td>
<td>Present study</td>
</tr>
<tr>
<td>\textit{Mycetomoellerius relicthus}</td>
<td>20 (10)</td>
<td>20M</td>
<td>Minas Gerais State, Brazil</td>
<td>Barros et al. (2013b)</td>
</tr>
<tr>
<td>\textit{Trachymyrmex septentrionalis}</td>
<td>20 (10)</td>
<td>20M</td>
<td>Barro Colorado Island, Panama</td>
<td>Murakami et al. (1998)</td>
</tr>
<tr>
<td>“\textit{Trachymyrmex}” sp. 1</td>
<td>12 (6)</td>
<td>12M</td>
<td>Barro Colorado Island, Panama</td>
<td>Murakami et al. (1998)</td>
</tr>
<tr>
<td>“\textit{Trachymyrmex}” sp. 2</td>
<td>18 (9)</td>
<td>18M</td>
<td>Barro Colorado Island, Panama</td>
<td>Murakami et al. (1998)</td>
</tr>
<tr>
<td>“\textit{Trachymyrmex}” sp. 3</td>
<td>22 (11)</td>
<td>18M + 4SM</td>
<td>Minas Gerais State, Brazil</td>
<td>Barros et al. (2013b)</td>
</tr>
</tbody>
</table>

$^*$ current junior synonym of \textit{M. urichi}. 

Ricardo Micolino et al. / Comparative Cytogenetics 14: 197–210 (2020)
*Mycetomoellerius iheringi* Emery, 1888, the type species of the genus, is a species endemic to South America, and it occurs mainly in the southern regions. The exclusive characteristic of *M. iheringi* is the finely striated discal area of the mandibles, which sets it apart from the congeneric species *Mycetomoellerius kempfi* Fowler, 1982 (Mayhé-Nunes and Brandão 2005). A feature of *M. iheringi* biology that facilitates field identification is the subterranean nest in the sand with a slim opening (Mayhé-Nunes and Brandão 2005). Some groups have been identified by morphological similarities within the former “*Trachymyrmex*”, including the *Iheringi* group that also includes *Mycetomoellerius holmgreni* Wheeler, 1925 whose karyotype has been already described (Mayhé-Nunes and Brandão 2005; Barros et al. 2018). This fact allows cytogenetic comparisons with *M. iheringi*. However, the phylogenetic position of *M. iheringi* has not yet been described; only the relationship between its fungal cultivars has been reported (see Solomon et al. 2019).

Here, we have described the *M. iheringi* karyotype on the basis of karyomorphometric analysis and fluorescence *in situ* hybridization (FISH) with a telomeric probe. In addition, we identified the phylogenetic position of *M. iheringi* and examined its relationship with other species of the genus. We have discussed our results in the light of chromosomal evolution among fungus-farming ants.

**Material and methods**

**Colony sampling**

Colonies of *M. iheringi* were collected from the Restinga environment of the Brazilian Atlantic coast at Joaquina Beach, Florianópolis, Santa Catarina State, Brazil (27°37'44"S; 48°26'52"W). A total of five distantly spaced colonies were sampled. Such colonies were maintained *in vivo* at the Laboratório de Genética Evolutiva e de Populações, Universidade Federal de Ouro Preto, Brazil, according to the protocol established by Cardoso et al. (2011).

**Chromosome preparation and FISH mapping**

Metaphase chromosomes from the brain ganglia of pre-pupal larvae were obtained using the method of Imai et al. (1988). The ganglia were dissected under a stereomicroscope and incubated in hypotonic solution containing 1% sodium citrate and 0.005% colchicine for 60 min, and consecutively dissociated and fixed on stereoscopic microscope slides in acetic acid: ethanol: distilled water (3:3:4) and acetic acid: ethanol (1:1). Subsequently, the metaphase chromosomes were examined under a phase-contrast microscope and stained with 4% Giemsa stain dissolved in Sorensen’s buffer, pH 6.8, to determine the chromosome number and morphology. We classified the chromosomes according to the nomenclature proposed by Levan et al. (1964), which is based on the ratio of the chromosomal arms (*r*), given by centromere posi-
tion. The chromosomes were classified into metacentric \((r = 1.0–1.7)\), submetacentric \((r = 1.7–3.0)\), subtelocentric \((r = 3.0–7.0)\), and acrocentric \((r > 7.0)\) categories, as modified by Crozier (1970). The metaphase chromosomes were measured using IMAGE-PRO PLUS software (Media Cybernetics, LP, USA), and the values were calibrated by the scale bar and transferred to EXCEL (Microsoft, Redmond, WA, USA). In addition, the degree of variation and karyotype measurement were validated using statistical tests, according to Cristiano et al. (2017).

FISH experiments were performed as previously described by Kubat et al. (2008), with detailed modifications for ants by Micolino et al. (2019a). For the hybridizations, we used the TTAGG\(_6\) telomeric motif, which has fine conservation in most insects and the advantage of being able to detect chromosomal rearrangements such as telomere-related inversions and fusions. The TTAGG\(_6\) probe was directly labeled with Cy3 at the 5\(^\prime\) terminal during synthesis (Sigma, St. Louis, MO, USA). The summarized technique involves several saline washes, alcohol dehydration, and formamide denaturation, until hybridization with the probe. For visualization, the metaphase chromosomes were stained with 4’,6-diamidino-2-phenylindole (DAPI Fluoroshield, Sigma-Aldrich) in an antifade solution. The metaphase chromosomes were analyzed under an OLYMPUS BX53 epifluorescence microscope with OLYMPUS CELLSENS IMAGING software (Olympus American, Inc., Center Valley, PA, USA), using WU (330–385 nm) and WG (510–550 nm) filters for DAPI and rhodamine, respectively. About 10–20 metaphases were analyzed in both cytogenetic analyses, and the images were edited with ADOBE PHOTOSHOP CC software.

**DNA extraction, sequencing, and phylogenetic analysis**

We extracted the DNA from *M. iheringi* ant workers, according to the standard CTAB/chloroform technique (Sambrook and Russell 2001). We sequenced the fragments of four nuclear genes, *elongation factor 1-alpha-F1* (EF1\(\alpha\)-F1), *elongation factor 1-alpha-F2* (EF1\(\alpha\)-F2), *wingless* (Wg), and *long-wavelength rhodopsin* (LWRh), and one mitochondrial gene, *cytochrome c oxidase I* (COI) (GenBank accession numbers: MT174160–MT174169). The primers used to generate the sequence data are listed in Table 2. Polymerase chain reaction was performed using a final volume of 25 μL, according to the manufacturer’s instructions (Promega, Madison, WI, USA). The amplification conditions and sequencing were based on the methodology outlined in previous studies (see Schultz and Brady 2008, Cardoso et al. 2015a, b, Ward et al. 2015).

The gene fragments were aligned and concatenated using MEGA7 software (Kumar et al. 2016) and incorporated into the dataset of Solomon et al. (2019). The phylogeny was inferred using the maximum likelihood criterion in RAxML (Stamatakis 2014) by using the simultaneous best-tree search and rapid bootstrapping analysis (1000 replicates) with the GTR + G model of evolution. The generated tree and branch labels were visualized using FIGTREE software (Rambaut 2009).
Results

Cytogenetic data

The karyotype of *M. iheringi* has 2n = 20 chromosomes (Fig. 1). Our karyomorphometric analysis revealed that this karyotype consists of nine metacentric pairs and one submetacentric pair; the karyotype formula is 2K = 18M + 2SM, and the fundamental number is FN = 40. The total average length of all chromosomes (i.e., of the diploid karyotype) was estimated to be 82.51 ± 0.52 μm. The average chromosome length ranged from 5.77 ± 0.91 μm to 3.37 ± 0.4 μm (Table 3). The telomere distribution of the TTAGG(6) motif was displayed at both ends of all *M. iheringi* chromosomes (Fig. 2a). No signals for interstitial telomeric sites (ITS) were detected using this probe. Moreover, DAPI staining revealed that both arms of all chromosomes were completely labeled, i.e., mostly A-T rich, whereas the centromeric region showed no labeling for this fluorochrome (Fig. 2b).

Molecular data

The maximum likelihood phylogeny showed *M. iheringi* as the sister species of a lineage defined as *Mycetomoellerius* n.sp. nr *cirratus* (see Solomon et al. 2019) (bootstrap value, PB = 90). The clade composed of *M. cirratus* Mayhé-Nunes & Brandão, 2005 + *M. kempfi* (PB = 98) forms the sister group of *M. iheringi* + *M. n.sp. nr *cirratus* (PB = 88). The species *M. holmgreni* previously diverged from the aforementioned clades (PB = 89), and *M. papulatus* Santschi, 1922 was estimated to be the most basal of the “Iheringi group” (PB = 93) (Fig. 3).

Discussion

Here, we have provided the karyotypic description of the fungus-farming ant *Mycetomoellerius iheringi*, which has 2n = 20 chromosomes; we presented its phylogenetic
**Figure 1.** Mitotic metaphase of *Mycetomoellerius iheringi* with 2n = 20 chromosomes and its karyotypic morphology. M: metacentric chromosomes; SM: submetacentric chromosomes. Scale bar: 5 μm.

**Table 3.** Karyomorphometric analysis of the chromosomes of *Mycetomoellerius iheringi*. TL: total length; L: long arm length; S: short arm length; RL: relative length; r: arm ratio (= L/S); ∑: total average length of all chromosomes or Karyotype length (KL).

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>TL     ±</th>
<th>L     ±</th>
<th>S     ±</th>
<th>RL     ±</th>
<th>r     ±</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.77±0.91</td>
<td>3.03±0.48</td>
<td>2.74±0.43</td>
<td>6.97±0.34</td>
<td>1.1±0.05</td>
<td>Metacentric</td>
</tr>
<tr>
<td>2</td>
<td>5.46±0.75</td>
<td>2.86±0.46</td>
<td>2.6±0.32</td>
<td>6.61±0.24</td>
<td>1.1±0.08</td>
<td>Metacentric</td>
</tr>
<tr>
<td>3</td>
<td>5.09±0.66</td>
<td>3.02±0.41</td>
<td>2.08±0.27</td>
<td>6.17±0.29</td>
<td>1.46±0.09</td>
<td>Metacentric</td>
</tr>
<tr>
<td>4</td>
<td>4.71±0.53</td>
<td>2.67±0.29</td>
<td>2.04±0.28</td>
<td>5.72±0.34</td>
<td>1.32±0.12</td>
<td>Metacentric</td>
</tr>
<tr>
<td>5</td>
<td>4.38±0.49</td>
<td>2.38±0.29</td>
<td>1.99±0.29</td>
<td>5.31±0.2</td>
<td>1.21±0.18</td>
<td>Metacentric</td>
</tr>
<tr>
<td>6</td>
<td>4.2±0.46</td>
<td>2.3±0.23</td>
<td>1.91±0.27</td>
<td>5.1±0.15</td>
<td>1.22±0.14</td>
<td>Metacentric</td>
</tr>
<tr>
<td>7</td>
<td>4.07±0.46</td>
<td>2.24±0.2</td>
<td>1.83±0.33</td>
<td>4.94±0.16</td>
<td>1.26±0.21</td>
<td>Metacentric</td>
</tr>
<tr>
<td>8</td>
<td>4.01±0.44</td>
<td>2.3±0.26</td>
<td>1.72±0.26</td>
<td>4.87±0.16</td>
<td>1.32±0.19</td>
<td>Metacentric</td>
</tr>
<tr>
<td>9</td>
<td>3.89±0.43</td>
<td>2.19±0.3</td>
<td>1.7±0.18</td>
<td>4.72±0.11</td>
<td>1.31±0.14</td>
<td>Metacentric</td>
</tr>
<tr>
<td>10</td>
<td>3.83±0.45</td>
<td>2.16±0.3</td>
<td>1.67±0.17</td>
<td>4.65±0.06</td>
<td>1.3±0.11</td>
<td>Metacentric</td>
</tr>
<tr>
<td>11</td>
<td>3.78±0.43</td>
<td>2.15±0.28</td>
<td>1.63±0.2</td>
<td>4.59±0.1</td>
<td>1.32±0.15</td>
<td>Metacentric</td>
</tr>
<tr>
<td>12</td>
<td>3.73±0.41</td>
<td>2.07±0.3</td>
<td>1.66±0.15</td>
<td>4.53±0.15</td>
<td>1.25±0.15</td>
<td>Metacentric</td>
</tr>
<tr>
<td>13</td>
<td>3.7±0.39</td>
<td>2.03±0.26</td>
<td>1.67±0.19</td>
<td>4.5±0.14</td>
<td>1.22±0.14</td>
<td>Metacentric</td>
</tr>
<tr>
<td>14</td>
<td>3.66±0.4</td>
<td>2.08±0.24</td>
<td>1.58±0.2</td>
<td>4.44±0.13</td>
<td>1.33±0.14</td>
<td>Metacentric</td>
</tr>
<tr>
<td>15</td>
<td>3.58±0.35</td>
<td>2.01±0.28</td>
<td>1.57±0.13</td>
<td>4.35±0.13</td>
<td>1.29±0.17</td>
<td>Metacentric</td>
</tr>
<tr>
<td>16</td>
<td>3.54±0.38</td>
<td>2.01±0.26</td>
<td>1.54±0.17</td>
<td>4.3±0.12</td>
<td>1.32±0.16</td>
<td>Metacentric</td>
</tr>
<tr>
<td>17</td>
<td>3.51±0.4</td>
<td>2.04±0.19</td>
<td>1.47±0.25</td>
<td>4.26±0.13</td>
<td>1.41±0.16</td>
<td>Metacentric</td>
</tr>
<tr>
<td>18</td>
<td>3.37±0.4</td>
<td>1.94±0.29</td>
<td>1.43±0.12</td>
<td>4.09±0.11</td>
<td>1.36±0.13</td>
<td>Metacentric</td>
</tr>
<tr>
<td>19</td>
<td>4.29±1.1</td>
<td>2.74±0.68</td>
<td>1.56±0.42</td>
<td>5.15±0.72</td>
<td>1.77±0.06</td>
<td>Submetacentric</td>
</tr>
<tr>
<td>20</td>
<td>3.94±0.59</td>
<td>2.51±0.37</td>
<td>1.43±0.22</td>
<td>4.76±0.25</td>
<td>1.76±0.03</td>
<td>Submetacentric</td>
</tr>
<tr>
<td>∑</td>
<td>82.51±0.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

position in the clade of the “Iheringi group”. Considering the cytogenetic data available from fungus-farming ants, we observed a numerical constancy among the karyotypes of the lineages that diverged most recently (i.e., leafcutter ants of the genera *Atta* and *Acromyrmex*), suggesting this karyotypic characteristic is shared by the relatively recent lineages. *Trachymyrmex septentrionalis*, a sister clade of leafcutter ants, has 2n = 20 metacentric chromosomes, equal to those of two *Mycetomoellerius* species, *M. holmgreni* and *M. relictus* Borgmeier, 1934 (see Table 1). All *Atta* species karyotyped to
Figure 2. DAPI-stained Mycetomoellerius iheringi chromosomal metaphases a FISH mapping of the TTAGG(6) telomeric motif on haploid metaphase b chromosomes uniformly stained with DAPI fluorochrome, except for the centromeric region. Scale bar: 5 μm.

Figure 3. Maximum-likelihood phylogeny of “higher” fungus-farming ants generated in RAxML. Mycetomoellerius iheringi is indicated in red. Node numbers represent the bootstrapping values after 1000 replications; values < 80 are not shown. Scale bar indicates nucleotide substitutions per site.
date have 2n = 22 chromosomes, and most *Acromyrmex* species have 2n = 38 (reviewed by Cardoso et al. 2018a). In other Hymenoptera species, such as stingless bees of the tribe Meliponini Lepeletier, 1836, this scenario can also be seen in the genera with a conserved chromosome number (Travenzoli et al. 2019).

In the new taxonomic status, *Mycetomoellerius* is composed of about 30 described species (Solomon et al. 2019), but only four have known karyotypes and, interestingly, a prevalence of metacentric chromosomes (see Table 1). The species *M. iheringi* and *M. holmgreni* are closely related morphologically (Mayhé-Nunes and Brandão 2005), and, as we have shown, *M. holmgreni* diverged previously from *M. iheringi*. Moreover, both species co-occur in southern Brazilian sand-dune habitats (Cardoso and Schoederer 2014). Importantly, the karyotypes of these two species are similar: they have analogous karyotype measurements and DAPI-staining pattern as well the chromosomal number 2n = 20, differing by only one pair of submetacentric chromosomes (Barros et al. 2018; Cardoso et al. 2018b). A likely, and the most parsimonious, scenario for explaining such cytogenetic differences would involve at least one major chromosomal rearrangement. Therefore, we suggest a pericentric inversion occurred in one of the larger *M. iheringi* chromosomes, resulting in the current karyotype morphology. Such chromosomal rearrangement could have occurred in any lineage of the clades underlying *M. holmgreni*; however, such lineages should be karyotyped to verify this hypothesis. The base chromosome number, defined as the haploid number present in the initial lineage of a monophyletic clade, may be directly related to the chromosomal variability within that clade (Guerra 2008). Thus, the assumption of this major inversion is attributable to the fact that *M. holmgreni* has a karyotype formed by only metacentric chromosomes, which becomes a putative ancestral characteristic of the underlying lineages, such as *M. iheringi*.

The application of classical and molecular cytogenetic techniques, such as chromosomal banding and FISH mapping, has increasingly contributed to comparative evolutionary studies. Because of new ant cytogenetic data, valuable information is being collected and correlated to their evolution and exceptional chromosomal diversity. For instance, fusion and fission rearrangements have been proposed to play a crucial role in the diversification of the fungus-farming ants of the genus *Mycetophylax* Emery, 1913 (Cardoso et al. 2014; Micolino et al. 2019b). Indeed, chromosomal changes may be directly related to the speciation process for a range of taxa (Rieseberg 2001; Faria and Navarro 2010). In particular, inversions are abundant in natural populations and can have several evolutionary implications, such as adaptation and divergence of lineages (Ayala and Coluzzi 2005; Wellenreuther and Bernatchez 2018). Inversion polymorphisms may contribute to speciation by reducing recombination and consequently protecting genomic regions from introgression (Hoffmann and Rieseberg 2008). Moreover, a model has predicted that closely related lineages that co-occur in a region could readily differ by one or more inversions because such lineages would persist longer in the face of gene flow than in the absence of these inversions (Noor et al. 2001). Our data support such a model, mainly because the species *M. iheringi* and *M. holmgreni* live sympatrically and are phylogenetically close.
The rich karyotypic diversity of ants deserves special attention. Inversion polymorphisms, for example, have been reported in many ant species. For example, intrapopulational polymorphism has been detected in the *Iridomyrmex gracilis* Lowne, 1865 complex. Such populations with the same chromosome number but distinct karyotype structures have led authors to propose that a pericentric inversion occurred in a metacentric chromosome, making it acrocentric (n = 6M + 1SM + 1A to n = 5M + 1SM + 2A) (Crozier 1968). The chromosome number and morphology of *Pachycondyla* Smith, 1858 are variable; their karyotypes show a predominance of submetacentric and acrocentric chromosomes, which allows the interpretation that fission and pericentric inversions (where metacentric chromosomes turn acrocentric or vice versa) would be the most frequent chromosomal rearrangements in the evolution of this genus and even contribute to the speciation processes (Mariano et al. 2012). The intraspecific chromosomal variability in social organization (monogyny vs. polygyny) found in the fire ant *Solenopsis invicta* Buren, 1972 can also be explained by at least one large inversion, which would account for a lack of recombination over more than half of the two heteromorphic “social chromosomes” (Wang et al. 2013).

Another interesting finding was reported in *Mycetomoellerius fuscus* Emery, 1894 (current junior synonym of *M. urichii* Forel, 1893, see Micolino et al. 2019a for discussion), a species with a geographic distribution similar to *M. iheringi* and *M. holmgreni* and found largely in southern South America (Brandão and Mayhé-Nunes 2007). They are phylogenetically closer than previously expected (Micolino et al. 2019a; Solomon et al. 2019). *Mycetomoellerius fuscus* has a chromosomal morphology of eight metacentric pairs and a submetacentric pair (2n = 18) (Barros et al. 2013a). As the submetacentric pair is the biggest chromosome of the karyotype, there could have been a Robertsonian fusion rearrangement, followed by a pericentric inversion, making it submetacentric. The other few species of “*Trachymyrmex*” with the described karyotype (see Table 1) do not allow us to picture a full scenario for the karyoevolution of the genera. Further, unidentified specimens vary relatively widely from 2n = 12 to 2n = 22. The karyotype 2n = 12 presented by Murakami et al. (1998) is quite intriguing, as this unidentified specimen could be a key piece to understanding the chromosomal evolution of the clade to which it belongs. We emphasize that specimens submitted for cytogenetic analysis should be taxonomically identified. The non-identification of a specific sample triggers a series of problems, such as in the comparison with sister groups and eventual karyoevolutionary trajectories.

Our karyomorphometric approach was used primarily to reveal the chromosomal morphology of *M. iheringi*. Besides, future karyomorphometric comparisons among populations or even closely related lineages may serve as a basis for a possible delimitation of incipient lineages. For example, populations of *M. holmgreni* distributed on a North/South continuum of its distribution area diverged significantly in the length of their chromosomes, and the results were supported by flow cytometry analyses of the genome size (Cardoso et al. 2018b). Further, those populations were later identified to differ in the proportion of repetitive DNA by using FISH with microsatellite probes (Micolino et al. 2019a) Thus, the authors demonstrated the importance of using a
standardized karyomorphometric approach coupled with genome size estimation to identify hidden chromosomal variations (see Cardoso et al. 2018b).

Finally, we used a FISH probe of the highly conserved TTAGG telomeric sequence in most insects (reviewed by Kuznetsova et al. 2020) to test the assumption that the putative inversion rearrangement occurred in *M. iheringi* and involved the telomere. However, we did not observe any signal for the probe at the interstitial telomeric sites, which would denote inversion involving the telomere. Indeed, the TTAGG sequence also seems to be fairly conserved in ants (Lorite et al. 2002), including fungus-farming ants such as *Acromyrmex striatus* Roger, 1863 (Pereira et al. 2018), *Mycetophylax* spp. (Micolino et al. 2019b), and *M. holmgreni* (Micolino et al. 2019a). In conclusion, we have described another ant species with the TTAGG sequence conserved in its telomeres, and we suggest a significant chromosomal mechanism, a major pericentric inversion, most likely occurred in *M. iheringi* and could have been involved in its diversification process.

**Acknowledgements**

We are grateful to many people who made this work possible. We thank all our colleagues at the Lab and Research Group of Genetics and Evolution of Ants (GEF-UFOP) for their help with the data. We are also grateful for the financial support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – MPC fellowship 309579/2018-0, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Estado do Paraná, and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). The sample collection was authorized by the “Instituto Chico Mendes de Conservação da Biodiversidade” – ICMBio (Special permit number 60019).

**References**


Cristiano MP, Pereira TTP, Simões LP, Sandoval-Gómez VE, Cardoso DC (2017) Reassessing the chromosome number and morphology of the turtle ant *Cephalotes pusillus* (Klug, 1824) using karyomorphometrical analysis and observations of new nesting behavior. Insects 8: 1–114. https://doi.org/10.3390/insects8040114


Induction and evaluation of colchitetraploids of two species of Tinospora Miers, 1851

Rakesh Kr. Thakur¹, Vijay Rani Rajpal², Satyawada Rama Rao³, Apekshita Singh¹, Lata Joshi¹, Pankaj Kaushal⁴, Soom Nath Raina¹

¹ Amity Institute of Biotechnology, Amity University, Noida, Uttar Pradesh, 201313, India ² Department of Botany, Hansraj College, University of Delhi, Delhi, 110007, India ³ Department of Biotechnology & Bioinformatics, North Eastern Hill University, Shillong, Meghalaya, 793022, India ⁴ ICAR- Indian Grassland and Fodder Research Institute, Jhansi, Uttar Pradesh, 284003, India

Corresponding author: Soom Nath Raina (soomr@yahoo.com)

Academic editor: E. Mikhailova | Received 26 January 2019 | Accepted 11 February 2020 | Published 20 May 2020

http://zoobank.org/180D107F-6923-4735-AF80-6D3472B1E3A4


Abstract

Autotetraploidy, both natural and/or induced, has potential for genetic improvement of various crop species including that of medicinal importance. Tinospora cordifolia (Willdenow, 1806) Miers, 1851 ex Hooker et Thomson, 1855 and T. sinensis (Loureiro, 1790) Merrill, 1934 are two diploid species, which are dioecious, deciduous and climbing shrubs with high medicinal importance. Among the three methods used for induction of polyploidy by colchicine treatment, it was cotton swab method which successfully induced the polyploidy in both species. The morphological and cytogenetical features of the synthetic tetraploids were compared with their diploid counterparts. The tetraploids were morphologically distinct from diploid plants. They exhibited larger organs, such as stem, leaves, inflorescence, fruits, flowers and seeds. The tetraploids were characterized by the presence of low quadrivalent frequency and high bivalent average. Unequal distribution of chromosomes at anaphase I was found in 60% cells. The present study provides important information on the superiority of autotetraploids as compared to diploid counterparts in both species.

Keywords
colchicine treatment, cytogenetics, flow cytometry, morphology, polyploidy, Tinospora cordifolia, Tinospora sinensis
Introduction

Polyploidy, the presence of more than two sets of chromosomes, has played a pivotal role in the diversity, evolution, genetic improvement and speciation of both wild and cultivated plants (Sattler et al. 2016). More than 70% angiosperms have polyploid ancestry (Masterson 1994, Soltis et al. 2014). Autopolyploidy involves multiplication of the same genome (Comai 2005) while allopolyploidy is the combination of the genomes of two or more taxonomically divergent species. Polyploidy directly impacts the nucleotype, morphology, physiology, genetics, and biochemistry of the plant (Raina et al. 1994, Hull-Sanders et al. 2009). The induction of polyploidy in the plant species by colchicine treatment has successfully been utilized to improve the yield and quality of some of the commercially important crops such as sugar beet, watermelon, red clover, rye, rye grass, grapes and several ornamental, horticultural and medicinal plants (Sattler et al. 2016). Due to increase in cell size, autopolyploidy is often associated with thicker and broader leaves, large flowers and seeds, making the plant appear robust and display characteristic features of gigantism (Levin 2002). Due to aberrant meiosis and resultant low seed set, induced autopolyploidy has been considered relatively more rewarding in such plants where vegetative or floral parts have commercial value and the plant propagates by vegetative means (Lavania 2005). Induced polyploidy may also lead to enhanced production and qualitative changes in secondary metabolites due to perceived increase in number of gene copies and probably the enzyme content of polyploids (Dhawan and Lavania 1996, Sattler et al. 2016).

Genus *Tinospora* includes 34 species distributed widely throughout the tropical and subtropical parts of Asia, Africa and Australia. Many of them are well known for their medicinal importance (Pathak et al. 1995, Chi et al. 2016). Three species are reported from India, *Tinospora cordifolia* (Willdenow, 1806) Miers, 1851 ex Hooker et Thomson, 1855, *T. sinensis* (Loureiro, 1790) Merrill, 1934 and *T. crispa* (Linnaeus, 1763) Hooker & Thomson, 1855. All of them are diploid (2n = 2x = 26), woody climbers and are dioecious. *Tinospora cordifolia*, commonly known as giloe, is a well-known medicinal plant species in ayurvedic and folk system of Indian medicine. *Tinospora cordifolia* has anticancer, antimalarial, antiglycemic, antioxidant, antipyretic, hepatoprotective, immunomodulator, anti-inflammatory, diuretic and hyperglycemic properties (Singh et al. 2003, Sinha et al. 2004, Mangal et al. 2012). *T. sinensis* has also immuno-modulator, anti-inflammatory, hyperglycemic and anti-leishmanial properties (Akram et al. 2014). Many herbal products from the species are available in the market (Mittal et al. 2014).

The present study deals with the induction, for the first time, of autotetraploidy in *T. cordifolia* and *T. sinensis* and their morphological and cytogenetical features in comparison to their diploid counterparts.

Material and methods

The stem cuttings and seeds of two plants (one male and one female) of *T. cordifolia* were collected from Central Institute of Aromatic and Medicinal Plants (CIMAP),
Induction and evaluation of colchitetraploids of two species of Tinospora Miers, 1851

213

Lucknow, India. The two plants (one male and one female) of T. sinensis were collected from surrounding forests of Shivaji University, Kolhapur, Maharashtra, India. The authenticity of the plant material of T. cordifolia and T. sinensis was duly verified by taxonomists at CIMAP and Department of Botany, Shivaji University, respectively. The voucher specimens were deposited in herbarium of Department of Botany, North Eastern Hill University, Shillong, India and accession numbers were obtained. The accession numbers allocated by the herbarium are NEHU-12091, NEHU-12092 for T. cordifolia and NEHU-12093 and NEHU-12094 for T. sinensis.

Colchicine treatment

Colchicine treatment was given to 2600 seeds/seedlings/vegetative buds of T. cordifolia and T. sinensis (Table 1). Three methods of colchicinization were employed with slight modifications in the protocols of Srivastav and Raina (1981) and Kushwah et al. (2018).

a. Seed treatment: Seeds of T. cordifolia and T. sinensis were immersed in 0.1% and 0.15% aqueous colchicine (Sigma-Aldrich) for 12 h and 24 h. After the treatment, the seeds were thoroughly washed in double distilled water and sown in pots with soil.

b. Vegetative bud treatment: Sterilized cotton balls immersed either in 0.1 or 0.15, or 0.2 % colchicine were placed on the growing buds of T. cordifolia and T. sinensis of ~ one year old rooted stem cuttings for 6 h each for 3 consecutive days.

c. Cotton swab method: Seeds were germinated in pots containing loamy soil and the protruding apical meristem tips between two cotyledonary leaves of ~ 5 days old seedlings were immersed in 0.1 or 0.15, or 0.2 % colchicine with the help of cotton swab soaked in colchicine, for 6 h each for 2, 3, 4 or 5 consecutive days. The colchicine solution was intermittently dropped on the swab to maintain the same colchicine concentration.

The colchicine treatment in all the three methods were carried out in growth chamber maintained at 27 °C, 60% humidity and photoperiod of 12 h duration. Treatment with distilled water of seeds/buds/apical meristem served as control. The pots containing treated and control seedlings/stem cuttings were transferred to glass house one month after treatment.

Stomatal analysis

Stomatal analysis was conducted in 633 plants of T. cordifolia and T. sinensis which survived after treatment and were transferred to glass house. Lower epidermal peel of the control and colchicine treated plants were mounted side by side on the same slide in drops of water and covered with coverslips (24 mm × 24 mm). Stomata cells of the control and the treated plants were observed under a microscope for obtaining data on the comparative size and number of stomata per unit area by Q CAPTURE PRO
5.0 software (QImaging, Surrey, Canada). Initially, the treated plants with distinct increase in size of stomata and low number of stomata per unit area were earmarked as tetraploids (Table 2). The treated plants which showed no change in size and number of stomata per unit area compared to control were considered as diploids.

In *T. cordifolia*, 14 plants which showed distinct increase in stomatal size and 41 randomly chosen treated plants which had no change in the stomata size, as well as 20 control plants after 45 days in glass house were transferred to experimental field containing loamy soil. In *T. sinensis*, 8 plants with distinct increase in stomatal size and 7 treated plants with no change in stomatal size, along with 10 control plants were transferred to experimental field.

**Flow cytometry**

The material for which flow cytometric analysis was carried out was used as a diploid control for colchicine treated (70) plants transferred to experimental field. Healthy young leaves (ca. 2 cm²) each from the sample and internal standard were chopped together with sharp razor blade for isolation of nuclei, stained in extraction and staining buffer (2 ml) containing 100 mM Tris HCl, 85 mM NaCl, 5 mM MgCl₂, 0.1% Triton X 100 and 1μg/ml DAPI (4’,6-diamidino-2-phenylindole ) pH 7.0. The solution was filtered through 30 μm nylon mesh and analysed in flow cytometer (FCM) (BD FACS Canto 11, BD Biosciences, San Jose, CA) equipped with software CA3 2.14/2004. Minimum 3000 nuclei were analysed per run. Coefficient of variation of G₀/G₁ peak up to about 4% was only accepted. Each sample was repeated at least thrice for ploidy estimation. *Pennisetum squamulatum* Fresenius, 1837 (2C = 7.26 pg) (Kaushal et al. 2010) was used as internal standard for relative DNA content measurement of the sample plants. FCM histograms were visualized in linear phase for the comparison between peak positions of the standard and the samples.

**Morphological analysis**

The data for morphological analysis was taken two years after field transplant of the control and colchicine treated plants. As mentioned before, at the time of colchicine treatment, the seedlings treated with distilled water instead of aqueous colchicine were grown to maturity. They served as control plants. Six control and 14 tetraploid plants of *T. cordifolia* and six control and 8 tetraploid plants of *T. sinensis* were evaluated for morphological features (Table 2). All these plants at the time of taking morphological data were fully matured bearing flowers and seeds. The data for each phenotypic trait among the control and corresponding tetraploid plants were averaged and standard error (SE) calculated (Table 2). The thickness of the stem was measured 90 cm above the ground. The sixth to tenth (five in number) fully expanded leaf counting from the tip of fifth side branch from the top of the main stem were measured for various leaf characters for each of the diploid (control) and colchitetraploid plants (tetraploidy was induced by colchicine treatment).
Table 1. Frequency of induced tetraploidy by colchicine treatment in *Tinospora cordifolia* and *T. sinensis*.

<table>
<thead>
<tr>
<th>Species</th>
<th>Seed treatment method</th>
<th>Bud treatment method</th>
<th>Cotton swab method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Concentration of colchicine (%)</td>
<td>No. of seeds treated</td>
<td>Duration of treatment (in h)</td>
</tr>
<tr>
<td><em>T. cordifolia</em></td>
<td>0.10</td>
<td>150</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>150</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total</td>
<td>300</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td><em>T. sinensis</em></td>
<td>0.10</td>
<td>150</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>150</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total</td>
<td>300</td>
<td>0</td>
<td>300</td>
</tr>
</tbody>
</table>

* Duration treatment spread equally for each day

** Identified following stomatal, flow cytometry and meiosis analysis
Table 2. Comparison of average morphological/micro and macroscopic characters of diploid and colchitetraploids of *Tinospora cordifolia* and *T. sinensis*.

<table>
<thead>
<tr>
<th>Characters</th>
<th><em>Tinospora cordifolia</em></th>
<th><em>Tinospora sinensis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Diploid (2n=2x=26)</td>
<td>Colchitetraploid (2n=4x=52)</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>No. of plants</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Thickness of stem (cm, circumference, 90 cm above the ground)</td>
<td>2.45 ± 0.13&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4.2 ± 0.45&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Length of leaf (cm)</td>
<td>4.81 ± 0.20&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4.57 ± 0.08&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Width of leaf (cm)</td>
<td>5.67 ± 0.22&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5.27 ± 0.21&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Length of petiole (cm)</td>
<td>4.28 ± 0.31&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.6 ± 0.33&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Number of stomata per unit area (/mm&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>75.14 ± 11.76&lt;sup&gt;a&lt;/sup&gt;</td>
<td>62.00 ± 4.65&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Length of stomata (μm)</td>
<td>23.82 ± 1.09&lt;sup&gt;a&lt;/sup&gt;</td>
<td>23.03 ± 0.40&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Width of stomata (μm)</td>
<td>21.10 ± 0.73&lt;sup&gt;a&lt;/sup&gt;</td>
<td>18.58 ± 0.76&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Length of Inflorescence (cm)</td>
<td>2.97 ± 0.30&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.55 ± 0.16&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Flowering period</td>
<td>February–March</td>
<td>February–March</td>
</tr>
<tr>
<td>Number of fruits per inflorescence</td>
<td>–</td>
<td>13.5 ± 1.40&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Fruit size (mm)</td>
<td>–</td>
<td>2.59 ± 0.13&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Seed weight (g/10 seeds)</td>
<td>–</td>
<td>0.45 ± 0.03&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Pollen grain size (μm)</td>
<td>16.22 ± 0.66&lt;sup&gt;a&lt;/sup&gt;</td>
<td>–</td>
</tr>
<tr>
<td>Pollen stainability %</td>
<td>90</td>
<td>60</td>
</tr>
<tr>
<td>Seed germination %</td>
<td>–</td>
<td>50</td>
</tr>
</tbody>
</table>

<sup>a</sup> denotes significant (p < 0.05) morphological variation between corresponding diploid male/female and colchitetraploid male/female plants. The corresponding values without <sup>a</sup> denote no significant variation.
Male meiosis

For meiotic studies, young flower buds of appropriate size were fixed at least for 24 h in freshly prepared acetic-ethanol (1:3) mordanted with saturated FeCl₃ solution. A saturated solution of FeCl₃ was prepared by dissolving substantial amount of FeCl₃ in 10 ml of distilled water. A small drop of FeCl₃ solution was added to 100 ml of acetic-ethanol mixture. The acetocarmine moderated with FeCl₃ increases the intensity of the stain in chromosomes. Before the anthers of appropriate size were used for meiotic analysis, they were hydrolysed in 1N HCl at 60 °C for 10 min and then stained in Feulgen solution. The stained anthers were subsequently squashed in 1% iron-acetocarmine to observe various stages of male meiosis. Photomicrographs were taken using Olympus CX40 Microscope fitted with 01-GO-3, QIMAGING camera. Twenty five meiocytes each showing metaphase I and anaphase I stages were analysed in each of the two diploid *T. cordifolia* and the two *T. sinensis* plants. The same number of meiocytes were analysed in three colchitetraploids each of *T. cordifolia* and *T. sinensis*.

Pollen fertility judged by its stainability

For pollen stainability, pollen grains about to dehisce anthers of the diploid and confirmed autotetraploids were separately immersed in a drop of 1:1 ratio of 1% acetocarmine and glycerine on the microslide and covered with a cover slip (22 mm × 22 mm). They were kept as such for 2 h at room temperature. The slide was then observed under the microscope for the number of pollen grains with intense stain and pollen grains with no stain or less stain. Those pollen grains which were intensely stained and circular were taken as fertile pollen, and those with less stain and crinkled shape were considered sterile. Approximately 500 pollen grains both for diploid and autotetraploid plants were analysed for pollen stainability for each species.

Statistical analysis

The SPSS ver. 22 statistical software (IBM SPSS Amos™ 22; IBM Corp. Released 2013) was used to assess the variation of phenotypic traits within and between the populations of diploid and colchitetraploid using t-test and one-way ANOVA.

Results

Efficiency of colchicine treatment

Thirteen hundred seeds/seedlings/vegetative buds each of *T. cordifolia* and *T. sinensis* were treated with three different concentrations (0.1, 0.15 and 0.2%) of aqueous col-
chicine for 6 or 12 h each for 2, 3, 4 or 5 days (Table 1). As is clear from Table 1, not a single seed/vegetative bud survived after colchicine treatment. On the other hand, several seedlings treated by means of cotton swab method survived till maturity and among these some were found to be tetraploids. Further, 0.2% colchicine treatment for more than 2 days proved to be lethal. 0.15% colchicine treatment for 18 h, spread over three days, was found to be the most effective method for induction of polyploidy in *T. cordifolia* and *T. sinensis*. Out of 700 seedlings each in *T. cordifolia* and *T. sinensis* treated by cotton swab method, 349 and 284 seedlings survived (Table 1) and out of these, based on flow cytometry and male meiosis, 14 (~4%) and 8 (~2.8%) were found to be colchitetraploid plants, respectively.

**Flow Cytometry in relation to stomatal analysis**

Fourteen plants in *T. cordifolia* and 8 plants in *T. sinensis* which were given colchicine treatment, and which exhibited distinct increase in the size of stomata (Figs 1a, b; 2a, b) had twice the DNA amount compared with the diploid control (Fig. 3 a-d). This clearly indicated the induction of autotetraploidy in these plants. The chromosome counts of these plants made at metaphase I and anaphase I confirmed that these plants were indeed tetraploids with 2n = 52 (Figs 4, 5). The 48 plants of *T. cordifolia* and *T. sinensis* treated with colchicine but with no change in the size of stomata were found to have DNA amount equivalent to the diploid control indicating thereby the induction of polyploidy was not successful in these plants.

**Morphology**

The characteristic feature of all the apical meristems of buds/seedlings treated with colchicine was stunted growth in initial stages and leathery thicker first leaves. After first 3–4 leaves the subsequent leaves in the seedlings showed either normal or thicker, darker and larger leaves. The plants with latter condition on further study were found to be tetraploids. Following cotton swab method, the same morphological condition (normal or thicker, darker and larger leaves) as above was observed in all the colchicine concentrations and duration of treatment.

The colchitetraploids compared to diploid plants were morphologically distinct in several characters (Figs 1a–i; 2a–f; Table 2). The variation between the diploid and colchitetraploid counterparts in various characters was either significant (p < 0.05) or not significant (Table 2). The commercially most important phenotypic traits like thickness of stem, length and width of leaves, and length of petiole (only in female) showed significant (p < 0.05) increase in size in male colchitetraploid compared to male diploid, and female colchitetraploid in comparison to female diploid *T. cordifolia*. The interesting feature about the length of petiole in diploid compared to tetraploid *T. cordifolia* male was reduced length in tetraploid plants and this variation was significant
Figure 1. Comparison between diploid (left) and colchitetraploid (right) *T. cordifolia* for a, b stomata c leaf d seed e male inflorescence f female inflorescence g, h pollen and i fruit. Scale bars: 10 μm.
Figure 2. Comparison between diploid (left) and colchitetraploid (right) *T. sinensis* for a, b stomata c leaf d male inflorescence and e, f pollen. Scale bars: 10 μm.

(p < 0.05). In *T. sinensis*, since no female colchitetraploid plant could be recovered, the comparison was made only for male diploid and male colchitetraploid plants. Between the two, there were significant (p < 0.05) differences in length and width of leaves. Thickness of stem and length of petiole did not show significant differences. As expected, the determinate organs, stomata and pollen grains, exhibited significant (p < 0.05) variation between respective sexes for diploid and colchitetraploid plants of *T. cordifolia* and *T. sinensis*. Barring determinate organs (stomata and pollen grains), the eleven male colchitetraploid plants of *T. cordifolia* showed significant (p < 0.05) differences in the remaining phenotypic traits. The stomata size, number of stomata per unit
area and pollen grain size did not show significant variation between 11 plants. Similar observation was made in relation to female colchitetraploid plants of *T. cordifolia* as well as male colchitetraploid plants of *T. sinensis*. All the tetraploid plants were, however, individually distinct from their diploid counterparts.

**Male meiosis study**

The data pertaining to meiotic analysis of diploids and colchitetraploids of two species *T. cordifolia* and *T. sinensis* is given in Tables 3 and 4. The chromosome preparations of different stages of meiosis are illustrated in Figs 4(a–g) and 5(a–f).

**Tinospora cordifolia**

**Diploid (2n = 2x = 26):** In majority of the PMCs observed at metaphase I, thirteen bivalents were regularly observed to occur. Few cells had a mix of both bivalents and univalents. On an average the PMC had 12.44 bivalents and 1.12 univalents. All the cells analysed at anaphase I were characterized by equal distribution (13:13) of chromosomes at two poles.

**Colchitetraploid (2n = 4x = 52):** The PMCs were characterized by the presence of quadrivalents, trivalents, bivalents and univalents at metaphase I. On an average per cell each PMC had 5.88 IV + 0.16 III+ 12.48 II and 4.16 I. Equal (26:26) distribution of chromosomes at anaphase I was found only in 40% of cells followed by unequal [27:25, 28:24 and 24:4U (Univalents):24] distribution of chromosomes in 60% cells.

**Tinospora sinensis**

**Diploid (2n = 2x = 26):** Most of the PMCs observed at metaphase I had thirteen bivalents. A few cells had both bivalents and univalents. The average frequency per cell of chromosome associations was 12.24 II+1.52 I. The presence of univalents in the diploid *T. cordifolia* and *T. sinensis* could be due to precocious separation of rod bivalents (Verma and Raina 1980). All the cells analysed at anaphase I were characteristic in having equal (13:13) distribution of chromosomes.
Figure 4. Metaphase I and anaphase I in a, b diploid (2n=2x=26) and c–g tetraploid (2n=4x=52) T. cordifolia. Note a 13 II and b 13:13 distribution of chromosomes at anaphase I. Note quadrivalents, trivalents, bivalents and univalents in c (5IV+13II+6I) d (10IV+1III+3II+3I) and e–g 26:26 distribution of chromosomes at anaphase I. Scale bar: 10 μm.
Figure 5. Metaphase I and anaphase I in a, b diploid (2n=2x=26) and c–f tetraploid (2n=4x=52) T. sinensis. Note a 13 II and b 13:13 distribution of chromosomes at anaphase I. Note quadrivalents, trivalents, bivalents and univalents in c (5IV+1III+10II+9I) d (10IV+5II+2I) and e, f 26:26 distribution of chromosomes at anaphase I. Scale bar: 10 μm.
**Table 3.** Average number and range of chromosome associations at metaphase I in the diploid (2x) and colchitetraploids (4x) *Tinospora cordifolia* and *T. sinensis*.

<table>
<thead>
<tr>
<th>Species</th>
<th>Ploidy</th>
<th>No. of cells analysed</th>
<th>Quadrivalents</th>
<th>Trivalents</th>
<th>Bivalents</th>
<th>Univalents</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Tinospora</em></td>
<td>2x</td>
<td>25</td>
<td>12.44; 10–13</td>
<td>1.12; 0–6</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>cordifolia</em></td>
<td>4x</td>
<td>25</td>
<td>5.88; 0–10</td>
<td>0.16; 0–1</td>
<td>12.48; 5–24</td>
<td>4.16; 0–16</td>
</tr>
<tr>
<td><em>T. sinensis</em></td>
<td>2x</td>
<td>25</td>
<td>12.24; 10–13</td>
<td>1.52; 0–6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4x</td>
<td>25</td>
<td>6.32; 3–10</td>
<td>0.24; 0–1</td>
<td>11.52; 5–20</td>
<td>3.28; 0–7</td>
</tr>
</tbody>
</table>

**Table 4.** Anaphase I distribution in diploid and colchitetraploids of *Tinospora cordifolia* and *T. sinensis*.

<table>
<thead>
<tr>
<th>Species</th>
<th>Ploidy</th>
<th>No. of Cells analysed</th>
<th>Chromosome distribution at anaphase I</th>
<th>No of cells (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Tinospora</em></td>
<td>2x</td>
<td>25</td>
<td>13:13</td>
<td>25(100)</td>
</tr>
<tr>
<td><em>cordifolia</em></td>
<td>4x</td>
<td>25</td>
<td>26:26; 27:25; 28:24; 24:4U:24</td>
<td>10(40)</td>
</tr>
<tr>
<td><em>T. sinensis</em></td>
<td>2x</td>
<td>25</td>
<td>13:13</td>
<td>25(100)</td>
</tr>
<tr>
<td></td>
<td>4x</td>
<td>25</td>
<td>26:26; 27:25; 28:24; 26:2U:24</td>
<td>10(40)</td>
</tr>
</tbody>
</table>

Univalents are indicated as U
The values in brackets denote fraction of cells

**Colchitetraploid (2n = 4x = 52):** Most of the PMCs had a mix of quadrivalents, trivalents, bivalents and univalents at metaphase I. On an average, each PMC had 6.32 IV + 0.24 III + 11.52 II and 3.28 I. Equal distribution (26:26) of chromosomes at anaphase I was recorded only in 40% of cells. The remaining 60% of the PMCs analysed had unequal (27:25, 28:24 and 26:2U:24) distribution of chromosomes.

**Discussion**

Among several protocols that have been developed for polyploidy induction, it is the colchicine treatment which has been the most successful procedure for last several decades. However, the induction of polyploidy by colchicine has been most successful in annuals rather than in perennial plants. There are hardly few among vast number of papers published on polyploid induction wherein successful induction in trees, shrubs and perennial climbers such as dioecious *Tinospora cordifolia* and *T. sinensis* has been reported (Lavania et al. 2012, Ramsey and Ramsey 2014, Sattler et al. 2016). The reasons for this aspect are unknown.

The success in induction of polyploidy in plants depends on many factors such as, treatment method, concentration of colchicine solution and duration of the treatment.
One could see on perusal of earlier literature that optimum colchicine concentration and duration of treatment differs from one species to other (Glowacka et al. 2009, Sarathum et al. 2010). In the present study, therefore, we took most widely used range of colchicine concentration and duration of treatment in three methods of colchicine treatment. Induction of tetraploidy in *T. cordifolia* and *T. sinensis* (first report) was successfully achieved only in cotton swab method when 0.15%/0.20% colchicine was applied for 12 h/18 h/24 h spread over 6 h each day. Twelve (55%) out of 22 tetraploids were recovered after treating the apical meristem with 0.15% colchicine for 18 h. Because colchicine treatment of certain concentration and duration in cotton swab method was effective in inducing polyploids in *T. cordifolia* and *T. sinensis*, it should also be effective in producing tetraploids in other medicinally important *Tinospora* species. The present study also indicated that compared to seed and growing vegetative bud treatment by colchicine, it is only the cotton swab method which was successful in polyploidy induction. The seed treatment method, possibly due to partial or complete check on root development and (or) germination (Liu et al. 2007), resulted in complete lethality. Similarly, none of the vegetative buds survived few days after the treatment. It is possible that the present combinations of concentration of colchicine and treatment duration inhibited further growth of vegetative buds.

There is a body of evidence to support that autopolyploidization leads to enhancement of morphological parameters (Zhang et al. 2008, Xu et al. 2010, Lin et al. 2011, Wu et al. 2012, Sattler et al. 2016) due to increase in cell size. There are also reports, though less in number, that increase in cell size does not always lead to increased size of the whole plant or its organs (Gaikwad et al. 2009, Cohen and Tel Zur 2012, Sattler et al. 2016). Our results regarding the morphological features of polyploidization in male and female *T. cordifolia* are in line with the published work that reports distinct larger organs compared to their diploid counterparts such as stem, leaf, inflorescence and seed. In *T. sinensis*, only male colchitetraploid plants were recovered. They had larger leaves and inflorescences. The thickness of stem did not show significant variation. The higher level of heterozygosity in autotetraploids of *T. cordifolia* and *T. sinensis* not only due to polysomic inheritance but also due to the species being dioecious leading to cross pollination will ensure better vigour increment in the tetraploids of both species. In several crop plants higher level of heterozygosity in autotetraploids has been positively correlated to vigour increment (Mendoza and Haynes 1974, Katepa-mupondwa et al. 2002).

The reduction in seed fertility in autotetraploids of *T. cordifolia* is of little consequence since the species is vegetatively propagated by stem cuttings. The multiplication through seed is rare almost non-existent. The increase in fruit size in autotetraploids, could be due to polyploidy induction and (or) reduce fruit load per plant. What is most important is that it is vegetative organs especially, stem and leaves, and not seeds which are medicinally important. Due to larger vegetative organs such as stem and leaves, the overall secondary metabolites production per unit area will substantially improve in autotetraploids of *T. cordifolia* and *T. sinensis*. Further, autotetraploids may positively affect the tolerance to some stresses such as nutrient deficiency, water deficit, temperature, drought, pests and pathogens (Levin 2002). On the face of it, therefore, *T. cordifolia*
and \textit{T. sinensis} are likely to outperform their diploid counterparts from the commercial point of view. Moreover, tetraploids obtained by chromosome doubling provide a wide platform for interploidy hybridization (Gmitter and Ling 1991, Zlesak et al. 2005). For example, tetraploids can be utilized in raising autotriploids which often exhibit heterotic effect. The tetraploids may also be important bridges for genetic transfer between \textit{T. cordifolia} and \textit{T. sinensis} in which direct crosses at diploid level may not be successful.

In autotetraploids due to occurrence of sets of 4 homologous chromosomes instead of 2 in diploids, all chromosome associations are expected to be of quadrivalent configuration. That is not, however, always the case in neoautotetraploids. The average number of quadrivalents per cell in \textit{T. cordifolia} and \textit{T. sinensis} was 5.88 and 6.32, respectively. The average number of bivalents in \textit{T. cordifolia} (12.48) and \textit{T. sinensis} (11.52) outnumbered the frequency of quadrivalents in the two tetraploid species. Such behaviour as in other neoautotetraploids, could be attributed to small size of chromosomes, cryptic structural hybridity and genetic control and (or) points of pairing initiation (Sybenga 1966, 1967, 1972, Srivastav and Raina 1987).

Conclusions

In conclusion, the present results demonstrate that cotton swab method was the best method for inducing polyploidy in the diploid \textit{Tinospora cordifolia} and \textit{T. sinensis}. Autopolyploidy of other \textit{Tinospora} species with medicinal potential may also be induced by this method. The autotetraploids of both species have many morphological features which would establish them as increasingly improved plant materials. The tetraploids can also be utilized for the production of triploids which usually offer heterotic advantage over its parents.

All authors declare that there is no conflict of interests exists. All the authors have contributed substantially to the manuscript and approved the submission.

Acknowledgements

The authors would like to thank National Medicinal Plants Board (NMPB), Ministry of AYUSH, Government of India for financial support. We also thank the anonymous reviewers and subject editor for comments on the ms and helpful suggestions.

References


Cytogenetic markers as a tool for characterization of hybrids of Astyanax Baird & Girard, 1854 and Hyphessobrycon Eigenmann, 1907

Caio Augusto Gomes Goes¹, Sandro Natal Daniel¹, Lucas Henrique Piva², George Shigueki Yasui², Roberto Ferreira Artoni³, Diogo Teruo Hashimoto⁴, Fausto Foresti⁵, Fábio Porto-Foresti¹

¹ Universidade Estadual Paulista (UNESP) “Júlio de Mesquita Filho”, Faculdade de Ciências, Edmundo Carrijo Coube Avenue, Bauru, SP, Brazil ² Centro nacional de Pesquisa e Conservação da Biota Aquática Continental (CEPTA-ICMBIO), Prefeito Euberto Nemésio Pereira Godói Highway, Pirassununga, SP, Brazil ³ Universidade Estadual de Ponta Grossa, Setor de Ciências Biológicas e da Saúde, Santos Andrade Square, Ponta Grossa, PR, Brazil ⁴ Universidade Estadual Paulista (UNESP) “Júlio de Mesquita Filho”, Centro de Aquicultura da UNESP, Prof. Paulo Donato Castelane Acess way, Jaboticabal, SP, Brazil ⁵ Universidade Estadual Paulista (UNESP) “Júlio de Mesquita Filho”, Instituto de Biociências, Prof. Montenegro Avenue, Botucatu, SP, Brazil

Corresponding author: Fábio Porto-Foresti (fp.foresti@unesp.br)

Abstract

Astyanax Baird et Girard, 1854, is one of the largest genera in the family Characidae and comprises 177 valid species. This genus has been the focus of cytogenetic studies primarily owing to the presence of B chromosomes and high karyotypic diversity among different populations. The intense genetic variability in Astyanax is one of the factors responsible for the occurrence of species complexes, which are groups (1) with certain difficulties in establishing common genetic pools or (2) belonging to different cryptic species. To evaluate cytogenetic marker inheritance and the possibility of the identification of these hybrids, this study aimed to describe cytogenetic hybrids from three strains of species of the genera Astyanax and Hyphessobrycon Eigenmann, 1908. A. lacustris Lürken, 1875, A. schubarti Britski, 1964, A. fasciatus Cuvier, 1819, and H. anisitsi Eigenmann, 1907 were used to generate three hybrid lineages. The diploid number, heterochromatin sites, and ribosomal genes (18S and 5S rDNA) of the parental strains and the hybrids were analyzed. The results indicated that the three hybrid lineages had cytogenetic markers of both par-
ents, presenting Mendelian inheritance. However, differences in distribution of heterochromatic blocks were observed between the hybrids and the parent strains. Our results allowed the identification of the hybrid strains based on the cytogenetic markers applied, reinforcing the efficiency of cytogenetic markers as tools for identification and indicating that such events may increase the karyotypic diversity in the genera *Astyanax* and *Hyphessobrycon*.

**Keywords**
neotropical fishes, B chromosomes, chromosome polymorphism, repetitive DNAs, species complex

**Introduction**

Interspecific hybridization is the union of distinct genetic pools, the progenies of which are usually individuals possessing intermediate taxonomic characteristics of both parental species (Mayr 1963). In fishes, hybridization is facilitated by reproductive peculiarities, such as external fertilization and sharing of spawning sites, which may eventually facilitate the occurrence of cross-fertilization and the emergence of hybrid strains (Hubbs 1955). Of note, sporadic cases of natural hybrids occur in Neotropical fish species (Artoni et al. 2006; Porto-Foresti et al. 2013; Hashimoto et al. 2014; Prado et al. 2017).

*Astyanax* Baird et Girard, 1854, belonging to the family Characidae, is one of the most species-rich genus and currently comprises 177 valid species (Eschmeyer and Fong 2020), known as tetras. The genus *Astyanax* is characterized by high phenotypic plasticity and a capacity to adapt to diverse habitats (Ornelas-Garcia et al. 2008). Cytogenetic data available for this genus reveal wide karyotypic diversity with exclusive chromosomal features of some species and populations, such as the presence of heterochromatin polymorphisms and distinct patterns of repetitive DNA dispersion (Mantovani et al. 2000; Almeida-Toledo et al. 2002; Kantek et al. 2009; Hashimoto and Porto-foresti 2010; Hashimoto et al. 2011; Utsunomia et al. 2017). These intense genetic polymorphisms result in several “species complexes,” described as a cluster of closely related populations, the individuals of which may represent more than one species (Fegan and Prior 2005). In the genus *Astyanax*, species complexes have been described in at least four species: *A. scabripinnis* Jenyns, 1842 (Moreira-Filho 1991), *A. lacustris* Lütken, 1875 (Fernandes and Martins-Santos 2004), *A. fasciatus* Cuvier, 1819 (Artoni et al. 2006), and *A. bimaculatus* Linnaeus, 1758 (Garutti and Langeani 2009). In these cases, different natural isolated populations of individuals with similar morphology considered as a unique species may not share the same cytogenetic markers or diploid number. In these cases, it is very difficult to define whether they share the same gene pool or if they are different cryptic species. In addition to the intense chromosomal polymorphisms, the possibility of the occurrence of hybrids in the natural environment can increase karyotypic diversity and complicate the accurate identification of the animals.

There has been a report of interspecific hybridization among *Astyanax* species in the nature (Pazza et al. 2006). Thus, the occurrence of natural hybrids in *Astyanax* pop-
ulations is a factor to be considered in the cytogenetic studies concerning this genus. Considering the importance of using efficient tools in the identification of hybrids, the objective of this study was to, for the first time, cytogenetically describe the hybrids of two strains between species of Astyanax and a strain between a species of Astyanax and a species of the genus Hyphessobrycon to observe the inheritance of cytogenetic markers from the parent stains. The study also aimed to verify the possibility of identifying a hybrid using cytogenetic markers, to contribute to the understanding of the evolutionary dynamics of the group.

**Material and methods**

The parent strains used in this study were obtained from the Instituto Chico Mendes de Conservação da Biodiversidade (CEPTA – ICMBIO/Pirassununga, SP, Brazil), where artificial crossing was performed. The crosses were directed using *A. lacustris* females and *A. fasciatus, A. schubarti* Britski, 1964, and *H. anisitsi* Eigenmann, 1907, males. Ovulation was induced in *A. lacustris* using the protocol established by Yasui et al. (2015), and spermatogenesis in males of the other species was induced with a single dose of carp pituitary gland (5 mg kg$^{-1}$). The gametes were collected by stripping, the oocytes were stripped on a plastic Petri dish and the sperm was collected using a 1000 μl micropipette and transferred to a tube containing 300 μl of Ringer solution (Piva et al. 2018). Oocytes fertilization was initiated in the Petri dish using 80μl of sperm from selected males, and gamete activation was achieved by adding 5ml of water followed by immediate mixing via gentle hand movements.

The hybrids were identified and deposited in the Laboratório de Genética de Peixes, Bauru, São Paulo, Brazil, under the accession numbers LGP8291–LGP8382. Fifty-nine animals were anesthetized using 1% benzocaine. Mitosis stimulation was performed using the method described by Oliveira et al. (1988). Subsequently, mitotic chromosomes were obtained from kidney tissue using protocols described by Foresti et al. (1981) and Foresti et al. (1993). Seventeen hybrids of *A. lacustris* × *A. fasciatus*, 10 of *A. lacustris* × *A. schubarti*, and 32 of *A. lacustris* × *H. anisitsi* were analyzed. C-positive heterochromatin was detected using the barium hydroxide method (Sumner 1972). Chromosomes were classified as metacentric (m), submetacentric (sm), subtelocentric (st), and acrocentric (a) according to their arm ratios (Levan et al. 1964).

5S (two different bands: 255 and 525 bp) and 18S (one band: 600 pb) rDNA probes were obtained using polymerase chain reaction with the primers 5S A (5’-TCAAC-CAACCACAAAGACATTGGCAC-3’) and 5S B (5’-TAGACTTCTGGGTGGGC-CAAGGATAATAC-3’) for the 5S gene (Pendás et al. 1994) and 18S A (5’-TACCCTCCTGGGATATCAGGCAAGGAATATATGGAATCA-3’) and 18S B (5’-CAGGCTGGTATGGCCGGAAGGC-3’) for the 18S gene (Utsunomia et al. 2016). For fluorescence in situ hybridization, chromosomes were treated following the protocol described by Pinkel et al. (1986). The probes were labeled using biotin-14-dATP and digoxigenin-11dUTP (Roche Applied Science) and the signals were detected using avidin-fluorescein conjugate (FITC)
and anti-digoxigenin-rhodamine, respectively. Images were captured using Olympus QColor coupled to a fluorescence photomicroscope (BX50, Olympus), and the images were processed using the CellSens Standard Software.

Results

All parent strains displayed stable diploid chromosome numbers; *A. lacustris* displayed 2n = 50 (6m+12sm+14st+18a) chromosomes; *A. fasciatus*, 2n = 48 (10m+12sm+12st+14a); *A. schubarti*, 2n = 36 (10m+10sm+10st+6a); and *H. anisitsi*, 2n = 50 (10m+2sm+20st+18a) (Fig. 1). The *A. lacustris × A. fasciatus* progeny displayed 49 chromosomes (8m+12sm+13st+16a) and the *A. lacustris × A. schubarti* progeny displayed 43 chromosomes (8m+11sm+12st+12a); the *A. lacustris × H. anisitsi* progeny displayed chromosome number variation, with some individuals showing 50 or 51 chromosomes (Fig. 2). Importantly, this extra chromosome (from individuals showing 51 chromosomes) was C-band positive, different from the regular set of chromosomes (Fig. 3).

**Figure 1.** Karyotypes of the parental individuals analyzed: *Astyanax lacustris* (3m+6sm+7st+9a), *Hyphessobrycon anisitsi* (5m+1sm+10st+9a), *A. fasciatus* (5m+6sm+6st+7a), and *A. schubarti* (5m+5sm+5st+3a). Scale bar: 5 μm.

**Figure 2.** Karyotypes of three hybrids of species of the genus *Astyanax*: *A. lacustris × Hyphessobrycon anisitsi* (8m+7sm+17st+18a), *A. lacustris × A. fasciatus* (8m+7sm+17st+18a), and *A. lacustris × A. schubarti* (8m+11sm+12st+12a). Scale bar: 5 μm.
Hybridization between genders in fish

Figure 3. Heterochromatic markers obtained by C-banding on metaphase plates of Astyanax lacustris (a), A. fasciatus (b), and A. schubarti (c), and Hyphessobrycon anisitsi (d) and hybrids A. lacustris × A. fasciatus (e), A. lacustris × A. schubarti (f), and A. lacustris × H. anisitsi (g, h) after C-banding. The arrows indicate heterochromatic markers. In h, a metaphase with 51 chromosomes, the chromosome being completely heterochromatic, can be observed. Scale bar: 5μm.

The results of C-positive heterochromatin revealed some interesting features. Astyanax lacustris and A. schubartii hybrids showed regular heterochromatic blocks inherited from both parent strains. The terminal heterochromatic blocks in subtelocentric/acrocentric chromosomes of A. fasciatus and the typical location of As51 satellite DNA were not detected in the hybrids (Figure 3); furthermore, the A. lacustris × H. anisitsi hybrids displayed a conspicuous heterochromatic block in the p arm of the large subtelocentric chromosome, and this was not detected in any parent strain (Fig. 3).
The ribosomal sites showed Mendelian inheritance, as revealed in Figure 4. *Astyanax lacustris* and *A. schubarti* displayed four sites of 18s rDNA and two sites of 5s rDNA. *Astyanax fasciatus* showed four sites of both markers, and *Hyphessobrycon anisitsi* showed intense dispersion of 18s rDNA, with 10 sites of this marker. This species demonstrated four sites of 5s rDNA, one of them syntenic with 18s rDNA. In general, the hybrids demonstrated the inheritance of cytogenetic markers as expected, with some inconsistency in the *A. lacustris × A. schubarti* hybrid, as indicated by the observation of three sites of 5s rDNA instead of just two and a bi-telomeric site of 18S rDNA in an acrocentric chromosome. All cytogenetic analysis is resumed in ideograms of parent (Fig. 5) and hybrid (Fig. 6) strains.

**Figure 4.** Fluorescence *in situ* hybridization with the probes DNA r 5S (green) and 18S (red). The results are labeled as: *Astyanax lacustris* (a, d, g), *A. schubarti* (b), hybrid *A. lacustris × A. schubarti* (c), *A. fasciatus* (e), hybrid *A. lacustris × A. fasciatus* (f), *Hyphessobrycon anisitsi* (h), and hybrid *A. lacustris × H. anisitsi* (i). Arrows and arrowheads indicate chromosomes bearing 18S and 5S rDNA clusters: arrows, *A. lacustris*; arrowheads, other species in the cross. Scale bar: 5µm.
Hybridization between genders in fish

Figure 5. Ideogram of parental strains.

Figure 6. Ideogram of hybrid strains.

Discussion

The genus Astyanax is rich in chromosomal polymorphisms (Moreira-Filho 1991; Fernandes and Martins-Santos 2004; Artoni et al. 2006; Garutti and Langeani 2009). Moreover, the results of the present study are consistent with the literature regarding diploid numbers and the distribution of cytogenetic markers in the species used as parent strains (Mantovani et al. 2000; Almeida-Toledo et al. 2002; Kantek et al. 2009; Hashimoto and Porto-foresti 2010; Hashimoto et al. 2011). As expected, the resulting hybrids showed typical karyotypic features, inherited from the distinct parental strains.

Hybridization between different fish species can generate individuals that diverge from simple diploids with equal parental contribution (Toledo-Filho et al. 1994); andro or gynogenetic offspring as well as haploid, triploid, or tetraploid animals can be obtained. In the present study, it was possible to characterize all the
strains as single diploid offspring because we identified the haploid sets from both parent strains involved in the crossing, resulting in diploid numbers intermediate to those of the parent strains.

The C-banding patterns revealed interesting features, as conspicuous heterochromatic blocks did not appear to be regularly inherited in some cases, indicating some degree of chromatin remodeling, similar to that in plant and mammal hybrids (O’Neill et al. 1998; Comai et al. 2003). In both cases, heterochromatin expansion occurred through hypomethylation of genomic regions containing transposable elements, allowing for expansion of these mobile sequences. Considering the heterochromatic areas of tetras are mainly composed of transposable elements (Vicari et al. 2008; Silva et al. 2013; Barbosa et al. 2017), it can be hypothesized that hybridization affects these regions within a single generation. Some inconsistencies were detected in the analysis of rDNA: an additional 5S rDNA site and a bi-telomeric 18S rDNA site in the *A. lacustris × A. schubarti* hybrid, synteny of the 5S and 18S genes in only one *H. anisitsi* chromosome [also observed in the *A. lacustris × H. anisitsi* hybrid, likely due to an intraspecific polymorphism of 18S rDNA distribution in *H. anisitsi* (Fig. 4), and an extra and totally heterochromatic chromosome in two *A. lacustris × H. anisitsi* hybrids (present in approximately 50% of analyzed cells). A case of B chromosomes totally heterochromatic from interspecific hybridization has been reported in fishes (Schartl et al. 1995); however, more studies are necessary to verify the hypothesis of this aneuploidy being a B chromosome.

Fertile hybrids have been described for different Neotropical fish species such as hybrids of the catfishes “cachapinta” and “pintachara,” *Pseudoplatystoma corruscans* (Spix et Agassiz, 1829) and *P. reticulatum* Eigenmann et Eigenmann, 1889, (Hashimoto et al. 2013; Prado et al. 2017) and those involving the Characiformes species *Piaractus mesopotamicus* (Holmberg, 1887), *Colossoma macropomum* (Cuvier, 1816), and *Piaractus brachypomus* (Cuvier, 1818) (Hashimoto et al. 2014). The fertility of the hybrids is a problematic issue owing to the extensive production of hybrids in Brazilian aquaculture and the recurrent escapes of these individuals to the nature, which threatens the maintenance of natural populations that are susceptible to back-crossing; contamination of their gene pools is also possible. In a recent study, using the same brood stock analyzed herein, Piva et al. (2018) stated that a complete sterile offspring was restricted to *A. lacustris × A. fasciatus* crossing. Surprisingly, offspring from distinct genera (*A. lacustris × H. anisitsi*) and those displaying highly differentiated karyotypes (*A. lacustris × A. schubarti*) showed normal gametogenesis. However, the possibility of viable gamete formation by these individuals and consequently their effective fertility can be affected owing to the unstable diploid number in some of the hybrid strains observed in this study, such as the *A. lacustris × A. schubarti* (2n = 43) hybrid, unlike other fertile natural hybrids resulting from parent strains with the same diploid number, as observed in hybrids of the catfishes “pintachara” and “cachapinta” (Prado et al. 2012).
Conclusion

The cytogenetic markers applied to the hybrid strains analyzed in this study were efficient in terms of identification based on the known karyotype of the parent strains, which differentiates the hybrids involving species of the genus Astyanax from other hybrids of Neotropical fish, which keeps its cytotypes conserved (Prado et al. 2012). In this sense, the diploid number was especially helpful in detecting hybrids. In cases wherein the hybrid had the same diploid number as the parent strains, 5s rDNA was the best marker. This study describes, for the first time, three hybrid strains involving species of the genera Astyanax and Hyphessobrycon and shows the efficiency of cytogenetic markers in their identification. The results presented herein will contribute to future cytogenetic and evolutionary studies involving these genera aimed at karyotypic diversity and species complex formation; the present study also highlights the possibility of the use of cytogenetic markers in the identification of hybrids.

Acknowledgements

The authors thank the Brazilian funding agencies Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: Processo n° 2015/12902-4), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for providing financial support. We would like to thank Editage (www.editage.com) for English language editing. This work was supported by the Fundação de Amparo à Pesquisa (FAPESP) under Grant (no. 2015/12902-4)

References


Kantek DLZ, Vicari MR, Peres WAM, Cestari MM, Artoni RF, Bertollo LAC, Moreira-Filho O (2009) Chromosomal location and distribution of As51 satellite DNA in five species
Hybridization between genders in fish


Prado FD, Nunes TL, Senhorini JA, Bortolozzi J, Foresti F, Porto-Foresti F (2012) Cytogenetic characterization of F1, F2 and backcross hybrids of the Neotropical Catfish spe-


Comparative molecular cytogenetic characterization of five wild Vigna species (Fabaceae)

Chao-Wen She\textsuperscript{1,2,3}, Ying Mao\textsuperscript{3}, Xiang-Hui Jiang\textsuperscript{1,2,3}, Chun-Ping He\textsuperscript{4}

\textsuperscript{1} Key Laboratory of Research and Utilization of Ethnomedicinal Plant Resources of Hunan Province, Huaihua University, Huaihua, Hunan, 418008, China \textsuperscript{2} Key Laboratory of Xiangxi Medicinal Plant and Ethnobotany of Hunan Higher Education, Huaihua University, Huaihua, Hunan, 418008, China \textsuperscript{3} College of Biological and Food Engineering, Huaihua University, Huaihua, Hunan, 418008, China \textsuperscript{4} College of Chemistry and Material Engineering, Huaihua University, Huaihua, Hunan, 418008, China

Corresponding author: Chao-Wen She (shechaowen@aliyun.com)

Abstract
To extend our knowledge on karyotype variation of the genus \textit{Vigna} Savi, 1824, the chromosomal organization of rRNA genes and fluorochrome banding patterns of five wild \textit{Vigna} species were studied. Sequential combined PI (propidium iodide) and DAPI (4',6-diamidino-2-phenylindole) (CPD) staining and fluorescence \textit{in situ} hybridization (FISH) with 5S and 45S rDNA probes were used to analyze the karyotypes of \textit{V. luteola} (Jacquin, 1771) Bentham, 1959, \textit{V. vexillata} (Linnaeus, 1753) A. Richard, 1845, \textit{V. minima} (Roxburgh, 1832) Ohwi & H. Ohashi, 1969, \textit{V. trilobata} (Linnaeus, 1753) Verdcourt, 1968, and \textit{V. caracalla} (Linnaeus, 1753) Verdcourt, 1970. For further phylogenetic analysis, genomic \textit{in situ} hybridization (GISH) with the genomic DNA of \textit{V. umbellata} (Thunberg, 1794) Ohwi & H.Ohashi, 1969 onto the chromosomes of five wild \textit{Vigna} species was also performed. Detailed karyotypes were established for the first time using chromosome measurements, fluorochrome bands, and rDNA-FISH signals. All species had chromosome number $2n = 2x = 22$, and symmetrical karyotypes that composed of only metacentric or metacentric and submetacentric chromosomes. CPD staining revealed all 45S rDNA sites in the five species analyzed, (peri)centromeric GC-rich heterochromatin in \textit{V. luteola}, \textit{V. trilobata} and \textit{V. caracalla}, interstitial GC-rich and pericentromeric AT-rich heterochromatin in \textit{V. caracalla}. rDNA-FISH
revealed two 5S loci in *V. caracalla* and one 5S locus in the other four species; one 45S locus in *V. luteola* and *V. caracalla*, two 45S loci in *V. vexillata* and *V. trilobata*, and five 45S loci in *V. minima*. The karyotypes of the studied species could be clearly distinguished by the karyotypic parameters, and the patterns of the fluorochrome bands and the rDNA sites, which revealed high interspecific variation among the five species. The *V. umbellata* genomic DNA probe produced weak signals in all proximal regions of *V. luteola* and all (peri)centromeric regions of *V. trilobata*. The combined data demonstrate that distinct genome differentiation has occurred among the five species during evolution. The phylogenetic relationships between the five wild species and related cultivated species of *Vigna* are discussed based on our present and previous molecular cytogenetic data.

**Keywords**

*Vigna* species, karyotype, fluorochrome banding, fluorescence in situ hybridization (FISH), ribosomal RNA gene (rDNA)

**Introduction**

The genus *Vigna* Savi, 1824, belonging to the tribe Phaseoleae of the family Fabaceae, includes over 100 species distributed throughout the Old and New Worlds (Schrire 2005). Taxonomically, this genus was divided into seven subgenera by Maréchal et al. (1981), among which subg. *Vigna* Savi, 1876 and subg. *Ceratotropis* (Piper) Verdcourt, 1969 includes the seven economically important crop species, *V. unguiculata* (Linnaeus, 1753) Walp, 1842, *V. subterranea* (Linnaeus, 1753) Verdcourt, 1980, *V. aconitifolia* (Jacquin, 1771) Maréchal, 1969, *V. angularis* (Willdenow, 1800) Ohwi & H. Ohashi, 1969, *V. mungo* (Linnaeus, 1753) Hepper, 1956, *V. radiata* (Linnaeus, 1753) R. Wilczek, 1954, and *V. umbellata* (Thunberg, 1794) Ohwi & H. Ohashi, 1969 (Smartt 1990). An understanding of the phylogenetic relationships among the cultigens and their wild relatives is helpful for developing crop improvement tools and gene transfer strategies. A lot of DNA-level studies, such as analyses of the internal transcribed spacers (ITS) of rDNA (Doi et al. 2002; Goel et al. 2002; Saini et al. 2008; Delgado-Salinas et al. 2011; She et al. 2015; Raveenadar et al. 2018), the intergenic spacer (IGS) of 5S rDNA (Saini and Jawali 2009), plastid DNA sequences (Doi et al. 2002; Tun and Yamaguchi 2007; Javadi et al. 2011; Delgado-Salinas et al. 2011; Raveenadar et al. 2018), and DNA amplification fingerprinting (Simon et al. 2007), have been conducted to reveal the phylogenetic relationships among *Vigna* species. A molecular cytogenetic investigation has also been performed to help clarify the phylogenetic relationships among the seven cultivated *Vigna* species (She et al. 2015). However, comparative molecular cytogenetic study on the phylogenetic relationships between the cultivated *Vigna* species and closely related wild species has not been conducted till now.

The chromosomes of *Vigna* species were rather small in size and poorly morphologically differentiated (Guerra et al. 1996), resulting in the difficulty of distinguish-
ing chromosomes. To date, only about twenty wild Vigna species were cytogenetically studied, and these studies were mostly restricted to chromosome counts and karyomorphological descriptions (Sen and Bhowal 1960; Joseph and Bouwkamp 1978; Rao and Chandel 1991; Galasso et al. 1993, 1996; Venora and Saccardo 1993; Venora et al. 1999; Shamurailatpam et al. 2012, 2015, 2016), which could not provide reliable information on genome evolution among related species. Although many molecular cytogenetic studies have been conducted for the cultivated Vigna species using fluorescence in situ hybridization (FISH) with 5S and 45S ribosomal genes (rDNAs; Galasso et al. 1995, 1998; Guerra et al. 1996; Khattak et al. 2007; de A Bortoleti et al. 2012; Choi et al. 2013; She et al. 2015), but only one wild Vigna species has been molecular-cytogenetically investigated so far (Choi et al. 2013).

FISH mapping of repetitive DNA sequences such as 5S and 45S rDNAs can not only generate useful landmarks for chromosome identification but can also provide valuable information on the evolutionary relationships between related species (e.g. Moscone et al. 1999; Zhang and Sang 1999; Hasterok et al. 2001; de Moraes et al. 2007; Hamon et al. 2009; Robledo et al. 2009; Wolny and Hasterok 2009; She et al. 2015; Li et al. 2016; Amosova et al. 2017; Maragheh et al. 2019). To date, the number and position of rDNA loci have been determined in more than 1600 plant species with FISH (Garcia et al. 2014). These studies showed that the number and position of the 5S and 45S rDNAs were usually characteristics of a given species or genus (e.g. Moscone et al. 1999; Hasterok et al. 2001; Chung et al. 2008; Hamon et al. 2009; Robledo et al. 2009; Wolny and Hasterok 2009; She et al. 2015; Li et al. 2016; Maragheh et al. 2019). Fluorochrome banding techniques using double fluorescent dyes such as CMA3 (chromomycin A3) /DAPI (4',6-diamidino-2-phenylindole) staining, and PI (propidium iodide)/ DAPI staining (called CPD staining) was used to localize the chromosome regions that are rich in GC and AT base pairs simultaneously, providing effective identifying markers for chromosomes, and revealing characteristic heterochromatin distribution along chromosomes (She et al. 2006; de Moraes et al. 2007; de A Bortoleti et al. 2012; She and Jiang 2015; She et al. 2015, 2017; Tang et al. 2019).

Detailed karyotypes can be constructed using the dataset of rDNA-FISH signals, fluorochrome bands and chromosome measurements, which reveals the genome organization of a plant species at chromosome level and is valuable in investigating the evolutionary relationships between related species (e.g. Moscone et al. 1999; de Moraes et al. 2007; Hamon et al. 2009; Robledo et al. 2009; Mondin and Aguiar-Perecin 2011; She and Jiang 2015; She et al. 2015, 2017; Zhang et al. 2015; Amosova et al. 2017; Tang et al. 2019) and helpful to integrate the genetic and physical maps of a plant species (Fuchs et al. 1998; Fonsêca et al. 2010). Comparative genomic in situ hybridization (cGISH) is a modification of the GISH technology in which the labelled total genomic DNA of one species is hybridized to the chromosomes of another species without the competitive DNA. It generates hybridization signals in the chromosomal
regions of conserved repetitive DNA sequences. Therefore, it can directly identify the genome relationships among related species (Falistocco et al. 2002; Wolny and Hasterok 2009; She et al. 2015, 2017; Amosova et al. 2017).

In the present study, molecular cytogenetic characterization of five wild Vigna species, *V. luteola*, *V. vexillata*, *V. minima*, *V. trilobata* and *V. caracalla* was conducted using sequential CPD staining and dual color FISH with 5S and 45S rDNA probes. Detailed karyotypes of the five species were established using a combination of chromosome measurements, fluorochrome bands, and rDNA-FISH signals. Six different parameters of karyotype asymmetry were calculated for the elucidation of karyotype variation among these species. cGISH with *V. umbellata* genomic DNA probe onto the somatic chromosomes of the five species, the method that was applied in the molecular-cytogenetic study on the seven cultivated Vigna species (She et al. 2015), was also performed. The datasets were assessed to gain insights into the genome differentiation and phylogenetic relationships among the five wild and seven cultivated Vigna species.

**Material and methods**

**Plant materials and DNA extraction**

Seeds of *V. luteola* (Jacquin, 1771) Bentham, 1959 (PI 406329), *V. vexillata* (Linnaeus, 1753) A.Richard, 1845 (PI 406428, Origin traced to PI 225934), *V. minima* (Roxburgh, 1832) Ohwi & H. Ohashi, 1969 (PI 483081), *V. trilobata* (Linnaeus, 1753) Verdcourt, 1968 (PI 286306), *V. caracalla* (Linnaeus, 1753) Verdcourt, 1970 (Synonym of Cochliasanthus caracalla (Linnaeus, 1753) Trew, 1764; PI 146800), and *V. umbellata* (Thunberg, 1794) Ohwi & H. Ohashi, 1969 (PI 208460) were obtained from the U.S. National Plant Germplasm System. Genomic DNA of *V. umbellata* was isolated from young leaves using Rapid Plant Genomic DNA Isolation Kit (Sangon Biotech, Shanghai, China).

**Chromosome preparation**

Mitotic metaphase chromosome spreads were prepared as previously described with minor modification (She et al. 2006). In brief, seeds were germinated on moistened filter paper in the dark at 28 °C. Root tips were harvested and treated in saturated α-bromonaphthalene at 28 °C for 2.0 h, and then fixed in methanol-glacial acetic acid (3:1) at 4 °C. The fixed root tips were thoroughly rinsed in double-distilled water and digested in an enzymatic solution composed of 1% cellulase RS (Yakult Pharmaceutical Industry Co., Ltd. Tokyo, Japan), 1% pectolyase Y23 (Yakult Pharmaceutical Industry Co., Ltd. Tokyo, Japan) in citric buffer (0.01 mM citric acid-sodium citrate, pH 4.5) at 28 °C for 100–120 mins. The digested root tips were gently placed on a glass slide with methanol-glacial acetic acid (3:1) and dissected thoroughly by using fine-pointed forceps. Then, the slides were flame-
dried. The slides with well-spread somatic metaphase chromosomes were screened under phase contrast microscope and stored at -20 °C until used.

**CPD staining**

CPD staining followed the procedure described by She et al. (2006). Briefly, chromosome preparations were treated with RNase A and pepsin and then stained with a mixture of 0.6 μg·ml⁻¹ PI and 3 μg·ml⁻¹ DAPI in a 30% (v/v) solution of Vectashield H100 (Vector Laboratories, Burlingame, US) for at least 30 min in the dark at room temperature. Slides were examined under an Olympus BX60 epifluorescence microscope. Separate images from UV and green filters were captured using a cooled CCD camera (CoolSNAP EZ; Photometrics, Tucson, US) controlled using METAMORPH software (Molecular Devices, California, US). DAPI and PI grey scale images of the same plate were merged to produce a CPD image. Final images were optimized for contrast and brightness using ADOBE PHOTOSHOP version 8.01.

**Probe DNA labelling**

A 45S rDNA clone containing a 9.04-kb tomato 45S rDNA insert (Perry and Pallukaitis 1990) and a pTa794 clone containing a 410-bp BamHI fragment of wheat 5S rDNA (Gerlach and Bedbrook 1979) were used as probes to localize the two ribosomal RNA genes. The 45S clone was labeled with biotin-16-dUTP, and the 5S clone and the *V. umbellata* genomic DNA were labeled with digoxigenin-11-dUTP, using Nick Translation Kit (Roche Diagnostics, Mannheim, Germany).

**Fluorescence in situ hybridization**

FISH with the 5S and 45S rDNA probes, and cGISH with *V. umbellata* genomic DNA probe were performed after CPD staining on the same slides. The slides previously stained by CPD were washed in 2x SSC, twice for 15 min each, dehydrated through an ethanol series (70%, 90%, and 100%, 5 min each) and then used for hybridization. The *in situ* hybridization methodology followed the protocol described by She et al. (2015). The biotin-labelled probe was detected using Fluorescein Avidin D (Vector Laboratories, Burlingame, USA). The digoxigenin-labeled probe was detected by anti-digoxigenin-rhodamine (Roche Diagnostics, Mannheim, Germany). The preparations were counterstained and mounted with 3 μg ml⁻¹ DAPI in 30% (v/v) Vectashield H-1000 and examined under the epifluorescence microscope mentioned above. Grey-scale images were digitally captured using METAMORPH software with UV, blue and green filters for DAPI, fluorescein, and rhodamine, respectively. The images were then merged and edited with ADOBE PHOTOSHOP version 8.01.
Karyotype analysis

The karyotyping methodology followed that described by She et al. (2015). Five metaphase plates of each species were measured using ADOBE PHOTOSHOP version 8.01. The chromosome relative lengths (RL, % of haploid complement), arm ratios (AR = long arm/short arm), size of the fluorochrome band, and percent distance from the centromere to the rDNA site were calculated. The total length of the haploid complement (TCL; i.e. the karyotype length) was measured using five metaphase cells with the highest condensation degree. The arm ratio was used to classify the chromosomes according to the system described by Levan et al. (1964). Idiograms were drawn based on measurements, fluorochrome bands, and rDNA-FISH signals. The chromosomes were organized in decreasing order. Karyotype asymmetry was determined using the mean centromeric index (CI), the intrachromosomal asymmetry index (A1), the interchromosomal asymmetry index (A2) (Romero Zarco 1986), the ratio of long arm length in chromosome set to total chromosome length in set (As K%) (Arano 1963), the asymmetry index (AI) (Paszko 2006), and the categories of Stebbins (1971).

Results

General karyotype features

Representative mitotic chromosomes of the five species studied are shown in Figure 1. The karyotypic parameters are listed in Table 1. The chromosome measurements for the five species are given in Suppl. material 1: Table S1. Idiograms displaying the chromosome measurements, position and size of the CPD bands and rDNA-FISH signals are illustrated in Figure 2.

All the five Vigna species studied have diploid chromosome number 2n = 2x = 22. The metaphase chromosomes were small, with a mean chromosome length between 2.33 μm (V. vexillata) and 4.24 μm (V. caracalla). The total length of the haploid complement (TCL) ranged from 25.67 μm to 46.62 μm, and the mean centromeric index (CI) of the complements varied between 42.15 ± 3.87 (V. trilobata) and 44.55 ± 2.03 (V. minima). V. caracalla exhibited the most variation in chromosome length, and V. trilobata was characterized by the highest level of variation in the centromeric index.

The karyotypes of V. luteola, V. vexillata, V. minima were composed of metacentric (m) chromosomes only, while those of V. trilobata and V. caracalla were composed of metacentric and submetacentric (sm) chromosomes (Table 1, Suppl. material 1: Table S1; Fig. 2). In V. caracalla, the first chromosome pair had a satellite with secondary constriction (SC) that located at the distal position of the short arm (Figs 11, 2E). All the karyotypes were quite symmetrical, falling into the Stebbins’ categories 1A or 1B (Table 1). The ranges of intrachromosomal asymmetry index (A1) and the interchromosomal asymmetry index (A2) were as follows: A1 = 0.19–0.27,
Molecular cytogenetics of five wild Vigna species

Figure 1. Mitotic chromosomes from *V. luteola* (A, B), *V. vexillata* (C, D), *V. minima* (E, F), *V. trilobata* (G, H), and *V. caracalla* (I–L) stained using CPD method and sequential dual-colour FISH with digoxigenin-labelled 5S and biotin-labelled 45S rDNA probes. A, C, E, G, I are the chromosomes stained using CPD. The chromosome numbers are designated by karyotyping. B, D, F, H, J are the chromosomes displaying the 5S (red) and 45S rDNA (green) signals. The total DNA was counterstained using DAPI (blue). K, L are DAPI and PI grey scale images of the *V. caracalla* chromosomes stained using CPD, respectively. The images are converted to reverse images with Photoshop software. Arrows and arrowheads in I indicate the satellites and interstitial CPD bands, respectively. Scale bars: 10 μm.

and $A_2 = 0.14–0.21$. The As K% ranged from 55.53 to 58.00, and the asymmetry index (AI) ranged from 0.66 to 1.76. According to the AI values, the karyotype of *V. minima* was the most symmetrical and that of *V. trilobata* was the most asymmetrical among the five taxa.
Table 1. Karyotypic parameters of the five wild *Vigna* species (all, 2n = 2x = 22).

<table>
<thead>
<tr>
<th>Species</th>
<th>KF</th>
<th>TCL ± SE (μm)</th>
<th>C (μm)</th>
<th>RRL</th>
<th>CI ± SE</th>
<th>A1 ± SE</th>
<th>A2 ± SE</th>
<th>AsK (%)</th>
<th>AI</th>
<th>Stebbins’ types</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>V. luteola</em></td>
<td>11m</td>
<td>33.81 ± 1.56</td>
<td>3.07</td>
<td>6.88–12.40</td>
<td>44.35 ± 2.45</td>
<td>0.20</td>
<td>0.21</td>
<td>55.97</td>
<td>1.15</td>
<td>1A</td>
</tr>
<tr>
<td><em>V. vexillata</em></td>
<td>11m</td>
<td>25.67 ± 2.02</td>
<td>2.33</td>
<td>6.99–12.66</td>
<td>43.24 ± 3.45</td>
<td>0.23</td>
<td>0.19</td>
<td>57.01</td>
<td>1.52</td>
<td>1A</td>
</tr>
<tr>
<td><em>V. minima</em></td>
<td>11m</td>
<td>38.29 ± 1.04</td>
<td>3.48</td>
<td>7.37–12.14</td>
<td>44.55 ± 2.03</td>
<td>0.19</td>
<td>0.14</td>
<td>55.93</td>
<td>0.66</td>
<td>1A</td>
</tr>
<tr>
<td><em>V. trilobata</em></td>
<td>9m + 2sm</td>
<td>36.56 ± 2.73</td>
<td>3.32</td>
<td>7.20–13.48</td>
<td>42.15 ± 3.87</td>
<td>0.27</td>
<td>0.19</td>
<td>58.00</td>
<td>1.76</td>
<td>1A</td>
</tr>
<tr>
<td><em>V. caracalla</em></td>
<td>10m (1SAT) + 1sm</td>
<td>46.62 ± 1.71</td>
<td>4.24</td>
<td>5.61–12.80</td>
<td>44.37 ± 3.13</td>
<td>0.20</td>
<td>0.20</td>
<td>59.39</td>
<td>1.41</td>
<td>1B</td>
</tr>
</tbody>
</table>

Notes: KF, Karyotype formula of haploid; TCL, total length of the haploid complement (i.e. karyotype length); C, mean chromosome length; SAT, satellite chromosome; RRL, ranges of chromosome relative length; CI, mean centromeric index; A1 and A2, the intrachromosomal asymmetry index and the interchromosomal asymmetry index of Romero Zarco (1986), respectively; AsK%, the ratio of length of all long arms in chromosome set to total chromosome length in set of Arano (1963); AI, the karyotype asymmetry index of Paszko (2006); Stebbins’ types, the karyotype asymmetry category of Stebbins (1971).

Fluorochrome banding patterns

CPD staining revealed distinct heterochromatin differentiation among the five species studied (Figs 1–3; Table 2). Red CPD bands were shown in all species, but blue-fluorescent DAPI+ bands were shown only in *V. caracalla* (Figs 1I, 3H). The CPD bands were shown to be reverse PI-DAPI bands resulting from the intensity of the contrast between the PI (red) and DAPI (blue) fluorescence (Fig. 1I, K, L). In each species, all the chromosomal regions corresponding to the 45S rDNA sites, which were demonstrated by sequential FISH with rDNA probes, displayed CPD bands (Fig. 1A, C, E, G, I). All (peri) centromeric regions in *V. luteola*, *V. trilobata* and *V. caracalla* showed CPD bands (Figs 1A, G, I, 3A, F, H), while those in *V. vexillata* and *V. minima* did not show CPD bands (Figs 1C, E, 3D). In particular, the 5S rDNA sites in *V. minima* (Fig. 1E, F), and three pairs of interstitial sites (located in both short and long arms of chromosome pair 4, and the long arms of chromosome pair 5, respectively) in *V. caracalla* displayed CPD bands (Figs 1I, 3H). *V. caracalla* showed eight pairs of DAPI+ bands that occurred in the pericentromeric regions of the short arms of chromosome pairs 2, 3, 4 and 5, and the pericentromeric regions of the long arms of chromosome pairs 4, 5, 6 and 8 (Figs 1I, K, 3H). These DAPI+ bands were also shown in the DAPI-counterstained chromosomes after the FISH procedure (Figs 1J, 3I). The total amount of non-rDNA CPD bands in *V. luteola*, *V. trilobata* and *V. caracalla* were 29.19%, 20.04%, and 21.68% of the karyotype length, respectively (Tables 2, Suppl. material 1: Table S1). The size of non-rDNA CPD bands varied between the chromosome pairs in each species (Fig. 2; Suppl. material 1: Table S1). The total amount of DAPI+ bands in relation to the karyotype length was 8.19% in *V. caracalla* (Fig. 2; Suppl. material 1: Table S1).

FISH patterns of 5S and 45S rDNA sites

FISH results of the 5S and 45S rDNA probes to the CPD-stained mitotic chromosomes are presented in Figure 1. The number and position of the rDNA sites are summarized in Table 2 and illustrated in Figure 2.
Figure 2. Idiograms of the five Vigna species that display the chromosome measurements, and the position and size of the fluorochrome bands and rDNA-FISH signals. A–E indicate V. luteola, V. vexillata, V. mini- ma, V. trilobata, and V. caracalla, respectively. The ordinate scale on the left indicates the relative length of the chromosomes (i.e. % of haploid complement). The numbers at the top indicate the chromosomes 1 to 11.
Table 2. The distribution of fluorochrome bands and rDNA sites in the five wild *Vigna* species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Fluorochrome bands</th>
<th>Number (pairs) and location of rDNA sites†</th>
<th>Amount (%)‡</th>
<th>Band size (mean)§</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type</td>
<td>Distribution†</td>
<td>5S</td>
<td>45S</td>
</tr>
<tr>
<td><em>V. luteola</em></td>
<td>CPD</td>
<td>all CENs, PCENs and 45S sites</td>
<td>29.19</td>
<td>1.98–3.21 (2.65)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>one (11S-PCEN (16.55%))</td>
<td>one (3S)</td>
<td></td>
</tr>
<tr>
<td><em>V. vexillata</em></td>
<td>CPD</td>
<td>all 45S sites</td>
<td>one (8L-INT (52.29%))</td>
<td>two (2S-TER (20.53%), 35S-TER (16.73%))</td>
</tr>
<tr>
<td><em>V. minima</em></td>
<td>CPD</td>
<td>all 45S and 5S sites</td>
<td>one (2S-INT (50.80%))</td>
<td>five (2L-TER (58.64%), 4L-TER (59.42%), 6S-TER (38.91%), 7S-TER (50.74%), 9S-TER (67.94%))</td>
</tr>
<tr>
<td><em>V. trilobata</em></td>
<td>CPD</td>
<td>all CENs, PCENs and 45S sites</td>
<td>20.04</td>
<td>2.73–1.12 (1.82)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>one (4L-PCEN(14.95%))</td>
<td>two (6S-PCEN (25.05%), 7S)</td>
<td></td>
</tr>
<tr>
<td><em>V. caracalla</em></td>
<td>CPD</td>
<td>all CENs, PCENs and 45S sites, 4S-, 4L-</td>
<td>21.68*</td>
<td>0.89–2.63 (1.55)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5L-INTs</td>
<td>two [2L-INT(34.32%), 5S-INT (36.7%)]</td>
<td></td>
</tr>
<tr>
<td>DAPI</td>
<td>2, 3, 4, 5S-PCENs;</td>
<td></td>
<td>one (1S)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4, 5, 6, 8L-PCENs</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† S and L represent short and long arms, respectively; CEN, PCEN, INT and TER represent centromeric, pericentromeric, interstitial, terminal position, respectively; figures ahead of the positions are the designations of the chromosome pair involved.

‡Amount of bands in the genome expressed as percentage of the karyotype length (rDNA CPD bands are excluded).

§The percentage of the size of the bands of each chromosome pair in relation to the karyotype length.

The FISH patterns of 5S and 45S rDNAs of the five species displayed conspicuous interspecific variation. Among the five taxa, *V. luteola*, *V. vexillata*, *V. minima* and *V. trilobata* had a single pair of 5S rDNA sites, while *V. caracalla* had two pairs of 5S sites (Figs 1B, D, F, H, J, 2A–E; Table 2). The 5S loci in *V. luteola* and *V. trilobata* were located in the pericentromeric regions of the relevant chromosome short or long arms, while those in *V. vexillata*, *V. minima* and *V. caracalla* were distributed in the interstitial regions of the short arms or long arms of the respective chromosomes. The 5S locus in *V. trilobata* was colocalized with a portion of the pericentromeric CPD bands (Figs 1G, H, 2D). With regard to length, the chromosome pair bearing the 5S locus in *V. luteola* was the shortest in the complement, and the 5S-bearing chromosome pairs in other four species were of an intermediate size.

For the 45S rDNA sites, there was considerable variation in number, size and position among the five taxa analyzed (Table 2). The number of 45S rDNA loci varied as follows: one in *V. luteola* and *V. caracalla*, two in *V. vexillata* and *V. trilobata* and five in *V. minima* (Figs 1B, D, F, H, J, 2A–E). In *V. luteola*, the single 45S locus comprised the entire short arms of chromosome pair 3 (Figs 1A, B, 2A). The two 45S loci in *V. vexillata* were terminally located on the short arms of pairs 2 and 3 and occupied the majority of the arms (Figs 1C, D, 2B). All the five 45S loci in *V. minima*, four major and one minor loci, were terminally located on the long or short arms of chromosome pairs 2, 4, 6, 7 and 9, among which the locus on pair 2 was syntenic to the 5S locus (Figs 1E, F, 2C). In *V. trilobata*, one major locus occupied the entire short arms of pair 7, and one minor locus was pericentromERICALLY placed on the short arms of pair 6 (Figs 1G, H, 2D). The single 45S locus in *V. caracalla* occupied the entire short arms...
Molecular cytogenetics of five wild Vigna species

Figure 3. GISH with Vigna umbellata genomic DNA probe (red) to the chromosomes of V. luteola (A, B), V. vexillata (C), V. minima (D, E), V. trilobata (F, G), and V. caracalla (H, I). A, D, F, H CPD banded chromosomes before the hybridization procedure. The chromosomes shown in C is the same spread shown in Figure 1C, D. The chromosomes showing interstitial CPD bands and pericentromeric DAPI+ bands in H are numbered according to the karyotype of this species. Arrowheads in B, C, E, G, I indicate the signals corresponding to the 45S rDNA sites. Arrows in H indicate the interstitial CPD bands. Scale bars: 10 μm.

of pair 1 except the satellites (Figs 1I, J, 2E). The 45S rDNA chromatins of the terminal loci in the five taxa accounted for 32.06–100% of the arm length (calculated from the size of the rDNA CPD bands; Table 2).

cGISH signal patterns

Comparative genomic in situ hybridization with V. umbellata genomic DNA probe was employed to reveal the homology of repetitive DNA sequences between V. umbellata and the five wild Vigna species (Fig. 3). The genomic probe produced 45S rDNA signals in all species and non-rDNA signals in V. luteola and V. trilobata (Fig. 3B, G). All 45S sites were strongly labeled by the genomic DNA probe in the five species (Fig. 3B,
In V. luteola, except for the 45S signals, weak signals were generated in the proximal regions of the two arms of each chromosome (Fig. 3A, B), while in V. trilobata, weak non-rDNA signals were mainly concentrated in all (peri)centromeric regions, which basically corresponded to the (peri)centromeric CPD bands (Fig. 3F, G).

**Discussion**

**Karyotype variation**

In the current study, detailed karyotypes of V. luteola, V. vexillata, V. minima, V. trilobata and V. caracalla are established using a dataset of chromosome measurements, fluorochrome bands, and rDNA-FISH signals, thus providing the first primary molecular cytogenetic characterization of these wild Vigna species. Although FISH mapping of rDNAs in V. vexillata var. tsusimensis Matsumura, 1902 has been conducted (Chio et al. 2013), but the detailed karyotype of this species has not yet been established. Our results reveal that the karyotypic parameters and patterns of the fluorochrome bands and rDNA sites vary among the five Vigna species studied, enabling an accurate distinction between individual genomes.

This study identifies the chromosome number of all the five species as 2n = 22, in accordance with that reported previously by other authors (Sen and Bhowal 1960; Joseph and Bouwkamp 1978; Rao and Chandel 1991; Galasso et al. 1993; Venora and Saccardo 1993; Venora et al. 1999; Shamurailatpam et al. 2012, 2016; Choi et al. 2013). The conventional karyotypes of the five species studied here have been reported by earlier workers (Joseph and Bouwkamp 1978; Rao and Chandel 1991; Venora et al. 1999; Shamurailatpam et al. 2016). However, the published karyotype formulae of V. minima (Shamurailatpam et al. 2016), V. trilobata (Rao and Chandel 1991) and V. caracalla (Joseph and Bouwkamp 1978) were not comparable because the chromosomes were not classified according to the system of Levan et al. (1964). The current karyotypes of V. luteola and V. vexillata, n = 11m, are more symmetric than the karyotypes reported by Venora et al. (1999), which were comprised of both metacentric and submetacentric chromosomes. This discrepancy is probably due to difference in the accessions analysed, and difficulty in identifying chromosomes using the classical staining technique in the previous studies.

The results reveal significant variation in karyotype length (TCL) among the five taxa studied. For example, the TCL of V. caracalla was 1.82 times longer than that of V. vexillata. Except V. caracalla, the TCLs of the other four wild species were much shorter than those of the seven cultivated Vigna species obtained previously by us (She et al. 2015). With respect to the karyotype asymmetry (according to the AI values), among the five wild and seven cultivated Vigna species that has been studied using molecular cytogenetic method, V. minima and V. subterranea have the lowest asymmetry; V. radiata, V. mungo var. mungo and V. aconitifolia have the most asymmetric; V. luteola, V. vexillata, V. trilobata, V. caracalla, V. unguiculata ssp. sesquipedalis, V. angularis and V. umbellata are intermediately asymmetric (She et al. 2015).
Heterochromatin differentiation

The significant variation in CPD and DAPI+ bands, with regard to appearance, position and size, reflects distinct GC-rich and AT-rich heterochromatin differentiation among the five wild Vigna species (She et al. 2006; She and Jiang 2015). Similar heterochromatin differentiation has been observed among the seven cultivated Vigna species (She et al. 2015). As we know, heterochromatic blocks are chromosomal regions that contain a high density of satellite DNA and transposable elements (Heslop-Harrison and Schwarzacher 2011). These facts indicate that alterations in repeated DNA sequences have contributed to the karyotypic differentiation during the diversification of Vigna species (de Moraes et al. 2007; Hamon et al. 2009; Robledo et al. 2009; Mondin and Aguiar-Perecin 2011; She et al. 2015; Amosova et al. 2017).

With the exception of the rDNA CPD bands, V. luteola, V. trilobata, and V. caracalla also displayed centromeric and pericentromeric non-rDNA CPD bands. Especially, V. caracalla possessed interstitial non-rDNA CPD bands, which have not been observed in other Vigna species (She et al. 2015). Centromeric, pericentromeric or proximal GC-rich heterochromatin without colocalization with rDNA sites have been observed by using CPD or CMA/DAPI staining on the chromosomes of the seven cultivated Vigna species (de A Bortoleti et al. 2012; She et al. 2015) as well as many other Phaseoloid species such as the two cultivated Canavalia (Adanson, 1763) species (She et al. 2017), Crotalaria (Linnaeus, 1753) species of Calycinae and Crotalaria sections (Mondin and Aguiar-Perecin 2011), Lablab purpureus (Linnaeus, 1753) Sweet, 1826 (She and Jiang 2015), the four cultivated Phaseolus (Linnaeus, 1753) species (Bonifácio et al. 2012) and Psophocarpus tetragonolobus (Linnaeus, 1753) Candolle, 1825 (Chawen et al. 2004). These facts suggest that the existence of (peri)centromeric GC-rich heterochromatin is an ancestral genome feature that occurred before the divergence of the Phaseoloid clade of the subfamily Papilionoideae (LPWG 2013). However, the inexistence of non-rDNA GC-rich heterochromatin in V. vexillata and V. minima seems to be in contradiction with this speculation. A reasonable explanation is that the non-rDNA GC-rich heterochromatin of these two species has undergone a reduction of GC content after speciation, resulting in the disappearance of red CPD bands (She et al. 2006). The changes of non-rDNA CPD bands in amount, distribution, and GC content have been observed among the seven cultivated Vigna species. For example, in V. radiata, non-rDNA GC-rich heterochromatin blocks disappeared from five pairs of chromosomes; in V. mungo, non-rDNA GC-rich heterochromatin blocks occurred only in the proximal regions of the long arms of eight pairs of chromosomes (She et al. 2015). As for the GC-rich regions corresponded to the 5S rDNA sites that observed in V. minima, the variation in the base composition of the non-transcribed spacer (NTS) of the 5S rDNA repeats or the interspersion of other GC-rich repeated DNAs with the 5S rDNA repeats may explain it (Cabra et al. 2006; Hamon et al. 2009).

The occurrence of the pericentromeric DAPI+ bands in V. caracalla was another conspicuous heterochromatic differentiation of this species. Among the Vigna species previously analyzed by fluorochrome banding technique, AT-rich heterochromatin blocks have been observed in the pericentromeric regions of several chromosome pairs
of *V. radiata* (de A Bortoleti et al. 2012; She et al. 2015). The AT-rich heterochromatin in *V. radiata* and *V. caracalla* should arise after the divergence of *Vigna* species because of its non-universality.

**Variation of rDNA loci**

To date, FISH mapping of rDNA sites has been reported only for *V. vexillata* var. *tsusimensis* among the wild species within the genus *Vigna* (Choi et al. 2013). Regarding the number and position of rDNA loci of this species, our findings is significantly different from the previous report, in which three pairs of 45S loci and two pairs of 5S loci were observed (Choi et al. 2013). The identified divergence could be due to the difference in the accessions analysed.

Our rDNA-FISH results reveal considerable variations in number, position and even size of both 45S and 5S rDNA sites among the five wild *Vigna* species studied. Similarly, wide interspecific differences in the pattern of rDNA sites were observed among the seven cultivated *Vigna* species (She et al. 2015). Inferring from the rDNA-FISH data of the twelve *Vigna* species investigated by us, the FISH patterns of the 45S rDNA sites in species of this genus were more polymorphic than those of the 5S rDNA. This phenomenon has been reported in many different plant genera such as *Phaseolus* Linnaeus, 1753 (Moscone et al. 1999), *Paeonia* Linnaeus, 1753 (Zhang and Sang 1999), *Brassica* Linnaeus, 1753 (Hasterok et al. 2001), *Oryza* Linnaeus, 1753 (Chung et al. 2008), *Coffea* Linnaeus, 1753 (Hamon et al. 2009), *Brachypodium* P. Beauvois, 1812 (Wolny and Hasterok 2009), *Citrullus* Schrader ex Ecklon & Zey-her, 1836 (Li et al. 2016) and *Allium* Linnaeus, 1753 (Maragheh et al. 2019). The interspecies and intraspecific variations in the number and location of rDNA sites has been attributed to various mechanisms such as transposon-mediated transposition, homologous and/or non-homologous unequal crossing over, inversion, translocation and locus duplication/deletion (Moscone et al. 1999; Zhang and Sang 1999; Datson and Murray 2006; Pedrosa-Harand et al. 2006; Chung et al. 2008; Raskina et al. 2008; Weiss-Schneeweiss et al. 2008). The differentiation in the chromosomal organization of rDNA clusters between plant species was generally correlated with the chromosome evolution during speciation (Datson and Murray 2006; Moscone et al. 2007; Raskina et al. 2008; Weiss-Schneeweiss et al. 2008). Among the five taxa studied the number of 5S loci is rather conserved: four species had a single 5S locus located in pericentromeric or interstitial regions. Similar, five of the seven cultivated *Vigna* species had only one 5S locus that was located in the proximal, interstitial, pericentromeric or centromeric regions (She et al. 2015). Furthermore, among the twelve species that were investigated using molecular cytogenetic approaches by us, the single 5S locus in *V. luteola*, *V. umbellata* and *V. aconitifolia* and one 5S locus in *V. radiata* were located in the pericentromeric, centromeric, or proximal regions of the short arms of the shortest chromosome pair (She et al. 2015). These facts suggest that the ancestral progenitor of the genus *Vigna* bear a single 5S locus that is located on the short arms of the shortest chromosomes
in the complement. Chromosome rearrangements such as inversion and translocation may change the position of the 5S locus or produce longer 5S-bearing chromosomes (Moscone et al. 2007; Chung et al. 2008; Weiss-Schneweiss et al. 2008; She et al. 2015). The increased number of 5S loci in V. caracalla probably originated from the transposition of the 5S rDNA (Raskina et al. 2008). As for 45S site, one, two, three, four and five loci were identified in the twelve Vigna species studied by us, respectively (She et al. 2015). A total of thirty-one 45S loci were detected in the twelve species, among which twenty-four were terminal and seven were pericentromeric. Considering that V. aconitifolia and V. luteola had a single terminal 45S locus and the Aconitifoliae section was the ancestral section within the subgenus Ceratotropis (Doi et al. 2002), the ancestral progenitor genome of Vigna species might bear a single terminal 45S locus. Another terminal 45S locus in V. vexillata, and the other four terminal 45S loci in V. minima might result from one or more non-homologous unequal crossing over between the terminal chromosomal regions (Zhang and Sang 1999; Pedrosa-Harand et al. 2006). The pericentromeric 45S locus in V. trilobata, like the pericentromeric locus in V. unguiculata subsp. sesquipedalis (Linnaeus, 1753) Verdcourt 1970, and three pericentromeric 45S loci in V. umbellata (She et al. 2015), might originate from transposition of the terminal 45S rDNA cluster (Datson and Murray 2006; Chung et al. 2008; Raskina et al. 2008).

**Phylogenetic relationships**

In the early time, the Vigna genus was divided into seven subgenera (Maréchal et al. 1981). Delgado-Salinas et al. (2011) proposed, based on phylogenetic analysis of cpDNA trnK and nuclear ribosomal ITS/5.8S (ITS) sequence variation, a new circumscription of Vigna Savi sensu stricto, which includes five subgenera, Ceratotropis, Haydonia, Lasiospron, Plectrotropis, and Vigna, of the seven recognized by Maréchal et al. (1981). The Vigna subg. Sigmoidotropis of Maréchal et al. (1981), in which V. caracalla was previously placed, is now divided into six genera, Ancistrotropis A. Delgado, 2011, Cochliasanthus Trew, 1764, Condylostylis Piper, 1926, Leptospron (Benth. and Hook.f., 1865) A. Delgado, 2011, Helicotropis A. Delgado, 2011, and Sigmoidotropis (Piper, 1926) A. Delgado, 2011 (Delgado-Salinas et al. 2011). V. caracalla is transferred to the monotypic genus Cochliasanthus, and named as Cochliasanthus caracalla. Our molecular cytogenetic karyotyping data revealed that this species had several distinct characteristics compared to the other eleven Vigna species studied by us: existence of several interstitial CPD bands, pericentromeric DAPI bands, as well as satellites associated with the short arms that consist of 45S rDNA clusters (She et al. 2015). These facts indicate that V. caracalla significantly differentiates from other Vigna species at chromosome level, supporting the taxonomic separation of V. caracalla from the genus Vigna (Delgado-Salinas et al. 2011).

Among the remaining four wild Vigna species analyzed, both V. luteola and V. vexillata are of African origin being categorized into Vigna subg. Vigna and subg.
Haydonia, respectively (Delgado-Salinas et al. 2011), while both V. minima and V. trilobata are Asiatic Vigna (subg. Ceratotropis) species, and belong to Section Angulares and Section Aconitifoliae, respectively (Doi et al. 2002; Goel et al. 2002; Javadi et al. 2011). The molecular phylogeny of Vigna has been investigated intensively using sequence data from the rDNA ITS, the IGS of 5S rDNA, and chloroplast DNA (Doi et al. 2002; Goel et al. 2002; Tün and Yamaguchi 2007; Saini et al. 2008; Saini and Jawali 2009; Delgado-Salinas et al. 2011; Javadi et al. 2011; She et al. 2015; Raveenadar et al. 2018). Here the molecular phylogenies revealed by other authors and the molecular cytogenetic data obtained by us are combined to analyze the phylogenetic relationships among the wild and cultivated Vigna species studied molecular-cytogenetically by us. The molecular phylogenetic trees inferred from cpDNA trnK and nrDNA ITS sequence by Delgado-Salinas et al. (2011) revealed that V. luteola and V. subterranea were included within the same group of one African Vigna subclade and belonged to different subgroups, while V. vexillata and V. unguiculata were included within the same group of another African Vigna subclade and placed at different subgroups; V. minima, V. umbellata and V. angularis were included within one subclade of the subg. Ceratotropis clade and clustered into three different subgroups, while V. trilobata and V. aconitifolia were included within another subclade of subg. Ceratotropis clade and clustered into different subgroups. Similar phylogenetic relationships among these species mentioned above were also revealed using the IGS of 5S rDNA (Saini and Jawali 2009), and the sequences of rbcL + psbA-trnH + ITS2 + matK region (Raveenadar et al. 2018). Our previous rDNA-FISH revealed that V. subterranea had two terminal and one pericentromeric 45S loci, and a single interstitial 5S locus located on a medium-sized chromosome pair (She et al. 2015), being significantly different from the rDNA distribution pattern of V. luteola. Especially, non-rDNA cGISH signals of V. umbellata genomic DNA probe were produced in V. luteola but not in V. subterranea. These facts suggest that there is significant genome differentiation between V. luteola and V. subterranea, in disagreement with the molecular phylogeny. Specially must point out in here, the production of non-rDNA cGISH signals on the chromosomes of V. luteola with V. umbellata genomic DNA probe was perplexing because V. luteola and V. umbellata belong to different subgenera and should be relatively distantly related (Delgado-Salinas et al. 2011). To solve this puzzling problem, more V. luteola accessions need to be studied using FISH. Our molecular cytogenetic data also revealed prominent differentiation between V. vexillata and V. unguiculata because, compared to V. unguiculata, V. vexillata lacked (peri)centromeric GC-rich regions and had less number of 45S and 5S loci (de A Bortoleti et al. 2012; She et al. 2015). The reported molecular phylogenies showed that V. minima and V. umbellata, V. trilobata and V. aconitifolia were closely related, respectively (Doi et al. 2002; Goel et al. 2002; Saini and Jawali 2009; Delgado-Salinas et al. 2011). Our molecular cytogenetic data support the close relationship between V. trilobata and V. aconitifolia because both of them had (peri)centromeric CPD bands, similar 45S-bearing chromosome pair (pair 7 and pair 4 in V. trilobata and V. aconitifolia, respectively), and pericentromeric cGISH signals of V. umbellata genomic DNA probe (She et al. 2015). However, the close
relationship between *V. minima* and *V. umbellata* was not confirmed by the molecular cytogenetic data because *V. minima* lacked cGISH signals of *V. umbellata* genomic DNA probe, and (peri)centromeric CPD bands which existed in all (peri)centromeric regions of *V. umbellata* (She et al. 2015). In summary, our molecular cytogenetic data not only partially support the molecular phylogenetic relationships between related *Vigna* species, but also reveal considerable genome differentiation between the *Vigna* species that have been proved to be closely related by molecular phylogenetic analysis. It is necessary to clarify the conflicts between the molecular phylogenies and molecular cytogenetic data by performing integrated study of molecular phylogenetic and molecular cytogenetic analyses using more accessions of related *Vigna* species.

### Conclusions

Molecular cytogenetic karyotypes of five wild *Vigna* species, *V. luteola*, *V. vexillata*, *V. minima*, *V. trilobata* and *V. caracalla* are established for the first time using fluorochrome banding and rDNA-FISH techniques. Comparative molecular cytogenetic karyotyping reveals distinct variations in the karyotypic parameters, and the patterns of the fluorochrome bands and rDNA sites among species, enabling an accurate distinction between individual genomes. The molecular cytogenetic data of the five species is helpful to clarify the phylogenetic relationships among related *Vigna* species.

### Acknowledgements

This work was supported by the Natural Science Foundation of Hunan Province, China (No. 09JJ3063 and No. 2019JJ40231) and the Foundation of Hunan Double First-rate Discipline Construction Projects, China.

### References


Mondin M, Aguiar-Perecin ML (2011) Heterochromatin patterns and ribosomal DNA loci distribution in diploid and polyploid Crotalaria species (Leguminosae, Papilionoideae), and inferences on karyotype evolution. Genome 54(9): 718–726. https://doi.org/10.1139/g11-034


**Supplementary material 1**

**Table S1. Chromosome measurements of the five wild Vigna species obtained from five metaphases per species**

Authors: Chao-Wen She, Ying Mao, Xiang-Hui Jiang, Chun-Ping He

Data type: species data

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/CompCytogen.v14i2.51154.suppl1
The epigenetic regulation of centromeres and telomeres in plants and animals

Magdalena Achrem¹,², Izabela Szućko¹,², Anna Kalinka¹,²

¹ Institute of Biology, University of Szczecin, Szczecin, Poland ² Molecular Biology and Biotechnology Center, University of Szczecin, Szczecin, Poland

Corresponding author: Magdalena Achrem (magdalena.achrem@usz.edu.pl)

Abstract

The centromere is a chromosomal region where the kinetochore is formed, which is the attachment point of spindle fibers. Thus, it is responsible for the correct chromosome segregation during cell division. Telomeres protect chromosome ends against enzymatic degradation and fusions, and localize chromosomes in the cell nucleus. For this reason, centromeres and telomeres are parts of each linear chromosome that are necessary for their proper functioning. More and more research results show that the identity and functions of these chromosomal regions are epigenetically determined. Telomeres and centromeres are both usually described as highly condensed heterochromatin regions. However, the epigenetic nature of centromeres and telomeres is unique, as epigenetic modifications characteristic of both eu- and heterochromatin have been found in these areas. This specificity allows for the proper functioning of both regions, thereby affecting chromosome homeostasis. This review focuses on demonstrating the role of epigenetic mechanisms in the functioning of centromeres and telomeres in plants and animals.

Keywords

cytosine methylation, histone code, non-coding RNA, pericentromeric, subtelomeric
Introduction

The term epigenetics refers to a variety of processes that change gene expression independently of DNA sequence. An important feature of the epigenetic pattern is that it is stable and inherited through cell divisions, although it can be reversible (John and Rougeulle 2018). Epigenetics is crucial for the proper development, differentiation and functioning of cells. The epigenome may change under the influence of various environmental conditions and stimuli from inside the cell (Shi et al. 2017). This epigenome diversity is provided by numerous epigenetic mechanisms, including DNA methylation, post-translational histone modifications, chromatin remodeling, histone variants and ncRNA (non-coding RNA) interaction (Kabesch et al. 2010).

DNA methylation is of great importance among the epigenetic mechanisms that regulate gene expression in plants and animals. DNA methylation is associated with gene silencing (Kumar et al. 2018). Methylcytosine (5-mC) is the most common among the modified bases in the eukaryotic genome and is often referred to as the fifth DNA base. Methylation of cytosine in DNA involves the covalent attachment of a methyl group at position 5 of the cytosine pyrimidine ring (5-mC). Analysis of the DNA methylation profile of the human genome showed that mainly cytosines in CpG dinucleotides are modified. In plants, cytosine methylation in DNA occurs in the CHG sequential contexts (H = C, A, T) and asymmetrically in CHH (Zhang et al. 2008). Cytosine methylation in DNA is catalyzed by DNA methyltransferases. In mammalian cells, DNA methyltransferase (DNMT1) is responsible for maintaining the methylation pattern during replication, DNMT3A (DNA methyltransferase 3A) and DNMT3B (DNA methyltransferase 3B) for de novo methylation. In plants, MET1 (methyltransferase 1), DDM1 (decrease in DNA methylation 1), CMT1 (chromomethylase 1) and DRM2 (domain rearranged methyltransferase 2) DNA methyltransferases are necessary to maintain the correct methylation pattern (Ogrocká et al. 2014, Zhang et al. 2018).

Chromatin remodeling results from the action of ATP-dependent complexes that change the association of DNA with core histones and from modifications of histone proteins, affecting the availability of DNA (Kang et al. 2020). The remodeling complexes change the structure of chromatin by repositioning, evicting or restructuring the nucleosome. Some complexes are involved in the formation of condensed chromatin, others promote the binding of transcription factors to DNA. They are therefore involved in such important processes as DNA transcription and replication, DNA repair and DNA recombination (Clapier and Cairns 2009). Chromatin remodeling factors are involved in the development and differentiation of cells in plants and animals. Chromatin remodelers include several sub-families of ATP-dependent enzymes. Each of these subfamilies has a specific composition of domains and subunits that are involved in histone exchange, assembly and repositioning of nucleosomes (Kang et al. 2020).

Post-translational modifications of histone proteins are another important epigenetic mechanism. Histones (H2A, H2B, H3 and H4) are the basic protein component of the nucleosome that forms the core around which a DNA strand of about 146 bp is wrapped (Luger et al. 1997). The N- or C-terminal tails of histones undergo
The epigenetic regulation of centromeres and telomeres in plants and animals

Post-translational modifications. These modifications include arginine (R) methylation, methylation, acetylation, ubiquitination and sumoylation of lysine (K) as well as phosphorylation of serine (S) and threonine (T). The pattern of these modifications creates a histone code, which shows the transcription potential of this genomic region (Kabesch et al. 2010). Appropriate histone modifications are necessary for the proper course of such important cellular processes as: DNA repair, replication, mitosis, apoptosis and gametogenesis. Histones, through post-translational modifications, participate in the regulation of DNA packaging, affecting the availability of chromatin for transcription factors (Quina et al. 2006). Histone modifications can change the structure of chromatin by changing the physical properties of individual nucleosomes. This affects the interaction between the DNA molecule and histone and creates an open chromatin structure that is available for many protein factors, or a higher order chromatin structure that prevents these factors from binding. These modifications are strengthened by protein complexes that do not participate in chromatin modifications, but by influencing its remodeling, they are of great importance for the epigenetic gene regulation (Kim et al. 2012). An important role in regulating the structure of chromatin is also played by histone variants, which differ from canonical histones by the amino acid sequence. The presence of specific histone variants affects transcription regulation, chromosome segregation, DNA repair, cell cycle regulation and apoptosis (reviewed in Henikoff and Smith 2015).

Epigenetic regulators also include non-coding RNA (ncRNA). In epigenetic processes, the most important role among non-coding RNAs is played by those molecules that act in the RNAi (RNA interference) pathway and certain IncRNA (long non-coding RNAs, over 200 nt in length) (Kurokawa et al. 2009). Detailed studies of biogenesis and function of ncRNA have elucidated their activity at many levels, forming an integrated interacting network in the cell. They can regulate expression at both the gene and chromosome level (Amaral and Mattick 2008) and can act at transcriptional and post-transcriptional levels by interacting with promoters, enhancers or chromatin remodeling complexes (Kurokawa et al. 2009). However, their influence is not limited to the euchromatin, as exemplified by centromeric sequences, where ncRNAs are necessary for the assembly and proper functioning of both centromere and kinetochore (Bobkov et al. 2018).

Most of the presented epigenetic mechanisms are closely associated with each other to ensure stabilization and transmission of epigenetic patterns from cell to cell during cell divisions. They interact with each other in different ways. DNA methylation can promote changes in histone modification and vice versa. However, they can also change accidentally under the influence of stimuli coming from the internal and external environment (Kabesch et al. 2010). Epigenetic mechanisms do not act solely at the level of gene expression regulation. They also play a key role in maintaining genomic stability. They are involved in the regulation of centromeres, telomeres and silencing transposable elements (TE), which enables proper chromosome segregation, reduces excessive recombination between repetitive elements, and prevents TE transposition (Dupont et al. 2009).

However, there is a fairly close connection between epigenetic regulators and the spatial structure of the cell nucleus due to the fact that the organization of chromatin is
epigenetically determined. In turn, the organization of chromatin influence the spatial structure of the cell nucleus. Based on the studies of the nucleus of mammalian cells, chromatin was divided into following compartments A – euchromatin, B – facultative heterochromatin (Solovei et al. 2016) and C – pericentromeric constitutive heterochromatin (Falk et al. 2019). It was shown that attractions between heterochromatic regions play crucial role in separation of the active from inactive parts of the genome in the nucleus. Constitutive heterochromatin, enriched with tandem repetitive sequences and transposable elements, located in the centromeric, pericentromeric or subtelomeric areas is the most enigmatic fraction of chromatin. Most of the heterochromatic regions remains unassembled due to their enrichment with the tandem repetitive sequences. The majority of the assembled mammalian genomes contain a 3 Mb Golden Path Gap (GPG) empty region around each centromere. However, gradually, more and more data on the composition of the sequence of constitutive heterochromatin regions are becoming available (Ostromyshenskii et al. 2018). Constitutive heterochromatin turns out to be surprisingly heterogeneous, characterized by plasticity, and its epigenetic regulators depend on the genomic context in which it is present. Although constitutive heterochromatin is gene-poor, its role turns out to be very significant (Saksouk et al. 2015).

The epigenetic nature of both centromeric and telomeric regions is not clearly defined. This is because these are regions built from repetitive sequences, which makes it difficult to accurately show epigenetic modifications of centromeres and telomeres. This review focuses on demonstrating how epigenetic mechanisms affect the functioning of centromeric and telomeric regions, taking into account differences in plants and animals.

**Centromere and pericentromere**

The centromere was first described by Walther Flemming (1882), who observed that there was one region in the chromosome that was smaller in diameter than the remaining portion of the chromosome. Cytogenetic and molecular analyses demonstrated centromeres as heterochromatin chromosomal domains that control the formation of the kinetochore, a protein structure that interacts with the mitotic spindle, ensuring proper segregation of chromosomes (reviewed in Cleveland et al. 2003, Allshire and Karpen 2008, Salmon and Bloom 2017).

The simplest centromere with a length of 125 bp is found in *Saccharomyces cerevisiae* (Meyen, 1883). This simple, small centromere contains a single cenH3 (centromere specific histone 3) nucleosome, which binds a single microtubule during cell division, which is why this centromere type is called the “point centromere” (Pluta et al. 1995, Furuyama and Biggins 2007). Numerous studies have shown that not all eukaryotic organisms have monocentric chromosomes characterized by the presence of the primary constriction. In some species, microtubules of the mitotic spindle attach to the chromosome along its entire length (White 1973). Thus, two types of chromosomes are distinguished: monocentric chromosomes that connect to the microtubules of the
spindle in a single region, and holocentric chromosomes, characterized by the presence of dispersed kinetochores that bind to spindle microtubules over their entire length (Wrench et al. 1994, Mandrioli and Manicardi 2012).

Holocentric chromosomes have been found in some plants (e.g. the genus *Luzula* Candolle and Lamarck, 1805), animals (several arthropods and nematodes) and Rhizaria (Cavalier-Smith, 2002) (Allshire and Karpen 2008, Heckmann et al. 2013). It is believed that holocentromeres have been evolved from monocentromeres at least 13 times independently, and their organization varies among taxa (Melters et al. 2012). The type of DNA sequence responsible for the formation of dispersed centromeres is not yet fully elucidated. The sequences located in the holocentromeres are very diverse, including those that directly bind cenH3. In *Rhynchospora pubera* (Linnaeus, 1872) holocentromeres are enriched in specific satellite DNA sequences (Tyba) (which bind CENH3) and retrotransposons (Ribeiro et al. 2017). In *Caenorhabditis elegans* (Mau pas, 1900) specific satDNA (satellite DNA) sequences that bind CENH3 are dispersed all over the genome (Subirana et al. 2018). In turn, no centromere-specific sequences were found in *Luzula elegans* (Lowe, 1838) (Heckmann et al. 2013). Hence, cenH3 probably binds not to specific sequences but to chromatin of appropriate status, indicating epigenetic regulation of holocentromers. The unusual structure of holokinetic chromosomes is also associated with the specific course of meiosis. Three types of meiosis can be distinguished in different species characterized by holocentric chromosomes: ‘chromosome remodeling’, ‘functional monocentricity’ and ‘inverted chromatid segregation’ (Heckmann et al. 2014, Lukhtanov et al. 2018). In *C. elegans* chromosome remodeling ensure chromosomes segregation typical for monocentric chromosomes. Other species have developed functional monocentricity, i.e. attachment of microtubules to one terminus of the chromosome, thus, holocentric chromosomes act as monocentric. These adaptations allow for a course of meiosis similar to canonical meiosis. In the first meiotic division, homologous chromosomes segregate, while sister chromatids are separated during the second meiotic division. However, many species with holokinetic chromosomes have developed an inverted meiosis, in which the order of major meiosis events is reversed, i.e. the sister chromatids are separated first (which results, among others, from the inability to maintain cohesion of sister chromatids up to AII (anaphase II) in holocentric chromosomes), followed by segregation of homologues (Heckmann et al. 2014, Lukhtanov et al. 2018).

In monocentric chromosomes of animals and plants the centromere region constitutes a segment from several kb to Mb in size, that contains satellite DNA with repeating monomers of ~100–400 bp (Melters et al. 2013). In general, chromosome centromeres in one species are characterized by the occurrence of a single family of sequence repeats (Zhong et al. 2002, Nagaki et al. 2003, Henikoff et al. 2015). This type of centromere restricted to a certain region is referred to as the regional centromere (Melters et al. 2013, Liu et al. 2015, Kursel and Malik 2016).

In plants, the centromeric region is composed of alternating tandem repeats and retrotransposons. For example, sequencing of maize centromeric DNA revealed two types of repetitive sequences in this region: satellite CentC (156 bp monomer) and
retrotransposon CRM (centromeric \textit{retrotransposon} of maize) sequences (Ananiev et al. 1998, Zhong et al. 2002, Birchler and Han 2009). In B chromosome of maize, an additional sequence was identified in this region known as B-repeat (Alfenito and Birchler 1993), flanked and interspersed with typical maize centromeric sequences, i.e. CentC and CRM (Jin et al. 2005, Lamb et al. 2005). A similar organization of sequences is found in rice centromeres, where the CentO satellite repetitive sequence (155 bp monomer) as well as the CRR (centromeric retrotransposon of rice ) retrotransposon sequence are distinguished (Cheng et al. 2002); other examples are pBV repetitive sequences and \textit{r} retrotransposon of the beetle family in \textit{Beta vulgaris} (Linnaeus, 1753) (Zakrzewski et al. 2013). A combination of satellite repeats in association with retrotransposons in the centromere region was also detected in \textit{Hordeum} (Linnaeus, 1753) (Houben et al. 2007), \textit{Saccharum officinarum} (Linnaeus, 1753) (Nagaki and Murata 2005), \textit{Brassica} (Linnaeus, 1753) (Wang et al. 2011), \textit{Raphanus sativus} (Linnaeus, 1753) (He et al. 2015) and \textit{Glycine} (Linnaeus, 1753) (Tek et al. 2010).

Human centromeres are characterized by the presence of satellite tandem repeats of -171 bp in size, arranged “head-to-tail”, that are further arranged in higher order repeats (HOR). Individual monomers share 50–70% sequence identity, but HORs have 95–98% similarity (Warburton et al. 1996, Alcan et al. 2007). The functional core of the centromere is composed of highly homogeneous HORs, and, depending on the chromosome, spans a region from 0.5 to 5 Mb (Altemose et al. 2014), flanked by 500-kb segments, containing \textit{L1} (\textit{LINE1}, long interspersed nuclear \textit{elements}) mobile elements (Schueler et al. 2001, Aldrup-MacDonald and Sullivan 2014). Within the human centromere, in the \textit{a} satellite DNA sequences, 17-bp sequence motifs occur, referred to as the CENP-B box, which are recognized by centromere protein B (CENP-B) (Masumoto et al. 1993). This protein has an important role in maintaining stability and in the proper arrangement of centromere nucleosomes, because it binds with N-terminus of CENP-A (centromere protein A) and CENP-C (centromere protein C) (Fachinetti et al. 2015, Fujita et al. 2015). Human Y chromosome (Choo 2001) or neocentromeres (Fachinetti et al. 2015) are an exception, as the CENP-B box sequences and CENP-B proteins were not detected, while other centromeric proteins were present. It is known, however, that the lack of the CENP-B box in \textit{α}-satellite sequences or mutations in these regions do not allow the formation of artificial chromosomes (Zhang et al. 2010). This suggests that CENP-B is not necessary for the centromere function, however, it contributes to its stabilization and maintenance (Schalch and Steiner 2017).

Centromeric DNA sequences are evolving relatively fast (Melters et al. 2013), which seems surprising considering the conservative function of the centromere (Henikoff et al. 2001, Rosin and Mellone 2017). Large differences in centromere sequences among wild \textit{Oryza} species (Linnaeus, 1753) (Lee et al. 2005), cultivated \textit{Cannalvia} (Adanson, 1763) species (She et al. 2017), between related species of \textit{Solanum tuberosum} (Linnaeus, 1753) and \textit{S. verrucosum} (Schlechtendal, 1839) (Zhang et al. 2014), or within one species of \textit{Pisum sativum} (Linnaeus, 1753) (Macas et al. 2007), can serve as examples. Hence, it is presumed that centromeres are not genetically determined by the occurrence of a specific DNA sequence, but they are rather epigenetically
defined by characteristic modifications (Simon et al. 2015). The confirmation of this fact are neocentromeres, which act as centromeres at the new chromosomal site even if satellite sequences are not present there (Williams et al. 1998, Marshall et al. 2008). Although satellite DNA is an inherent element of centromeres, it is not required for the functioning of these regions (Willard 1990, Csink and Henikoff 1998). Nevertheless, repeated DNA is the preferred DNA environment for centromere formation, and if the neocentromere is formed in a region devoid of repetitive sequences, then they begin to gradually accumulate there (Han et al. 2009, Plohl et al. 2014).

The centromeric core, which provides the kinetochore attachment site, is flanked by pericentromeric regions. Pericentromeric chromatin stabilizes the centromeric core, inhibiting internal recombination between core repeat sequences (Hettr and Allis 2005), and is responsible for the attachment of sister chromatids during cell division (Schalch and Steiner 2017), promoting bidirectionality and creating tension between them (Bernard et al. 2001, Sakuno et al. 2009, Yamagishi et al. 2010, Yi et al. 2018).

Pericentromeres, like the core centromere, mainly consist of repetitive sequences. Among the sequences included in pericentromeric DNA, there are satellite sequences, as well as transposons, LTR and non-LTR retrotransposons (Smurova and Wulf 2018). Typically, these regions are described as genetically inactive, although some of the sequences found in these regions, such as 5S rRNA genes are highly transcribed (Cloix et al. 2002, Simon et al. 2015). Pericentromeric sequences show both inter- and intraspecific variation (Charlesworth et al. 1994, Plohl et al. 2008).

Epigenetic regulation of centromeres and pericentromeres

As previously mentioned, it is believed that satellite DNA is not essential for maintaining centromere structure and function. The term “centromere paradox” defines the fact that centromere sequences are very variable, while centromere function is conservatively maintained. However, as it turns out, centromere functionality does not result from the composition of the relevant DNA sequences, but the epigenetic mechanisms are responsible for it (Allshire and Karpen 2008). Epigenetic mechanisms play an important role in the establishment, maintenance and functioning of centromeres (Allshire and Karpen 2008) (Table 1). Centromere can be inactivated (Sullivan and Schwartz 1995, Han et al. 2006, Zhang et al. 2010), but also can switch from the inactive to active state, enabling transcription of ncRNA, which plays a role in the proper functioning of the centromere (Han et al. 2009). Centromeric ncRNAs interact with many proteins i.a. CENP-A (Rošić et al. 2014), CENP-B (Carone et al. 2009), CENP-C (Du et al. 2010), HJURP (Quénet and Dalal 2014) and AURORA B (Ferri et al. 2009). For example, centromere inactivation in dicentric chromosomes or activation of neocentromeres in non-centromeric regions were reported (Williams et al. 1998, Nasuda et al. 2005, Marshall et al. 2008, Topp et al. 2009).

The results of studies on the epigenetic regulation of centromeric regions are ambiguous. The difficulty in studying these regions is caused by the fact that centromeres
in most multicellular eukaryotes are formed of numerous copies of repetitive sequences (Henikoff et al. 2001). Identification of individual epigenetic modifications is particularly difficult if the sequences of the same family of repeats have different epigenetic markers. For this reason, many studies do not present unequivocal results. There is also a limitation in the selection of methods to study these regions. For example, standard methods used to map DNA methylation, including high-throughput techniques based on microarrays and WGBS sequencing (bisulfite sequencing-based platforms), do not allow to assess methylation within highly repetitive DNA sequences. Therefore, in this case, immunofluorescence (IF) analysis is often used in combination with FISH (fluorescence in situ hybridization) on stretched DNA fibers (Koo et al. 2011).

Many studies on centromere chromatin in *Arabidopsis thaliana* (Linnaeus, 1753) have shown that it forms chromocentres in the interphase nuclei, it is rich in H3K9me2, characterized by DNA hypermethylation and enrichment in histone variant H2A.W (Probst et al. 2003, Stroud et al. 2013, Yelagandula et al. 2014). However, comprehensive IF studies using anti-5-methylcytosine antibody showed that the DNA in centromeric region is unmethylated. IF on the stretched fibers of the early pachytene chromosomes confirmed these observations, indicating that DNA sequences (178 bp tandem repeats) in the core regions with CENH3 were differently methylated than in the flanking pericentric regions. Regions in which CENH3 is present, and directly adjacent regions, are

### Table 1. Epigenetic modifications of centromeric regions and their functions in plants and animals.

<table>
<thead>
<tr>
<th>Epigenetic modification</th>
<th>Region</th>
<th>Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>histone variant CENH3</td>
<td>centromeric</td>
<td>specifies centromere location</td>
<td>Gieni et al. 2008</td>
</tr>
<tr>
<td>CENP-A</td>
<td></td>
<td>essential for kinetochore assembly</td>
<td></td>
</tr>
<tr>
<td>H3K4me1, H3K4me2, H3K36me2,</td>
<td>centromeric</td>
<td>maintenance of centromere stability</td>
<td>Yan et al. 2005 Foltz et al. 2009</td>
</tr>
<tr>
<td>H3K36me3</td>
<td></td>
<td>RNA II pol activity recruitment of HJURP</td>
<td></td>
</tr>
<tr>
<td>proteins CENP-A</td>
<td></td>
<td>deposition</td>
<td></td>
</tr>
<tr>
<td>H4K5ac and H4K12ac</td>
<td>centromeric</td>
<td>CENP-A deposition</td>
<td>Shang et al. 2016</td>
</tr>
<tr>
<td>H4K20ac</td>
<td>centromeric</td>
<td>required for transcriptional activity</td>
<td>Shang et al. 2016</td>
</tr>
<tr>
<td>required for kinetochore</td>
<td></td>
<td>formation in human and <em>Gallus</em> cells</td>
<td></td>
</tr>
<tr>
<td>H2AT133ph H2AT120ph</td>
<td>centromeric</td>
<td>recruitment of Shugoshin (Sgo1) protein</td>
<td>Kawashima et al. 2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prevents precocious separation of sister</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>chromatids</td>
<td></td>
</tr>
<tr>
<td>monoubiquitinated H2B</td>
<td>centromeric</td>
<td>required for transcriptional activity</td>
<td>Sadeghi et al. 2014</td>
</tr>
<tr>
<td>(H2Bub1)</td>
<td></td>
<td>provides structural integrity required for</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>proper chromosome segregation</td>
<td></td>
</tr>
<tr>
<td>H3K9me</td>
<td>pericentromeric</td>
<td>chromatin condensation ensures</td>
<td>Gieni et al. 2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>chromatid cohesion provides</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>structural integrity</td>
<td></td>
</tr>
<tr>
<td>H4K20me</td>
<td>pericentromeric</td>
<td>chromatin condensation provides</td>
<td>Gieni et al. 2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>structural integrity</td>
<td>Hori et al. 2014</td>
</tr>
<tr>
<td>H3K27me</td>
<td>pericentromeric</td>
<td>transcriptional repression of transposable</td>
<td>Jacob et al. 2010 Feng et al. 2017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>elements</td>
<td></td>
</tr>
<tr>
<td>H3 and H4 lysine residues</td>
<td>pericentromeric and</td>
<td>increase in chromatin compaction</td>
<td>Gieni et al. 2008</td>
</tr>
<tr>
<td>acetylation</td>
<td>centromeric</td>
<td>heterochromatin integrity</td>
<td></td>
</tr>
<tr>
<td>Cytosine methylation of DNA</td>
<td>pericentromeric and</td>
<td>chromatin condensation provides</td>
<td>Gieni et al. 2008</td>
</tr>
<tr>
<td></td>
<td>centromeric</td>
<td>structural integrity</td>
<td>Song et al. 2013</td>
</tr>
</tbody>
</table>
The epigenetic regulation of centromeres and telomeres in plants and animals

unmethylated or significantly less methylated, while the remaining 178 bp repeats are highly methylated. Thus, DNA sequences in centromeric chromatin are hypomethylated compared to the sequences found in the flanking pericentric chromatin (Zhang et al. 2008). In addition, a correlation was found in Arabidopsis between the occurrence of 5mC and H3K9me2 in centromeric regions. Similar results were obtained while studying centromeric regions in maize. The methylation status of centromeric CentC repeats in maize is variable, whereby, similarly to Arabidopsis, DNA sequences associated with CENH3 in maize are hypomethylated (Koo et al. 2011).

In contrast, studies on centromeres in rice have shown that DNA sequences in a functional centromere can be both hypo- and hypermethylated. DNA methylation patterns appear to be correlated with specific sequence motifs (CG, CHG, CHH) in centromeric DNA (Yan et al. 2010). Detailed studies of the centromeric maize region have shown that there is a tendency of increased DNA methylation in CG and CHG motifs towards the centromere and decreased towards the chromosomal arms. This was also observed in Populus trichocarpa (Torrey et Grey, 1851) (Feng et al. 2010, Zemach et al. 2010). In turn, CHH methylation was relatively similar in different maize chromosomal domains, which was also confirmed by studies concerning rice centromere (Feng et al. 2010). Although general methylation level was similar in centromeres and pericentromeres, a slight increase in CG methylation and a decrease in CHG was observed in the centromeric core, with a marked difference between centromeres (Gent and Dawe 2012). This variation may result from the relative differences in the size of CentC sequence stretches in the individual centromeres (Jin et al. 2005).

Research on the level of DNA methylation in medaka fish (Oryzias latipes Temminck et Schlegel, 1846) demonstrated that centromeres are mainly hypermethylated, but have hypomethylated subregions (Ichikawa et al. 2017). It was found that DNA methylation patterns in centromeres were not correlated with the phylogensis of centromeric sequences, but the hypo-/hypermethylated regions in individual chromosomes evolved independently by acquiring a unique sequence composition. In turn, examining methylation level in mouse cells, it was found that it depended on the type of tissue being tested. The highest level was observed in somatic cells, intermediate in sperm and the lowest in egg cells (Yamagata et al. 2007).

Centromeric chromatin (CEN) is characterized by the presence of specific histone H3 variant – cenH3 (CENP-A in mammals, CID (centromere identifier) in Drosophila melanogaster (Fallén, 1823), CENH3 in plants) (Steiner and Henikoff 2015). In multicellular eukaryotes, centromeres consist of alternating blocks of nucleosomes containing H3 or cenH3 (Blower et al. 2002, Sullivan and Karpen 2004, Alonso et al. 2007). The cenH3 nucleosomes recruit complexes that directly bind to cenH3, which in turn allows the attachment of numerous centromeric proteins termed CCAN (constitutive centromere-associated network) (Foltz et al. 2006, Carroll et al. 2009) (Fig. 1). The HJURP chaperone protein (Holliday junction recognition protein) is involved in the process of CENP-A deposition and complex formation between CENP-A and H4 (Shuaib et al. 2010). The structure of human CENP-A differs from canonical H3 histone, inter alia, by loop 1, which contains two additional amino acid residues (Arg80 and Gly81), affecting centromere chromatin stabilization (Tachiwana et al. 2011,
González-Barrios et al. 2012). CENP-A shows only 50% homology to H3 amino acid sequence. There is also variation in length and sequence of N- and C-termini among these proteins (Malik and Henikoff 2003), simultaneously the C-terminus retains the...
hydrophobic region necessary for interaction with CENP-C (Kato et al. 2013). Moreover, it was shown that around the nucleosome containing CENP-A only 121 bp of the DNA is wrapped, 13 bp from both DNA ends are invisible in the crystal structure suggesting highly flexible ends (Tachiwana et al. 2011, Roulland et al. 2016). This structure disrupts the binding of histones H1 with the nucleosomes, allowing a more open configuration of the chromatin, which in turn enables the attachment of the CCAN complex (Roulland et al. 2016). Studies have shown that there are structural differences between CENP-A/H4 and H3/H4 heterotetramers (reviewed in Verdaasdonk and Bloom 2011). The presence of the CENP-A protein in the nucleosome ensures its more compact and rigid structure (Black et al. 2007). Similarly to CENP-A, plant centromeric CENH3 is characterized by significant variability between species (Malik and Henikoff 2009). CENH3 has a conserved histone-fold domain (HFD), instead the most significant differences in the structure of this protein in relation to H3 occur at the N-terminus (Ravi et al. 2010; Lermontova et al. 2014). This may be due to the fact that the C-terminus of CENH3 is responsible for histone H4 binding, which allows the formation of stable nucleosomes (Feng et al. 2019).

In human CEN chromatin, nucleosomes containing the CENP-A variant alternate with nucleosomes with the canonical histone H3. Histones H3 in this region undergo methylation at lysine positions 4 and 36 (H3K4me1, H3K4me2, H3K36me2, H3K36me3), characteristic of transcriptionally active chromatin. They affect RNA pol II (RNA polymerase II) activity and play an important role in the recruitment of HJURP proteins that participate in the CENP-A deposition (Bergmann et al. 2011, Duda et al. 2017). The absence of H3K4me2 in the centromere of artificial human chromosomes resulted in the inactivation of this centromere (Bergmann et al. 2011), which shows a functional link between epigenetic modification of CEN chromatin and maintaining centromere stability. Similarly, in plants, dimethylation of histone H3 at lysine 4 (H3K4me2) is a common modification in the centromeric H3 subdomains (Wu et al. 2011), which was not observed, for example, in the CENH3 subdomains of rice. It has even been hypothesized that the transcribed sequences located in the rice centromere can be a barrier preventing the introduction of CENH3 into the region of H3 subdomains. This separation of the CENH3 and H3 subdomains in the centromere core may be necessary for the formation of three-dimensional structure and functioning of rice centromere (Wu et al. 2011).

Interestingly, CEN is not usually associated with the presence of H3K9me2 or H3K4me3 heterochromatin markers, although H3K9me3 modification has been shown in this region to be associated with transcription repression (Bergmann et al. 2012). This illustrates that CEN chromatin can be both silenced heterochromatin as well as active euchromatin (Sullivan and Karpen 2004), however, it is important that the balance between them is preserved. Introduction of repressors or activators of transcription in artificial chromosomes disrupts the balance between modifications such as H3K4me2 and H3K9me3, which leads to the loss of the centromere function (Nakano et al. 2008).

In maize centromeres, the presence of histone post-translational modifications associated with transcriptional activity, such as histone H4 acetylation and H3K4me2, has been revealed. It was indicated that centromeres in this species are organized as eu-
chromatin regions flanked by pericentromeric H3K9me2-enriched heterochromatin (Yan et al. 2005). Histone H4 acetylation (H4K5ac and H4K12ac) was also detected in Gallus (Brisson, 1760) cells as a modification necessary for CENP-A deposition (Shang et al. 2016). It was shown that H4K20ac is essential for transcription of ncRNA, which is necessary for the deposition of CENP-A and kinetochores assembly in human and Gallus cells (Sullivan and Karpen 2004, Wang et al. 2008, Bergmann et al. 2011, Hori et al. 2014). Moreover, for the transcription of centromeric DNA monoubiquitination of lysine 119 in histone H2B (H2BK119ub1) must occur (Zhu et al. 2011, Sadeghi et al. 2014). It is mediated by the ubiquitin ligase E3 RNF20 (ring finger protein 20) in humans or Brl1 in Schizosaccharomyces pombe (Lindner, 1893) (Sadeghi et al. 2014). The H2BK119ub1 modification interacts with many proteins such as RNA pol II and SWI/SNF (switch/sucrose non-fermentable) protein complexes (Shema-Yaacoby et al. 2013), which contributes to the formation and maintenance of transcriptionally active chromatin. This modification also affects centromere integrity and accurate chromosome segregation. It has been shown that the decrease in RNF20 level results in H2B-K119ub1 deficiency in this region, which in turn causes heterochromatin formation, thereby reducing the transcription of the centromeric DNA sequence and resulting in an abnormal chromosome segregation in human and S. pombe (Lindner, 1893) cells (Sadeghi et al. 2014, Zhang et al. 2017).

CENP-A is less likely to undergo post-translational modification than canonical histone H3 (Fig. 2). This is due to, inter alia, the lower lysine content in CENP-A. In histone H3, up to 17 different types of post-translational modifications were found (Xu et al. 2014), whereas only four modifications were detected in CENP-A: methylation, acetylation, phosphorylation and ubiquitination (Srivastava and Foltz 2018). The most characteristic CENP-A modifications are Gly1 trimethylation, Ser 7, 16, 18 and 68 phosphorylation and monomethylation, acetylation and ubiquitination of lysine 124. These CENP-A-specific modifications, play an important role in chromosome segregation during cell division, because they regulate CENP-A deposition in centromeric chromatin and participate in CCAN recruitment (Srivastava and Foltz 2018).

It has long been believed that centromeric chromatin is transcriptionally inactive because it is formed mainly by satellite sequences. It is now known that CEN transcription is mediated by RNA pol II, which was detected in centromeric regions in both S. pombe, Drosophila, mouse, human, Zea (Linnaeus, 1753), Oryza (Linnaeus, 1753) and neocentromeres, as well as in CEN of human artificial chromosomes (HAC) (Chueh et al. 2009, Ferri et al. 2009, Ohkuni and Kitagawa 2011, Chan and Wong 2012, Podgornaya et al. 2013, Quénet and Dalal 2014, Rošić et al. 2014). The important role of transcription in centromere integrity was shown by numerous studies on its inhibition, which resulted in the loss of centromere function (Quénet and Dalal 2014, Rošić et al. 2014, Sadeghi et al. 2014). Many genes have been identified in the centromeric regions of various plants, including rice (Jiang 2013) and A. thaliana (May et al. 2005). Transcribed centromeric elements can activate the process of RNAi by forming siRNA (small interfering RNA) and affecting both DNA and histone modifications in the centromeric region (Lippman and Martienssen 2004).
Studies also showed transcriptional activity of centromeric retrotransposons that affect the formation, stabilization and functioning of centromeres (Jiang et al. 2003, Topp et al. 2004). An example is the CRM transcript in maize, which contributes to the stabilization of centromere chromatin (Topp et al. 2004) or the CRR transcript in rice that is involved in the formation and maintenance of centromeres through RNAi pathway (Neumann et al. 2007). The additional evidence, that transcription of centromeric DNA is common, is the presence of H3K4me2 modification in this region of many plants (onion, rice, Arabidopsis, maize). Maintaining CEN chromatin in the active state and its transcription is also necessary for the replacement of histone H3 with cenH3 (Quénet and Dalal 2014, Bobkov et al. 2018). The lack of centromeric transcripts leads to disturbances during mitosis (Quénet and Dalal 2014). Centromeric chromatin is transcriptionally active even during mitotic division (Chan et al. 2012), which ensures stability of kinetochores and coherence of centromeres (Liu et al. 2015). Phosphorylation of centromeric histone H2A (H2AT120ph in animals, H2AT133ph in plants) by the Bub1 (budding uninhibited by benzimidazoles 1) kinase is required for the recruitment of the Shugoshin protein (Sgo1). This protein ensures chromatid

Figure 2. Epigenetic modifications in centromeric and pericentromeric chromatin. Centromeres consist of alternating blocks of nucleosomes containing H3 or cenH3. At pericentric sites, only H3-containing nucleosomes are present. Epigenetic markers in centromere and pericentromere regions characteristic for both plants and animals are marked with black color, only for plants with violet color, only for animals with rose color. (+) epigenetic marker always present; (-/+ epigenetic modification present or absent.

Studies also showed transcriptional activity of centromeric retrotransposons that affect the formation, stabilization and functioning of centromeres (Jiang et al. 2003, Topp et al. 2004). An example is the CRM transcript in maize, which contributes to the stabilization of centromere chromatin (Topp et al. 2004) or the CRR transcript in rice that is involved in the formation and maintenance of centromeres through RNAi pathway (Neumann et al. 2007). The additional evidence, that transcription of centromeric DNA is common, is the presence of H3K4me2 modification in this region of many plants (onion, rice, Arabidopsis, maize). Maintaining CEN chromatin in the active state and its transcription is also necessary for the replacement of histone H3 with cenH3 (Quénet and Dalal 2014, Bobkov et al. 2018). The lack of centromeric transcripts leads to disturbances during mitosis (Quénet and Dalal 2014). Centromeric chromatin is transcriptionally active even during mitotic division (Chan et al. 2012), which ensures stability of kinetochores and coherence of centromeres (Liu et al. 2015). Phosphorylation of centromeric histone H2A (H2AT120ph in animals, H2AT133ph in plants) by the Bub1 (budding uninhibited by benzimidazoles 1) kinase is required for the recruitment of the Shugoshin protein (Sgo1). This protein ensures chromatid
coherence in internal centromeres (Kawashima et al. 2010). Sgo1 interacts with RNA pol II and is directed to the inner centromere between two sister chromatids. The open chromatin structure in the centromeric region allows binding of the Sgo1 protein to cohesin and provides protection against premature chromatid separation (Kang et al. 2011, Liu et al. 2015). Initiation of centromeric DNA transcription must be preceded by chromatin remodeling. An important factor in this process is a histone chaperone, FACT (facilitates chromatin transcription) (reviewed in Reinberg and Sims 2006). FACT allows transcription through the destabilization of nucleosomes, allowing polymerase to access DNA (Belotserkovskaya et al. 2003). After polymerase passes, it allows a return to the earlier chromatin structure (Jamai et al. 2009).

It has also been proven that the region directly adjacent to the centromere plays a role in sister chromatid cohesion (Bernard et al. 2001, Steiner and Henikoff 2015). Between the prophase and anaphase, sister chromatids are kept together in pericentromeres after cohesins are removed from other chromosome regions (Nasmyth and Haering 2009). There are known various epigenetic mechanisms associated with chromatin silencing that provide cohesion maintenance in pericentromeres (HP1-heterochromatin protein 1, H3K9me3, RNAi) (Mosch et al. 2011). Changes in this region may lead to impairment of proper chromosome segregation (Allshire et al. 1995, Steiner and Henikoff 2015). However, there are hypotheses that this heterochromatin region is necessary to establish the centromere, but is not required to retain it (Folco et al. 2008). In addition, studies on neocentromeres, which can form in euchromatin areas, indicate that pericentromeric heterochromatin (PHC) is not necessary for the proper functioning of the centromere (Shang et al. 2013). Nevertheless, it is believed, that pericentromeric heterochromatin regions may play a role in preventing the centromere from spreading to adjacent regions (Sullivan 2002). From an epigenetic point of view, pericentromeres show a greater similarity to centromeres than to other chromosomal regions. This is reflected in siRNA transcription, DNA methylation and some post-translational modifications of histones. Although there is evidence that centromeres may function independently of pericentromeres, as found, for example, in studies conducted on *S. cerevisiae* (Weber et al. 2004), there is a strong interdependence of these two regions (Han et al. 2006).

Histones in pericentric chromatin are mostly hypoacetylated, which causes chromatin condensation. Pericentromeric areas are characterized by the presence of histone variants H3.3 and H2A.Z (Drané et al. 2010, Santenard et al. 2010), modifications of histones such as mono-, di- and trimethylation of H3K9, H3K27 and H4K20 (Feng et al. 2017) and a high level of 5-mC in DNA (Song et al. 2013). These modifications are characteristic of transcriptionally inactive chromatin and play a role in the silencing of genetic mobile elements occurring abundantly in these chromosomal regions (Roudier et al. 2011, Rose and Klose 2014, Feng et al. 2017). For example, monomethylation of lysine 27 in histone H3 is associated with constitutive repression of transcription. This was confirmed by the study of pericentromeric regions of polytene chromosomes of *Drosophila*. They correspond to green – inactive (the division of chromatin into the following shades: red, yellow, blue, green and black; according to Filon et al. 2010) or ruby chromatin (the division of chromatin into the following shades: aquamarin,
lazurite, malachit and ruby; according to Zhimulev et al. 2014), which is characterized by H3K27 methylation as well as SU(VAR)3-9 and HP1 presence (Boldyreva et al. 2017). Loss of H3K27 methylation in the pericentromeric regions causes transposons reactivation (Jacob et al. 2010). This may result in a cancer or other diseases such as ICF (immunodeficiency, centromere instability, facial anomalies). ICF is a rare autosomal recessive disease characterized by a lack of DNMT3B activity. DNA methylation depletion results in the loss of repressive histone modifications (often H3K27me3) and the appearance of modifications characteristic of euchromatin (H3K9ac, H3K4me), which further leads to reactivation of transposons (Jin et al. 2008).

A characteristic protein of this region is HP1 or its homologs (Guenatri et al. 2004, Cam et al. 2005), which affect the stabilization and maintenance of the heterochromatin state (Saksouk et al. 2015) of pericentromeric regions. The HP1 protein interacts with the Suv39h histone methylation kinase, which catalyzes the trimethylation of lysine 9 in H3 (Aagaard et al. 1999, Grewal and Jia 2007). In mice, it has been found that Suv39h deficiency results in a lack of H3K9me3, disrupting the occurrence of HP1 in the pericentromeric heterochromatin, which in turn translates into abnormal chromosomal segregation (Peters et al. 2001, Maison et al. 2002). The heterochromatic nature of the pericentromeric region is also confirmed by the analysis of marker gene expression. Inserted into the pericentromeric region, they are transcriptionally silenced, while the insertion of the same genes into the CEN region shows a significantly weaker silencing effect (Allshire et al. 1995).

The analysis of human neocentromeres that showed centromere functioning without satellite repeats (although they had a slightly higher AT content, from 59.9 to 66.1% compared to genomic average of 59%). The acquisition of centromeric function by a chromatin region without changing the DNA sequence was called the “centromerization” phenomenon (Choo 2000). Such neocentromeres, formed outside the centromeric regions, while maintaining the characteristics of the original centromere without the underlying centromere DNA, were also observed in animals and plants (Gallus (Bisson, 1760), Equus (Linnaeus, 1758), Solanum (Linnaeus, 1753), Hordeum (Linnaeus, 1753), Avena (Linnaeus, 1753) and Zea (Linnaeus, 1753)) (Nasuda et al. 2005, Ishii et al. 2008, Kagansky et al. 2009, Topp et al. 2009, Piras et al. 2010, Gong et al. 2012, Fu et al. 2013, Shang et al. 2013). The existence of neocentromeres and rapid evolution of centromeric DNA suggest that these are epigenetic mechanisms, rather than DNA sequence itself, that determine centromere functions (Piras et al. 2010).

Studies on dicentric chromosomes also support this fact. Dicentric chromosomes are the result of genomic rearrangements placing two active centromeres on the same chromosome. Most dicentric chromosomes are unstable and only due to epigenetic mechanisms, which deactivate one of the centromeres, monocentric chromosomes can be formed that normally segregate during cell division (Sullivan and Schwartz 1995, Chiatante et al. 2017). If one of the centromeres is not turned off, the chromosome breaks during division. DNA sequences of the active and inactive centromeres of dicentric chromosomes are almost identical, but the centromere activity states are completely different. Centromere inactivation on the dicentric chromosome is carried out
by H3K27me2 and H3K27me3. Smaller centromeres appear to be inactivated more frequently than the larger ones (Han et al. 2009). It was confirmed by analyses of dicentric chromosomes in plants e.g. Zea mays (Linnaeus, 1753), (Han et al. 2006), Oryza sativa (Linnaeus, 1753) (Wang et al. 2013) and in humans. This explains some processes regarding the formation and maintenance of neocentromeres in human, because neocentromeres are always smaller than the native ones. If small centromeres are more susceptible to inactivation compared to larger ones, then most of the newly formed neocentromeres will be inactivated during subsequent cell divisions (Zhang et al. 2010).

Evolutionary repositioning or shift of the centromere along the chromosome with its function, leading to the formation of new evolutionary centromeres (ENCs), is another phenomenon that shows the epigenetic nature of these structures. This phenomenon was observed in primate chromosomes, other placental, marsupials and birds (Montefalcone et al. 1999, Ventura et al. 2007, Piras et al. 2010, Zlotina et al. 2012). The beginning of repositioning causes the loss of the function of the original centromere, followed by epigenetic changes in the non-centromeric position, leading to the formation of a new functional centromere in the chromosome region devoid of satellite DNA (Montefalcone et al. 1999). The resulting neocentromere may gradually accumulate repetitive DNA sequences through recombination mechanisms during evolution (Piras et al. 2010). Accumulation of these sequences probably ensures the stabilization of the centromere during cell division (Marshall et al. 2008), facilitates incorporation of histone cenH3 (Steiner and Henikoff 2015) and the accuracy of chromosomal segregation (Piras et al. 2010). All these reports shed more light on the role of satellite sequences. Despite their heterogeneity between species, a common pattern of structural DNA motifs required for centromere specification begins to be noticed (Black and Giunta 2018, Oliveira and Torres 2018). This hypothesis is supported by the fact that de novo chromosome formation revealed preferential centromere occurrence in areas built of tandem repeats (Grimes et al. 2002, Masumoto et al. 2004, Nagaki et al. 2004, Han et al. 2009).

**Telomere and subtelomere**

Telomeres are specialized structures located at the ends of linear eukaryotic chromosomes. Their function is to protect the ends of chromosomes from inappropriate enzymatic degradation. They are also responsible for chromosome localization in the cell nucleus and transcription regulation of genes located near telomeres (Deng et al. 2008, Fojtová and Fajkus 2014). Telomeres also protect chromosomes from fusions, formation of dicentric chromosomes and homologous recombination (Artandi and DePinho 2010). While telomere function has been well known for a long time, the role of the subtelomeric region is still being investigated. It is indicated that subtelomeres support telomeres in their function, because they may affect processes such as cell cycle regulation, cell aging, motility and chromosomal localization in the nucleus (Riethman et al. 2005).

Due to the important functions they perform in the cell, telomeres are evolutionarily conserved regions and their structure is only slightly different in individual species.
However, the length of telomeric sequences shows individual, tissue and cellular variability (Marión and Blasco 2010). Telomeres contain a double-stranded region composed of tandem DNA repeats, which can be described by the following formula: $5' - T_x(A)G_y - 3'$ ($x, y$ – number of repeats) and single-stranded free $3'$ end rich in guanine (G-overhang) (Wang and Zakian 1990, Smogorzewska and de Lange 2004), whose length varies from 16 to 200 nt depending on the species (Kazda et al. 2012). There are, however, exceptions from the above formula for telomere monomers, e.g. in *Allium cepa* (Linnaeus, 1758) this is the $(CTCGGTTATGGG)_n$ sequence (Fajkus et al. 2016), in *Genlisea* (Bentham and Hooker, 1883) two sequence variants TTCAGG and TTTCAGG (Tran et al. 2015) and in *Ascaris lumbricoides* (Linnaeus, 1758) – TTAGGC (Müller et al. 1991). In general, however, it is assumed that this sequence in vertebrates consists of $(TTAGGG)_n$ tandem repeats (Moyzis et al. 1988), $(TTAGG)_n$ in arthropods (Kuznetsova et al. 2015), and in most plants – $(TTTAGGG)_n$ (Richards and Ausubel 1988). The telomere sequence is usually very homogeneous, particularly in contrast to the subtelomeric sequences constituting a border region between the telomere and the region where genes are located. The subtelomeric regions include a fragment of about 500 kb (Macina et al. 1994) and similarly as telomeres, it consists of repetitive DNA sequences. However, the presence of genes and CpG islands has not been found in telomeres, while the subtelomers are characterized by the presence of a small number of genes and CpG islands (Blasco 2007). The common feature of the subtelomeric regions of various eukaryotic organisms is the presence of long arrays of tandem repetitive (TR) sequences or duplicated DNA fragments, which also include telomeric sequence motifs (Torres et al. 2011).

In mammals, the DNA stretch comprising a telomere is terminated with single-stranded free G-overhangs of varying, species-specific length (Kazda et al. 2012). G-overhangs are important for telomere maintenance, acting as a primer for telomerase (Lingner and Cech 1996). These $3'$ ends form a spatial structure called the G-quadruplex (G4-DNA), which protects the telomere from exonucleases, thereby protecting the DNA strand against degradation (Sen and Gilbert 1988), and also inhibits telomerase activity (Zahler et al. 1991).

Telomeric chromatin has a typical organization, forming the nucleosome fiber at the basal level. This structure may be different only in regions where there are telomere-specific proteins (Pisano et al. 2007). Telomere structure is formed with the participation of a protein complex called shelterin (Fig. 3). The complex consists of six proteins: TRF1 and TRF2 (telomere repeat-binding factor 1 and 2) (Zhong et al. 1992, Chong et al. 1995, Bilaud et al. 1997), RAP1 (repressor/activator protein 1), TIN2 (TRF1-interacting nuclear factor 2) (Kim et al. 1999, Li et al. 2000), TPP1 (TINT1/PTOP/PIP1 protein) (Houghtaling et al. 2004) and POT1 (protection of telomeres 1) (Baumann and Cech 2001). TRF1 and TRF2 proteins bind to telomere double-stranded DNA, while other proteins stabilize the structure of the shelterin complex. The interaction between telomere DNA and shelterin proteins first of all protects and stabilizes telomere structure, and secondly, regulates the access of proteins involved in DNA repair and elongation (de Lange 2005). Double-stranded telomeric sequence, due to interactions with shelterin proteins, folds and closes forming a larger T-loop.
In turn, the free 3’ overhang at the end of the chromosome in the T-loop binds to the double-stranded telomere fragment to form a smaller D-loop. It has been found that the T-loops are characteristic of eukaryotic organism telomeres, although it is not certain whether they are present in all of them (de Lange 2004).

*D. melanogaster* telomeres have yet another structure. Three following retrotransposons have been identified in the telomere sequence: HeT-A, TART and TAHRE (HTT). At the ends of telomeres, there are numerous copies of HTT retrotransposon, while in the most proximal region, there are sequences called TAS (telomere associated sequence). The ends of telomeres are protected and stabilized by a protein complex. An important role is played by the heterochromatin 1 (HP1) protein, which binds to dimethyl lysine 9 in histone H3 (H3K9me2) (Vermaak and Malik 2009). Its absence contributes to the fusion of *Drosophila* chromosomes (Fanti et al. 1998).

In plants, telomeres are usually several kbs in size (*A. thaliana* – 2–9 kb), although they may be longer in some plants, e.g. tobacco telomeres may have a size of up to 150 kb (Richards and Ausubel 1988, Fajkus et al. 1995). G-overhang size may be 20–30 nt, however, it may not be present in all telomeres (Riha et al. 2000). Studies have shown that several proteins bind to telomeric dsDNA (double stranded DNA) as well as G-rich ssDNA (*single stranded DNA*), but they are not fully characterized.
Two proteins are known that bind to single-stranded telomeric sequences: GTBP1 (G-strand specific single stranded telomere-binding protein 1) and STEP1 (single stranded telomere-binding Protein 1) (Kwon and Chung 2004, Lee and Kim 2011). Homologs of the POT1 protein, which forms a heterodimer with the TPP1 protein have been also detected (Wang et al. 2007). Studies of the function of these proteins in *A. thaliana* showed that the POT1a homologue binds telomerase and is involved in the synthesis of telomere repeats, while the POT1b and POT1c homologs are involved in the protection of chromosome termini (Shakirov et al. 2005, Kobayashi et al. 2019). In *A. thaliana*, TRB proteins (telomere repeat-binding factors) were also identified (Mozgová et al. 2008), containing a conserved domain similar to the telobox-type Myb (short telomeric motif, Myb-related DNA-binding domain) (Peška et al. 2011), through which they bind to telomeric dsDNA. This domain is typical for mammalian TRF1 and TRF2 proteins, although differently located. In TRB proteins, it is present at the N-terminus and in TRF, at the C-terminus. In addition, TRB proteins were found to possess a histone-like domain (H1/5) that plays a role in DNA-protein reactions and interaction with POT1b (Schrumpfova et al. 2008).

**Epigenetic regulation of telomere and subtelomere regions**

The epigenetic nature of telomeres and subtelomeres remains controversial (Vaquero-Sedas and Vega-Palas 2011, Galati et al. 2013, Ichikawa et al. 2015, Adamusová et al. 2019). In the classic model, animal and plant telomeres were interpreted as heterochromatic structures (Kavi et al. 2005, Postepska-Igielska et al. 2013). However, more and more data indicate their dual character, showing modifications of histones characteristic of both the eu- and heterochromatin fraction (Vrbsky et al. 2010) (Fig. 4). Some studies even indicate that telomeres may exhibit mainly euchromatin traits, while subtelomeres – heterochromatin features (Vaquero-Sedas et al. 2011). However, this is not definitively established, especially that even the level and occurrence of DNA methylation within telomeres remains unexplained (Blasco 2007, Vrbsky et al. 2010, Vaquero-Sedas et al. 2012, Ogrocká et al. 2014).

The variety of information regarding telomere regions may partly result from experimental limitations, but also due to the epigenetic diversity of animal (Cubiles et al. 2018) and plant cells (Majerová et al. 2014). Difficulty in determining the epigenetic state of telomeric chromatin also results from the presence of interstitial telomere repeats (ITRs) within the internal regions of chromosomes. Most of the ITRs were found within or adjacent to the constitutive heterochromatin (Meyne et al. 1990, Rodionov et al. 2002, Galkina et al. 2005, Vaquero-Sedas and Vega-Palas 2011). ITR sequences differ from typical telomere sequences in that they are heterogeneous, degenerate and contain other sequence types in addition to telomere sequence repeats (Lin and Yan 2008, Vega-Vaquero et al. 2016).

Telomeric and subtelomeric chromatin studies in mouse showed the presence of histone modifications characteristic of the heterochromatin fraction (Garcia-Cao et al. 2004, Gonzalo et al. 2006). Telomeres in vertebrates, as well as in *D. melanogaster*,
are rich in H3K9me3 (Peters et al. 2001, Garcia-Cao et al. 2004). This modification is recognized by heterochromatic protein 1 (HP1), which can recruit histone methyltransferases (HMTase) such as SuM4-20h1 and SuM4-20h2, which methylate H4 at lysine 20 (H4K20me3) (Nakayama et al. 2001, Benetti et al. 2007). In telomeres, Dot1L HMTase mediates methylation of lysine 79 in H3 (H3K79me2) (Shanower et al. 2005) and methylates lysine 20 in H4 (H4K20me3) (Jones et al. 2008). In addition, histones H3 and H4 are not strongly acetylated in telomeres (Benetti et al. 2007). In human telomeres that lack SIRT6 deacetylase, a higher level of H3K9 acetylation is observed, which usually leads to telomere dysfunction (Michishita et al. 2008).

However, in mouse cells, telomeres are enriched in modifications specific to heterochromatin (H3K9me3) and euchromatin (H3K4me3). Although the H3K4me3 modification was at a lower level compared to H3K9me3 (Cao et al. 2009). ChIP-seq analysis of telomeres of various human cells has shown that they are characterized by low levels of H3K9me3, typical of heterochromatic regions, while they are enriched with euchromatin H4K20me1 and H3K27ac modifications (Rosenfeld et al. 2009, O’Sullivan et al. 2010, Cubiles et al. 2018).

Similar results were obtained in studies on plant telomeres. In Arabidopsis, heterochromatin modifications, such as H3K9me2 and H3K27me3, as well as euchromatin H3K4me3 modification have been reported (Vrbsky et al. 2010, Majerová et al. 2014, Adamusová et al. 2019). This occurrence of both heterochromatin and euchromatin modifications in the Arabidopsis telomere region was defined as the presence of an “intermediate” heterochromatin (Vrbsky et al. 2010, Majerová et al. 2014). Subsequent studies have shown that histones in telomeres have modifications typical of euchro-

Figure 4. Epigenetic modifications in telomere and subtelomere chromatin and adjacent euchromatin. Epigenetic markers in telomere and subtelomere regions characteristic for both plants and animals are marked with black color, only for plants with violet colour, only for animals with rose color.
matin, while histones within ITR regions possess modifications typical of condensed chromatin (Vaquero-Sedas et al. 2012). In the case of Ballantinia antipoda (Mueller, 1974), the H3K9me2 heterochromatin modification occurred mainly in telomeres, and H3K4me3 was found at a lower level, whereas only the H3K9me2 modification was present in the ITR region. Thus, it can be concluded that the chromatin of telomeres has both euchromatin and heterochromatin epigenetic markers, while the ITR regions are mainly heterochromatic (Majerová et al. 2014). In A. thaliana (Vrbsky et al. 2010) and Nicotiana tabacum (Linnaeus, 1753) telomeres, in addition to H3K9me2 and H3K4me3 modifications, the presence of H3K27me3 modifications was found, typical for heterochromatin, and it also occurs in human telomeres (Boros et al. 2014), although it is absent in mouse telomeres (Saksouk et al. 2014). Recent studies of human telomeres revealed that the PRC 2 (Polycomb 2) complex is responsible for the occurrence of H3K27me3, which affects the H3K9me3 heterochromatin modification to recruit HP1 to heterochromatin (Boros et al. 2014). It was also found that the TERRA transcript (TELomeric Repeat-containing RNA) is necessary for telomeric heterochromatin formation, the amount of modifications such as H3K9me3, H4K20me3 and H3K27me3 depends on the level of the TERRA transcript (Montero et al. 2018). It was found that lower levels of this transcript were associated with a decrease in the level of heterochromatin modifications in telomeres, H3K9m3 in particular (Deng et al. 2009).

Studies on telomere DNA methylation have not found so many discrepancies. Telomeres in mammalian cells are deprived of CpG dinucleotides, and therefore do not undergo DNA methylation (Draskovic and Londono-Vallejo 2013). Methylation studies of telomere sequences in plants have yielded conflicting results. Cytosine methylation in telomere CCCTAAA repeats was found in A. thaliana (Cokus et al. 2008), N. tabacum (Majerová et al. 2011), as well as in some other plants (Majerová et al. 2014). In turn, other studies on A. thaliana telomere DNA revealed low or no methylation (Vega-Vaquero et al. 2016). Detailed studies have shown that ITR sequences and sequences at the border of the telomere/subtelomere region are characterized by high levels of cytosine methylation (Cokus et al. 2008, Vrbsky et al. 2010, Vaquero-Sedas et al. 2012, Ogrocká et al. 2014). Very low level of genomic DNA methylation caused disturbances in telomere homeostasis in A. thaliana (Ogrocká et al. 2014, Xie and Shippen 2018), while no such changes were observed in N. tabacum (Majerová et al. 2011). This shows the differences in the role of DNA methylation in the regulation of telomere homeostasis in various plants (Fojtová and Fajkus 2014, Procházková-Schrumpfová et al. 2019).

While there is great controversy about the heterochromatic nature of telomeres, most studies show that this chromatin fraction is characteristic of subtelomeric regions. In animal and human cells, the subtelomeric regions are characterized by high CpG methylation and trimethylation of lysine 9 in histone H3 (H3K9me3) (Gonzalo et al. 2006). They can have a silencing effect on the expression of adjacent genes, as well as TERRA transcription. This silencing is defined as the telomere position effect (Azzalin et al. 2007, Cubiles et al. 2018). The analysis of most plant subtelomeric regions has also shown a high level of DNA methylation (Majerová et al. 2014, Ogrocká et al. 2014).

The heterochromatic state plays an important role in telomere biology, suggesting that the integrity of the subtelomeric heterochromatin may be important for the proper func-
tional of telomeres. A correlation was found between changes in the level of DNA methylation in the subtelomeric region and regulation of telomere length (Garcia-Cao et al. 2004, Gonzalo et al. 2006). In Arabidopsis, the subtelomeric region regulates the telomere length homeostasis. Genome hypomethylation in A. thaliana caused shortening of telomeres, although it was not so extensive to lead to genomic or chromosomal instability (Fajkus et al. 1995, Ogrocká et al. 2014). It has also been shown that post-translational modifications of histones have no effect on telomere length in N. tabacum (Majerová et al. 2011).

In budding yeasts, heterochromatinization of the subtelomeric region positively regulates telomere length (Nislow et al. 1997). For animals the opposite is true, a decrease in the occurrence of heterochromatin markers, including DNA methylation in the subtelomeric region, correlates with telomere elongation and increased recombination (Gonzalo et al. 2006, Benetti et al. 2007, Blasco 2007, Ng et al. 2009). An example is the research by Gonzalo et al. (2006), showing elongated telomeres with reduced methylation of the subtelomeric regions. Mouse mutants lacking DNA methyltransferases DNMT1 or DNMT3A and DNMT3B have very long telomeres and exhibit ALT (alternative lengthening of telomeres) characteristics, i.e. an increased rate of T-SCE (telomeric sister chromatin exchange) and the presence of APB (ALT-associated PML body) (Gonzalo et al. 2006).

Surprisingly, different reports have indicated that the length of telomeres does not change in epigenetic mutants (Roberts et al. 2011), or shown the association of very short telomeres with hypomethylation of subtelomeric regions (Benetti et al. 2007) or global hypomethylation (Pucci et al. 2013). In addition, telomere elongation has been linked to DNMT3A targeting to subtelomeric regions, resulting in increased DNA methylation (Cubiles et al. 2018).

For a long time, telomeres were perceived as silenced, transcriptionally inactive chromosome segments. This fact is negated by the presence of telomeric RNAs containing UUAGGG repeats, called TERRA, which are transcribed from the subtelomeric regions towards the ends of the chromosome by RNA pol II in yeasts, vertebrates and plants (Azzalin et al. 2007, Luke et al. 2008). The prevalence of these transcripts suggests that this is a conservative trait associated with an important function in telomere biology (Azzalin et al. 2007, Luke et al. 2008). Two classes of TERRA promoters were found in the chromosomes, and their expression is regulated by CTCF (CCCTC-binding factor) and RAD21 cohesin (radiation-sensitive 21) (Deng et al. 2012, Porro et al. 2014, Bettin et al. 2019). Absence or decrease in RAD21 or CTCF levels results in the loss of RNA pol II binding to TERRA promoters, resulting in the reduction in TERRA expression regions, therefore, an increase in DNA methylation in this region is associated with a decrease in the expression level (Yehezkel et al. 2008, Nergadze et al. 2009, Farnung et al. 2012). The correlation was shown between inhibition of TERRA transcription and the presence of H3K9me3, H4K20me3 and DNA methylation in telomeric and subtelomeric regions (Schoeftner and Blasco 2008, Nergadze et al. 2009, Farnung et al. 2012). Moreover, it turned out that histone acetylation and DNA hypomethylation positively affect the TERRA transcription process (Azzalin and Lingner 2008). Hypomethylation of subtelomeric sequences in mammalian cells lacking DNA
methyltransferases leads to TERRA overexpression. In mouse, TERRA transcript level in cell lines with deficiency of Suv3-9h and Suv4-20h HMTase is elevated compared to wild-type mouse cells. The level of epigenetic modifications characteristic for heterochromatin also regulates TERRA transcription in yeasts (Cusanelli and Chartrand 2014). In yeast, TERRA transcripts are maintained at a low level by Rat1 (Luke et al. 2008), the Sir2/Sir3/Sir4 sirtuin complex (histone deacetylases) and Rif1 and Rif2 (Rap1-interacting factor 1 and factor 2) (Iglesias et al. 2011). These results suggest that TERRA expression depends on the epigenetic status of subtelomeres and telomeres (Iglesias et al. 2011, Arnoult et al. 2012).

Binding of the TERRA transcripts to telomeres seems to be crucial for their structure and function (Luke et al. 2008). TERRA transcripts can negatively impact telomeres elongation. TERRA is believed to bind to the telomere region and regulate the length of telomeres by negatively controlling telomerase activity (Azzalin et al. 2007, Ng et al. 2009). Cells with active telomerase show a high level of TERRA promoter methylation, in contrast to those where the presence of this enzyme is not detected (Ng et al. 2009). This is probably because TERRA telomere repeats are complementary to the RNA template of telomerase and it is inhibited by competitive base pairing (Bisoffi et al. 1998). TERRA transcripts are involved in the formation of heterochromatin at chromosome ends interacting with the HP1 proteins and H3K9me3, as well as with HMTase Suv39H1 or Polycomb Repressive Complex 2 (PRC2) (Montero et al. 2018).

The interaction of TERRA transcript with TRF1 and TRF2 proteins can facilitate the binding of TERRA to the ends of chromosomes. Due to the fact that TRF1 and TRF2 can interact with chromosomes also in different regions (especially with ITR) (Simonet et al. 2011), TERRA transcripts can also bind non-telomeric sites (Cusanelli et al. 2013). TERRA, therefore, can regulate the expression of many genes (Chu et al. 2017). TERRA forms a complex with TRF2 and ORC1 (origin recognition complex 1), which facilitates DNA replication in telomeres (Deng et al. 2009). In addition, TERRA transcription itself, by the relaxation of chromatin, influences the initiation of DNA replication in this region during the S phase of the cell cycle (Bettin et al. 2019). It has been demonstrated that the expression level of TERRA depends on the phase of the cell cycle. It is high during the transition from the G1 to S phase, it is very high in the initial S phase, while it is reduced during the transition from the G2 phase to mitosis (Porro et al. 2010).

TERRA transcripts can promote homologous recombination between telomeres by creating RNA-DNA heteroduplex (R loops) at the ends of chromosomes (Chawla and Azzalin 2008). R loops can also block replication fork progression, cause double-strand breaks, delay cell aging and maintain genomic instability (Cusanelli and Chartrand 2015, Sollier and Cimprich 2015). For example, in the cells of the ICF syndrome, no methylation of the subtelomeric DNA was found, due to mutations in the DNMT3B gene. This results in a high level of the TERRA transcript, which forms telomeric R-loops, which in turn causes telomere dysfunctions (Cubiles et al. 2018). In addition, TERRA transcripts play a role in DNA damage response (DDR) caused by dysfunctional telomeres (Cusanelli and Chartrand 2015). Decrease in TERRA levels...
resulting from either the action of siRNA (Deng et al. 2009) or ASO-LNA (antisense oligonucleotides – locked nucleic acid) (Chu et al. 2017) as well as their incorrect localization leads to many chromosome abnormalities. Depletion of TERRA transcripts activates DDR at the ends of the chromosomes, which leads to the formation of the “telomere dysfunction-induced foci” (TIF) (Lopez de Silanes et al. 2010). Hence, proper expression and localization of TERRA is required to maintain telomeres and chromosomal stability (reviewed in Bettin et al. 2019).

Histone substitution with their variants is another epigenetic mechanism that plays a role in the functioning of telomeres. In human and mouse cells, histone H3.3 variant was correlated with TERRA transcriptional repression in telomeres and subtelomeres (Law et al. 2010). Telomeric histone H3.3 variant is deposited through the ATRX (alpha thalassemia/mental retardation syndrome x-linked)-DAXX (death-domain associated protein) complex. The loss of the function of this complex results in the reduction of modifications characteristic of heterochromatin fractions in telomeric regions, also associated with lower H3.3 levels. It has the destabilizing effect through increased homologous recombination of telomeres, which facilitates ALT (Heaphy et al. 2011). MacroH2A1.2 histone variant involvement in ALT has also been demonstrated. MacroH2A1.2 is present in telomeres, especially in ALT cells, being a mediator of homologous recombination and response to replication stress (Kim et al. 2019). H2A.Z is another histone variant that occurs in telomeres. In S. cerevisiae H2A.Z variants hinder the spread of the heterochromatin (Grunstein and Gasser 2013). A strong anticorrelation was found between this histone variant deposition and DNA methylation (Zilberman et al. 2008, Kobor and Lorincz 2009). Higher levels of the histone H2A.Z variant were observed in A. thaliana mutants with reduced DNA methylation. Thus, it can be pointed out that H2A.Z deposition somehow protects the genome against DNA methylation (Zilberman et al. 2008). The study of the Trypanosoma brucei (Plimmer and Bradford, 1899) chromatin showed the presence of the H3V (histone H3 variant) protein in the telomeres. It has been found that H3V has several features common to CenH3, however, its absence does not disrupt chromosomal segregation (Lowell and Cross 2004). Another example of the histone variant is sperm-specific spH2B. This variant of H2B forms a specific complex with DNA in vitro, which may indicate its role in the recognition of telomeric DNA. It is also believed that this protein may be involved in the attachment of telomeres to the nuclear envelope (Gineitis et al. 2000).

Conclusions

Centromeres and telomeres are indispensable elements of every functional chromosome in Eukaryota. Considering the conservative role, their structure should be similar, not only in the context of the DNA nucleotide sequence, but also at the level of chromatin organization. Whereas in the case of telomeres this can be seen, in centromeres the similarity is observed mainly at the level of epigenetic modifications, with
a great diversity of nucleotide sequences. Although microscopic analysis indicates that they are heterochromatin elements, they should now be considered as specific regions of the so-called intermediate heterochromatin, i.e. having epigenetic features of both euchromatin and heterochromatin. Undoubtedly, epigenetic status plays an extremely important role in regulating both telomeres and centromeres. For it is the specific structure of chromatin, and not just the DNA sequence itself, that ensures the proper functioning of these regions during the entire cell cycle. Many analyses have been carried out, the results of which were often contradictory, hindering an unambiguous determination of epigenetic markers of centromeric and telomeric regions.

However, these analyses have allowed us to perceive the epigenetic nature of telomeres and centromeres as very complex systems, precisely regulated at many levels. Disorders of this regulation can lead to destabilization of the entire genome. It also turned out that adjacent regions, i.e. subtelomeres and pericentromeres, often no less important than key elements, were thought for a long time to be heterochromatin boundary areas. Currently, it seems that maintaining their epigenetic status affects the structure and functioning of telomeres and centromeres. There is a need for further research on other species that will allow better understanding of telomere and centromere regulation systems in all their complexity.

References


The epigenetic regulation of centromeres and telomeres in plants and animals


The epigenetic regulation of centromeres and telomeres in plants and animals


The epigenetic regulation of centromeres and telomeres in plants and animals


is essential for mammalian development and heterochromatin structure. PLoS Genetics 4:e1000190. https://doi.org/10.1371/journal.pgen.1000190


Lopez de Silanes I, Stagno d’Alcontres M, Blasco MA (2010) TERRA transcripts are bound by a complex array of RNA-binding proteins. Nature Communications 1: 33. https://doi.org/10.1038/ncomms1032


and Aneuploidy. NATO ASI Series (Series H: Cell Biology), Springer, Berlin, Heidelberg, 31–43. https://doi.org/10.1007/978-3-642-84938-1_3


The epigenetic regulation of centromeres and telomeres in plants and animals


