

RESEARCH ARTICLE

Open Access

Organization and evolution of a novel cervid satellite DNA with yeast CDEI-like repeats

Lei-Jiau Hsieh^{1†}, Ya-Ming Cheng^{2†}, Yi-Chun Wang^{3,4}, Chyi-Chyang Lin^{5,6} and Yueh-Chun Li^{1,3,4*}

Abstract

Background: It has been proposed that pericentromeric satellite DNA arises from the progressive proximal expansion of ancient centromeric DNA. In an attempt to recover putative ancestral centromeric DNA, we microdissected the pericentromeric/centromeric DNA from the chromosome X + 3 of Indian muntjac (*Muntiacus muntjak vaginalis*) and constructed a microclone-library of the X + 3 centromeric DNA.

Results: A new cervid satellite DNA element, designated as satellite VI, was isolated from this library. Fluorescence in situ hybridization (FISH) studies revealed that satellite VI is predominately located on the distal pericentromeric region of the Indian muntjac chromosome X + 3 and on the pericentromeres of several Old World deer species studied. Its sequence is organized as 11-bp monomeric (ATCACGTGGGA) tandem repeats. Further sequencing on a BAC clone of Indian muntjac harboring this repeat showed that an array of this repeat stretches over approximately 5 kb followed by approximately 3 kb of interspersed repetitive sequences, such as long interspersed elements (LINEs), short interspersed elements (SINEs), and long terminal repeats (LTRs).

Conclusions: Based on the chromosomal localization, genomic and sequence organization, and copy numbers of satellite VI in deer species studied, we postulate that this newly found satellite DNA could be a putative ancient cervidic centromeric DNA that is still preserved in some Old World deer. Interestingly, the first eight nucleotides of the 11-bp monomeric consensus sequences are highly conserved and identical to the CDEI element in the centromere of the budding yeast *Saccharomyces cerevisiae*. The centromeric/pericentromeric satellite DNA harboring abundant copies of CDEI sequences is the first found in a mammalian species. Several zipper-like d (GGGA)₂ motifs were also found in the (ATCACGTGGGA)_n repeat of satellite VI DNA. Whether the satellite VI is structurally and functionally correlated with the CDEI of centromere of budding yeast and whether a zipper-like structure forms in satellite VI require further studies.

Keywords: Indian muntjac; Centromeric satellite DNA; Ancient centromeric DNA; Concerted evolution; CDEI element; Chromosome microdissection

Background

A pericentromeric satellite DNA has been proposed to have originated from the progressive proximal expansion of ancient centromeric DNA (Schueler and Sullivan 2006). This hypothesis was supported by the following lines of evidence: (1) The sequence organization of centromeric domains on the primate X chromosomes is physically symmetrical (Schueler et al. 2005), (2) organization of

monomeric units in the pericentromeric regions is different from that of higher order repeat units in the centromeric region (Willard 1985; Alexandrov et al. 1993; Puechberty et al. 1999; Horvath et al. 2000; Schueler et al. 2001; Rudd and Willard 2004; Schueler et al. 2005; Rudd et al. 2006), (3) the genomes of lower primates show existing monomeric alpha satellites and no higher-order alpha satellites (Goldberg et al. 1996; Alexandrov et al. 2001), (4) human pericentromeric α -satellite monomers are frequently interrupted by LINE (long interspersed element), SINE (short interspersed element), and LTR (long terminal repeat) retrotransposons (Schueler et al. 2005). In addition, the older L1P elements (primate-specific LINE interspersed repeats) exist in the more distal pericentromeric α -satellite

* Correspondence: ycl@csmu.edu.tw

†Equal contributors

¹Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan

³Department of Biomedical Sciences, Chung Shan Medical University, Taichung, Taiwan

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

monomers (Smit et al. 1995; Schueler et al. 2001). (5) The centromeric higher order units of orthologous chromosomes from different primate species are more divergent than the pericentric monomer units (Rudd et al. 2006). These findings suggested that the distal pericentromere is the region that may contain 'paleontological record' of ancient satellite arrays and could represent the functional centromeric regions in ancestral primates (Bayes and Malik 2008). In other words, those studies hint to the possibility that even more ancient centromeric DNA might be preserved in the more distal pericentromeric region.

What is the ancient centromeric DNA of mammals? The question remains unanswered in most cases because of considerably diverse centromeric DNA among species from lower to higher organisms and uncompleted sequencing of the centromeric/pericentromeric regions. Additionally, the centromeric region of chromosomes in most mammals is usually too small to dissect the chromosome localization of centromeric and pericentromeric DNA by fluorescence in situ hybridization (FISH) study. The centromeric/pericentromeric region of the Indian muntjac (*Muntiacus muntjak vaginalis*; barking deer) chromosome X + 3 is known to be exceptionally huge and compound in nature (Brinkley et al. 1984) and may be composed of many different satellite DNA families (including functional centromeric and non-functional centromeric satellite DNAs as well as new and old satellite DNAs). Indeed, cervid satellite I (an old cervid satellite DNA) (Lee et al. 1994; Lin et al. 1991), cervid satellite II (a functional centromeric satellite DNA) (Vafa et al. 1999; Li et al. 2000b), and cervid satellite IV (Li et al. 2002) had been found in the X + 3 centromere region of Indian muntjac.

In this study, we made use of Indian muntjac with the very large centromeric/pericentromeric region of chromosome X + 3 for our investigation into the possible existence of ancient centromeric DNAs. We microdissected the centromeric/pericentromeric region of the Indian muntjac chromosome X + 3 and microcloned a novel cervid satellite DNA family from the dissected pericentromeric region. The newly isolated satellite DNA (designated as cervid satellite VI) is organized as 11-bp monomer tandem repeats. Most monomers harbor the same eight nucleotides, ATCACGTG, which is identical to the centromeric element CDEI of the budding yeast *Saccharomyces cerevisiae*. The monomeric, LINE-, and SINE-interrupted organization and distal pericentromeric location of satellite VI in Indian muntjac imply that this novel satellite DNA could be a vestige of an ancient cervid centromeric DNA element.

Methods

Cell lines, chromosome preparations, and DNA isolation

Metaphase chromosome spreads and genomic DNAs were prepared from the fibroblast cell lines of the following species: Indian muntjac (*Muntiacus muntjak vaginalis*) (male

cell line, CCL-157, American Type Culture Collection; female cell line, kindly supplied by Dr. Andrew P. Feinberg, School of Medicine, Johns Hopkins University, Baltimore, MD, USA); Formosan muntjac (*M. reevesi micrurus*) (Chiang et al. 2004); caribou (*Rangifer tarandus caribou*) (Lin et al. 1991; Lee et al. 1994; Lin et al. 2004); male black-tailed deer (*Odocoileus hemionus hemionus*) (CRL-6193, American Type Culture Collection); female Chinese water deer (*Hydropotes inermis*) (kindly provided by Dr. F. Yang, Centre for Veterinary Science, University of Cambridge, Cambridge, UK); Roe deer (*Capreolus capreolus capreolus*) cell line (from San Diego Zoo, San Diego, CA, USA) and male Formosan sambar deer (*Cervus unicolor swinhoei*) (Li et al. 2005). Metaphase chromosome spreads and genomic DNAs of Formosan sika (*Cervus nippon taioanus*) and Asian red deer (*Cervus elaphus*) were prepared from blood samples provided by a certified deer farm in Taiwan. The fixed cells and genomic DNAs of other mammals (goat, bull, boar, man, and rat) are available in our laboratory. The deer species selected above including Old World deer and New World deer (Hernandez Fernandez and Vrba 2005; Pitra et al. 2004). The detailed protocols for chromosome preparation and DNA isolation have been described elsewhere (Li et al. 2000a).

Microdissection and microcloning of the chromosome X + 3 centromeric DNA

Briefly, the centromeric/pericentromeric region of female Indian muntjac chromosome X + 3 was scraped from metaphase spreads under an inverted microscope (Olympus X-81, Tokyo, Japan) with a siliconized glass needle attached to a mechanical micromanipulator (Narishige). The microdissected centromeric elements were collected in 20 μ l of ddH₂O for the subsequent degenerate oligonucleotide-primed polymerase chain reaction (DOP-PCR) experiment. The DNA of 20 microdissected centromeric elements were amplified in 25 μ l of DOP-PCR reaction volumes with 2 μ mol/l of DOP primer (5'-CCGACTCGAGNNNNNNATGTGG-3') and 12.5 μ l of 2 \times DOP-PCR master mix (400 μ mol/l of each deoxynucleotide triphosphate, 20 mmol/l Tris-HCl, 100 mmol/l KCl, 3 mmol/l MgCl₂, and 5U Taq polymerase) (DOP-PCR Master kit purchased from Roche, Basel, Switzerland). The amplified PCR products were ligated into pSMART[®] GC HK (Lucigen, Middleton, WI, USA). The recombinant plasmids (designated as microclones) were used to transform XL1-Blue *Escherichia coli* competent cells. The transformation mixture was plated onto Luria-Bertani (LB) agar plates containing 100 μ g/ml of kanamycin, 40 μ g/ml of X-gal, and 0.05 mmol/l IPTG to construct a X + 3 centromere mini-library (designated as the pIMCentX + 3). The detailed protocol used for

chromosome microdissection and DOP-PCR was referred to Li et al. (2005).

Isolation of a novel cervid centromeric DNA clone

Microclone DNAs from the X+3 centromere mini-library pIMCentX+3 were screened by the colony hybridization method. Briefly, 576 microclones were orderly duplicated onto 6 new agar plates and then incubated until each colony had a size of 2 mm in diameter. All duplicated microclones were lifted onto nylon membranes and denatured in an alkali solution. A probe mixture containing ³²P-labeled satellite I (C5; (Lin et al. 1991)), satellite II (Mmv-0.7; (Li et al. 2000b)), satellite IV (Mmv-1.0; (Li et al. 2002)), and satellite V (Mmv-0.32#1; (Li et al. 2005)) was used to hybridize the membranes for the first screening. A ³²P-labeled DOP-PCR amplified X+3-microdissected DNA probe was used to hybridize the membranes for the second screening. The conditions for filter hybridization and washing were similar to those mentioned in the Southern blot analysis section (see below). The microclones with strong signals for the X+3-microdissected DNA probe and negative signals for satellite DNA I, II, IV, and V probes were picked up for further characterization.

DNA sequencing of microclones

The inserts of microclones were sequenced using the ABI BigDye Terminator Cycle Sequencing kit (PE Applied Biosystems) and the ABI 3730 Genetic Analyzer (Applied Biosystems, Tokyo, Japan). Twelve microclones shared the highly similar sequences (80% to 100%); therefore, sequences of three representative microclones (pIMCentX+3-1C5, pIMCentX+3-1C6, and pIMCentX+3-1G1) were deposited into the NCBI GenBank database (accession numbers JN798609, JN798610, JN798611). Due to sequence novelty, the insert of microclones was designated as a cervid satellite VI element.

Isolation and full sequencing of IM-BAC DNA containing satellite VI DNA element

In order to isolate a larger DNA element that may contain numerous monomers of satellite VI DNA, ³²P-labeled pIMCentX+3-1C5 (designate as satVI-1C5) was used to probe four BAC clones that had been known to be localized on Xp11.1 in our previous data (Lin et al. 2008). A positive BAC clone (1249A1) was fully sequenced by hierarchical sequencing. Briefly, the 1249A1 was digested with *EcoRI* and the digested five fragments were subcloned into a pBluescript II SK (-) plasmid vector. Five subclones (designated E1, E3, E4, E5, and E34) representing five different digested fragments were obtained. The subclones (E1, E3, E5, and E34) with an insert <2.0 kb and BAC-1249A1 were directly sequenced from both M13-F and M13-R of the vector using the

ABI BigDye v3.1 Sequencing Kit (PE Biosystems, Chiba, Japan) and then read by an ABI 3730 Genetic Analyzer (Applied Biosystems, Tokyo, Japan). The subclone E4 with an insert >2.0 kb was sequenced through the *ExoIII*-nested deletion clones. Briefly, the M13-F side of the insert of the subclone was cleaved with *HindIII* and *KpnI* at 37°C for 1 h. A linearized DNA with a 5'-protruding and 3'-protruding end was digested from the 5'-end by *ExonucleaseIII* at 37°C (theoretically, the digestion rate is about 300 bp/min). The nested deleted DNAs at three given digestion time intervals (2 min/4 min/8 min) were subcloned into pBluescript II SK (-) plasmid vectors. The exodeleted subclones with the appropriate insert size were picked up for further DNA sequencing from the M13-F end of the vector. The detailed protocol was based on the Erase-a-Base System (Promega) protocol. We use the BLAST2 program to assemble each nested deleted E4 subclones with over 99% identity. The order and direction of E1, E3, E5, and E4 was determined by primer walking and BAC-end sequencing using the BAC-1249A1 as the template. The order of E34 can be determined while its direction cannot be decided. We assumed the sequence direction of E34 as same as that of E4 in the BAC-1249A1 clone. Sequences of the BAC-1249A1 clone were deposited into the NCBI GenBank database as IM04-1249A1 with accession numbers JN798612.

DNA sequences analysis

We compared the similarity of sequences with all nucleotide sequences in the NCBI GenBank database using the nucleotide-nucleotide BLAST program. The suspected repetitive sequences were analyzed by comparing with a database of repetitive elements using RepeatMasker software (Smit et al., unpublished work). The size, consensus sequence, and copy number of monomer of repetitive sequences were analyzed by Tandem repeat finder software (Benson 1999). The conserved frequency of each nucleotide of monomeric consensus sequence was determined by WebLogo (Crooks et al. 2004).

Southern blot, zoo blot, and dot blot analyses

Southern blot and zoo blot analyses were conducted under similar conditions. In brief, a 10-μg aliquot of genomic DNA from each species was digested with *EcoRI* for zoo blot analyses. For Southern blot analyses, each 10-μg aliquot of genomic DNA of Indian muntjac was digested with one of five different restriction endonucleases (*BamHI*, *EcoRI*, *HaeIII*, *HindIII*, and *XhoI*). The digests were electrophoretically fractionated on 1% agarose gel, transferred to a nylon membrane (Biodyne), and hybridized with ³²P-labeled satVI-1C5 DNA. The conditions used for hybridization, filter washing, and autoradiography have been described previously (Li et al. 2000a).

For the copy number estimation of repeated DNA elements in the genome, we dot blot 1 μ l of a series 0.5X dilution of the genomic DNA and the satVI-1C5 plasmid DNA (start from 100 to 0.0487 ng) for each on the nylon membrane. We used the 32 P-labeled insert fragment of satVI-1C5 to probe the nylon membrane. The dot blot hybridization procedure has also been described elsewhere (Lee et al. 1994; Li et al. 2000a). A copy number was roughly estimated from the signal intensity of a control plasmid satVI-1C5 DNA (0.195 ng; approximately 0.054 ng of insert) that showed similar signal intensity as the genomic DNA (3.125 ng for male Indian muntjac and male Formosan muntjac each; 100 ng for male Sambar deer). Hence, the 0.78-kb insert of satVI-1C5 comprises some 1.72% of male Indian muntjac genome and male Formosan muntjac genome as well as 0.054% of male Sambar deer. This is only a rough estimation since signal intensity is not linear across the entire intensity curve. No dot blot hybridization signal can be detected in other deer species (male caribou, male white tailed deer, and female Chinese water deer) each with 100 ng of genomic DNA. Assuming that the haploid genomic size of Indian muntjac is approximately 2.2×10^9 bp (Li et al. 2000a; Li et al. 2000b), 3.44×10^6 copies of the 11-bp repetitive unit are estimated for a male Indian muntjac haploid genome.

Fluorescence *in situ* hybridization and reverse FISH

Metaphase chromosome spreads were made and aging at 65°C for 3 h and at room temperature for 3 days. For reverse FISH experiment, the microdissected DNA was DOP-PCR amplified and then labeled with digoxigenin-dUTP as a probe. For FISH study, the satVI-1C5 was labeled with either biotin-dUTP or digoxigenin-dUTP (Roche, Basel, Switzerland). Cervid satellite I DNA (C5) (Lin et al. 1991) was labeled with either biotin-dUTP or digoxigenin-dUTP. Cervid satellite II DNA (Mmv-0.7) (Li et al. 2000b) was directly labeled with Cy3-dUTP

(presented as a pseudo-orange color). IM04-526B9, IM04-50A1, and IM04-121A1 BAC clones (Lin et al. 2008) were labeled with digoxigenin-dUTP for identifying the specific chromosome of the respective deer species based on the comparative map of Indian muntjac-Chinese muntjac (Formosan muntjac) (Fronicke et al. 1997; Li et al. 2000b; Murmann et al. 2008) and Indian muntjac-Formosan sambar-Formosan sika (unpublished data). FISH probes or reverse FISH probes were hybridized to the aged metaphase chromosome spreads. For a single- or dual-color FISH experiment, biotin-labeled probes were detected with AF568-conjugated avidin (red) and digoxigenin-labeled probes were detected with FITC-conjugated antibodies (green). For a triple-color FISH experiment, biotin-labeled probes were detected with Cy5-conjugated avidin (presented as a pseudo-green color), digoxigenin-labeled probes were detected with FITC-conjugated antibodies (presented as a pseudo-red color) and the directly labeled Cy3 probes did not need detection *via* antibodies (presented as a pseudo-orange color). The procedures for denaturation, hybridization, post hybridization washing, and signals detection were described in detail previously (Li et al. 2000a; Lin and Li 2006). Fluorescence signals were captured by a Leica ALM fluorescence microscope (Wetzlar, Germany) equipped with appropriate filter sets and a cooled charge-coupled device (CCD) camera. The images were normalized and enhanced using the FISH software (Applied Spectral Imaging, Israel), and processed in Photoshop (Adobe, San Jose, CA, USA).

Results

Molecular cloning of a novel satellite DNA from the pericentromere of Indian muntjac chromosome X + 3

A X + 3 centromere mini-library pIMCentX + 3 was constructed from DOP-PCR products, which had been amplified from 20 microdissected pericentromeric/centromeric regions of the Indian muntjac chromosome X + 3 (Figure 1). The reverse FISH study showed that the positive FITC

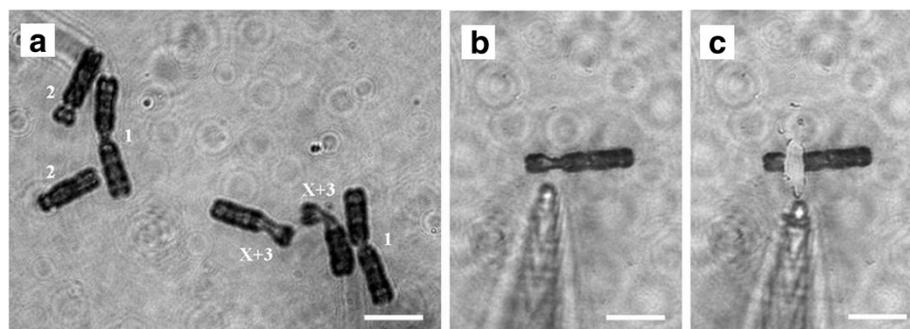


Figure 1 Microdissection of the pericentromeric/centromeric region of the Indian muntjac chromosome X + 3. (a) The metaphase spread of a female Indian muntjac. Chromosomes 1, 2, and X + 3 are indicated. (b) The centromeric/pericentromeric region of an Indian muntjac chromosome X + 3 is shown before microdissection. (c) The same chromosome as in panel (b) is shown after the centromeric/pericentromeric region is scraped by microdissection. Scale bar = 10 μ m.

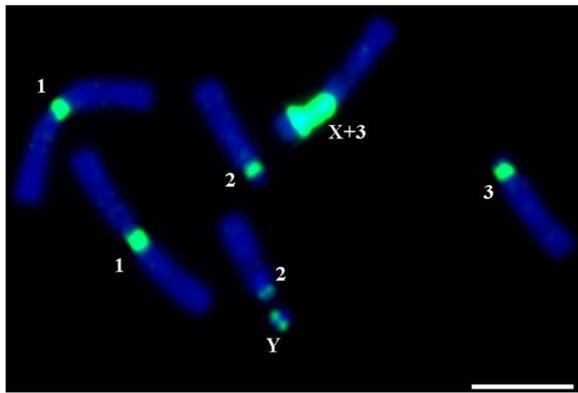


Figure 2 Reverse FISH analysis of microdissected centromeric DNA. The DOP-PCR products from the microdissected DNA of peri-centromeric/centromeric region of chromosome X + 3 were labeled with DIG-dUTP, hybridized to male Indian muntjac chromosome spreads and detected with FITC-conjugated anti-DIG antibody (green). The centromeric heterochromatin regions of all chromosomes show positive FITC hybridization signals. Chromosomes were counterstained with DAPI (blue). Scale bar = 10 μ m.

signals of digoxigenin-labeled DOP-PCR products were localized to all centromeres of male Indian muntjac chromosomes (Figure 2). This indicated that the DOP-PCR products contain a mixture of centromeric DNA satellite DNA families. A total of 576 microclones of pIMCentX + 3 were screened using the amplified X + 3-microdissected DNA probe and four different cervid satellite DNA (satellites I, II, IV, and V) probes. We detected 30 microclones with strong signals for the microdissected X + 3 DNA probe and negative signals for satellite DNA I, II, IV, and V probes. One of these microclones (designated as pIMCentX + 3-1C5 or satVI-1C5) was randomly picked up as a probe to screen the abovementioned 30 microclones. There were 14 microclones (including satVI-1C5) with positive satVI-1C5 signals and 12 out of 14 microclones for further sequencing (two microclones failed to grow). The size of insert of each of 12 microclones was from 708 to 168 bp. Two out of 12 microclones having the same size of insert and sharing completely identical sequence should be isolated from the same fragment of the Indian muntjac chromosome X + 3. The sequences of 11 microclones with the various size of insert were

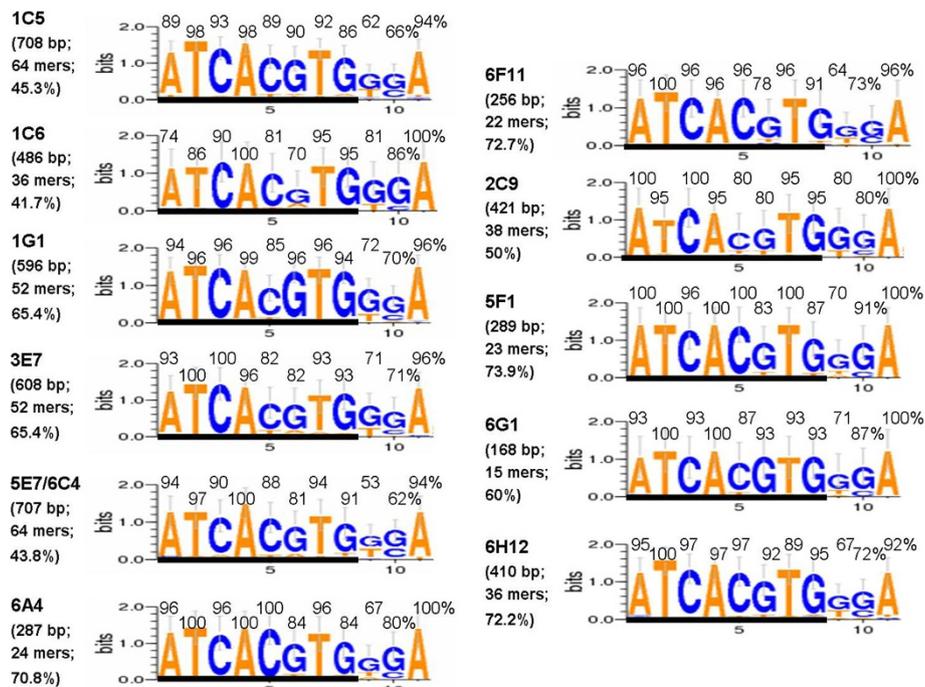


Figure 3 The monomeric consensus sequence of 11 different microclones. The sequence of 11 different microclones is organized in tandem repeats of 11-bp monomer units. The conserved frequency of each nucleotide of consensus sequence was indicated by the size of base and digits above each base. The first eight nucleotides of consensus sequence (underline) are identical with the functional CDEI element of the budding yeast centromere. The assigned number of each microclone was indicated in the left-hand side of the monomeric consensus sequence. 5E7 and 6C4 were the microclones with the identical nucleotide sequences and same size. The ratio of monomers, which contain the completely identical ATCACGTG sequence, was indicated in the bracket under the assigned number of a microclone. The size of insert and the total numbers of monomers for each microclone were also indicated in this bracket.

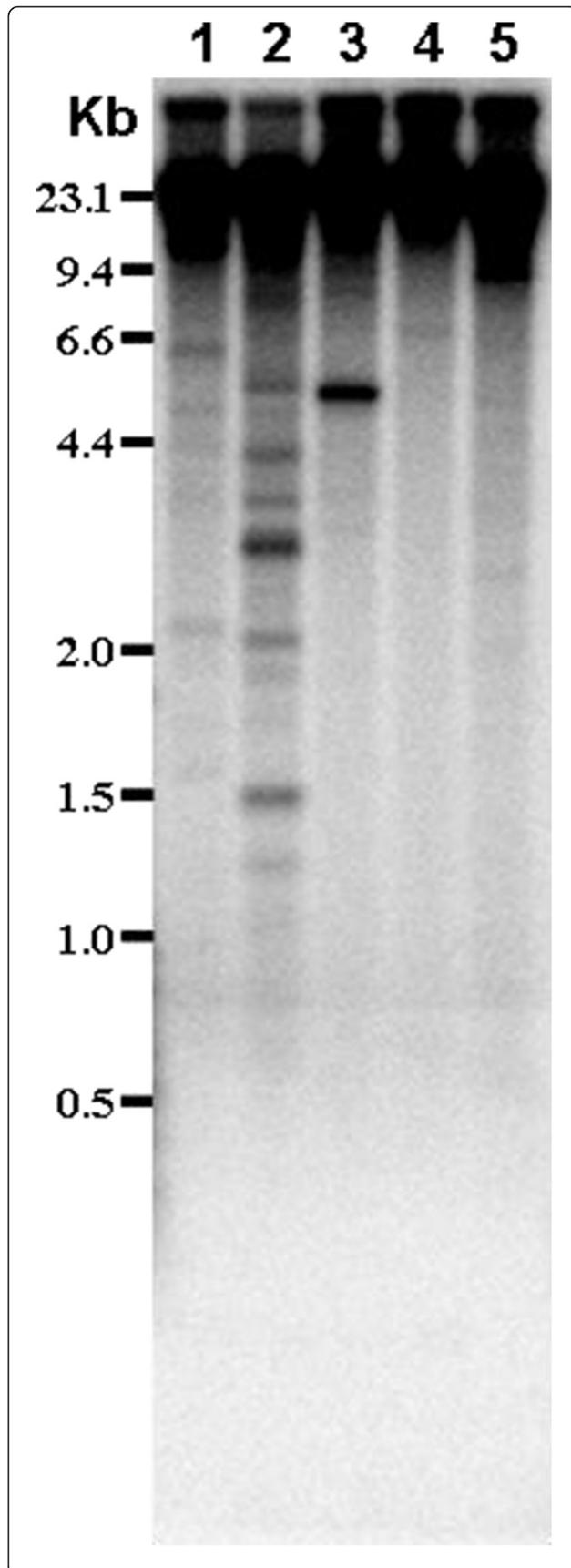


Figure 4 Southern blot hybridization. Genomic DNA of a female Indian muntjac was digested with *Bam*HI (lane 1), *Eco*RI (lane 2), *Hae*III (lane 3), *Hind*III (lane 4), and *Xho*I (lane 5) and probed with ³²P-labeled satVI-1C5. Multiple hybridization bands with irregular patterns are observed in each digest. More apparent hybridization bands are seen in the *Eco*RI digest than in other digests. An approximately 5.0-kb strong hybridization band is also observed in the *Hae*III digest. Size markers are shown on the left side of the blot.

aligned and shared 80% to 93% sequence identity with each other. Therefore, these 11 microclones should be isolated from the different fragments of the Indian muntjac chromosome X + 3.

Characterization of a new satellite VI DNA in Indian muntjac

DNA sequences of the 11 different microclones were compared with all nucleotide sequences in the NCBI GenBank database and no similarities to any of the sequences deposited in the database were found. The tandem repeat analysis showed that the insert of each of the 11 microclones had a monomer size of 11 bp and shared the same monomeric consensus sequence, ATCACGTGGGA (Figure 3). These analyses indicated that the insert fragments of the 11 microclones are likely derived from the same, newly discovered cervid satellite DNA designated as cervid satellite VI. Moreover, the first 8 nucleotides of the consensus monomers from the 11 different microclones had a considerably conserved frequency compared to the 9th and 10th nucleotide. There were 41.7% to 73.9% of monomers from the 11 different microclones containing the completely identical ATCACGTG sequence. Southern blot hybridization with the satVI-1C5 probe to female Indian muntjac genomic DNA produced multiple hybridization bands with irregular patterns in the *Eco*RI digest (Figure 4). Multiple faint bands were observed in the other digests with exception of the *Hae*III digest, in which an approximately 5.0-kb strong hybridization band was seen. In each digest, a strong hybridization band at 23 kb was also observed. This result suggests that this novel satellite DNA family is a highly conserved tandemly monomeric repeat and is not organized in a higher order repeat fashion in the genome of Indian muntjac. Copy number estimation revealed that there are approximately 3.44×10^6 copies (1.72%) of the 11-bp repetitive unit in the genome of male Indian muntjac. A FISH study was performed with satVI-1C5 as a probe to investigate the chromosomal location of this satellite VI DNA. The satellite VI signal was localized at the pericentromeric region Xq11 of the chromosome X + 3 and was more distally located than the satellite I signal was, whereas the satellite II signal was located in the center of the centromeric region (Figure 5d,e). Weaker satellite VI signals were observed at the interstitial sites (1p32, 1p17, 1p14, 1q21,

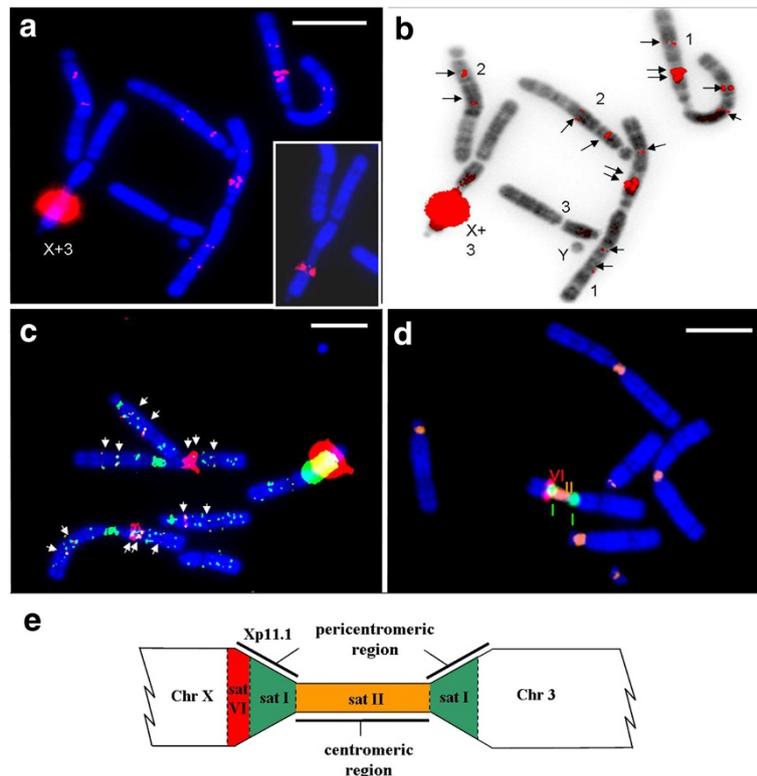


Figure 5 FISH analysis of satellite VI in the metaphase spreads of a male Indian muntjac. (a) A strong satellite VI DNA signal (red) was detected at the pericentric X region of chromosome X + 3 and 14 interstitial hybridization signals were observed after longer exposure. Inset: a clear satellite VI signal was observed at the Xp11.1 side of the X + 3 pericentromere after short exposure. (b) The G-banded like chromosomes are presented from the inverted DAPI image of panel (a). Red interstitial satellite VI signals were indicated by arrows. (c) All interstitial satellite VI fluorescence signals (red; indicated by arrows) are co-localized with satellite I (C5 as a probe; green). (d) Triple-color FISH study with biotinylated satellite I (presented as a pseudo-green color), Cy3 direct-labeled satellite II (Mmv-0.7 as a probe; presented as a pseudo-orange color) and digoxigenin-labeled satellite VI (presented as a pseudo-red color) cohybridized to the metaphase chromosomes. The satellite II signal is located in the centromeric region of the chromosome X + 3. The satellite I signal is located in both pericentromere regions of the X + 3. The satellite VI signal is located in the Xp11.1 of the X + 3 pericentromere region and is more distal to the centromere than the satellite I signal. The FISH image was captured under low exposure time (scale bar = 10 μ m). (e) Schematic illustration of the chromosome X + 3 in the panel (d) which shows the organization of cervid satellite I, II, and VI in the centromeric and pericentromeric regions of chromosome X + 3. The exact boundary of each satellite DNA could not be exactly defined and was indicated by dotted lines.

1q27, 2q14, and 2q27) of the Indian muntjac chromosomes (Figure 5a,b). Every interstitial signal of satellite VI DNA was co-localized with an interstitial signal of satellite I DNA (Figure 5c). No signals were detected in the centromeric regions of other Indian muntjac chromosomes.

Genomic organization of a large satellite VI array in Indian muntjac

We further examined an Indian muntjac BAC clone (IM04-1249A1) that showed positive satellite VI signals to better understand the genomic organization of this

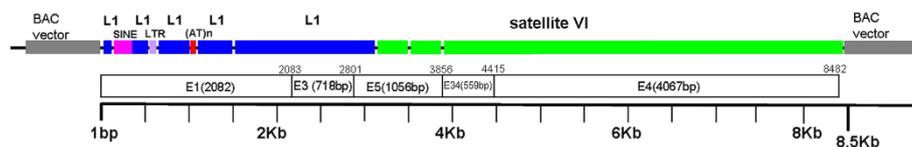


Figure 6 The genomic organization of the Indian muntjac BAC clone IM04-1249A1. DNA sequences from five cloned *Eco*RI fragments of IM04-1249A1 were assembled and shown as colored bars. The positions of five fragments (E1, E3, E4, E5, and E34) are shown as open boxes under the equivalent sequence domain: Cervid satellite VI (green bars), LINES (L1, blue bars), SINEs (BovA; pink bar), LTR elements (ERVK; purple bar), (AT) n simple repeat (red bar).

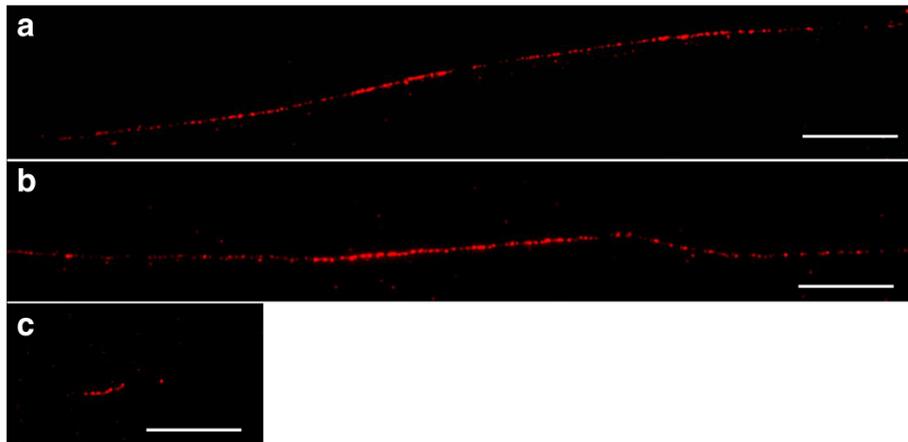


Figure 7 The array length of satellite VI in genomes of three species by genomic fiber-FISH experiment. The relatively contiguous 'beads-on-a-string' hybridization signals (red) of satellite VI array were organized in a highly extended chromatin fiber of (a) Indian muntjac (approximately 85 μm), (b) Formosan muntjac (approximately 96 μm), and (c) Sambar deer (approximately 5.6 μm). The biotinylated satVI-1C5 probe was hybridized to the highly stretched chromatin fiber of the tested deer species and the FISH signal was detected by AF568 conjugated avidin (red fluorescence). The picture of each fiber FISH is composed of several overlapping images to show the total length of a satellite VI array. Scale bar = 10 μm .

novel satellite DNA array in Indian muntjac. The 8,482-bp-long insert of IM04-1249A1 was fully sequenced. The sequences revealed a 5,463-bp satellite VI array that adjoins a 3,019-bp interspersed repeats including 5 LINEs (L1 family), 1 SINE (BovA), 1 LTR (ERVK), and a stretch of 79-bp AT repeats (Figure 6). The L1 (1,229 to 2,998 nt) showed 78% identity with 119 to 1,958 nt of L1-1_Ttr from Dolphin (*Tursiops truncatus*). The satellite VI array in IM04-1249A1 BAC clone is organized as 496 monomers, 48.4% of which contains the identical ATCACGTG sequence. All restriction enzyme digests of the part of satellite VI array were analyzed in silico, they did not show a typical type-A pattern that represented a higher order repeat organization of satellite DNA. A genomic fiber FISH study showed that a relatively contiguous satellite VI signal arrays on the highly extended chromatin fiber in approximately 85- μm length (Figure 7a). Based on the previous studies (Haaf and Ward 1994; Li et al. 2000a), the resolution of a 'beads-on-a-string'-type chromatin fiber was about 10 kb/ μm ; therefore, the tandemly repeated units of the satellite VI array correspond to approximately 850 kb of DNA.

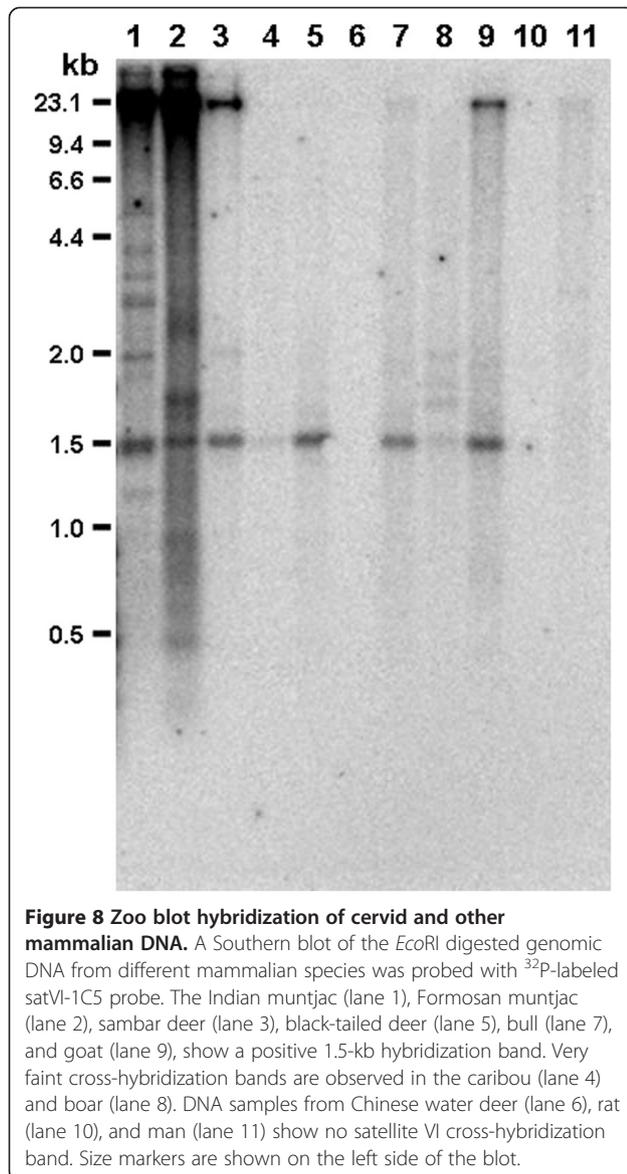
Distribution of the satellite VI element in the genome of other mammalian species

The zoo blot analysis (Figure 8) showed that satellite VI predominantly exists in the genomes of male Indian muntjac and male Formosan muntjac. The genome of male Formosan sambar deer and male goat both had 1.5- and 23.1-kb satellite VI hybridization bands. A 1.5-kb satellite VI hybridization band was also seen in the

genome of male black-tailed deer and bull. A very faint 1.5-kb satellite VI hybridization band was seen in male caribou and boar. No hybridization band was observed in female Chinese water deer, male rat, or human. This observation indicates that satellite VI DNA is not species-specific. Moreover, the hybridization patterns revealed a 1.5-kb *Eco*RI fragment in all tested Pecoran (Artiodactyla) species with the exception of Chinese water deer. Copy number analysis estimated that the satellite VI DNA element comprises 1.72% of the genome of male Formosan muntjac and 0.054% of the genome of male Sambar deer. The satellite VI DNA element makes up less than 0.001% of the genomes of other deer species (male caribou, male black-tailed deer, and female Chinese water deer) (data not shown).

Chromosomal localization of satellite VI elements in other mammalian species

We further performed FISH experiments to identify the chromosomal location of the satellite VI element in other deer species and some other mammalian species. Significant FISH signals were observed in Formosan muntjac, Formosan Sambar deer, Formosan sika deer, and Asian red deer. In the Formosan muntjac, FISH signals of satellite VI were detected on pericentromeric/centromeric regions of 35 autosomes and the X chromosome. No pericentromeric/centromeric FISH signals were detected in an autosome 4 or in autosomal pairs of 2, 3, 18, 19, or the Y chromosome. One interstitial FISH signal of satellite VI was found on chromosomes 3 of the species (Figure 9a). In the Formosan Sambar deer,



satellite VI FISH signals were observed on the pericentromeric/centromeric regions of 32 autosomes and the X chromosome (Figure 9b). The satellite VI FISH signals were detected on the pericentromeric/centromeric regions of 28 autosomes of Asian red deer (Figure 9c) and on 38 autosomes and the X chromosome of Formosan sika deer (Figure 9d). The FISH signals of satellite VI could not be detected on any of the metacentric chromosomes or Y chromosome of Formosan Sambar deer, Asian red deer, or Formosan sika deer. Interestingly, we observed that only one homologous autosome 4 had a positive FISH signal of satellite VI in Formosan muntjac (Figure 9a) and that the strongest FISH signal was present on one of the autosomes in Formosan Sambar deer as well as in Formosan sika deer (Figure 9b,d). No

significant fluorescence signal was detected in black-tailed deer, caribou, Chinese water deer, roe deer, bull, goat, or human (data not shown). These findings indicate that the quantity of the satellite VI DNA element has a species-specific difference in the pericentromeric/centromeric region of the abovementioned deer species.

Discussions

The cervid satellite VI DNA could be the vestige of an ancient centromeric DNA

We microdissected and microcloned a novel cervid satellite family designated as satellite VI from the pericentromeric/centromeric region of the chromosome X + 3 of Indian muntjac by chromosome microdissection and colony hybridization. A FISH study showed that the novel satellite VI was mainly present in the distal pericentromeric region of the Indian muntjac chromosome X + 3. It was more distal than cervid satellite I, which has been considered to be an old satellite DNA (Buntjer et al. 1998). This newly found satellite VI DNA is organized in monomeric repeats each comprising an 11-bp unit with the consensus sequence ATCACGTGGGA. The sequence of Indian muntjac BAC clone showed that an approximately 5 kb array of satellite VI DNA adjoined with approximately 3 kb of interspersed repetitive sequences, such as LINES, SINES, and LTRs. Based on the model of progressive proximal expansion, more ancient centromeric DNA might exist in more distal pericentromeric regions and is most likely organized in monomeric repeats interrupted by transposons (Schueler and Sullivan 2006; Schueler et al. 2005). Therefore, we suggest that the novel pericentromeric satellite VI is likely the vestige of an ancient centromeric DNA. Furthermore, we detected a 1.5-kb satellite VI hybridization band in the genomes of deer, goat, and bull but few faint hybridization bands in the genome of boar using zoo blot analysis. The observed differential intensity of hybridization signals agrees that deer (Cervidae) is phylogenetically more close to goat and bull (Bovidae) than boar (Suidae) (Price et al. 2005). Also, the result of zoo blot suggested that the satellite VI may already have existed in the ancestor of Artiodactyla family. Additionally, the flanking L1 elements of satellite VI arrays shared 78% identity with the L1-1_Ttr found in the genome of Dolphin (Jurka 2008). Unfortunately, the Dolphin DNA is not available in our laboratory; we cannot perform a zoo blot study to verify whether the satellite VI DNA is presented in Dolphin species (Cetartiodactyla family) as well.

The concerted evolution of cervid satellite VI in deer species

Based on the results of zoo blot, FISH, and copy number analysis, we observed that the considerably differential amount of satellite VI element presents in the genomes

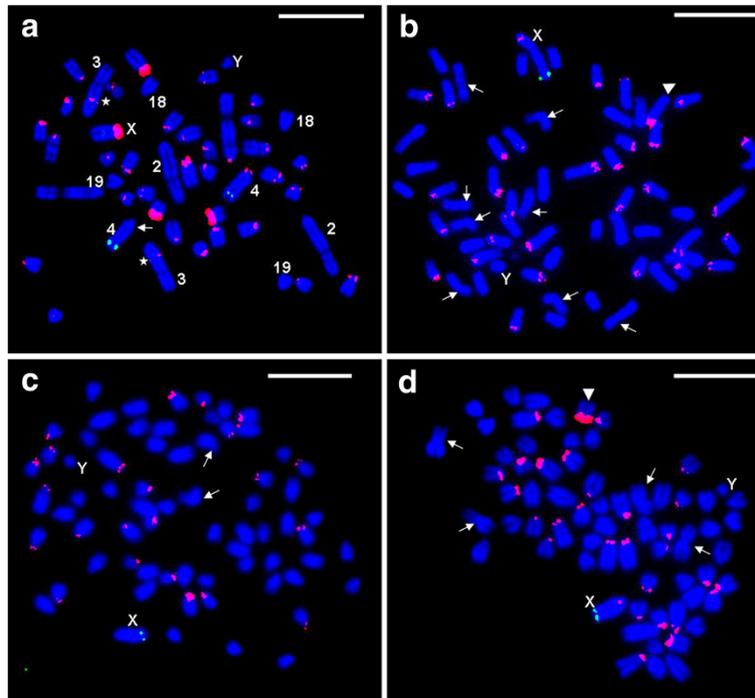


Figure 9 FISH analysis of satellite VI in Formosan muntjac, Sambar deer, Formosan sika deer, and Asian red deer. The biotinylated satVI-1C5 probe was hybridized to the metaphase chromosomes of the tested deer species and the FISH signal was detected by AF568 conjugated avidin (red fluorescence). The metaphase chromosomes were counterstained with DAPI (blue fluorescence). **(a)** The metaphase of Formosan muntjac shows the chromosomal location of satellite VI. Chromosomes 2, 3, 18, 19, and Y lack a centromeric FISH signal. Asterisks (*) indicate the interstitial FISH signal on each chromosome 3. An arrow indicates the one chromosome 4 that lacks a satellite VI FISH signal. Both chromosomes 4 were identified by a digoxigenin-labeled IM04-526B9 Indian muntjac BAC DNA (green signals) which maps to 1q25 of the Indian muntjac equivalent to chromosome 4 of the Chinese muntjac (Fronicke et al. 1997; Lin et al. 2008; Murmann et al. 2008). The metaphase spreads of the **(b)** Formosan Sambar deer, **(c)** Asian red deer, and **(d)** Formosan sika deer show the chromosome location of satellite VI. Arrows indicate metacentric autosomes of the respective species in panels (b,c,d). Arrowheads indicate the chromosomes with the strongest FISH signals of satellite VI in panels (b,d). The digoxigenin-labeled probe IM04-50A1 (green signal in the panels (b,c)) and IM04-121A1 (green signal in the panel (d)) was used for identifying the X chromosome of the respective species.

of Indian muntjac, Formosan muntjac, Formosan Sambar, Formosan sika, and Asian red deer (all belong to Old World deer) while it is less abundant or even eliminated in roe deer, black tailed deer, caribou, and Chinese water deer (New World deer) (Pitra et al. 2004; Hernandez Fernandez and Vrba 2005). The characterization of copy number variants of satellite VI in the related deer species agrees with the satellite DNA library model as the consequence of concerted evolution (Fry and Salsler 1977; Ugarkovic and Plohl 2002). The variable FISH signals of satellite VI were observed between some homologous chromosomes in Formosan muntjac, Formosan Sambar deer, and Formosan sika deer. Such different FISH signal strengths/intensities of satellite VI between the homologous chromosomes were most likely due to unequal crossing-over events. It has been suggested that a short monomer size, such as the 11-bp monomer size of satellite VI in this study, offers ample opportunities for unequal crossing-over during meiosis and results in copy number

variation between homologous chromosomes (Smith 1976; Ugarkovic and Plohl 2002).

The characteristic sequence of satellite VI DNA

All 11 microclones and the BAC clone IM04-1249A1 carried the consensus monomer ATCACGTGGGA in their repeat arrays. Interestingly, the first eight nucleotides of this consensus monomer are highly conserved and completely identical with the ATCACGTG of the budding yeast centromeric CDEI element. There were 41.7% to 73.9% of monomers from the 11 different microclones and the studied BAC clone containing the completely identical ATCACGTG sequence. Although we found the highly conserved ATCACGTG sequence existing abundantly in satellite VI element, it is not possible to demonstrate the structural and functional relation of satellite VI and CDEI element of budding yeast centromere at present. It is also noted that an interesting central 5'-GGGA-3' tetranucleotide was presented when

the 11-bp consensus monomeric sequences was written as ACGTGGGAATC. It has been reported that the central 5'-GGGA-3' tetranucleotide played a key role in stabilizing the fold-back structures of *Drosophila* pericentromeric dodeca satellite DNA (Ferrer et al. 1995) and a similar 5'-GGA-3' trinucleotide adopted a fold-back structure of human centromeric satellite III (TGGA)_n repeat (Chou et al. 1994). Furthermore, the (GPuA)₂ tract and the (GPuPuA)₂ tract adopted a similar zipper-like structure of human satellite III DNA and *Drosophila* dodeca satellite DNA, respectively (Chou et al. 1994; Ferrer et al. 1995; Chou and Chin 2001). Chou and Chin (2001) proposed that the zipper-like interdigitated motifs of centromeric satellite DNA may serve as common cores in organizing the eukaryotic centromere structure. In this study, the sequences of satellite VI sharing the (GPuPuA)₂ tract in (ACGTGGGAATC)_n repeat imply a possible zipper-like structure forming in the pericentromeric region. However, it still requires more studies to verify the possible zipper-like structure of satellite VI.

Conclusions

In conclusion, we have isolated a novel cervid satellite VI DNA from the genome of Indian muntjac. Because the distal pericentromeric localization and monomeric and LINE-, SINE-interrupted organization of satellite VI DNA in Indian muntjac, we postulate that this pericentromeric satellite VI DNA could be a vestige of an ancient cervid centromeric DNA. The species-specific copy number profile of satellite VI varied dramatically among deer species studies. Such as Old World deer species had abundant satellite DNA VI while it was less or even loss in New World deer species. This could be the result of concerted evolution of satellite VI DNA (Ugarkovic and Pohl 2002). We further speculated that this satellite VI DNA may already be preserved in the ancestor of the Artiodactyla family because it was found in Cervidae, Bovidae and Suidae. Interestingly, the high percentage of monomers of satellite VI has a highly conserved sequence identical to the CDEI sequence of the budding yeast centromere. Moreover, several characteristic structural feature (GPuPuA)₂ tracts, that adopted a similar zipper-like structure in organizing a common core of a eukaryotic centromere (Chou and Chin 2001), were found in satellite VI DNA element. However, whether the satellite VI plays a role in the centromeric structure and function requires further studies.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

The work presented here was carried out in collaboration among all authors. LJH carried out chromosome microdissection, FISH experiments, microcloning, nested deletion subcloning and sequence alignment. YMC established the microclone library and carried out the Southern and dot blot

experiment. YCW carried out the sequence analysis. CCL made comments and suggestions in the experimental design and drafting of the manuscript. YCL designed the study, executed the study and drafted the manuscript. All authors made comments on the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This research was supported by grants from National Sciences Council, Taiwan (NSC-97-2311-B-040-003-MY3 to YCL and NSC-99-2314-B-039-003-MY2 to CCL). We acknowledge the Instrument Center of Chung Shan Medical University supported by National Science Council, Ministry of Education and Chung Shan Medical University for the DNA sequencing services.

Author details

¹Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan. ²Department of Agronomy, National Chung Hsing University, Taichung, Taiwan. ³Department of Biomedical Sciences, Chung Shan Medical University, Taichung, Taiwan. ⁴Department of Medical Research, Chung Shan Medical University Hospital, Taichung, Taiwan. ⁵Graduate Institute of Basic Medical Science, China Medical University, Taichung, Taiwan. ⁶Department of Medical Research, China Medical University Hospital, Taichung, Taiwan.

Received: 18 January 2014 Accepted: 21 April 2014

Published online: 04 June 2014

References

- Alexandrov IA, Medvedev LI, Mashkova TD, Kisselev LL, Romanova LY, Yurov YB (1993) Definition of a new alpha satellite suprachromosomal family characterized by monomeric organization. *Nucleic Acids Res* 21(9):2209–2215
- Alexandrov I, Kazakov A, Tumeneva I, Shepelev V, Yurov Y (2001) Alpha-satellite DNA of primates: old and new families. *Chromosoma* 110(4):253–266
- Bayes JJ, Malik HS (2008) The evolution of centromeric DNA sequences. *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd, Chichester
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27(2):573–580
- Brinkley BR, Valdivia MM, Tousson A, Brenner SL (1984) Compound kinetochores of the Indian muntjac: evolution by linear fusion of unit kinetochores. *Chromosoma* 91(1):1–11
- Buntjer JB, Nijman IJ, Zijlstra C, Lenstra JA (1998) A satellite DNA element specific for roe deer (*Capreolus capreolus*). *Chromosoma* 107(1):1–5
- Chiang PY, Lin CC, Liao SJ, Hsieh LJ, Li SY, Chao MC, Li YC (2004) Genetic analysis for two subspecies of the Reeve's muntjac (Cervidae: *Muntiacus reevesi*) by karyotyping and satellite DNA analyses. *Zool Stud* 43:9
- Chou SH, Chin KH (2001) Quadruple intercalated G-6 stack: a possible motif in the fold-back structure of the *Drosophila* centromeric dodeca-satellite? *J Mol Biol* 314(1):139–152, doi:10.1006/jmbi.2001.5131
- Chou SH, Zhu L, Reid BR (1994) The unusual structure of the human centromere (GGA)₂ motif: unpaired guanosine residues stacked between sheared G.A pairs. *J Mol Biol* 244(3):259–268, doi:10.1006/jmbi.1994.1727
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188–1190, doi:10.1101/gr.849004
- Ferrer N, Azorin F, Villasante A, Gutierrez C, Abad JP (1995) Centromeric dodeca-satellite DNA sequences form fold-back structures. *J Mol Biol* 245(1):8–21
- Fronick L, Chowdhary BP, Scherthan H (1997) Segmental homology among cattle (*Bos taurus*), Indian muntjac (*Muntiacus muntjak vaginalis*), and Chinese muntjac (*M. reevesi*) karyotypes. *Cytogenet Cell Genet* 77(3–4):223–227
- Fry K, Salsler W (1977) Nucleotide sequences of HS-alpha satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* 12(4):1069–1084
- Goldberg IG, Sawhney H, Pluta AF, Warburton PE, Earnshaw WC (1996) Surprising deficiency of CENP-B binding sites in African green monkey alpha-satellite DNA: implications for CENP-B function at centromeres. *Mol Cell Biol* 16(9):5156–5168
- Haaf T, Ward DC (1994) Structural analysis of alpha-satellite DNA and centromere proteins using extended chromatin and chromosomes. *Hum Mol Genet* 3(5):697–709
- Hernandez Fernandez M, Vrba ES (2005) A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant relatives. *Biol Rev Camb Philos Soc* 80(2):269–302

- Horvath JE, Viggiano L, Loftus BJ, Adams MD, Archidiacono N, Rocchi M, Eichler EE (2000) Molecular structure and evolution of an alpha satellite/non-alpha satellite junction at 16p11. *Hum Mol Genet* 9(1):113–123
- Jurka J (2008) LINE1 repeats from dolphin. *Rebase Reports* 8:1
- Lee C, Ritchie DB, Lin CC (1994) A tandemly repetitive, centromeric DNA sequence from the Canadian woodland caribou (*Rangifer tarandus caribou*): its conservation and evolution in several deer species. *Chromosome Res* 2(4):293–306
- Li YC, Lee C, Hseu TH, Li SY, Lin CC (2000a) Direct visualization of the genomic distribution and organization of two cervid centromeric satellite DNA families. *Cytogenet Cell Genet* 89(3–4):192–198
- Li YC, Lee C, Sanoudou D, Hseu TH, Li SY, Lin CC (2000b) Interstitial colocalization of two cervid satellite DNAs involved in the genesis of the Indian muntjac karyotype. *Chromosome Res* 8(5):363–373
- Li YC, Lee C, Chang WS, Li SY, Lin CC (2002) Isolation and identification of a novel satellite DNA family highly conserved in several Cervidae species. *Chromosoma* 111(3):176–183, doi:10.1007/s00412-002-0200-x
- Li YC, Cheng YM, Hsieh LJ, Ryder OA, Yang F, Liao SJ, Hsiao KM, Tsai CH, Lin CC (2005) Karyotypic evolution of a novel cervid satellite DNA family isolated by microdissection from the Indian muntjac Y-chromosome. *Chromosoma* 114(1):28–38, doi:10.1007/s00412-005-0335-7
- Lin CC, Li YC (2006) Chromosomal distribution and organization of three cervid satellite DNAs in Chinese water deer (*Hydropotes inermis*). *Cytogenet Genome Res* 114(2):147–154, doi:10.1159/000093331
- Lin CC, Sasi R, Fan YS, Chen ZQ (1991) New evidence for tandem chromosome fusions in the karyotypic evolution of Asian muntjacs. *Chromosoma* 101(1):19–24
- Lin CC, Chiang PY, Hsieh LJ, Liao SJ, Chao MC, Li YC (2004) Cloning, characterization and physical mapping of three cervid satellite DNA families in the genome of the Formosan muntjac (*Muntiacus reevesi micrurus*). *Cytogenet Genome Res* 105(1):100–106, doi:10.1159/000078015
- Lin CC, Hsu PC, Li TS, Liao SJ, Cheng YM, Hsieh LJ, Li YC (2008) Construction of an Indian Muntjac BAC library and production of the most highly dense FISH map of the species. *Zool Stud* 47:11
- Murmann AE, Mincheva A, Scheuermann MO, Gautier M, Yang F, Buitkamp J, Strissel PL, Strick R, Rowley JD, Lichter P (2008) Comparative gene mapping in cattle, Indian muntjac, and Chinese muntjac by fluorescence in situ hybridization. *Genetica* 134(3):345–351, doi:10.1007/s10709-008-9242-1
- Pitra C, Fickel J, Meijaard E, Groves PC (2004) Evolution and phylogeny of old world deer. *Mol Phylogenet Evol* 33(3):880–895, doi:10.1016/j.ympev.2004.07.013
- Price SA, Bininda-Emonds OR, Gittleman JL (2005) A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (Cetartiodactyla). *Biol Rev Camb Philos Soc* 80(3):445–473
- Puechberty J, Laurent AM, Gimenez S, Billault A, Brun-Laurent ME, Calenda A, Marcas B, Prades C, Ioannou P, Yurov Y, Roizes G (1999) Genetic and physical analyses of the centromeric and pericentromeric regions of human chromosome 5: recombination across 5cen. *Genomics* 56(3):274–287, doi:10.1006/geno.1999.5742
- Rudd MK, Willard HF (2004) Analysis of the centromeric regions of the human genome assembly. *Trends Genet* 20(11):529–533, doi:10.1016/j.tig.2004.08.008
- Rudd MK, Wray GA, Willard HF (2006) The evolutionary dynamics of alpha-satellite. *Genome Res* 16(1):88–96, doi:10.1101/gr.3810906
- Schueler MG, Sullivan BA (2006) Structural and functional dynamics of human centromeric chromatin. *Annu Rev Genomics Hum Genet* 7:301–313, doi:10.1146/annurev.genom.7.080505.115613
- Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. *Science* 294(5540):109–115, doi:10.1126/science.1065042
- Schueler MG, Dunn JM, Bird CP, Ross MT, Viggiano L, Rocchi M, Willard HF, Green ED (2005) Progressive proximal expansion of the primate X chromosome centromere. *Proc Natl Acad Sci U S A* 102(30):10563–10568, doi:10.1073/pnas.0503346102
- Smit AF, Toth G, Riggs AD, Jurka J (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J Mol Biol* 246(3):401–417, doi:10.1006/jmbi.1994.0095
- Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* 191(4227):528–535
- Ugarkovic D, Plohl M (2002) Variation in satellite DNA profiles—causes and effects. *EMBO J* 21(22):5955–5959
- Vafa O, Shelby RD, Sullivan KF (1999) CENP-A associated complex satellite DNA in the kinetochore of the Indian muntjac. *Chromosoma* 108(6):367–374
- Willard HF (1985) Chromosome-specific organization of human alpha satellite DNA. *Am J Hum Genet* 37(3):524–532

doi:10.1186/s40555-014-0025-3

Cite this article as: Hsieh et al.: Organization and evolution of a novel cervid satellite DNA with yeast CDEI-like repeats. *Zoological Studies* 2014 **53**:25.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com