

# SCIENTIFIC REPORTS



OPEN

## Higher-order repeat structure in alpha satellite DNA occurs in New World monkeys and is not confined to hominoids

Received: 29 October 2014

Accepted: 25 March 2015

Published: 14 May 2015

Penporn Sujiwattanarat<sup>1,2</sup>, Watcharaporn Thapana<sup>1,2</sup>, Kornson Srikulnath<sup>2</sup>, Yuriko Hirai<sup>1</sup>, Hirohisa Hirai<sup>1</sup> & Akihiko Koga<sup>1</sup>

Centromeres usually contain large amounts of tandem repeat DNA. Alpha satellite DNA (AS) is the most abundant tandem repeat DNA found in the centromeres of simian primates. The AS of humans contains sequences organized into higher-order repeat (HOR) structures, which are tandem arrays of larger repeat units consisting of multiple basic repeat units. HOR-carrying AS also occurs in other hominoids, but results reported to date for phylogenetically more remote taxa have been negative. Here we show direct evidence for clear HOR structures in AS of the owl monkey and common marmoset. These monkeys are New World monkey species that are located phylogenetically outside of hominoids. It is currently postulated that the presence of HOR structures in AS is unique to hominoids. Our results suggest that this view must be modified. A plausible explanation is that generation of HOR structures is a general event that occurs occasionally or frequently in primate centromeres, and that, in humans, HOR-carrying AS became predominant in the central region of the centromere. It is often difficult to assemble sequence reads of tandem repeat DNAs into accurate contig sequences; our careful sequencing strategy allowed us to overcome this problem.

The centromere is part of a chromosome essential for correct chromosome segregation during cell division, serving as the point to which the spindle fiber attaches via the kinetochore. Centromeres of higher eukaryotes generally contain large amounts of tandem repeat DNA. Alpha satellite DNA (AS) is the most abundant tandem repeat DNA of primate centromeres<sup>1,2</sup>, although this may not be true of suborder Strepsirrhini, one member of which (the aye-aye, *Daubentonia madagascariensis*) is known to carry other unrelated tandem repeat DNA as main components of its centromeres<sup>3</sup>. Strepsirrhini is a taxon that diverged from other groups in an early stage of the primate evolution. The length of the repeat units of AS is approximately 170 bp in parvorder Catarrhini, which includes hominoids (superfamily Hominoidea; humans, great apes, and small apes) and Old World monkeys (family Cercopithecidae; macaques, baboons, and related monkeys found in Africa and Asia)<sup>4–6</sup>. In New World monkeys (parvorder Platyrrhini; monkeys inhabiting Central and South America), the length of the repeat units is approximately 340 bp<sup>6,7</sup>. Sequence analyses have suggested that this difference is due to an ancient dimeric structure of AS of New World monkeys<sup>6</sup>. The parvorders Catarrhini and Platyrrhini constitute the infraorder Simiiformes (simian primates).

The AS of humans, which has been extensively studied for its structural features, is known to contain sequences organized into higher-order repeat (HOR) structures, which are tandem arrays of larger repeat units that consist of multiple basic repeat units<sup>8,9</sup>. The larger repeat units that have so far been identified include those comprising 2, 4, 5, 6, 8, 11 and 13 basic repeat units<sup>9–15</sup>. The HOR structure also occurs in

<sup>1</sup>Primate Research Institute, Kyoto University, Inuyama City 484-8506, Japan. <sup>2</sup>Faculty of Science, Kasetsart University, Bangkok 10900, Thailand. Correspondence and requests for materials should be addressed to A.K. (email: koga.akihiko.5n@kyoto-u.ac.jp)

other hominoid species<sup>2,5</sup>, including gibbons<sup>16,17</sup>, but has not been reported in Old World monkeys or New World monkeys. For this reason, it is currently postulated that the HOR structure of AS is a unique attribute of hominoids<sup>6,16</sup>. In the present study, we asked whether the HOR structure occurs in a taxon that is located phylogenetically outside of hominoids. We raised this question because the repetitive DNA sequences that constitute aye-aye centromeres (called DMA1 and DMA2) exhibit the HOR structure<sup>3</sup>. The aye-aye is phylogenetically more remote from hominoids than Old World monkeys or New World monkeys are. It is possible that the HOR structure has not been identified in Old World monkeys or New World monkeys simply because the experimental methods used to date lack sufficient detection power or because the data currently available in sequence databases are of insufficient quality.

In our previous studies<sup>16,17</sup>, we identified the HOR structure in AS of gibbons by obtaining accurate, long sequences of AS-carrying genomic DNA clones. Our strategy consisted of two steps. First, we selected AS-carrying clones that appeared to contain the HOR structure. Second, we sequenced a long region within the identified candidate clone. By inserting a bacterial transposon into the candidate clone at various positions, we collected two kinds of information: the position of transposon insertion, and the sequences of the transposon-flanking regions. Using the former positional information to collate the latter sequence data, we constructed accurate contig sequences of the long regions. In the present study, we further improved our method and applied it to AS of Azara's owl monkey (*Aotus azarae*) and the common marmoset (*Callithrix jacchus*) and obtained clear evidence of the HOR structure outside the hominoid lineage.

## Results

**Screening of genomic library for AS-carrying clones.** In our previous study<sup>18</sup>, we identified two types of AS in the owl monkey, which we named OwlAlp1 (185-bp repeat units) and OwlAlp2 (344-bp repeat units). Our cloning method employed genomic hybridization against fosmid clones randomly selected from an owl monkey genomic library. In the present study, we conducted another round of library screening using the same method. Starting with 384 clones (contained in four 96-well plates) randomly selected from the genomic library, we picked 24 clones that exhibited the highest levels of signal intensities. One end (>750bp) of the insert fragments of the 24 clones was sequenced with a primer that represented one terminal region of the pCCIFOS vector. Six of the 24 clones were found to contain tandem repeat structures showing sequence identities of >90% with the consensus sequence of OwlAlp1, and another six clones showed sequence identities of >90% to the consensus sequence of OwlAlp2<sup>18</sup>. Of the remaining 12 clones, nine were found to contain OwlRep, another tandem repeat DNA present in large amounts in the owl monkey genome<sup>19</sup>.

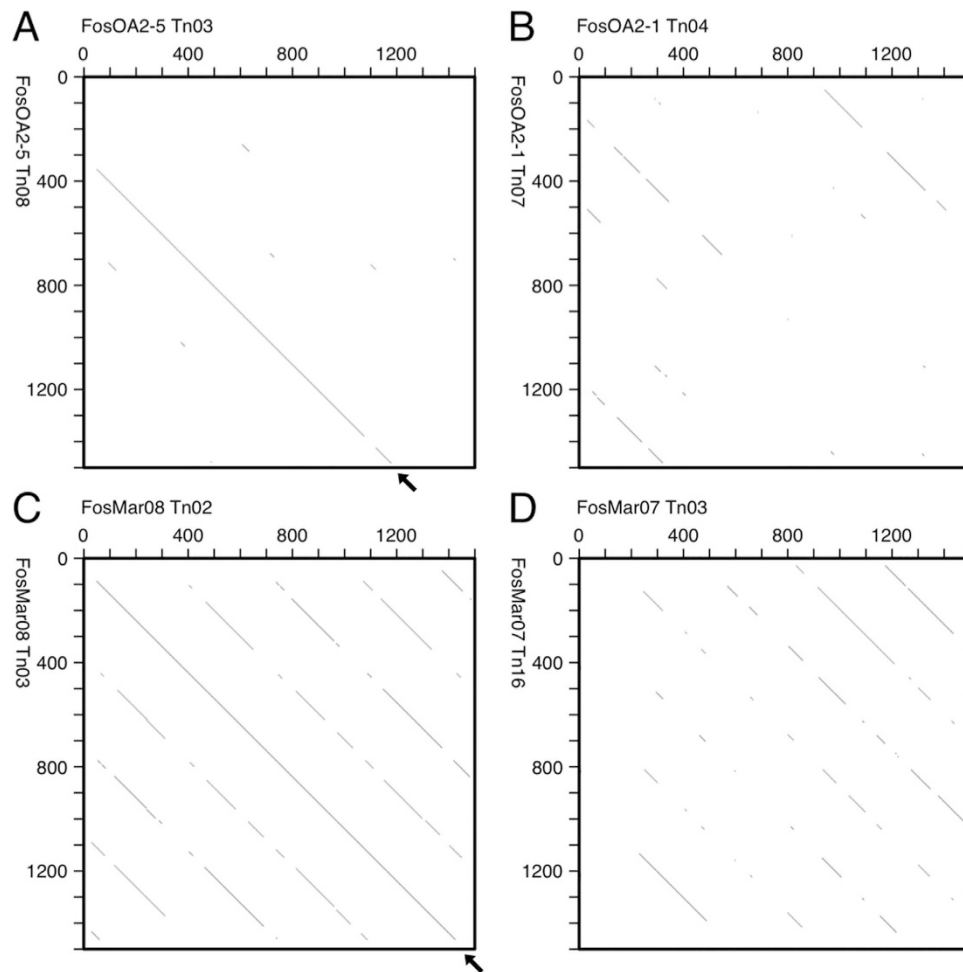
Using the same method, from the marmoset library, we collected 24 clones that were presumed to carry tandem repeat DNA. We examined the terminal regions of six of these 24 clones, and all exhibited a >90% sequence identity with the consensus sequence of the marmoset AS<sup>6</sup>.

**Detection of an HOR sign by partial sequencing.** We delivered a modified Tn5 transposon to the six OwlAlp2 and six marmoset AS clones, and collected those carrying the transposon in their insert portions. For each original clone, we sequenced three transposon-carrying clones for transposon-flanking regions. The three clones were selected so that their insertion points would be apart from one another by >3kb in order to ensure no overlap among their sequence reads. We then compared the sequence data with a dot matrix analysis for a sign of the HOR structure. The logic underlying this test has been described in our previous report<sup>11</sup>. Briefly, if a line spanning more than two repeat units on a dot matrix is observed, it can be regarded as an HOR sign.

We found signs of the HOR structure in two of the six owl monkey clones (FosOA2-2 and FosOA2-5) and one of the six marmoset clones (FosMar08), as shown in Fig. 1 (negative cases are also shown).

**Construction of long contig sequences.** We conducted further experiments, using FosOA2-5 and FosMar08, to obtain long contig sequences of the AS. The insert fragments of these clones were tandem repeat DNA and, in addition, were considered to contain identical sequence blocks at multiple locations because of the anticipated HOR structure. For this reason, we could not rely on the shotgun or primer-walking sequencing strategies. We used the same strategy, with a small modification, as that employed in our previous study<sup>17</sup>, in which multiple sequence reads were collected together with the location data of the sequence reads and they were assembled into a contig sequence based on their location. The modification that we made in the present study was to collect fosmid clones that carried the transposon at various locations and sequence these fosmds to obtain the transposon-flanking regions. In our previous method, an insert fragment was transferred once to a plasmid and a series of nested deletion clones were prepared with restriction endonucleases and an exonuclease. This modification was effective in reducing the number of clones to be sequenced because the length of each sequence stretch was doubled by reading it in both directions from a single point. The modification was also useful for reducing the possibility of artificial sequence rearrangements by removing the subcloning step<sup>20</sup>.

For the FosOA2-5 clone, we obtained a total of 18 partially overlapping sequence reads and assembled them into a 10829-bp contig sequence. For the FosMar08 clone, we collected 24 sequence reads, which were assembled into a 13124-bp sequence. We deposited these sequences in GenBank under the accession numbers LC002884 and LC030305, respectively.



**Figure 1.** Dot matrix analysis of sequence reads for a sign of the HOR structure. The criterion in the analysis was that a 49- or 50-nucleotide match should exist over a window of 50 nucleotides. A line spanning more than two basic repeat units (approximately 680 bp), if found, is regarded as a sign for an HOR structure. Examples of the results are shown. **A.** Fosmid clone FosOA2-5 of owl monkey that showed a sign. **B.** FosOA2-1 of owl monkey that was negative. **C.** FosMar08 of marmoset that showed a sign. **D.** FosMar07 of marmoset that was negative.

**Evidence for HOR structure from the contig sequences.** The contig sequences derived from the FosOA2-2 and FosMar08 clones contained 31 and 38 basic repeat units, respectively, for which the same boundaries of the repeat units as those defined in our previous study<sup>18</sup> were used. The alignment of these repeat units (Figs. 2 and 3) suggested the presence of the HOR structure, showing repetitions of several similar patterns along the sequence. Pairwise comparisons of the sequence identities between the repeat units (Figs. 4 and 5) provided clear evidence for the HOR structure: mutually parallel lines that consisted totally or mostly of red cells appeared against a background comprised mostly of yellow and white cells. The average pairwise identities among the basic repeat units are shown in Table 1. The distance (number of cells along the axis) between the parallel lines in Figs. 4 and 5 corresponds to the size of the larger repeat units, which was nine and 12 basic repeat units in FosOA2-2 and FosMar08, respectively.

## Discussion

The HOR structure in AS has been found in humans and other hominoid species, but has not been reported in phylogenetically more remote primate taxa, such as Old World monkeys or New World monkeys. For this reason, it is currently postulated that the HOR structure is a unique attribute of hominoids. In the present study, we observed clear HOR structures in AS of the owl monkey and marmoset, which are New World monkey species. Our results suggested that the HOR structure occurs in AS of a wide range of simian primates. We undertook this study to identify the HOR structure in AS of New World monkeys partly because the HOR structure has already been identified in the centromere-region repetitive DNA (DMA1 and DMA2) of the aye-aye<sup>3</sup>, a species phylogenetically located more remote from hominoids than New World monkeys. This raised the possibility that the HOR structure is common in AS of simian primates, and is not confined to hominoids. It now appears possible that the generation of

```

Con ACAGAGAAATATCTGCTTCTAAAACCAAACGGAGCCTCTFAACTGAATGGTGTCCATCTGTGCATTCAACTGACAGAGTAAATCAT
01 .....C.....
02 .....T.....A.....C.....
03 .....C.....T.....G.....A.....T.....A.....A.....
04 .....T.....G.....A.....A.....A.....A.....
05 .....A.....C.....A.....C.....G.....C.....
06 .....C.....A.....C.....A.....C.....G.....G.....
07 .....T.....A.....A.....A.....G.....A.....
08 .....C.....T.....A.....G.....A.....
09 .....T.....A.....A.....A.....C.....T.....A.....
10 .....C.....T.....A.....A.....C.....T.....A.....
11 .....T.....A.....A.....A.....C.....T.....A.....
12 .....C.....T.....G.....A.....A.....T.....A.....
13 .....T.....G.....A.....A.....A.....A.....
14 .....A.....C.....C.....A.....C.....G.....C.....
15 .....C.....A.....C.....A.....A.....C.....G.....
16 .....T.....A.....G.....A.....G.....A.....
17 .....C.....T.....A.....G.....A.....
18 .....T.....A.....A.....A.....C.....T.....A.....
19 .....T.....A.....A.....A.....C.....T.....A.....
20 .....C.....T.....G.....A.....A.....T.....A.....
21 .....C.....T.....G.....A.....A.....T.....A.....
22 .....T.....G.....A.....A.....T.....A.....
23 .....A.....C.....A.....C.....G.....C.....
24 .....C.....A.....C.....A.....C.....G.....G.....
25 .....T.....A.....A.....A.....G.....A.....
26 .....C.....T.....A.....G.....A.....
27 .....T.....A.....A.....A.....C.....T.....A.....
28 .....T.....A.....A.....A.....C.....T.....A.....
29 .....T.....A.....A.....A.....C.....T.....A.....
30 .....C.....T.....G.....A.....A.....T.....A.....
31 .....T.....G.....A.....A.....T.....A.....

Con GT-GTGGATTGAGCGGTGTGCAACCCCTTCTGTGCAGAAATCCGAAACCGGCATTTCAGGCCAATGCCGCATACGTAGGAACAGCA
01 C.....C.....T.....A.....G.....G.....G.....A.....A.....
02 .....C.....T.....A.....G.....A.....G.....T.....A.....T.....G.....
03 .....A.....C.....T.....A.....T.....G.....G.....A.....G.....
04 .....T.....A.....G.....A.....A.....A.....A.....A.....
05 .....A.....G.....A.....A.....A.....A.....C.....T.....
06 .....T.....T.....A.....A.....A.....G.....A.....
07 .....T.....A.....C.....T.....T.....C.....T.....
08 .....T.....A.....A.....A.....T.....
09 .....A.....T.....A.....T.....
10 .....C.....T.....A.....A.....G.....G.....A.....A.....
11 .....C.....T.....A.....A.....T.....G.....T.....T.....G.....
12 .....A.....C.....T.....A.....T.....G.....G.....A.....G.....
13 .....A.....G.....A.....A.....A.....A.....A.....
14 .....A.....T.....A.....A.....A.....C.....T.....
15 .....T.....T.....A.....A.....A.....G.....A.....
16 .....T.....A.....C.....T.....T.....C.....T.....
17 .....T.....A.....A.....A.....T.....
18 .....A.....T.....A.....T.....
19 .....C.....T.....A.....A.....G.....G.....A.....A.....
20 .....C.....T.....A.....A.....T.....G.....T.....T.....G.....
21 .....A.....C.....T.....A.....T.....G.....G.....A.....G.....
22 .....A.....G.....A.....A.....A.....A.....A.....
23 .....A.....G.....A.....A.....A.....C.....T.....
24 .....T.....T.....A.....A.....A.....G.....A.....
25 .....T.....A.....C.....T.....T.....C.....T.....
26 .....T.....A.....A.....A.....T.....
27 .....A.....G.....A.....A.....A.....T.....
28 .....C.....T.....A.....A.....G.....G.....A.....A.....
29 .....A.....C.....T.....A.....A.....G.....T.....T.....G.....
30 .....A.....A.....G.....T.....A.....A.....A.....G.....A.....
31 .....A.....G.....A.....A.....A.....A.....A.....G.....A.....

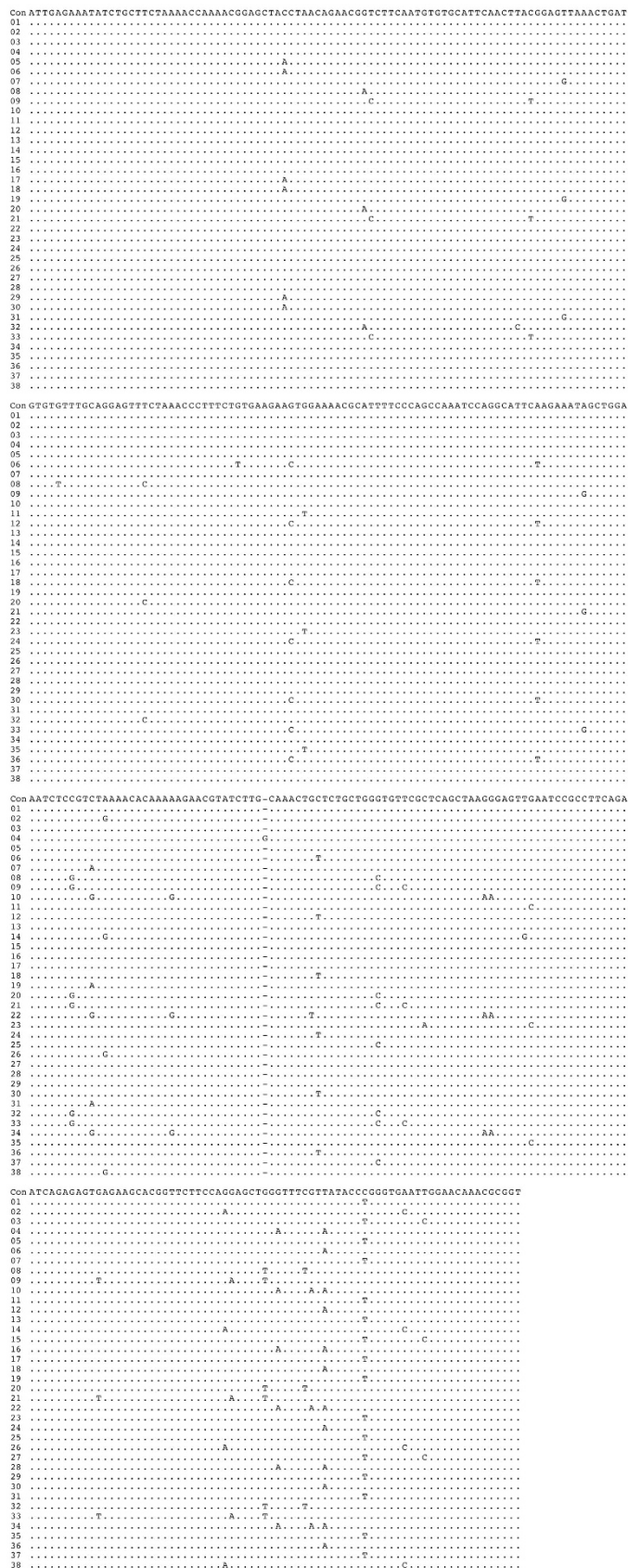
Con GGAGACTTCGCTTAAAACCCAGGAGAAACGTTTCCGTGAAACTGCTCTGTGGTTCGTTACCTAAGAGAGTTGAATCTATGTTGA
01 .....A.....C.....A.....C.....C.....C.....A.....G.....
02 .....T.....C.....T.....T.....C.....C.....A.....G.....
03 .....T.....A.....T.....T.....G.....A.....C.....
04 .....A.....A.....T.....T.....C.....G.....C.....
05 .....A.....A.....T.....C.....G.....C.....
06 .....T.....A.....A.....A.....A.....A.....
07 .....T.....A.....A.....A.....A.....A.....
08 .....C.....A.....A.....A.....A.....A.....
09 .....C.....A.....A.....A.....C.....A.....
10 .....C.....A.....A.....A.....C.....A.....
11 .....T.....C.....T.....T.....C.....A.....G.....
12 .....T.....A.....T.....T.....C.....G.....
13 .....A.....A.....T.....T.....C.....C.....
14 .....A.....A.....T.....C.....G.....C.....
15 .....T.....A.....A.....A.....A.....A.....
16 .....A.....A.....A.....A.....A.....
17 .....C.....A.....A.....A.....A.....A.....
18 .....C.....A.....A.....A.....A.....A.....
19 .....A.....C.....A.....A.....C.....C.....A.....
20 .....T.....C.....T.....T.....C.....C.....A.....G.....
21 .....A.....T.....T.....A.....C.....G.....C.....
22 .....A.....T.....T.....T.....C.....A.....C.....
23 .....A.....A.....T.....C.....G.....C.....
24 .....T.....A.....A.....A.....A.....A.....
25 .....A.....A.....A.....C.....A.....A.....
26 .....C.....A.....A.....A.....A.....A.....
27 .....A.....C.....A.....A.....A.....A.....
28 .....A.....C.....A.....A.....C.....C.....A.....
29 .....T.....C.....T.....T.....A.....C.....A.....G.....
30 .....A.....T.....T.....A.....C.....A.....G.....C.....
31 .....A.....T.....T.....A.....T.....C.....C.....

Con GATTACGGAGTTAGAACACGTTTCTCTAGGAGCT-GCGTGGGATATCCCGAGTGAATGGGACCATAGGCTGT
01 .....A.....A.....C.....CC.....TA.....T.....
02 .....G.....T.....A.....A.....C.....T.....
03 .....C.....A.....A.....C.....C.....C.....
04 .....A.....C.....G.....C.....T.....C.....T.....A.....C.....
05 .....A.....C.....G.....C.....T.....C.....T.....A.....G.....
06 .....A.....C.....G.....T.....C.....T.....T.....
07 .....A.....C.....T.....C.....T.....T.....T.....
08 .....A.....T.....T.....C.....A.....A.....G.....G.....
09 .....A.....A.....A.....C.....CC.....TA.....A.....
10 .....A.....A.....A.....C.....CC.....TA.....T.....
11 .....G.....T.....A.....A.....A.....C.....T.....
12 .....C.....A.....A.....C.....C.....C.....
13 .....A.....C.....G.....C.....T.....C.....T.....C.....
14 .....A.....C.....G.....C.....T.....C.....T.....A.....G.....
15 .....A.....C.....G.....T.....C.....T.....T.....
16 .....A.....C.....G.....T.....C.....T.....T.....
17 .....A.....T.....T.....C.....A.....A.....G.....G.....
18 .....A.....A.....A.....T.....C.....A.....A.....A.....
19 .....A.....A.....A.....C.....CC.....TA.....T.....
20 .....G.....T.....A.....A.....A.....C.....T.....
21 .....C.....A.....A.....C.....C.....C.....
22 .....A.....C.....G.....C.....T.....C.....T.....A.....C.....
23 .....A.....C.....G.....C.....T.....C.....T.....A.....G.....
24 .....A.....C.....G.....T.....C.....T.....T.....
25 .....C.....T.....T.....C.....A.....A.....G.....G.....
26 .....A.....T.....T.....C.....A.....A.....G.....G.....
27 .....A.....A.....A.....C.....CC.....TA.....A.....
28 .....A.....A.....A.....C.....CC.....TA.....T.....
29 .....G.....T.....A.....A.....A.....C.....T.....
30 .....C.....G.....T.....A.....A.....C.....C.....T.....
31 .....A.....C.....G.....T.....C.....T.....C.....

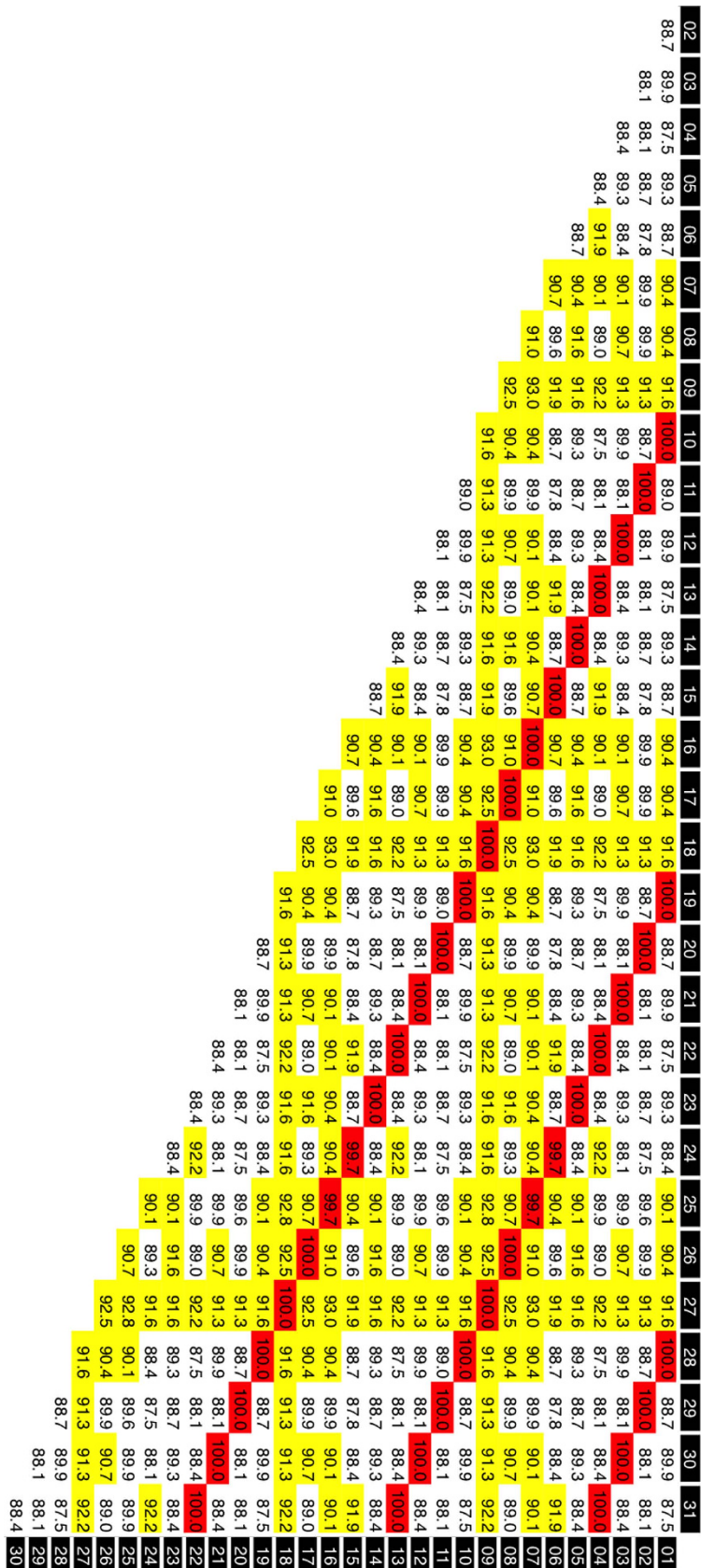
```

**Figure 2.** Alignment of the nucleotide sequences of the 31 basic repeat units in the FosOA2-5 clone. The alignment was performed with the MEGA6 program<sup>24</sup> under default settings. *Con* indicates the consensus sequence that was assembled by collecting the most common nucleotides at the respective sites. Nucleotide sites occupied by the base in the consensus sequence are indicated by dots. Nucleotide sites containing different bases are shown as the respective bases observed. The minus symbol indicates the absence of a nucleotide at that position.

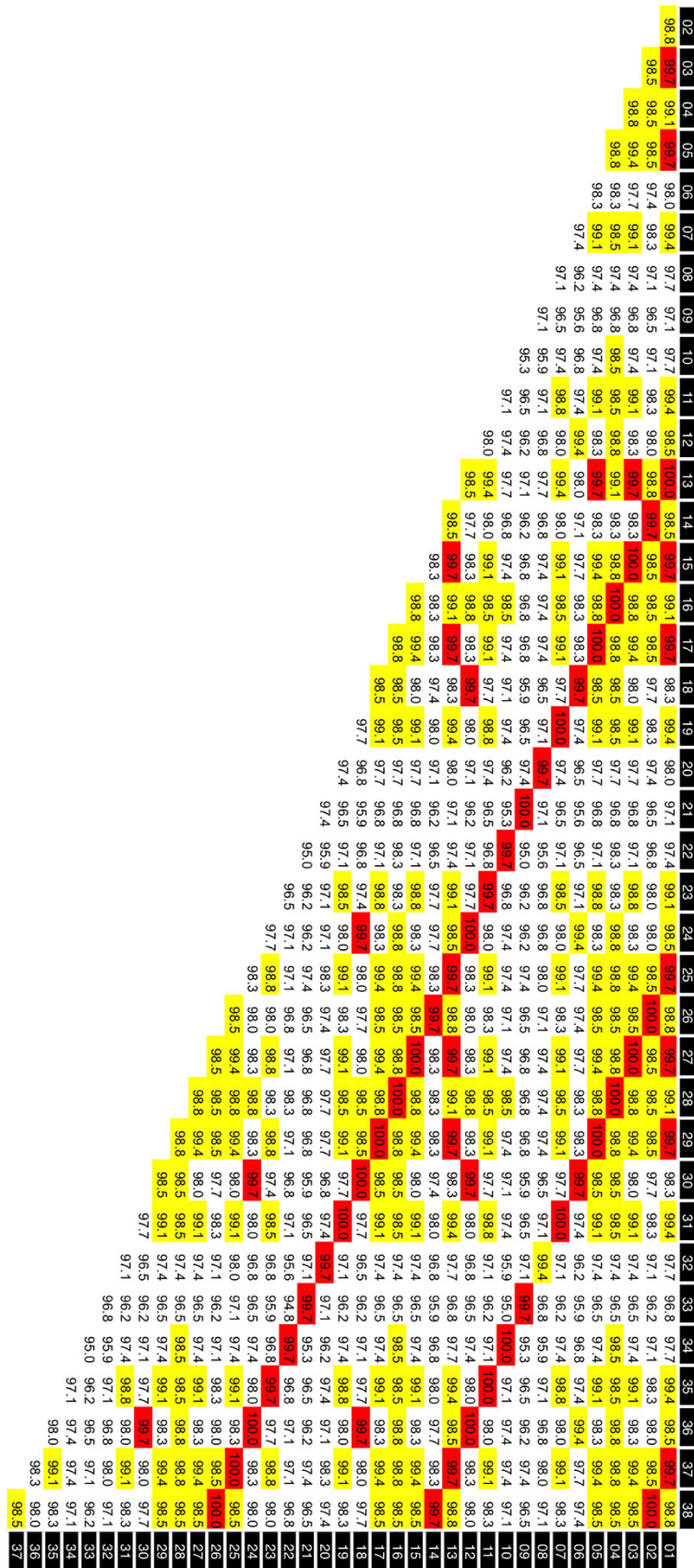




**Figure 3.** Alignment of the nucleotide sequences of the 38 basic repeat units in the FosMar08 clone. The sequence data obtained from the FosMar08 clone were treated by the same methods as those described in the legend to Fig. 2.



**Figure 4.** Pairwise comparisons of basic repeat unit sequences in the FosOA2-5 clone. The comparisons were made with the MEGA6 program<sup>24</sup> under default settings. The horizontal and vertical axes of each matrix represent a one-dimensional array of the basic repeat units. Each cell contains the nucleotide identity of the corresponding pair. The 31 basic repeat units in the FosOA2-5 clone. Yellow cells, nucleotide identities of 90.0–95.0; red cells, >95.0%.



**Figure 5.** Pairwise comparisons of basic repeat unit sequences in the FosMar08 clone. The sequence data obtained from the FosMar08 clone were treated by the same methods as those described in the legend to Fig. 4, except for the coloring criteria: yellow cells, nucleotide identities of 98.5–99.5%; red cells, >99.5%.



Clone	FosOA2-5	FosMar08
All pairs		
No. of pairs	$31 \times 30/2 = 465$	$38 \times 37/2 = 703$
Average identity	90.7%	98.0%
Pairs at identical positions <sup>a</sup>		
No. of pairs	$22 + 13 + 4 = 39$	$26 + 14 + 2 = 42$
Average identity	100%	99.9%
Pairs at nonidentical positions <sup>b</sup>		
No. of pairs	$465 - 39 = 426$	$703 - 42 = 661$
Average identity	89.8%	97.9%

**Table 1.** Average pairwise identities. <sup>a</sup>Pairs of basic repeat units that were located at identical positions in different larger repeat units (pairs on the mutually parallel lines that consist totally or mostly of red cells in Figs. 4 and 5). <sup>b</sup>Pairs of basic repeat units not bearing the above-mentioned relationship.

an HOR structure is a general event occurring occasionally or frequently in centromeres of all simian primates and that, in humans, HOR-carrying AS became predominant in the central region of the centromere by being associated with a significant centromere function, changes in the turnover mechanisms of AS, and/or other unknown mechanisms.

Detection of an HOR structure in AS of Old World monkeys and New World monkeys has been attempted using various methods, including restriction enzyme analysis of genomic DNA<sup>4,5,7,21</sup>, and computational analysis of shotgun sequence databases<sup>5,6</sup>. To our knowledge, however, no results have been positive. The restriction enzyme method has the potential to succeed only when the larger repeat units carry the recognition sites of the enzymes used. For the computational analyses tried to date, the original data to be analyzed (sequence reads generated by current sequencing machines) may not be long enough for detection of an HOR structure. For the marmoset AS in particular, the high nucleotide identity among repeat units may pose another obstacle to detection of an HOR structure. In the human AS, the average identities among basic repeat units located at identical and nonidentical positions (in different larger repeat units) have been reported to be 95–99% and 70–90%, respectively<sup>2,5,9</sup>. The HOR structures of gibbon AS that we identified in our previous studies exhibited values comparable with these<sup>16,17</sup>. In the marmoset AS, however, the nucleotide identities were 99.9% and 97.9%, respectively. The overall identity was as high as 98.0%.

The method we used in the present study was sufficiently powerful to detect the HOR structure. Some drawbacks of this method are that it requires much manual handling of clones and that it is time-consuming. Clones of repetitive DNA are often degraded easily when maintained or amplified in the host bacteria<sup>20</sup>. The marmoset AS clones were far more fragile than the other repetitive DNAs we have so far analyzed, probably due to their high sequence homogeneity, and they required more careful handling. Despite these drawbacks, our method is, to our knowledge, the only method to provide clear and direct evidence of the HOR structure in AS of New World monkeys.

## Methods

**Ethics statement.** All animal experiments in this study were approved by the Animal Care and Use Committee of Kyoto University Primate Research Institute (KUPRI), and were performed in accordance with the Guidelines for Care and Use of Nonhuman Primates (Version 3; June 2010) of KUPRI.

**Azara's owl monkey, genomic library, and library screening.** We used the same *A. azarae* genomic library as that used in our previous study<sup>16</sup>. The essential information is as follows: source of genomic DNA, cultured epithelial cells originating from an adult female; vector, fosmid pCC1FOS that is 8.1 kb in length and carries the chloramphenicol-resistance gene; and insert DNA, 40- to 44-kb fragments produced by mechanical shearing. We screened this library for repetitive sequences by the genomic hybridization technique, of which the strategy and methods have been described in our previous studies<sup>18,22,23</sup>. We used the AlkPhos Direct Labelling and Detection System (product of GE Healthcare) for hybridization and signal detection. The probe used was owl monkey genomic DNA that had been mechanically sheared to lengths of around 10 kb. The hybridization temperature was 59°C, at which a medium hybridization stringency was expected.

**Common marmoset.** Genomic DNA was extracted from cultured epithelial cells of a male common marmoset (bred at KUPRI; individual identification number 186). The method for the library



construction and screening was the same as that for the owl monkey. The probe used for the hybridization was marmoset genomic DNA sheared to lengths of around 10 kb.

**Preparation of fosmid clones carrying transposon insertions.** The bacterial transposon Tn5 was used for detection of signs of the HOR structure and sequencing of candidate clones. We induced an *in vitro* transposition reaction and collected secondary fosmids that carried the transposon on the original fosmid. The methods for these processes have been described in full previously, with schematic illustrations<sup>17</sup>. The Tn5 transposon was modified in advance to carry the kanamycin-resistance gene for the selection of insertion-carrying fosmids and some restriction enzyme sites for determining Tn5 insertion points by restriction mapping.

**Sequencing of terminal regions and transposon-flanking regions.** Fosmid DNAs were sequenced by Sanger's method with an Applied Biosystems 3730xl DNA Analyzer. To obtain sequences of the terminal regions of the insert fragment, we used primers that represented nucleotide blocks encompassing the insertion site of the pCCIFOS vector (nucleotides 276–305 of GenBank file EU140751). Each sequencing assay provided a sequence read of >1000 bp, from which we used the first 750 bp as an accurate sequence data. For the sequence data of the transposon-flanking regions, we used primers that represented the left and right terminal regions of the modified transposon and were oriented outwards (5'-GAATTTTGAATTCGGTACCATGCGGCCGCT-3' and 5'-TGAGCGGCCGCTAAAGCTTCTAGACCAACA-5'). By combining the pair of sequence reads at the transposon insertion breakpoint, we could obtain a sequence stretch of >2000 bp. We used its internal 1500-bp portion, 750 bp each from the junction point, as accurate sequence data.

The fosmid vector pCCIFOS is present as a single copy in bacterial cells because it is controlled by the replication origin *ori2*. This vector carries another replication origin, *oriV*, that leads to multiple copies by exposing bacterial cells to an induction reagent. The induction of multiple copies is usually applied for sequencing purposes. We, however, did not include multiple copy induction in any part of our experiments. The OwlAlp2 sequence is fragile when carried by a cloning vector, especially a multiple-copy vector<sup>20</sup>, and the FosMar08 sequence is even more fragile. To obtain accurate sequence data by preventing structural changes, we treated the bacterial culture so that the single-copy situation would be maintained. The reduction in the DNA amount that was caused by the no induction treatment was compensated for by increasing the culture volume from 1 ml to 16 ml per sequencing reaction.

The temperature of bacterium culturing is another significant factor for avoiding structural changes. The FosOA2-5 and FosMar08 clones were highly fragile when cultured at 37°C, but stable when cultured at 25°C and 22°C, respectively. We conducted bacterium culturing at these temperatures. Because the bacterium growth was slow at these low temperatures, the culturing time was extended, from regular overnight culturing, to 48 h.

## References

- Willard, H. F. Evolution of alpha satellite. *Curr. Opin. Genet. Dev.* **1**, 509–514 (1991).
- Rudd, M. K., Wray, G. A. & Willard, H. F. (2006) The evolutionary dynamics of alpha-satellite. *Genome Res.* **16**, 88–96.
- Lee, H. R., Hayden, K. E., & Willard, H. F. Organization and molecular evolution of CENP-A-associated satellite DNA families in a basal primate genome. *Genome Biol. Evol.* **3**, 1136–1149 (2011). doi: 10.1093/gbe/evr083.
- Pike, L. *et al.* Sequence and evolution of rhesus monkey alphoid DNA. *J. Mol. Evol.* **23**, 127–137 (1986).
- Alkan, C. *et al.* Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. *PLoS Comput. Biol.* **3**, 1807–1818 (2007). doi: 10.1371/journal.pcbi.0030181.
- Cellamare, A. *et al.* New insights into centromere organization and evolution from the white-cheeked gibbon and marmoset. *Mol. Biol. Evol.* **26**, 1889–1900 (2009). doi: 10.1093/molbev/msp101.
- Alves, G., Seuánez, H. N. & Fanning, T. Alpha satellite DNA in neotropical primates (Platyrrhini). *Chromosoma* **103**, 262–267 (1994).
- Willard, H. F. & Wayne, J. S. Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends Genet.* **3**, 192–198 (1987). doi:10.1016/0168-9525(87)90232-0.
- Warburton, P. E. & Willard, H. F. Genomic analysis of sequence variation in tandemly repeated DNA. Evidence for localized homogeneous sequence domains within arrays of alpha-satellite DNA. *J. Mol. Biol.* **216**, 3–16 (1990).
- Haaf, T. & Willard, H. F. Orangutan alpha-satellite monomers are closely related to the human consensus sequence. *Mamm. Genome* **9**, 440–447 (1998).
- Looijenga, L. H. *et al.* Alpha satellite DNAs on chromosomes 10 and 12 are both members of the dimeric suprachromosomal subfamily, but display little identity at the nucleotide sequence level. *Genomics* **13**, 1125–1132 (1992).
- Alexandrov, I. A. *et al.* Definition of a new alpha satellite suprachromosomal family characterized by monomeric organization. *Nucleic. Acids Res.* **11**, 2209–2215 (1993).
- Rosandić, M. *et al.* CENP-B box and p[alpha] sequence distribution in human alpha satellite higher-order repeats (HOR). *Chromosome Res.* **14**, 735–753 (2006).
- Waye, J. S. & Willard, H. F. Concerted evolution of alpha satellite DNA: evidence for species specificity and a general lack of sequence conservation among alphoid sequences of higher primates. *Chromosoma* **98**, 273–279 (1989).
- Greig, G. M., Warburton, P. E. & Willard, H. F. Organization and evolution of an alpha satellite DNA subset shared by human chromosomes 13 and 21. *J. Mol. Evol.* **37**, 464–475 (1993).
- Terada, S., Hirai, Y., Hirai, H. & Koga, A. Higher-order repeat structure in alpha satellite DNA is an attribute of hominoids rather than hominids. *J. Hum. Genet.* **58**, 752–754 (2013). doi: 10.1038/jhg.2013.87.
- Koga, A. *et al.* Evolutionary origin of higher-order repeat structure in alpha-satellite DNA of primate centromeres. *DNA Res.* **21**, 407–415 (2014). doi: 10.1093/dnares/dsu005.
- Prakhongcheep, O. *et al.* Two types of alpha satellite DNA in distinct chromosomal locations in Azara's owl monkey. *DNA Res.* **20**, 235–240 (2013).

19. Prakhongcheep, O. *et al.* Heterochromatin blocks constituting the entire short arms of acrocentric chromosomes of Azara's owl monkey: formation processes inferred from chromosomal locations. *DNA Res.* **20**: 461–470 (2013b). doi: 10.1093/dnares/dst023.
20. Thapana, W. *et al.* Reduction in the structural instability of cloned eukaryotic tandem-repeat DNA by low-temperature culturing of host bacteria. *Genet. Res. (Camb)* **96**, e13 (2014). doi: 10.1017/S0016672314000172.
21. Pieczarka, J. C. *et al.* Analysis of constitutive heterochromatin of Aotus (Cebidae, Primates) by restriction enzyme and fluorochrome bands. *Chromosome Res.* **6**, 77–83 (1998).
22. Hara, T. *et al.* Tandem repeat sequences evolutionarily related to SVA-type retrotransposons are expanded in the centromere region of the western hoolock gibbon, a small ape. *J. Hum. Genet.* **57**, 760–765 (2012). doi: 10.1038/jhg.2012.107.
23. Koga, A., Hirai, Y., Hara, T. & Hirai, H. Repetitive sequences originating from the centromere constitute large-scale heterochromatin in the telomere region in the siamang, a small ape. *Heredity* **109**, 180–187 (2012). doi: 10.1038/hdy.2012.28.
24. Tamura, K. *et al.* MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).

## Acknowledgements

We are grateful to Hiroshi Masumoto for helpful discussion and Yuki Enomoto for technical assistance. This work was supported by Grants-in-Aid (25650152 and 23470098 to A. K.) from the Japan Society for the Promotion of Science. P. S. was supported by the Royal Golden Jubilee Ph. D. Program of the Thai Government.

## Author Contributions

A. K. conceived and designed the experiments. P. S., W. T., and A. K. performed molecular biology experiments. Y. H. and H. H. conducted cytology experiments. K. S., H. H. and A. K. carried out data analysis.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Sujiwattanasat, P. *et al.* Higher-order repeat structure in alpha satellite DNA occurs in New World monkeys and is not confined to hominoids. *Sci. Rep.* **5**, 10315; doi: 10.1038/srep10315 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>