## SHORT COMMUNICATION

## Higher-order repeat structure in alpha satellite DNA is an attribute of hominoids rather than hominids

Shoko Terada, Yuriko Hirai, Hirohisa Hirai and Akihiko Koga

Alpha satellite DNA (AS), a major DNA component of primate centromeres, is composed of a tandem array of repeat units of approximately 170 bp. The AS of hominids (family Hominidae; humans and great apes) includes sequences organized into higher-order repeat (HOR) structures, with a periodic appearance of multiple copies of the basic repeat units. Here, we identified an HOR in AS of the siamang, a small ape phylogenetically distinct from hominids but included in hominoids (superfamily Hominoidea). We sequenced long stretches of genomic DNA, and found a repetition of blocks consisting of six and four basic repeat units. Thus, AS organization into HOR is an attribute of hominoids, rather than, as currently postulated, hominids. In addition to centromeres, siamangs carry AS in terminal heterochromatin blocks, and it cannot be determined at present whether these HOR-containing AS sequences originate from the centromere or from the terminal heterochromatin. Even if the latter is the case, these sequences might affect the composition of centromeric AS by being transferred to the centromere.

Journal of Human Genetics (2013) 58, 752-754; doi:10.1038/jhg.2013.87; published online 15 August 2013

Keywords: alpha satellite DNA; higher-order repeat; hominids; hominoids; tandem repeat

The satellite DNA now known as alpha satellite DNA (AS) was first described as a large-scale repetitive sequence in the African green monkey,<sup>1</sup> and was found to reside in the centromere.<sup>2</sup> AS is now known to be a major DNA component of centromeres in primates.<sup>3,4</sup> AS consists of tandem repeats of DNA sequences that are approximately 170 bp in length. Human AS consists of a large number of subfamilies, some of which are simple tandem repetitions of the basic repeat units, and others that are organized into higher-order repeats (HORs), where an HOR refers to a structure in which multiple copies of the basic repeat units appear periodically. AS containing HORs is known to be more important in regulating the centromere function.4,5 The chromosomespecific organization of AS was discovered approximately 30 years ago,<sup>6</sup> and chromosome-specific subfamilies are characterized by HOR structures as well as their monomer sequences. The present study deals with the evolutionary emergence of HORs in AS in the primate lineage.

HORs have also been observed in great apes, including the chimpanzee, gorilla and orangutan, with less variation observed in the periodicity within species, as compared with the variation observed in humans.<sup>4,7</sup> Humans and great apes belong to the family Hominidae (hominids). This family, along with the family Hylobatidae (gibbons; also called small apes), forms the superfamily Hominoidea (hominoids). Computational analyses of data generated

by shotgun sequencing of the Nomascus leucogenys (white-cheeked gibbon) genome suggested that HORs may be present in AS of members of the family Hylobatidae.<sup>8</sup> Subsequent experimental approaches by the same group, however, did not find direct evidence for the existence of HOR in AS of N. leucogenys, and the authors concluded that HORs are a peculiarity of hominids.<sup>9</sup> In the present study, we obtained direct evidence for the presence of HOR in AS of Symphalangus syndactylus (siamang), another species of the Hylobatidae family. We have previously reported that this species carries AS in the terminal regions of chromosomes, in addition to the centromeres.<sup>10</sup> To characterize AS, we had sequenced a 4101-bp region of a genomic DNA clone (GenBank accession number: AB678729).<sup>10</sup> Subsequent to the publication of this work, a re-analysis of this sequence (which contained 24 repeat units) with an alignment of the repeat units revealed a periodic variation in the nucleotide sequence. Because the analyzed sequence was not long enough to clarify the pattern of organization of the repeat units in detail, we sequenced another clone in the present study to obtain a longer sequence. We also intended to confirm the HOR presence in another clone.

The analysis of the pFosSia1 sequence suggested that its HOR is a mixture of two repeat intervals: an interval of four repeat units and one of six repeat units. For a nucleotide sequence of a block of 850  $(5 \times 170)$  consecutive nucleotides, if the block contains an HOR with

Primate Research Institute, Kyoto University, Inuyama, Japan

Correspondence: Dr A Koga, Primate Research Institute, Kyoto University, 41-2 Kanrin, Inuyama, Japan.

E-mail: koga.akihiko.5n@kyoto-u.ac.jp

Received 21 May 2013; revised 16 July 2013; accepted 24 July 2013; published online 15 August 2013

an interval of four repeat units, we can expect detection of a sign for this HOR by dot matrix analysis to compare the sequence with itself. One sequence assay of a fosmid clone, using a universal primer, provides sequences of 1000-1100 nucleotides, and the first 700-nucleotide region exhibits a significantly low frequency of sequencing error. In the next 200-nucleotide region, the error frequency is higher, but it still provides sequence information sufficient for a dot matrix analysis for signs of HOR. We cloned 19 additional fosmid clones (pSiaFos2 to 20) that exhibited strong signals by the method described in our previous study.<sup>10</sup> We then sequenced one end of the 19 clones, and 16 of them were found to contain AS. Dot matrix analysis of the respective sequences (Supplementary Figure 1) suggested that 3 (pFosSia7, 15, and 19) of the 16 sequences contain HORs with an interval of four repeat units. We selected one of these (pFosSia7) as a second clone to be sequenced.

The sequencing strategy used has been described previously,<sup>10</sup> and involves the preparation of deletion clones of different sizes, sequencing of these clones using a universal primer and assembly of sequence reads into a single stretch. However, we altered our protocol to use exonuclease III and mung bean nuclease<sup>11</sup> in place of restriction endonucleases, because the use of these enzymes permits more variety in the size of deletion clones, and leads to a higher efficiency in collecting sequencing samples. We sequenced a 9517-bp region of the pFosSia7 clone, deposited in GenBank with the accession number AB819921, and found 55 consecutive repeat units therein.

Sequence alignment of the 55 repeat units of the pFosSia7 clone is shown in Supplementary Figure 2. Sequence alignment of the 24 repeat units of the pFosSia1 clone has been previously published.<sup>10</sup> In both cases, a non-random distribution of variation is apparent at many points in the nucleotide sequence. Pairwise comparisons of the repeat units for sequence identities are shown in Supplementary Figure 3, in which cells representing identities of 90-95% and >95%are in yellow and red, respectively. In the figure, red cells form several step-like patterns, which are parallel to one another, indicating that multiple copies of the basic repeat units appear periodically. For further clarification of the HOR structure, we constructed a neighborjoining phylogenetic tree of the repeat units, assigned numbers to distinctive blocks, and examined the sequence of these numbers along the nucleotide sequences of the pSiaFos1 and pSiaFos7 clones (Figure 1). The most common patterns were '123456', '123478' and '1278', which we designated as  $\alpha$ ,  $\beta$  and  $\gamma$ , respectively. Thus, the sequenced AS contains an HOR structure, the most common repeat intervals being six and four. Repetition of a specific combination of the  $\alpha$ ,  $\beta$  and  $\gamma$  blocks was not observed within the present sequence data; this might appear if a further long region is sequenced.

The centromere protein B (CENP-B) box is a 17-bp sequence embedded in AS, and was first identified in human as an important signal for the centromere function.<sup>12</sup> CENP-B, a highly conserved centromere-associated protein, binds to this region. We examined the consensus sequence of pFosSia1 (Figure 4 of Koga *et al.*<sup>10</sup>) and that of pFosSia7 (Supplementary Figure 2 of the present study) for a CENP-B box, but we could not find it or a similar sequence block. However, this does not necessarily mean that the HOR-containing AS sequence is devoid of centromere function. The CENP-B box has been demonstrated to be present in AS of humans and great apes by hybridization experiments, but the same experiments did not detect it in gibbons or other primates examined.<sup>13</sup>

S. syndactylus carries large constitutive heterochromatin blocks in the terminal regions of its chromosomes,  $^{14}$  and these blocks are



**Figure 1** Structure of the higher-order repeat. (a) A neighbor-joining phylogenetic tree for the 79 repeat units (A01–A24 and B01–B55) was constructed using the MEGA5 program<sup>19</sup> under default settings. To the eight distinctive blocks indicated by green vertical bars, numbers (1–8) were assigned arbitrarily so that their sequence in **b** would become easy to recognize. (b) The block number sequences along the nucleotide sequences of the pSiaFos1 and pSiaFos7 clones are shown.  $\alpha$ ,  $\beta$  and  $\gamma$  are the three most common patterns of the block number appearance, formed as '123456', '123478' and '1278', respectively, and indicated by thin green bars.

composed mostly or solely of AS.<sup>10</sup> We performed fluorescence *in-situ* hybridization analysis with the intention to determine whether the HOR-containing AS originates from the centromere or from the terminal heterochromatin. The results from assays using pSiaFos1 as a probe have been shown in our previous reports.<sup>10,15</sup>

753

Hybridization signals appeared in both the centromere and telomere regions; we were therefore unable to determine the locations. Additional assays using pSiaFos7 and some other clones resulted in the same signal patterns (Supplementary Figure 4). Thus, it cannot be determined at present which region the HOR-containing AS originates from. Even if the origin is the telomere region, these sequences might affect the composition of centromeric AS by being transferred to the contromere. One possible form of transfer would be extrachromosomal circular DNA, which is known to often contain AS in human cells.<sup>16</sup> There is also an example of vigorous amplification of a tandem-repeat sequence integrated into the centromere in a gibbon.<sup>17</sup>

Apes of the Hylobatidae family can be divided into four genera: *Hoolock, Hylobates, Nomascus* and *Symphalangus*.<sup>18</sup> Of these, *Nomascus* and *Symphalangus* include species that have AS at chromosomal ends in addition to centromeres.<sup>9,10,15</sup> While studying *N. leucogenys*, Cellamare *et al.*<sup>9</sup> cloned numerous AS fragments from both locations and analyzed nucleotide sequences for HOR, and concluded that AS of this species does not have HOR. In the present study, we obtained evidence for HOR in AS of *S. syndactylus*. These contrasting results may be due to the differences in the species used, or due to differences in the principal methods employed (computer analysis of a collection of short sequences in the study by Cellamare *et al.*,<sup>9</sup> and traditional sequencing of long genomic clones in the present study).

It is widely postulated that HOR in AS is an attribute of hominids, but our results necessitate a modification of the current understanding: HOR is an attribute of hominoids.

## ACKNOWLEDGEMENTS

We are grateful to Hirakawa Zoo for providing a tissue sample of siamang through the Great Ape Information Network (GAIN) program. This study was supported by Grants-in-Aid (23657165 to AK, 23470098 to AK and 22247037 to HH) from the Japan Society for the Promotion of Science.

- Maio, J. J. DNA strand reassociation and polyribonucleotide binding in the African green monkey, *Cercopithecus aethiops. J. Mol. Biol.* 56, 579–595 (1971).
- Kurnit, D. M. & Maio, J. J. Subnuclear redistribution of DNA species in confluent and growing mammalian cells. *Chromosoma* 42, 23–36 (1973).
- Musich, P. R., Brown, F. L. & Maio, J. J. Highly repetitive component alpha and related alphoid DNAs in man and monkeys. *Chromosoma* 80, 331–348 (1980).
- Willard, H. F. Evolution of alpha satellite. *Curr. Opin. Genet. Dev.* 1, 509–514 (1991).
  Alexandrov, I., Kazakov, A., Tumeneva, I., Shepelev, V. & Yurov, Y. Alpha-satellite DNA
- of primates: old and new families. *Chromosoma* **110**, 253–266 (2001).
- 6 Willard, H. F. Chromosome-specific organization of human alpha satellite DNA. Am. J. Hum. Genet. 37, 524–532 (1985).
- 7 Haaf, T. & Willard, H. F. Orangutan alpha-satellite monomers are closely related to the human consensus sequence. *Mamm. Genome* 9, 440–447 (1998).
- 8 Alkan, C., Ventura, M., Archidiacono, N., Rocchi, M., Sahinalp, S. C. & Eichler, E. E. Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. *PLoS Comput. Biol.* **3**, e181 (2007).
- 9 Cellamare, A., Catacchio, C. R., Alkan, C., Giannuzzi, G., Antonacci, F., Cardone, M. F. et al. New insights into centromere organization and evolution from the white-cheeked gibbon and marmoset. *Mol. Biol. Evol.* 26, 1889–1900 (2009).
- 10 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).
- 11 Henikoff, S. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* 28, 351–359 (1984).
- 12 Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. & Okazaki, T. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J. Cell Biol. 109, 1963–1973 (1989).
- 13 Haaf, T., Mater, A. G., Wienberg, J. & Ward, D. C. Presence and abundance of CENP-B box sequences in great ape subsets of primate-specific alpha-satellite DNA. J. Mol. Evol. 41, 487–491 (1995).
- 14 Wijayanto, H., Hirai, Y., Kamanaka, Y., Katho, A., Sajuthi, D. & Hirai, H. Patterns of C-heterochromatin and telomeric DNA in two representative groups of small apes, the genera *Hylobates* and *Symphalangus. Chromosome Res.* **13**, 717–724 (2005).
- 15 Baicharoen, S., Arsaithamkul, V., Hirai, Y., Hara, T., Koga, A. & Hirai, H. In situ hybridization analysis of gibbon chromosomes suggests that amplification of alpha satellite DNA in the telomere region is confined to two of the four genera. *Genome* 55, 809–812 (2012).
- 16 Cohen, S., Agmon, N., Sobol, O. & Segal, D. Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mob. DNA* 1, 11 (2010).
- 17 Hara, T., Hirai, Y., Jahan, I., Hirai, H. & Koga, A. 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).
- 18 Brandon-Jones, D., Eudey, A. A., Geissmann, T., Groves, C. P., Melnick, D. J., Morales, J. C. et al. Asian primate classification. Int. J. Primatol. 25, 97–164 (2004).
- 19 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739 (2011).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)

754