ORIGINAL ARTICLE

Tandem repeat sequences evolutionarily related to SVA-type retrotransposons are expanded in the centromere region of the western hoolock gibbon, a small ape

Toru Hara, Yuriko Hirai, Israt Jahan, Hirohisa Hirai and Akihiko Koga

Hoolock hoolock (the western hoolock gibbon) is a species of the family Hylobatidae (small apes), which constitutes the superfamily Hominoidea (hominoids) together with Hominidae (great apes and human). Here, we report that centromeres or their vicinities in this gibbon species contain tandem repeat sequences that consist of 35–50-bp repeat units, and exhibit a sequence similarity with the variable number of tandem repeat (VNTR) region of the SVA, LAVA and PVA transposons. SVA is a composite retrotransposon thought to have been formed by fusion of three solo elements in the common ancestor of hominoids. LAVA and PVA are recently identified retrotransposons that have the same basic structure as SVA. Thus, the large-scale tandem repeats in the centromere region may have been derived from one or more of SVA-type transposons, including the three mentioned above and other yet unknown elements, or the repeat sequences could have served as a source for such elements. Amplification of VNTR-related sequences in another gibbon species, *Hoolock leuconedys* (eastern hoolock gibbon), has recently been reported, but it is yet to be examined whether the large-scale tandem repeats observed in the two species originated from a single event that occurred in their common ancestor. The repeat sequences in the western hoolock gibbon are mostly 40 kb or more in length, are present in 28 of the 38 chromosomes of the somatic cells, and are homozygous for chromosomal presence/absence.

INTRODUCTION

Centromeres and their vicinities, known as pericentromeric regions, typically contain large numbers of tandem repeat sequences that are packaged into heterochromatin. The most abundant component of human centromeres is alpha satellite DNA, as is the case in most or all primates,[1,2] which comprises tandem repeats of AT-rich units mainly 171 bp in length. Other tandem repeat sequences known to be present in the centromere regions of humans include satellite 1,[3] satellite 2,[4] beta satellite[5] and gamma satellite,[6] with typical repeat units of 42, 5, 68 and 220 bp, respectively. The origins of these repetitive sequences are mostly unknown, but it is noteworthy that some of them are not specific to centromere regions. For example, beta satellite is also present in the interstitial regions of some chromosomes.[7] Thus, one speculation about the origins is that any micro- or mini-satellite DNA that is located in the centromere region can possibly be amplified by innate centromeric mechanisms. The initial encounter of satellite DNA and a centromere may be the result of chromosomal reorganization, such as inversion and translocation, movement of a transposable element or virus or neocentromere formation at a place where repetitive sequences reside.

Comparative genomic hybridization (CGH) is an effective method for identifying the differences in the copy number of multicopy genes between strains (or species) or in transcript amounts between strains (or tissues). The target elements used in CGH experiments are usually oligonucleotides or cDNAs that represent a large number of genes. We modified this method using clones of large genomic DNA fragments as targets to identify DNA sequences that are highly repetitive in one species, but not in another. By applying this method to a gibbon (western hoolock gibbon *Hoolock hoolock*) and human, we found several clones that are highly repetitive only in the gibbon. Although our initial purpose was not directed to centromeres only (shown below), the obtained clones exhibited an interesting feature in relation to centromeres. On metaphase chromosome spreads, the clones produced strong hybridization signals in the centromere region, indicating that the repetitive sequences represented by these clones occupied substantial lengths in the gibbon centromere regions. The

clone exhibited a sequence similarity with the variable number of tandem repeat (VNTR) region of the SVA retrotransposon,[8] which was first identified in humans about 10 years ago, and the LAVA[9] and PVA[10] transposons which were recently identified in gibbons. In the present study, we characterized the newly identified repetitive sequences, and have discussed possible relationships between these sequences and SVA-type retrotransposons.

MATERIALS AND METHODS
Animals for collection of cells and DNA
We used animals belonging to the following five primate species: human (an adult male donor), chimpanzee (male, bred at Kyoto University), gorilla (male, bred at Kyoto City Zoo, Japan), western hoolock gibbon (female, bred at Bangabandhu Sheikh Mujib Safari Park, Bangladesh) and rhesus monkey (male, bred at Kyoto University).

Experiments involving DNA manipulations
A genomic library of the western hoolock gibbon was constructed, as described previously.[11] The vector was the 8.1-kb fosmid pCC1FOS and the insert was 40–44 kb of genomic DNA fragments that had been generated by mechanical shearing and isolated by gel electrophoresis and subsequent recovery from a gel piece. This library was screened by the modified CGH technique[8] for highly repetitive sequences. Other regular DNA manipulation experiments, such as cloning, sequencing and Southern hybridization, were conducted as described previously.[12–14] Fluorescent *in situ* hybridization (FISH) analysis of chromosomes was performed following the procedures described previously.[15,16] Specific conditions are explained in each case.

RESULTS
Cloning of highly repetitive sequences
Gibbons are known to have undergone frequent chromosomal reorganizations. For our initial purpose of elucidating the mechanisms that lead to frequent chromosomal reorganizations, we conducted experiments to identify DNA sequences that were highly repetitive in the genome of a gibbon, but not in that of a human. One of such sequences identified was a long tandem repeat of the western hoolock gibbon that exhibited a sequence similarity with the VNTR region of the SVA-type transposons (SVA, LAVA and PVA).

We first constructed the genomic library of the gibbon. Second, we spread, on agar plates, bacteria containing recombinant fosmids from the library and performed colony hybridization. We then picked up several colonies that exhibited relatively strong signals (Figure 1, upper panel). The probe used for this screening was genomic DNA of the gibbon. Strong signals therefore imply that the corresponding colonies contained DNA fragments that were highly repetitive in the gibbon genome. We then performed a secondary screening for clones exhibiting strong signals against the gibbon probe but weak or no signals against a human genomic DNA probe (Figure 1, lower panel). We obtained 12 such clones, starting with ∼4000 colonies for the initial screening. The 12 fosmid clones were designated pFosHho1–pFosHho12 (Fos for fosmid, Hho for *Hoolock hoolock*).

Identification of tandem repeat sequences
We determined the sequences of the terminal regions (500–800 nucleotides each) of the 12 clones. All 24 sequence reads were found to contain repetitive sequences consisting of 35–50-bp repeat units. We compared, by dot matrix analysis, the 24 sequence reads with the sequence of the VNTR region of a human SVA element. The results were essentially the same among the 24 sequence reads. Tandem repeat structures were clearly observed in the gibbon sequences as well as in the human sequence, and comparison between the species showed that their repeat structures shared similarities with each other.
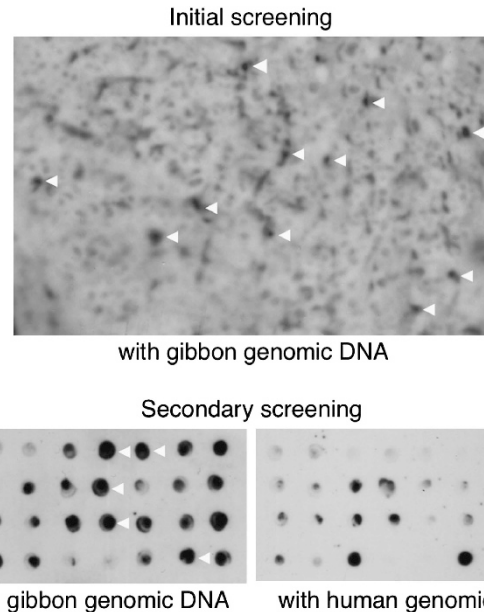
Initial screening

with gibbon genomic DNA

Secondary screening

with gibbon genomic DNA   with human genomic DNA

Figure 1 Detection of clones that are highly repetitive in the gibbon genome but not in the human genome. Approximately 4000 colonies from the gibbon genomic library were grown on agar plates, and colony hybridization was conducted as an initial screening. The probe used was gibbon genomic DNA labeled with alkaline phosphatase. Part of the autoradiogram obtained is shown here. Relatively strong signals (indicated by white arrowheads) were selected, and colonies responsible for these signals were picked up. For a secondary screening, the bacterial colonies collected were cultured separately in 96-well culture plates. Two nylon membranes were blotted with these bacterial cultures in duplicate. One membrane was hybridized with gibbon genomic DNA as a probe. Part of the autoradiogram obtained is shown (left panel). The other membrane was hybridized with human genomic DNA (right panel). Colonies that produced strong signals only when probed with gibbon genomic DNA (indicated by white arrowheads in the left panel) were selected.

We termed the newly found repetitive sequences of the gibbon as HhoRep (Rep for repeats). Figure 2 shows the results of comparison in which a longer HhoRep sequence (2.5-kb restriction fragment explained below; deposited in GenBank with accession number AB698821) was used. These results suggested that the complete insert portions (40–44 kb) of the gibbon clones were HhoRep sequences. We examined whether this was in fact true, by sequencing several different portions in one (pFosHho1) of the twelve clones. The pFosHho1 clone contained 10 recognition sites for restriction endonuclease *Sac*I. We cloned, into plasmid DNA, fragments generated by *Sac*I digestion of pFosHho1, and sequenced their terminal regions. We thereby obtained a total of 10 different sequence reads, and they all showed dot matrix patterns similar to those in Figure 2. This does not necessarily mean that the insert portion of the pFosHho1 clone consists only of HhoRep sequences, but does indicate that the major component of the insert portion is HhoRep. Thus, the gibbon genome contains one or more DNA regions that are 40 kb in length or longer, and consists mostly, or possibly solely, of HhoRep sequences.

Consensus sequences
We performed a quantitative analysis of the human VNTR sequence and the gibbon HhoRep sequence by comparing their consensus sequences, which were drawn by partitioning the entire sequences into
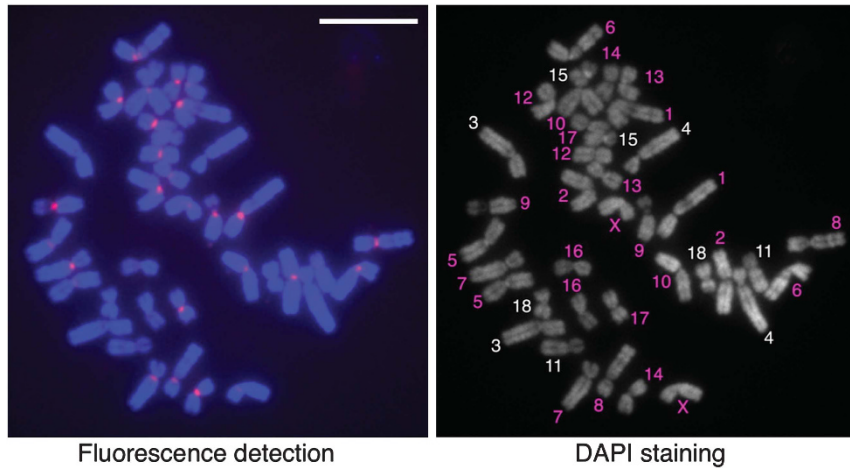
Figure 2 Dot matrix analysis of the VNTR and HhoRep sequences. 'VNTR' is the VNTR part of a relatively long SVA element that we chose from the human genome browser via the RepeatMasker program (1045 bp; chromosome 17: 26987696–26988741). 'HhoRep' is the sequence that we obtained by sequencing analysis of a 2.5-kb fragment that was generated by *Sac*I digestion of pFosHho1. These sequences were compared within and between species by dot matrices. The criterion for matching was a 70% match over a window of 10 nucleotides.

repeat units by the Tandem Repeats Finder program, (http://tandem.bu.edu/trf/trf.html),[17] and then aligning the units by the ClustalW2 program (http://www.ebi.ac.uk/Tools/msa/clustalw2/),[18] both with default settings. As shown in Figure 3, the consensus sequence lengths were 37 and 39 bp in VNTR and HhoRep, respectively, and the nucleotide identity (excluding the vacant VNTR sites) was 97% (36/37). These results, along with those of the dot matrix analysis (Figure 2), can be regarded as evidence that the two sequences originated from a common ancestor.

Chromosomal locations of HhoRep sequences

We conducted FISH analysis of gibbon chromosomes to determine the locations of HhoRep sequences, using pFosHho1 as the probe. The result was surprising in that strong signals were observed in centromere regions. Because the possibility that pFosHho1 contains sequences other than HhoRep could not be excluded, we conducted the analysis again using a smaller probe that had been confirmed to contain HhoRep only. The probe that was used the second time was a plasmid subclone of a 2.5-kb *Sac*I-restriction fragment from pFosHho1 (the clone used for comparison in Figures 2 and 3; GenBank accession number AB698821). We designated this probe ProHho. The FISH result obtained (Figure 4) was the same as that with the pFosHho1 probe: strong signals in the centromere regions of 28 chromosomes. The chromosome spread preparations were derived from white blood cells and somatic cells containing a total of 38

```
VNTR   CCTCTGCCCGGCCGCCC--CGTCTGGGAAGTGAGGAGCG
       ||||||||||||| ||| |||||||||||||||||||
HhoRep CCTCTGCCCGGCCACCCATCGTCTGGGAAGTGAGGAGCG
```

Figure 3 Consensus sequences of the VNTR and HhoRep repeat units. The respective consensus sequences are aligned. The vertical bar indicates a site where nucleotides are identical between the two sequences. The minus sign implies that there was no corresponding site in the consensus sequence.

chromosomes. Each chromosome can be identified by the length, shape and banding pattern,[19] and the chromosome numbers of all chromosomes are also shown in Figure 4. This chromosome identification revealed that the presence/absence of the signals was homozygous for all chromosomes. For example, both sister chromosomes of chromosome 2 exhibited signals, whereas both sister chromosomes of chromosome 3 were devoid of signals.

Comparison of sequence abundance among species

We conducted Southern blot analysis to compare the abundance of HhoRep/VNTR sequences among species. Prior to the analysis, we prepared an additional probe that contained a VNTR sequence from human genomic DNA, because there was a possibility that a slight sequence difference between humans and the gibbon might affect the intensity of signals, such as producing a stronger signal with its own probe. We conducted PCR against human genomic DNA with

Figure 4 FISH analysis of chromosomes to determine the HhoRep sequence locations. The 2.5-kb *Sac*I fragment (GenBank accession number AB698821) was labeled and used as probe. The left panel is an image of fluorescence detection. Strong signals were observed at the centromeres of 28 chromosomes, and not observed on 10 chromosomes. The bar represents 10 μm. The right panel represents DAPI staining of the same chromosome spread. The chromosomes producing signals are labeled with the chromosome numbers in red, and the chromosomes exhibiting no signals in white.
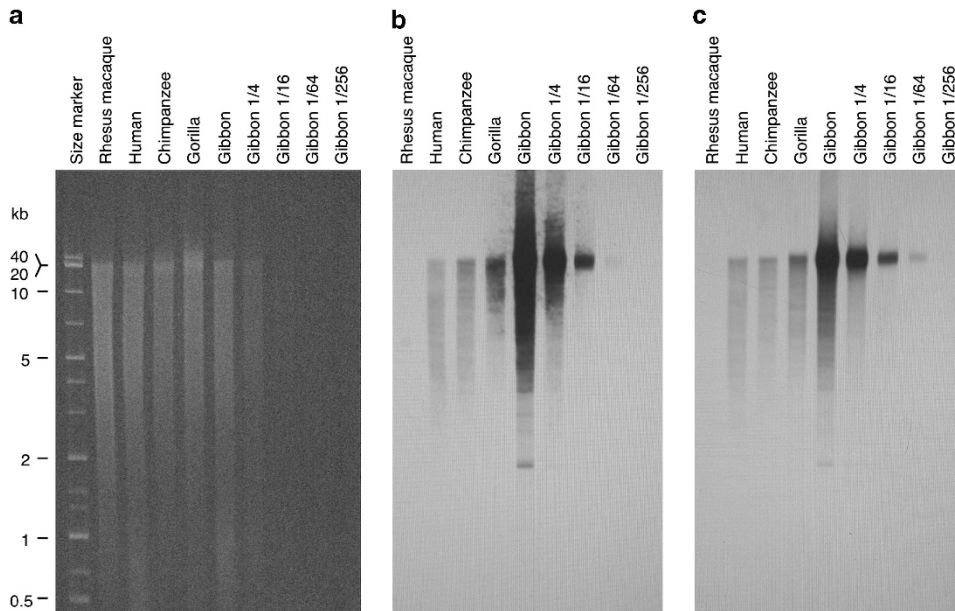
Figure 5 Southern hybridization analysis to compare the abundance of VNTR sequences among primate species. Genomic DNA of the five primate species indicated above the lanes was digested with excess amounts of the restriction endonuclease *Bgl*II (20 units for 1 μg), and 400 ng of each (unless otherwise noted) was applied to gel slots. The second to fifth lanes of the gibbon contained DNAs that corresponded to the indicated fraction of 400 ng of gibbon DNA. Two gels were prepared, and electrophoresed DNAs were transferred to nylon membranes. One membrane (**b**) was hybridized with ProHho (containing the gibbon HhoRep), and the other (**c**) with ProHum (containing the human VNTR). (**a**) A photograph of the gel from which DNA was transferred to the membrane shown in (**b**). The sizes of the marker DNA fragments are indicated along the left margin. The two fragments of sizes 20 and 40 kb overlapped with each other. A faint fragment exhibiting a slightly lower mobility than the 40-kb fragment was a BAC clone whose size had been estimated to be larger than 100 kb but has not been determined accurately.

primers just adjacent to the VNTR region of a human SVA element (nucleotides 333–362 and 1501–1472 of GenBank accession number L09706), and cloned a DNA fragment of the PCR product into a plasmid. This probe was designated ProHum.

Figure 5a shows the gel after electrophoresis and ethidium bromide staining of the DNA. There was no significant difference in the DNA amount among the five species used (except for the four lanes containing diluted gibbon DNA samples). In addition, among the five

species, there were no significant differences in the within-lane distribution pattern of DNA fragments, indicating that the DNAs had been digested to almost the same extent with the restriction enzyme *Bgl*II. This can be regarded as a complete digestion because we used excess units of the restriction enzyme. Figures 5b and c show the autoradiograms of hybridization with ProHum and ProHho, respectively. The signal patterns obtained using the two probes were similar, excluding the aforementioned possibility. The signal intensity

was not very different among the three hominid species, and the gibbon showed a more intense signal than the hominids. The signal intensity in the lane for a fourfold lower amount of gibbon DNA was stronger than that in the lane for human DNA, and that in the lane for a 16-fold lower amount of gibbon DNA was almost equal or weaker. If we assume that there is no significant difference in the genome size between the human and gibbon, this result indicates that the number of HhoRep sequences in the gibbon genome is roughly 10 times larger than the number of VNTR sequences in the human genome.

On the autoradiograms of Figures 5b and c, a significant difference in the size distribution of signal-producing fragments was observed between the gibbon and the three hominid species, as the gibbon peak size was much larger. This was consistent with our inference that HhoRep sequences are longer than the VNTR regions in SVA elements. The restriction enzyme BglII recognizes six consecutive nucleotides (AGATCT), and the expected average fragment size of completely digested DNA is $\sim 4.1\,\mathrm{kb}$ (4^6 bp) on the assumption of a random array of equal frequencies (25% each) of the four nucleotides and no methylation status effects. The consensus sequences (Figure 3) do not contain AGATCT or slightly different six nucleotide blocks. Thus, it is expected that the majority of BglII-digested fragments exhibiting signals have breakpoints not in the repeat region but rather in the flanking regions. Because the average size of human SVA elements has been estimated to be 0.8 kb,[20] the expected average size of signal-producing fragments is 4.9 kb (4.1 + 0.8 kb). The signal distribution patterns in the three hominid species are consistent with this expectation. In case of the gibbon HhoRep sequence, the majority of the signals were located at or around the position of the 40-kb size marker fragment. This is consistent with the results of our cloning and sequencing analyses (of HhoRep sequences at both ends of all the 12 clones examined).

DISCUSSION

The main findings of this study were as follows: (1) the genome of the western hoolock gibbon contains DNA regions, designated HhoRep, that share a sequence similarity with the VNTR region of the SVA-type transposons; (2) the lengths of the HhoRep sequences are more than 40 kb; (3) the HhoRep sequences are located in the centromere regions of 28 of the 38 chromosomes; (4) all HhoRep sequences are homozygous; and (5) the total number of HhoRep sequences is roughly 10 times larger than that of VNTRs in the human genome.

Long VNTR-related sequences in the centromere region have recently been reported in the eastern hoolock gibbon.[9] We have, however, independently identified the HhoRep sequences in the centromere region of the western hoolock gibbon, as evidenced by the registration date of GenBank AB698821. The differences in the main methods are of interest: those authors performed FISH analysis of chromosomes, whereas we conducted CGH experiments.

From the results of dot matrix analysis and comparison of consensus sequences, it is evident that the HhoRep sequences and VNTR region of the SVA-type transposons shared a common evolutionary origin. Three processes regarding the generation of these sequences can be postulated: (a) the common ancestor was neither in the centromere region nor in the SVA-type transposons, and HhoRep and the SVA-type transposons were derived independently from this common origin; (b) the SVA-type transposons retained the ancestral form, and HhoRep was derived from the SVA-type transposons; and (c) HhoRep retained the ancestral form and the SVA-type transposons were derived from HhoRep. In evolutionary biology, the number of events required to explain the current situation is often regarded as a

key factor; the smaller the number of events, the more likely the scenario. From this viewpoint, (a) is more difficult to support than (b) or (c). Figure 6 depicts the three scenarios with minimum numbers of events on evolutionary branches. Scenario (a) requires at least four events.

Scenario (c) requires at least two events, in which the second event required is extinction of HhoRep from all centromeres. The results of the FISH analysis appear to be evidence against the occurrence of such an event. All HhoRep sequences were shown to be homozygous for the presence/absence. This situation indicates that neither a gain of a new HhoRep sequence nor a loss of an existent HhoRep sequence has taken place, as the situation of the 14 homozygous sets arose in the gibbon lineage; otherwise one or more heterozygous (in a strict sense, hemizygous) HhoRep sequences are expected to be present. Thus, the extinction of the HhoRep sequence would be unlikely to occur even on a single chromosome, and therefore the extinction from all chromosomes would be even more unlikely. If scenario (c) is true, it may lead to new insights into the formation process of the SVA-type transposons. One suggested mechanism for VNTR acquisition by Alu is the encounter of SVA2 (or its ancestral element) and Alu, and subsequent mRNA splicing,[21] where SVA2 is a dispersed element consisting of VNTR and other sequences. The total length of HhoRep sequences is likely to far exceed that of SVA2s. Therefore, if the first encounter is an Alu transposition, it is expected that transposition into HhoRep or its vicinities would be more frequent than transposition into SVA2 or its vicinities.

Scenario (b) requires HhoRep formation (elongation of a VNTR sequence) in the gibbon lineage. If this is true, there may be the head and tail regions of an SVA-type transposon adjacent to HhoRep.
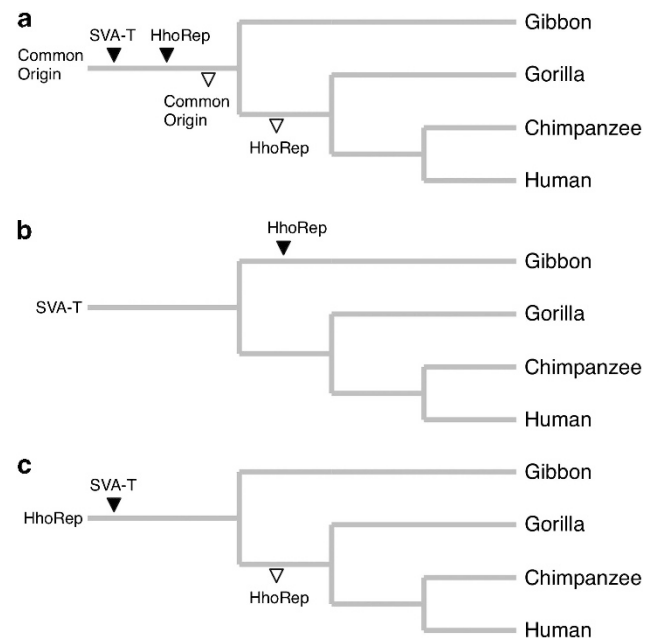

Figure 6 Possible scenarios to explain the current distribution of HhoRep sequences. 'SVA-T' indicates an SVA-type transposon. The assumptions are as follows: (a) HhoRep and SVA-T originate independently from an element of another form, (b) HhoRep was derived from SVA-T, and (c) SVA-T was derived from HhoRep. Black and white triangles indicate the generation and extinction of sequences, respectively. In each case, the scenario that requires the minimum number of events is shown. Other scenarios that involve more events are possible. In (a), the generation of SVA-T and HhoRep is interchangeable.

Detection of such a linkage would be a sufficient condition for scenario (b), but it is not a necessary condition because deletion of the head and/or tail region may occur after the integration of the transposon into the centromere region. If scenario (b) is true, it may be possible that an event similar to the HhoRep formation could also occur in humans, because humans have numerous SVA elements scattered throughout the genome. SVA transposition is not the only possible mechanism for the initial encounter of SVA and the centromere. Chromosome reorganization and neocentromere formation are also candidate mechanisms.

ACKNOWLEDGEMENTS

1 Rudd, M. K., Wray, G. A. & Willard, H. F. The evolutionary dynamics of alpha-satellite. *Genome Res.* **16,** 88–96 (2006).
2 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,** 1807–1818 (2007).
3 Kalitsis, P., Earle, E., Vissel, B., Shaffer, L. G. & Choo, K. H. A chromosome 13-specific human satellite I DNA subfamily with minor presence on chromosome 21: further studies on Robertsonian translocations. *Genomics* **16,** 104–112 (1993).
4 Grady, D. L., Ratliff, R. L., Robinson, D. L., McCanlies, E. C., Meyne, J. & Moyzis, R. K. Highly conserved repetitive DNA sequences are present at human centromeres. *Proc. Natl Acad. Sci. USA* **89,** 1695–1699 (1992).
5 Meneveri, R., Agresti, A., Della Valle, G., Talarico, D., Siccardi, A. G. & Ginelli, E. Identification of a human clustered G+C-rich DNA family of repeats (Sau3A family). *J Mol. Biol.* **186,** 483–489 (1985).
6 Lee, C., Li, X., Jabs, E. W., Court, D. & Lin, C. C. Human gamma X satellite DNA: an X chromosome specific centromeric DNA sequence. *Chromosoma* **104,** 103–112 (1995).
7 Hirai, H., Taguchi, T. & Godwin, A. K. Genomic differentiation of 18S ribosomal DNA and beta-satellite DNA in the hominoid and its evolutionary aspects. *Chromosome Res.* **7,** 531–540 (1999).
8 Ostertag, E. M., Goodier, J. L., Zhang, Y. & Kazazian, Jr H. H. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* **73,** 1444–1451 (2003).
9 Carbone, L., Harris, R. A., Mootnick, A. R., Milosavljevic, A., Martin, D. I., Rocchi, M. *et al.* Centromere remodeling in *Hoolock leuconedys* (Hylobatidae) by a new transposable element unique to the gibbons. *Genome Biol. Evol.* **4,** 648–658 (2012).
10 Hara, T., Hirai, Y., Baicharoen, S., Hayakawa, T., Hirai, H. & Koga, A. A novel composite retrotransposon derived from or generated independently of the SVA (SINE/VNTR/*Alu*) transposon has undergone proliferation in gibbon genomes. *Genes Genet. Syst.* **87,** 181–190 (2012).
11 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).
12 Koga, A., Shimada, A., Kuroki, T., Hori, H., Kusumi, J., Kyono-Hamaguchi, Y. *et al.* The *Tol1* transposable element of the medaka fish moves in human and mouse cells. *J. Hum. Genet.* **52,** 628–635 (2007).
13 Koga, A., Higashide, I., Hori, H., Wakamatsu, Y., Kyono-Hamaguchi, Y. & Hamaguchi, S. The *Tol1* element of medaka fish is transposed with only terminal regions and can deliver large DNA fragments into the chromosomes. *J. Hum. Genet.* **52,** 1026–1030 (2007).
14 Koga, A., Notohara, M. & Hirai, H. Evolution of subterminal satellite (StSat) repeats in hominids. *Genetica* **139,** 167–175 (2011).
15 Hirai, H., Hirai, Y., Kawamoto, Y., Endo, H., Kimura, J. & Rerkamnuaychoke, W. Cytogenetic differentiation of two sympatric tree shrew taxa found in the southern part of the Isthmus of Kra. *Chromosome Res.* **10,** 313–327 (2002).
16 Hirai, H., Matsubayashi, K., Kumazaki, K., Kato, A., Maeda, N. & Kim, H. S. Chimpanzee chromosomes: retrotransposable compound repeat DNA organization (RCRO) and its influence on meiotic prophase and crossing-over. *Cytogenet. Genome Res.* **108,** 248–254 (2005).
17 Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27,** 573–580 (1999).
18 Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGetligan, P. A., McWilliam, H. *et al.* ClustalW and ClustalX version 2. *Bioinformatics* **23,** 2947–2948 (2007).
19 Nie, W., Rens, W., Wang, J. & Yang, F. Conserved chromosome segments in *Hylobates hoolock* revealed by human and *H. leucogenys* paint probes. *Cytogenet. Cell Genet.* **92,** 248–253 (2001).
20 Wang, H., Xing, J., Grover, D., Hedges, D. J., Han, K., Walker, J. A. *et al.* SVA elements: a hominid-specific retroposon family. *J. Mol. Biol.* **354,** 994–1007 (2005).
21 Hancks, D. C. & Kazazian, Jr. H. H. SVA retrotransposons: evolution and genetic instability. *Semin. Cancer Biol.* **20,** 234–245 (2010).