Evolutionary Origin of OwlRep, a Megasatellite DNA Associated with Adaptation of Owl Monkeys to Nocturnal Lifestyle

Hidenori Nishihara[1], Roscoe Stanyon[2], Junko Kusumi[3], Hirohisa Hirai[4], and Akihiko Koga[4,*]

[1]Department of Life Science and Technology, Tokyo Institute of Technology, Yokohama, Japan

[2]Department of Biology, University of Florence, Italy

[3]Faculty of Social and Cultural Studies, Kyushu University, Fukuoka, Japan

[4]Primate Research Institute, Kyoto University, Inuyama, Japan

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

Accepted: December 23, 2017

Abstract

Rod cells of many nocturnal mammals have a "non-standard" nuclear architecture, which is called the inverted nuclear architecture. Heterochromatin localizes to the central region of the nucleus. This leads to an efficient light transmission to the outer segments of photoreceptors. Rod cells of diurnal mammals have the conventional nuclear architecture. Owl monkeys (genus *Aotus*) are the only taxon of simian primates that has a nocturnal or cathemeral lifestyle, and this adaptation is widely thought to be secondary. Their rod cells were shown to exhibit an intermediate chromatin distribution: a spherical heterochromatin block was found in the central region of the nucleus although it was less complete than that of typical nocturnal mammals. We recently demonstrated that the primary DNA component of this heterochromatin block was OwlRep, a megasatellite DNA consisting of 187-bp-long repeat units. However, the origin of OwlRep was not known. Here we show that OwlRep was derived from HSAT6, a simple repeat sequence found in the centromere regions of human chromosomes. HSAT6 occurs widely in primates, suggesting that it was already present in the last common ancestor of extant primates. Notably, Strepsirrhini and Tarsiformes apparently carry a single HSAT6 copy, whereas many species of Simiiformes contain multiple copies. Comparison of nucleotide sequences of these copies revealed the entire process of the OwlRep formation. HSAT6, with or without flanking sequences, was segmentally duplicated in New World monkeys. Then, in the owl monkey linage after its divergence from other New World monkeys, a copy of HSAT6 was tandemly amplified, eventually forming a megasatellite DNA.

Key words: primate, night vision, rod cell, duplication, repetitive DNA.

Introduction

Heterochromatin serves as a structural component of centromeres, telomeres, and associated (nearby) regions (Schoeftner and Blasco 2009; Steiner and Henikoff 2015). Heterochromatin is also known to play a role in controlling gene expression, DNA recombination, nuclear organization, and various other molecular biological processes in cells (Grewal and Jia 2007; Jost et al. 2012). However, there are few straight forward examples of how heterochromatin directly contributes to a specific phenotype and promotes organismal adaptation. One rare example is the role of large-scale heterochromatin in the adaptation of night vision in nocturnal mammals. Here heterochromatin takes part in the formation of a lens-like structure in the nucleus of rod cells (Solovei et al. 2009, 2013; Eberhart et al. 2013; Joffe et al. 2014). Rods and cones are photoreceptor cells of vertebrate eyes. These cells are elongated in shape and light passes through the nucleus before reaching the outer segments of the retina where it is absorbed by opsins. In rod cells of many nocturnal mammals, heterochromatin is localized in the central region of the nucleus and functions like a lens to send light efficiently to the outer segments (Solovei et al. 2009, 2013). This chromatin distribution in the internal space of the nucleus is called the "inverted nuclear architecture". In contrast, in the

"conventional nuclear architecture", heterochromatin is distributed mainly in the internal periphery of the nucleus. All cells of diurnal mammals and cells other than rod cells of nocturnal mammals normally exhibit the conventional nuclear architecture (Solovei et al. 2009).

The order Primates consists of two suborders, Strepsirrhini and Haplorrhini. Strepsirrhini contains lemurs, lorises, and galagos. Although extant lorises and galagos are all nocturnal, lemurs include nocturnal, diurnal, and cathemeral species. Analyses of activity patterns by the Bayesian phylogenetic methods suggested multiple, independent occurrences of shifts in the activity pattern among nocturnality, diurnality, and cathemerality in the lemur phylogeny (Griffin et al. 2012; Santini et al. 2015). Haplorrhini consists of tarsiers and simians. Tarsiers are nocturnal but simians are almost all diurnal. It is widely thought that the last common ancestor of simians was diurnal (Kay et al. 1997; Heesy and Ross 2001), and this is consistent with the results from the above-mentioned Bayesian phylogenetic approach (Griffin et al. 2012; Santini et al. 2015), which suggested a transition from nocturnality to diurnality at the root of the simian radiation. An exception among simian primates is owl monkeys (genus *Aotus*) (Ross and Hylander 2000; Heesy and Ross 2001), which inhabit tropical forest areas in Central and South America. This genus consists of nocturnal and cathemeral species (Dyer et al. 2009; Fleagle 2012). The Bayesian phylogenetic analyses (Griffin et al. 2012; Santini et al. 2015) suggested that their last common ancestor was nocturnal. Later, some species expanded their activity pattern to include extra day-time activity (Santini et al. 2015) to the point that such species are now classified as cathemeral. However, even so-called cathemeral species are more active at night than during daytime (Fernández-Duque et al. 2010).

Rod cells of owl monkeys were previously shown to have an intermediate nuclear architecture: a spherical heterochromatin block is formed in the central region of the nucleus on a background of the conventional nuclear architecture (Joffe et al. 2014). Thus, their inverted nuclear architecture appears immature when compared with that of typical nocturnal mammals such as mice, cats, and opossums (Solovei et al. 2009; Joffe et al. 2014). Apparently, owl monkey rod cells are in an early stage of a shift from the conventional to the inverted nuclear architecture. Owl monkeys are estimated to have diverged from other New World monkeys approximately 20 Mya (Perelman et al. 2011; Schneider and Sampaio 2013), and this time span is perhaps insufficient to establish the full inverted nuclear architecture. This situation is fortunate because it means that owl monkeys provide an ideal model for elucidating the genetic and evolutionary factors that interplay in the initial phases of adaptation to a nocturnal lifestyle. Thus, owl monkeys are a unique taxon for understanding adaptation to night vision because we know of no other species which has acquired nocturnality in as short a time span as the last 20 Myr. We have recently shown that the primary DNA component of the central heterochromatin block found in owl monkey rod cells is OwlRep, which is a satellite DNA present in large amounts in the owl monkey genome (Koga et al. 2017). OwlRep contains 187-bp-long repeat units, and consists of the entire short arms of acrocentric chromosomes (Prakhongcheep, Chaiprasertsri, et al. 2013). Our extensive BLAST searches, and Southern blot analyses, did not find similar sequences in organisms other than owl monkeys (Prakhongcheep, Chaiprasertsri, et al. 2013). OwlRep provides a unique example of a satellite DNA that has a clear biological function directly related to organismal adaptation: the primary component of the lens-like structures. However, two important, related questions remained unanswered. From what sequence did OwlRep originate and what were the processes by which the original sequence was amplified? Unfortunately, because of lack of information on similar or related sequences, this question was left unsolved.

Therefore, in the present study one goal was to determine the evolutionary origin of OwlRep. We conducted detailed analyses of genome sequence databases for similar sequences, accompanied by experimental verifications, and found that a simple repeat sequence called HSAT6 shares multiple features with OwlRep. HSAT6 was first described in humans. We found that sequences similar in sequence and structure to HSAT6 occur in a wide range of primates, including lemurs and galagos that diverged from humans and other Haplorhini primates 60–70 Mya (Springer et al. 2012; Finstermeier et al. 2013; Pozzi et al. 2014). In this report, we call these sequences HSAT6. Our main conclusion is that HSAT6 gave origin to OwlRep: an HSAT6 sequence was amplified tandemly and developed into a megasatellite DNA in the owl monkey linage after its divergence from other New World monkeys.

Materials and Methods

Data Sources for Repeat Annotation

RepeatMasker annotation was referred to search HSAT6 elements in the genomes of 17 primates. The RepeatMasker output files with the repeat library (ver. 20140131) were obtained from the RepeatMasker website for the following species (http://www.repeatmasker.org/genomicDatasets/RMGenomicDatasets.html; last accessed January 3, 2018): human (hg18), orangutan (ponAbe2), gibbon (nomLeu3), crab-eating macaque (macFas5), marmoset (calJac3), and bush baby (otoGar3), tree shrew (tupBel1), and mouse (mm10). In addition, the RepeatMasker output files were obtained from the UCSC Genome Browser database for the latest genome version of the following species: chimpanzee (panTro5), bonobo (panPan1), gorilla (gorGor5), rhesus (rheMac8), baboon (papAnu2), tarsier (tarSyr2), squirrel monkey (saiBol1; The Broad Institute), mouse lemur (micMur2), and Malayan flying lemur (galVar1; The Genome Institute at Washington University).

Owl Monkey Genome Sequences

The owl monkey genome (*Aotus nancymaae*), released from the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC; https://www.hgsc.bcm.edu/non-human-primates/owl-monkey-genome-project; last accessed January 3, 2018), was obtained from the NCBI Genome database. The MSTA + HSAT6 sequences were identified by a local Nucleotide BLAST search ($r = 2$, $G = 5$, $E = 2$, *e*-value cutoff of 1×10^{-10}) using the corresponding marmoset sequence as a query.

Phylogenetic Analysis

The MSTA + HSAT6 sequences (corresponding to such as chr16: 32232593–32233266 in human [hg38]) and the AluSc8 + MSTA + HSAT6 sequences (corresponding to such as the human chr16: 32230840–32233266) were separately used for the phylogenetic analyses. Based on the RepeatMasker annotation, the homologous sequences were obtained through the UCSC Table Browser. The sequences were aligned by using MAFFT 7.305 with the settings of –maxiterate 1000 and –localpair (Katoh and Standley 2013). The specific insertion of other transposable elements (i.e., denoted as grey arrows in fig. 2*A*) was removed from the alignment before phylogenetic reconstruction. Maximum likelihood trees were constructed using RAxML 8.1.16 (Stamatakis 2014) with the GTR + CAT model and with 1000 bootstrap replicates.

Cloning and Sequencing

To verify the sequence data derived from databases, we cloned DNA fragments by PCR from genomic DNA and sequenced these clones. We first selected pairs of 20-bp-long segments that were assumed to be unique to the respective HSAT6 sequences. Using oligomers representing these sequences as primers, we conducted PCR amplification from genomic DNA. The PCR conditions were: [94°C, 4 min], 36 cycles of [98°C, 10 s; 64°C, 15 s; 68°C, 2 min], and then [94°C, 2 min]. The product fragments were considered to be mixture of fragments originating from two alleles, and the two alleles could contain polymorphic nucleotide sites, which might result in ambiguity in sequence data to be obtained. To avoid this, we inserted the product fragments into a plasmid vector (pUC118) at its *Hin*cII digestion site, and selected a single plasmid clone that represented one of the two alleles. We then sequenced these plasmid clones, by using sequencing primers that represented part of the vector sequence close to the vector-insert breakpoints.

Genomic Southern Blot Analysis

The OwlRep consensus sequence carries a cutting site for the restriction enzyme *Nsp*I (RCATGY). Genomic DNA (10 μg) of human, marmoset, *A. azarae* and *A. nancymaae* was digested with *Nsp*I to a completion, run on a 1.4% agarose gel, and transferred to a nylon membrane. The entire samples (10 μg) of human and marmoset were loaded on gel slots. For *A. azarae* and *A. nancymaae*, the digested DNA was diluted with water and portions equivalent to 100 and 20 ng were loaded on gel slots. These samples, fixed on the membrane by UV irradiation, were hybridized with the OwlRep probe that had been chemically labeled. The reagent system for probe labeling, hybridization, and signal detection, and the experimental conditions for its use were the same as those previously described (Prakhongcheep et al. 2013; Prakhongcheep, Chaiprasertsri, et al. 2013).

Results

High Similarity between OwlRep and HSAT6 Satellite DNA

We first assumed that OwlRep was derived from another repetitive DNA in primates. The 187 bp consensus sequence of OwlRep (Prakhongcheep, Chaiprasertsri, et al. 2013) was compared with a collection of consensus sequences of all known repetitive elements in primates by using RepeatMasker. We found that HSAT6, a 126 bp-long human centromeric satellite DNA represents 82.5% sequence identity (104/126 bp) with OwlRep (fig. 1*A*). There are 22 different nucleotides between OwlRep and HSAT6 (fig. 1*A*), and the transition/transversion ratio of the 22 sites is 2.14 (15/7) which is significantly higher than 0.5 under an unbiased sequence comparison ($P < 0.05$, Fisher's exact test). The remaining 60 bp of OwlRep did not show any meaningful similarity with HSAT6 nor other repetitive elements in primates.

The sequence structure was also compared between HSAT6 and OwlRep (fig. 1*B*). HSAT6 consists of three tandem repeats of 42 bp units, and each of them contains a 13 bp direct repeat and a 17 bp partial palindrome (blue and yellow arrows in fig. 1*B*, respectively). Correspondingly, OwlRep also consists of three units containing a direct repeat and partial palindrome (fig. 1*B*). The high similarity in both the sequence and structure clearly suggests that OwlRep and HSAT6 share the same evolutionary origin. Considering the relatively long evolutionary distance between owl monkeys and humans, these findings also suggest that sequences similar to HSAT6 occur widely in primates. Therefore, we collectively call such sequences HSAT6.

Identification of HSAT6 in Marmoset

It was expected that OwlRep originated from HSAT6 specifically in the owl monkey lineage and that the original HSAT6 sequence is harbored in the owl monkey genome. To identify the owl monkey HSAT6, we attempted a BLAST search using the human HSAT6 consensus sequence as a query against the draft genome of *Aotus nancymaae* (Ma's night monkey) released by Human Genome Sequencing Center. However, HSAT6 was apparently indistinguishable from a large number of OwlRep and similar copies in the genome. Therefore, we then focused on the genome of the common marmoset

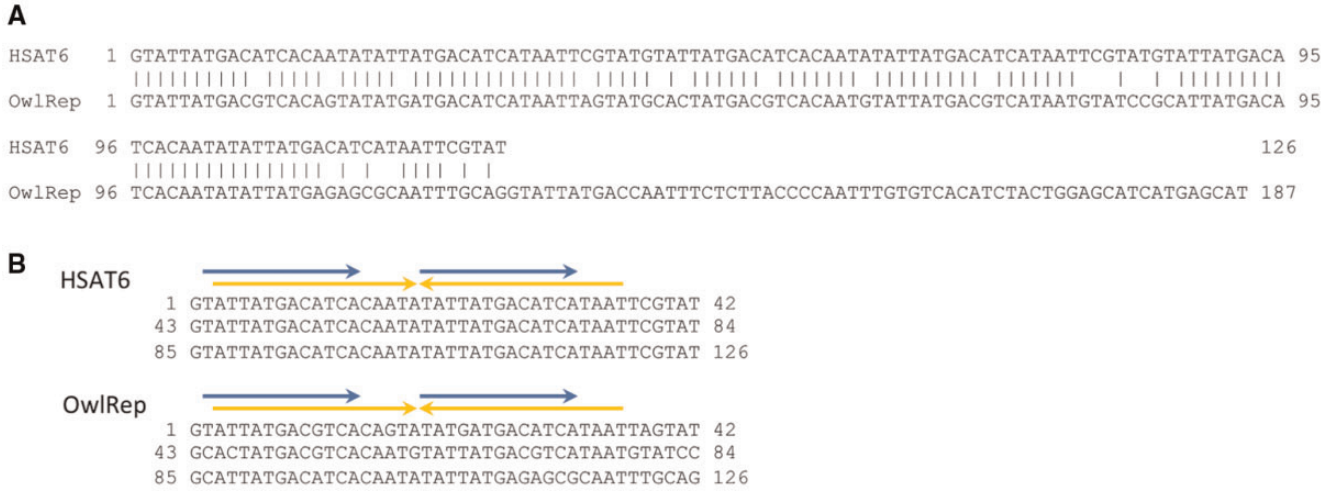

Fig. 1.—Sequence comparison between OwlRep and HSAT6. (*A*) Sequence comparison of the OwlRep consensus sequence retrieved from Prakhongcheep, Chaiprasertsri, et al. (2013) and HSAT6 obtained from Repbase. (*B*) Comparison of sequence structure between HSAT6 and OwlRep. Direct and inverted repeats within each of the three 42 bp units are shown with blue and yellow arrows, respectively.

(Marmoset Genome Sequencing and Analysis Consortium. 2014), another New World monkey. We investigated the RepeatMasker annotation data of marmoset obtained from the RepeatMasker website (calJac3, pre-masked with the repeat library 20140131; http://www.repeatmasker.org/species/calJac.html; last accessed January 3, 2018), and found that the marmoset genome harbors four HSAT6 loci.

HSAT6 is a repetitive DNA. It is widely thought that sequence data of repetitive DNA tend to contain more sequencing or assembly errors than those of non-repetitive regions. To provide our analyses with a firm basis, we conducted experimental verification of the HSAT6 sequences. Supplementary figure S1, Supplementary Material online shows, in an alignment format, the nucleotide sequences of the four marmoset HSAT6 loci and their flanking regions that we derived from the marmoset sequence database (calJac3). An HSAT6 copy of human we arbitrarily selected was also included in this test, which is shown in the figure separately from the marmoset alignment. These sequences were designated as H6Mar1 to H6Mar4, and H6Hum1. Supplementary figure S2, Supplementary Material online shows the sequence data obtained from the cloned fragments, which are aligned individually with the sequence derived from the database. The nucleotide identities were as follows: H6Mar1, 100% (1004/1004); H6Mar2, 99.9% (975/976); H6Mar3, 99.8% (1320/1322); H6Mar4, 100% (860/860); H6MHum1 100% (917/917). Thus, for all four marmoset HSAT6 sequences and the one human HSAT6 sequence, the sequence data derived from the databases were proved to reflect nucleotide blocks that exist in the marmoset and human genomes, respectively.

Wide Distribution of HSAT6 in Primates

It is notable that not only the four HSAT6 copies but also their flanking sequences are similar to each other. The flanking sequences contain the same transposable element set including AluSc8 and MSTA. These observations indicate that the four copies of HSAT6 resulted from repeated duplication events of a sequence cluster containing AluSc8 + MSTA + HSAT6 (fig. 2*A*). A search for the flanking repeat sequences in the primate genomes enabled us to trace the distribution and evolution of the HSAT6-containing clusters. We investigated all the HSAT6-flanking regions in human based on the RepeatMasker annotation. We found that 12 among the 17 HSAT6 loci are accompanied by MSTA and/or AluSc8 (fig. 2*A*). Most of the human HSAT6-containing clusters seem to be gathered near the centromeric region of chromosome 16. Likewise, the baboon genome (Old World monkey) has three HSAT6 along with MSTA and/or AluSc8 (fig. 2*A*). In the genomes of tarsier and Strepsirrhini such as bushbaby and mouse lemur, we found a single HSAT6 locus each. None was associated with MSTA or AluSc8. We found no HSAT6 sequence in the mouse or other non-primate mammals.

These findings suggest that the origin of HSAT6 can be traced back to the common ancestor of all extant primates, and the AluSc8 + MSTA + HSAT6 cluster originated in the common ancestor of simians. The current distribution of dozens of the MSTA + HSAT6 cluster (fig. 2*A*) probably resulted from a repeated segmental duplications of the cluster during simian evolution.

Because all of the marmoset HSAT6 belong to the AluSc8 + MSTA + HSAT6 cluster, it is likely that HSAT6 was present as a member of the cluster at the time of split of marmoset and owl monkey. To identify this cluster in the owl monkey genome, a BLAST search was conducted using the marmoset MSTA + HSAT6 sequence (650 bp) as a query. We found three HSAT6 copies in the owl monkey assembly which were also accompanied by MSTA and AluSc8 (fig. 2*A*). It is likely that OwlRep originated from at least one of the

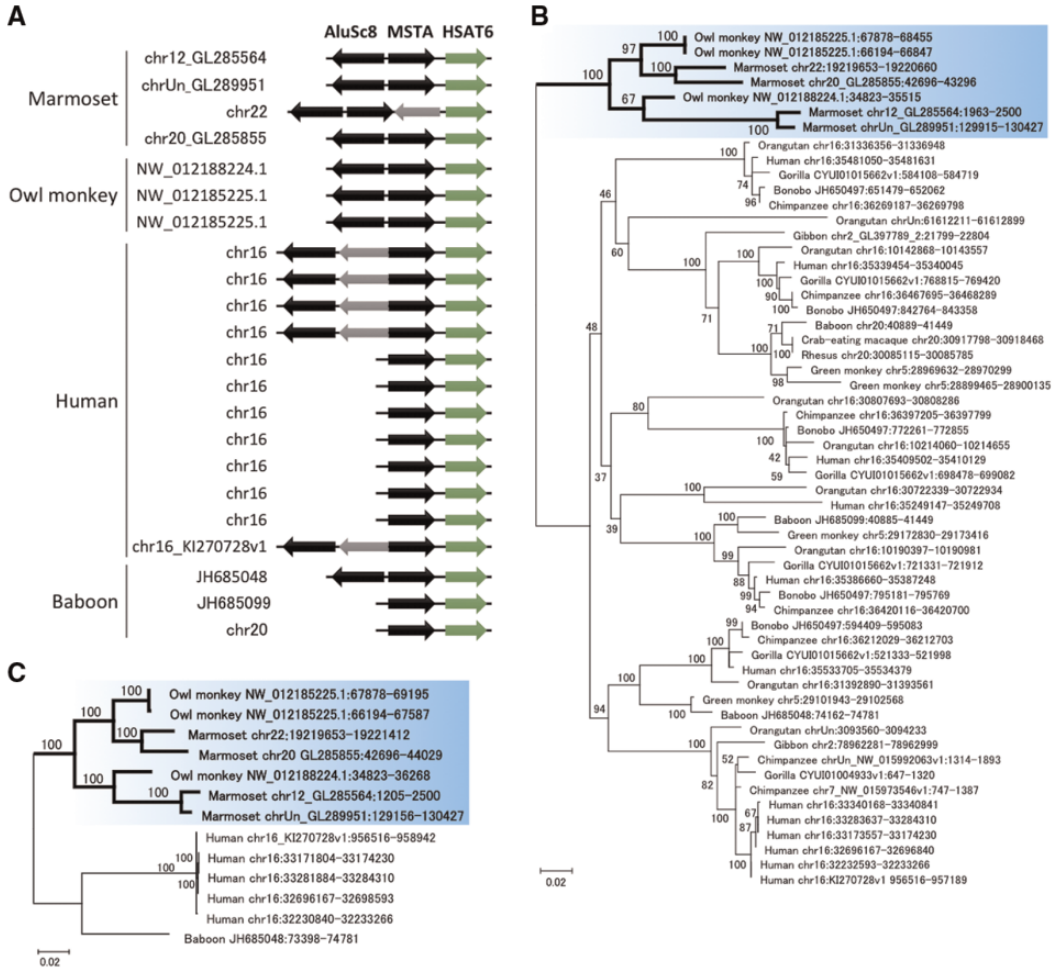

Fig. 2.—Evolution of the HSAT6-containing cluster. (*A*) Schematic representation of the HSAT6 clusters accompanied by MSTA and AluSc8 in primates. HSAT6, MSTA/AluSC8, and other transposable elements are shown with green, black, and grey arrows, respectively. (*B*) Maximum likelihood tree of all the MSTA + HSAT6 cluster sequences found in the primate genomes. (*C*) Maximum likelihood phylogeny of the AluSc8 + MSTA + HSAT6 cluster sequences from the owl monkey, marmoset, human, and baboon. For (*B*) and (*C*), the taxon names include species names and genomic regions used. Numbers at each node are bootstrap values. The New World monkey clade is shaded with light blue.

three HSAT6 sequences in the clusters. Indeed, the 3′ part of the OwlRep consensus sequence (nt 127–159) that does not share homology to HSAT6 (fig. 1*A*) shows 76% sequence identity to both the marmoset and owl monkey HSAT6 sequences (supplementary fig. S3, Supplementary Material online). This fact further reinforces our conclusion that OwlRep descended from these loci.

Phylogenetic Analysis of HSAT6-Containing Clusters in New World Monkeys

Based on the RepeatMasker annotations, we found at least one MSTA + HSAT6 cluster in six Hominoidea (human, chimpanzee, bonobo, gorilla, orangutan, and gibbon) and three Old World monkeys (rhesus macaque, crab-eating macaque, and baboon) (fig. 3*A*). To reveal the evolutionary history of the HSAT6 elements of the New World monkeys (owl monkey and marmoset), we reconstructed a maximum likelihood tree of the MSTA + HSAT6 sequences (0.6 kb) collected from the 12 primates (fig. 2*B*). Remarkably, all seven sequences from the New World monkeys form a monophyletic group with high node support (100% bootstrap probability [BP]). To validate the relationship within the New World monkeys, we further conducted a phylogenetic analysis of longer sequences (1.4 kb) including the AluSc8 + MSTA + HSAT6 elements of the owl monkey, marmoset, human, and baboon (fig. 2*C*). Two HSAT6-containing sequences in the scaffold NW_012185225.1 of the owl monkey form a monophyletic group, and it is a sister of the clade comprising two HSAT6 sequences in the marmoset (BP = 100). Also, the other HSAT6 in the scaffold NW_012188224.1 in owl monkey is a sister of the monophyletic group composed of the other two HSAT6 sequences of the marmoset (BP = 100, fig. 2*C*).

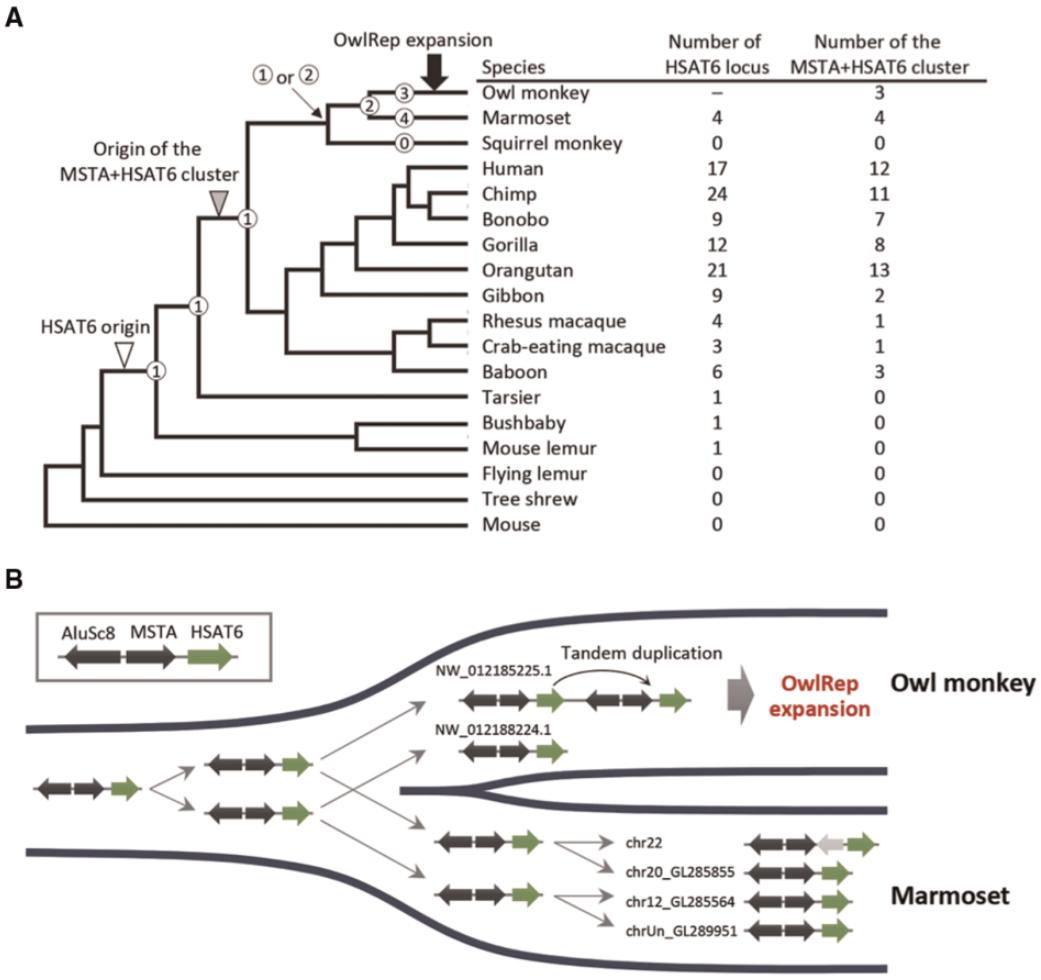

Fig. 3.—Evolution of HSAT6 in primates and New World monkeys. (*A*) Primate phylogeny as well as the number of HSAT6 loci and among them the number of the MSTA + HSAT6 clusters. The origin of HSAT6 and the MSTA + HSAT6 cluster is shown with open and grey triangles on the phylogeny. Circled numbers from a primate ancestor to the New World monkeys represent the estimated copy number of HSAT6 in the last common ancestors. (*B*) Evolution of the AluSc8 + MSTA + HSAT6 clusters in the owl monkey and marmoset lineages.

Based on the phylogenetic analyses, an evolutionary scenario of HSAT6 in New world monkeys can be developed (fig. 3). It is likely that HSAT6 existed as a single locus in the last common ancestor of primates, because only one HSAT6 locus was found in the genomes of the tarsier and Strepsirrhini (fig. 3*A*). HSAT6 apparently also existed as a single locus at the split between New World monkeys and Old World monkeys. This conclusion is supported by the phylogenetic trees that showed the monophyly of all the New World monkey sequences. Apparently, the number of HSAT6 copies increased independently in Old World and New World monkeys. The HSAT6 amplifications probably occurred mostly via segmental duplications involving the flanking MSTA and AluSc8 elements, because most of the simians (9 among the 12 species) harbor multiple numbers of the clusters (fig. 3*A*). In the ancestral New World monkey lineage, this cluster underwent a segmental duplication resulting in two copies which are found today in both the owl monkey and marmoset (fig. 3*B*). After their split, the HSAT6 clusters again underwent additional duplications independently in both lineages. In the owl monkey lineage, at least one of these HSAT6 sequences was presumably amplified and became a source of OwlRep, which later expanded throughout the genome and across all chromosomes.

Southern Blot Analysis for OwlRep Scale in Another Owl Monkey Species

OwlRep was first identified as a megasatellite DNA in *A. azarae* by genomic hybridization experiments (Prakhongcheep, Chaiprasertsri, et al. 2013). OwlRep was also found in a genome sequence database of *A. nancymaae*. In *A. nancymaae*, OwlRep is also likely to be a large-scale satellite DNA because a BLAST search resulted in hits for a large number of different contig sequence files. Further, we conducted Southern blot analysis of *A. nancymaae*, and the

results obtained confirmed that *A. nancymaae* contains OwlRep on a scale that is comparable to that found in *A. azarae*. Figure 4 shows the autoradiogram of this Southern blot analysis. Although the human and marmoset lanes did not exhibit any detectable signal, ladder signals were observed in the lane for *A. azarae* which contained 1/100 and 1/500 as much amount of genomic DNA as that of human and marmoset. The lanes for *A. nancymaae* contained the same amounts as those of *A. azarae*. The signals had a similar ladder pattern with similar intensities.

Discussion

OwlRep is a megasatellite DNA that was first identified in *A. azarae* (Prakhongcheep, Chaiprasertsri, et al. 2013). Its presence in the genome of *A. lemurinus* was revealed by chromosome FISH (fluorescence in situ hybridization) analysis (Koga et al. 2017). In addition, in the present study, we confirmed by Southern blot analysis that *A. nancymaae* carries OwlRep as a megasatellite DNA (fig. 4). In widely accepted phylogenetic trees of *Aotus* species, the genus *Aotus* comprises two species groups, with *A. azarae* and *A. nancymaae* in one and *A. lemurinus* in the other (Perelman et al. 2011; Fleagle 2012). The presence of OwlRep in these three species suggests that all extant *Aotus* species contain OwlRep as a megasatellite DNA in their genomes. Furthermore, OwlRep was not found in any other organisms we examined. We conclude that OwlRep was highly amplified in the *Aotus* lineage after its divergence from other New World monkeys, and before the divergence of extant owl monkey species. The primary goal of this study was to discover the evolutionary origin of OwlRep. We also wanted to determine the evolutionary events that led to the formation of OwlRep and identify key factors involved in these events. The similarity between HSAT6 and OwlRep at both levels of nucleotide sequence and complex repeat structure provides strong support to the hypothesis that OwlRep was derived from HSAT6.

Our results on the HSAT6 distribution in a wide array of primates, allows us to formulate the following evolutionary scenario: 1) HSAT6 originated in an early ancestor of primates, 2) an *Alu* element and an MSTA element were transposed into the neighboring region of the HSAT6 sequence in an early ancestor of simian primates, 3) segmental duplication then multiplied a chromosomal segment containing the Alu + MSTA + HSAT6 cluster, 4) part of an Alu + MSTA + HSAT6 cluster, containing HSAT6 and its flanking region, was subjected to tandem amplification in an ancestor of owl monkeys, and 5) this tandem repeat DNA was further amplified, eventually constituting the megasatellite DNA now called OwlRep, before the divergence of the lineages descending to *A. azarae*/*A. nancymaae* and *A. lemurinus*.

Once a small-scale tandem repeat DNA was formed in step (4), it is very likely that it could easily increase its scale through

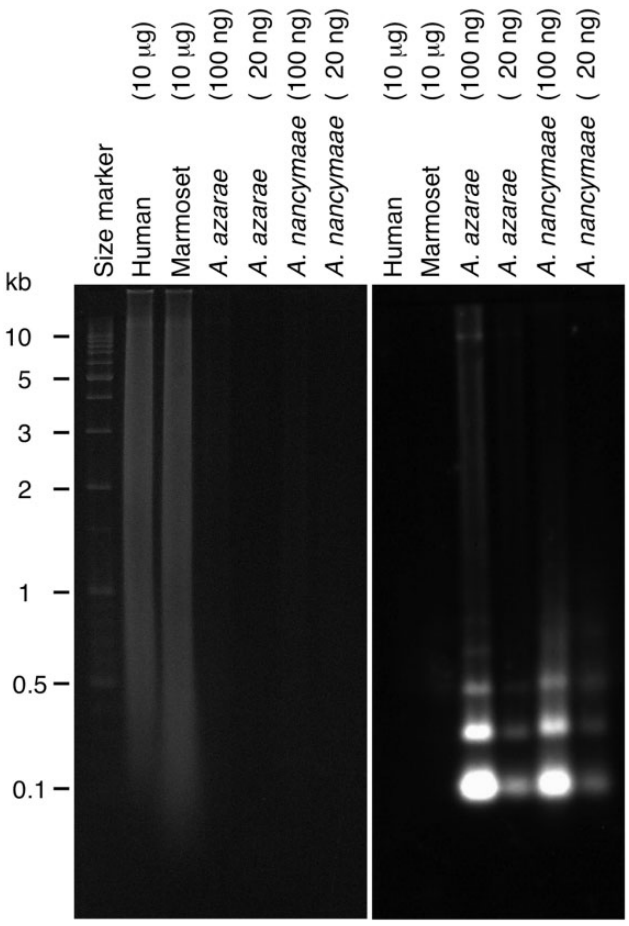

Fig. 4.—Southern blot analysis to compare the OwlRep scale. The left panel shows the agarose gel after electrophoreses, in which DNA was stained with ethidium bromide. The right pane is the autoradiogram after hybridization with the OwlRep probe.

various mechanisms, such as unequal crossing-over or rolling-circle replication. Transfer of this repetitive sequence to other chromosomes may have been mediated by the generation and migration of extrachromosomal circular DNA. This amplification to form the megasatellite DNA corresponds to step (5), and it is likely that the amplification was favored by natural selection because we know that the abundance of OwlRep is directly associated with adaptation to a nocturnal lifestyle (Koga et al. 2017).

Although it is likely that natural selection favored OwlRep in step (5), the formation of small-scale tandem repeats in step (4) is likely to have been a random event. However, the location of the original Alu + MSTA + HSAT6 cluster might have been a significant factor for the occurrence of this random event. We found a total of 17 HSAT6 loci in the human genome, and 16 of them were located in the centromere region. In chimpanzee, at least 12 HSAT6 loci among the 24 loci found were present in the chromosomal region homologous to the human, but the locations of most of the rest were unknown due to an insufficient assembly of the chimpanzee

genome. The locations of the HSAT6 loci in other primate species were also not clear because their genome sequence databases have not been as fully detailed as that of humans, especially in the centromere and telomere regions. From the situation in human, we assume that the original Alu + MSTA + HSAT6 cluster existed in the centromere region of the ancestral simian primate.

The centromere region of chromosomes usually contains large amounts of tandem repeat DNA, and tandem repeat DNA in this region is known to frequently fluctuate in copy number (Willard 1991). This fluctuation, combined with bottle-neck effects, often results in tandem multiplication and subsequent expansion of a variant-type sequence (Willard 1991; Rudd et al. 2006) or even a new, unrelated sequence (Hara et al. 2012). The tandem amplification of the HSAT6-containing sequence in an initial Alu + MSTA + HSAT6 cluster in step (4) may have been easily triggered by its location in the centromere region of an ancestral simian primate. If so, the HSAT6 sequence that first emerged may have been located in the centromere region.

The formation of the Alu + MSTA + HSAT6 cluster in step (2) may have been another key factor; for example, changes in sequence or structure caused by the annexation of the two transposons may have enabled or facilitated subsequent amplification in step (4) or (5). However, the results of the present study do not provide any clues about this possibility.

The HSAT6 in tarsier shows the signature of a unique translocation. The tarsier genome harbors a single HSAT6 locus, which may have resulted from the integration of HSAT6 via extrachromosomal circular DNA. This tarsier HSAT6 sequence showed a 74.6% identity to the human HSAT6 consensus sequence and is likely to have been integrated uniquely in tarsier (supplementary fig. S4, Supplementary Material online). Among all primates, this locus shares a ~23 bp sequence 87% similar to a part of HSAT6 (underlined in supplementary fig. S4, Supplementary Material online), which might have been a target for integration. Because HSAT6 represents a tandem repeat structure of a 42 bp sequence unit (fig. 1*B*), it can potentially translocate as an extrachromosomal circular DNA. Based on this observation, it is possible that an early OwlRep might have been translocated from one chromosomal place to another in the common ancestor of owl monkeys.

Next, we discuss the time span in which step (5) began and was completed. In widely accepted molecular phylogenetic trees of New World monkeys, the genus *Aotus* diverges from the lineage leading to Callitrichinae (marmosets and tamarins) approximately 20 Mya (Perelman et al. 2011; Schneider and Sampaio 2013). Superimposing the limited distribution of OwlRep on this divergence pattern suggests that OwlRep started its final amplification after the Callitrichinae and *Aotus* lineages began diverging. This divergence is estimated to have started at or after 20 Mya (Perelman et al. 2011). With regard to the completion of the amplification, the time point is estimated to have been at or before the divergence of the split of the two species groups (one containing *A. azarae*/*A. nancymaae* and the other containing *A. lemurinus*) because both groups carry large amounts of OwlRep in their genomes. The time point of this split has been estimated to be approximately 5 Mya (Perelman et al. 2011). Thus, step (5) probably occurred sometime between 20 and 5 Mya.

For over a decade, dozens of studies revealed that various repetitive elements, including transposable elements and short tandem repeats, have acquired functions by co-option during mammalian evolution and contributed to controlling the expression of neighboring genes (Bejerano et al. 2006; Sasaki et al. 2008; Chuong et al. 2016; Gymrek et al. 2016; Nishihara et al. 2016). Thus, repetitive sequences have made an impact on the gene regulatory alterations and morphological evolution by serving as enhancers or promoters. Aside from the co-option of individual repetitive elements, it is also possible that collective repeat sequences are involved in the compartmentalization of nuclear architecture because SINEs and LINEs/LTRs preferably locate in euchromatin and heterochromatin, respectively (Solovei et al 2016). This model suggests the possibility that repetitive sequences are also generally important in determining nuclear organization leading to cellular evolution, even if a molecular mechanism facilitating the mutual attraction between repetitive sequences is still largely unknown. In the case of owl monkeys, OwlRep was shown to be a contributing factor for compartmentalization of the inverted nuclear architecture in the rod cells (Koga et al. 2017), possibly by participating in the mutual attraction among repeats. Their inverted nuclear architecture is apparently immature (Joffe et al. 2014) and thought to be in an initial stage of transition from the conventional to inverted nuclear architecture. Therefore, the origin of OwlRep revealed in this study is a key clue to understanding the onset and process of the evolutionary change of the nuclear structure at the molecular level. The primary conclusion of the present study is that OwlRep originates from HSAT6. We also propose the evolutionary scenario of the OwlRep formation through the steps (1–5). In future research, it may be possible to reveal a higher resolution formation process by identifying a DNA segment of an intermediate type between HSAT6 and OwlRep. Such a sequence may be harbored in the owl monkey genome, and we intend to carry out an in-depth analysis of owl monkey genome sequences to discover this intermediate form. Successful identification of an intermediate type would more fully link the molecular evolution of repetitive DNA and changes in nuclear architecture of rod cells as an adaptation to a nocturnal lifestyle.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Author Contributions

A.K. conceived and designed the overall framework of this study. H.N. and J.K. performed the bioinformatics analyses. A.K. conducted the molecular biology experiments. H.H. and R.S. arranged for the animals and cell lines. H.N., A.K., and R.S. wrote the report. All authors approved the final version of the manuscript.

Acknowledgments

We thank Human Genome Sequencing Center, Baylor College of Medicine (https://www.hgsc.bcm.edu/non-human-primates/owl-monkey-genome-project; last accessed January 3, 2018) for the public release of the *Aotus nancymaae* (Ma's night monkey) genome. Computational analyses were partially performed on the supercomputer system of the Institute of Statistical Mathematics. This work was supported by Grants-in-Aid from the MEXT of Japan (grant numbers 15H04427 and 23114005 to A.K.; 26840117 and 17K19424 to H.N.) and a grant from the Italian Ministry of Universities and Research (2015RA7XZS_003 to R.S.).

Literature Cited

Bejerano G, et al. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441(7089):87–90.

Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science 351(6277):1083–1087.

Dyer MA, et al. 2009. Developmental sources of conservation and variation in the evolution of the primate eye. Proc Natl Acad Sci U S A. 106(22):8963–8968.

Eberhart A, et al. 2013. Epigenetics of eu- and heterochromatin in inverted and conventional nuclei from mouse retina. Chromosome Res. 21(5):535–554.

Fernández-Duque E, de la Iglesia H, Erkert HG. 2010. Moonstruck primates: owl monkeys (*Aotus*) need moonlight for nocturnal activity in their natural environment. PLoS One 5(9):e12572.

Fleagle JG. 2012. Perspectives in primate evolution. Evol Anthropol. 21(6):207.

Finstermeier K, et al. 2013. A mitogenomic phylogeny of living primates. PLoS One 8(7):e69504.

Grewal SI, Jia S. 2007. Heterochromatin revisited. Nat Rev Genet. 8(1):35–46.

Griffin RH, Matthews LJ, Nunn CL. 2012. Evolutionary disequilibrium and activity period in primates: a Bayesian phylogenetic approach. Am J Phys Anthropol. 147(3):409–416.

Gymrek M, et al. 2016. Abundant contribution of short tandem repeats to gene expression variation in humans. Nat Genet. 48(1):22–29.

Hara T, Hirai Y, Jahan I, Hirai H, Koga A. 2012. 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(12):760–765.

Heesy CP, Ross CF. 2001. Evolution of activity patterns and chromatic vision in primates: morphometrics, genetics and cladistics. J Hum Evol. 40(2):111–149.

Joffe B, Peichl L, Hendrickson A, Leonhardt H, Solovei I. 2014. Diurnality and nocturnality in primates: an analysis from the rod Photoreceptor nuclei perspective. Evol Biol. 41:1–11.

Jost KL, Bertulat B, Cardoso MC. 2012. Heterochromatin and gene positioning: inside, outside, any side?. Chromosoma 121(6):555–563.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Kay RF, Ross C, Williams BA. 1997. Anthropoid origins. Science 275(5301):797–804.

Koga A, et al. 2017. Co-opted megasatellite DNA drives evolution of secondary night vision in Azara's owl monkey. Genome Biol Evol. doi: 10.1093/gbe/evx142.

Marmoset Genome Sequencing and Analysis Consortium. 2014. The common marmoset genome provides insight into primate biology and evolution. Nat Genet. 46:850–857.

Nishihara H, et al. 2016. Coordinately co-opted multiple transposable elements constitute an enhancer for *wnt5a* expression in the mammalian secondary palate. PLoS Genet. 12(10):e1006380.

Perelman P, et al. 2011. A molecular phylogeny of living primates. PLoS Genet. 7(3):e1001342.

Pozzi L, et al. 2014. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. Mol Phylogenet Evol. 75:165–183.

Prakhongcheep O, et al. 2013. Two types of alpha satellite DNA in distinct chromosomal locations in Azara's owl monkey. DNA Res. 20(3):235–240.

Prakhongcheep O, Chaiprasertsri N, et al. 2013. 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.

Ross CF, Hylander W. 2000. Electromyography of the anterior temporalis and masseter muscles of owl monkeys (*Aotus trivirgatus*) and the function of the postorbital septum. Am J Phys Anthropol. 112(4):455–468.

Rudd MK, Wray GA, Willard HF. 2006. The evolutionary dynamics of alpha-satellite. Genome Res. 16(1):88–96.

Santini L, Rojas D, Donati G. 2015. Evolving through day and night: origin and diversification of activity pattern in modern primates. Behav Ecol. 26(3):789–796.

Sasaki T, et al. 2008. Possible involvement of SINEs in mammalian-specific brain formation. Proc Natl Acad Sci U S A. 105(11):4220–4225.

Schneider H, Sampaio I. 2013. The systematics and evolution of New World primates: a review. Mol Phylogenet Evol. 82:348–357.

Schoeftner S, Blasco MA. 2009. A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. EMBO J. 28(16):2323–2336.

Solovei I, et al. 2009. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. Cell 137(2):356–368.

Solovei I, et al. 2013. LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. Cell 152(3):584–598.

Solovei I, Thanisch K, Feodorova Y. 2016. How to rule the nucleus: divide et impera. Curr Opin Cell Biol. 40:47–59.

Springer MS, et al. 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. PLoS One 7(11):e49521.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.

Steiner FA, Henikoff S. 2015. Diversity in the organization of centromeric chromatin. Curr Opin Genet Dev. 31:28–35.

Willard HF. 1991. Evolution of alpha satellite. Curr Opin Genet Dev. 1(4):509–514.

Associate editor: Takashi Gojobori