Human Molecular Genetics, 2003, Vol. 12, No. 11 **1321–1328**
DOI: 10.1093/hmg/ddg138

Trans mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling

Yosuke Ejima[1,*] and Lichun Yang[2]

[1]Department of Radiological Sciences, Hiroshima Prefectural College of Health Sciences, 1-1 Gakuen-machi, Mihara, Hiroshima 723-0053, Japan and [2]Hiroshima Prefecture Collaboration of Regional Entities for the Advancement of Technological Excellence, JST, 3-10-32 Kagamiyama, Higashihiroshima, Hiroshima 739-0046, Japan

Received December 17, 2002; Revised and Accepted March 24, 2003

Exon shuffling, the juxtaposition and new combinations of exons from different genes, facilitates evolutionary changes by increasing protein diversity or by generating new function. Exon shuffling is generated as a consequence of segmental duplications. Long interspersed element (LINE)-1 (L1)-mediated 3′ transduction is a potential pathway for exon shuffling by which L1 associates 3′ flanking DNA in *cis* as a read-through transcript and carries the DNA to a new genomic location. In this pathway, however, the targets are limited to the regions located 3′ to L1s. Here we propose that the genomic DNA distant from L1 may be mobilized by an alternative (*trans*) action of L1. A partial *ATM* sequence containing a single exon and flanking introns has been retrotransposed to a new genomic location on chromosome 7. There was no L1 around the exon of the authentic *ATM* locus. An unusual feature that the poly(A) tail tagged to the transposed sequence oriented oppositely to the *ATM*'s transcriptional orientation suggests that a *trans* action of reverse transcriptase on antisense transcript has driven the duplication of genomic DNA without removing introns. Taking account of similar duplication events in previous studies, a certain class of segmental duplications in the human genome may be accounted for by the *trans* action of retrotransposon machinery.

INTRODUCTION

Segmental duplications involving the transfer of a partial genomic sequence from one chromosomal site to another are abundant in the human genome (1–3). Many of the segmental duplications have emerged during the past 35 million years, some of which may represent critical genetic changes associated with primate evolution (4–6). An important outcome of segmental duplications is the evolution of new function by the process of exon shuffling (7–9). A potential mechanism for exon shuffling is the long interspersed element (LINE)-1 (L1)-mediated 3′ transduction. Upon retrotransposition, L1 often associates 3′ flanking DNA as a read-through transcript and carries the non-L1 sequence to a new genomic location (10–12). L1 is the most abundant retrotransposon in the human genome, and it serves as a major source of reverse transcriptase activity (13–15). The 3′ transduction is likely to be an efficient mechanism because sequence homology or physical proximity between the donor and recipient DNA is not required. As this operates in copy-and-paste manner via RNA intermediates, the donor sequences remain unaffected. In this pathway, however, only those regions that are located 3′ to the active L1s are the targets for duplication.

L1 has a capability of mobilizing another class of non-L1 sequence by acting in *trans* to cellular RNA substrates. A major consequence is the formation of processed pseudogenes. They are generated by the reverse transcription of intronless mRNAs that have undergone RNA splicing, followed by the integration of resultant cDNAs to new genomic locations (16). The amplification of *Alu* elements, the most abundant non-autonomous short interspersed elements in the human genome, is also suspected to be due to a similar mechanism (17). The integration of processed pseudogenes occasionally leads to the generation of new genes (18,19), but the duplicated segments lack intron–exon structure. It would be plausible to suppose further that the L1 machinery can also act in *trans* to unspliced RNA intermediates, not intronless mRNAs, occasionally transcribed from antisense strand, and let the intron-containing cDNAs integrate to new genomic locations. Then possible targets to be mobilized by the L1-mediated mechanism would not be limited to the 3′ flanking

*To whom correspondence should be addressed. Tel: +81 848601171; Fax: +81 848601129; Email: ezima@hpc.ac.jp

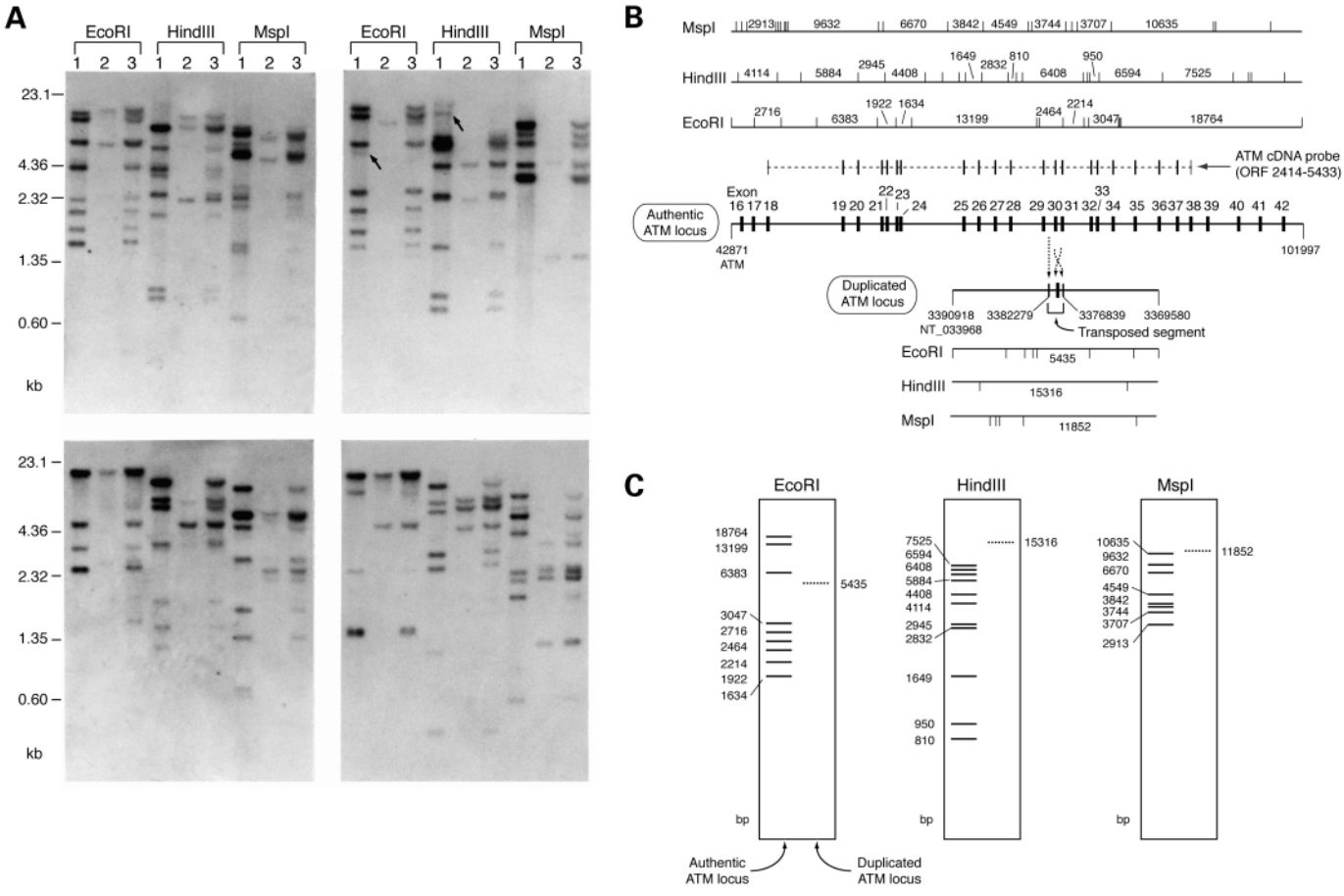

Figure 1. Southern blot results showing the presence of *ATM*-related sequence outside the *ATM* locus. (**A**) Southern blot results. Restriction enzyme-digested DNA from human diploid fibroblasts (lane 1), mouse cells (lane 2) or human chromosome-11-containing somatic cell hybrids (lane 3) was hybridized with *ATM* cDNA probes: *ATM* ORF 1-3033, corresponding to exons 4–22 (upper left panel), *ATM* ORF 2414–5433, exons 18–38 (upper right panel), *ATM* ORF 5062–8091, exons 36–57 (lower left panel), or *ATM* ORF 7453–9171, exons 52–65 (lower right panel). Arrows in the upper right panel indicate the bands seen in lane 1 but not in lane 3. (**B**) Restriction enzyme maps for the *ATM* locus and the chromosome-7 locus containing the *ATM* insert. Structures of the authentic *ATM* locus (59 kb region between intron 15 and intron 42), the duplicated *ATM* locus (21 kb region encompassing the *ATM* insert) and the *ATM* cDNA probe (ORF 2414–5433) are also shown. Numbers on the restriction maps represent the size of fragments detectable by the probe. (**C**) Southern blot diagram for the *ATM* cDNA probe (ORF 2414–5433). Solid bars represent expected products from the authentic *ATM* locus. Dotted bars represent expected products from the duplicated *ATM* locus.

regions but extended to the majority of the other genomic regions, providing the L1 machinery with an additional capability for exon shuffling. Here we propose that such a mechanism might have operated to generate a certain class of gene duplication events in the human genome.

RESULTS AND DISCUSSION

The *ATM* (AT, mutated) is the gene responsible for the human autosomal-recessive disorder ataxia–telangiectasia (20). The *ATM* is located at chromosome 11q22.3, spans the 146 kb genomic region, and contains 66 exons (21,22). The presence of a partial *ATM* sequence on chromosome 7 was first recognized through a Southern blot analysis. An *ATM* cDNA probe containing nucleotides 2414–5433 of *ATM* ORF (exons 18–38 of the *ATM* gene) detected two unusual bands that are present in human diploid fibroblasts but not in the human chromosome-11-containing somatic cell hybrids (Fig. 1A). We supposed these bands represent *ATM*-like sequence outside the *ATM* locus. Additional analyses revealed they originate from chromosome 7. We next screened the human chromosome-7-specific genomic library (LA07NS01, *Eco*RI-digested DNA cloned in Charon21A vector), and isolated a positive clone (LA07NS01-atmr). In the 5442 bp LA07NS01-atmr insert, a partial *ATM* sequence was embedded in the 1126 bp region located between the two disrupted L1ME ORF2 blocks (Fig. 2). This corresponds to the region containing exon 30 and flanking introns of the *ATM* locus ranging from nucleotide 75894 (intron 29, IVS30 −700) to 77075 (intron 30, IVS30 +295) of the *ATM* genomic sequence (GenBank accession no. U82828) (22). A difference from the authentic *ATM* locus was the occurrence of a 5′ inversion (the 454 bp region around exon 30, *ATM* 76622–77075). By a BLAST search, this sequence was mapped to chromosome 7p11.2 (GenBank accession no. NT_033968, nucleotides 3376839–3382279) located within the 280 kb region between the two hypothetical genes *LOC223000* and *LOC222001*. Because exon 30 was the only exon included in the partial *ATM* sequence on chromosome 7, the *ATM* cDNA probe (ORF 2414–5433) should detect a single

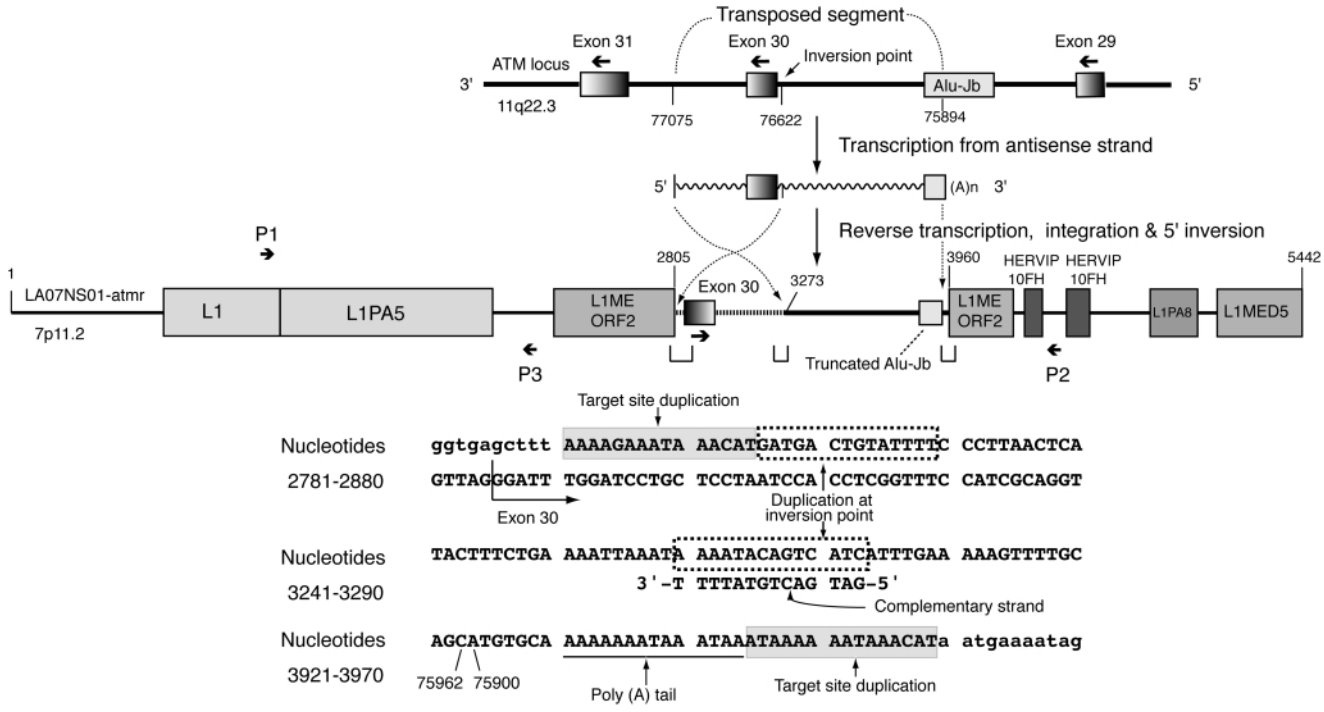

Figure 2. Retrotransposed *ATM* sequence on human chromosome 7. Structure of the retrotransposed *ATM* sequence, the corresponding region on the *ATM* locus, and a proposed mechanism for transposition are shown. Inverted region is indicated by a thick dotted line. Arrows on exons indicate *ATM*'s transcriptional orientation. $(A)_n$ denotes poly(A) tail. Numbers above the LA07NS01-atmr give the nucleotide positions within the phage insert. Numbers beneath the *ATM* give the nucleotide positions in the genomic *ATM* sequence (GenBank accession no. U82828). Brackets indicate the regions whose sequences are depicted below. Positions for PCR primers (P1–P3) are also shown. In the sequence data (below), *ATM* sequence is shown in uppercase letters. Note that the duplicated sequences associated with the inversion are complementary to each other. The position of 61 bp deletion (*ATM* 75901–75961) is also shown.

extra band (5435, 15 316 or 11 852 bp) for each restriction enzyme (*Eco*RI, *Hind*III or *Msp*I, respectively; Fig. 1B and C). Southern blot results (Fig. 1A, upper right panel) agreed with this assumption, although the 11 852 bp *Msp*I band was hardly discernible from the 10 635 bp band derived from the authentic *ATM* locus. Also, we performed a BLAST search of the whole *ATM* locus (91 909 nucleotides excluding interspersed repeats) against the draft human genome sequence, but we detected no additional duplication event involving another part of the *ATM* locus.

Molecular features suggested that this duplication be mediated by L1 retrotransposition (Fig. 2). The *ATM*-derived sequence was flanked by 15 bp target site duplications (TSDs). The sequence around the 5′ end of the 5′ TSD (TTAAAA) matched with the consensus sequence for L1 endonuclease cleavage site (23). There was found a poly(A) tail immediately preceding the 3′ TSD. No apparent polyadenylation signal was recognized. Instead, a 61 bp sequence had been lost from the 3′ terminus of the *ATM* insert. The deletion (*ATM* 75901–75961) involved the A-rich linker sequence connecting the two *Alu* monomers in the *Alu*-Jb element oriented oppositely to the *ATM* gene, where resides a sequence AATATAAA (complementary to *ATM* 75920–75927). Probably there once was a canonical polyadenylation signal AATAAA, which has degenerated by subsequent A/T substitution at position 3 or 5. Loss of the 3′ terminus polyadenylation signal is occasionally seen in naturally occurring or experimentally induced L1 retrotransposition (12,16). In one relevant example (12) (case of gi 4006838) where L1 associated with a partial *Alu*-Y element as a 3′-transduced segment, the region around the polyadenylation signal had been lost from the *Alu*-Y 3′ terminus, whereas the downstream poly(A) tail and 3′ TSD remained intact. The occurrence of inversion is also a hallmark of L1 retrotransposition. The inversion points associated with L1 retrotransposition are known to cluster within the region 722–922 bp from the 3′ end poly(A) tail (24). The distance between the inversion point and the 3′ end was 729 bp in our case, which is within the predicted range. Duplication of a short stretch of nucleotides at the inversion point (duplication of 14 nucleotides in our case) is also a feature characteristic to L1-associated inversions. Because no L1 element was present neither adjacent to the retrotransposed *ATM* insert nor around the authentic *ATM* exon 30, the mobilization of the *ATM* sequence to chromosome 7 cannot by explained simply by 3′ transduction. Conspicuously the orientation of poly(A) tail was opposite to the transcriptional orientation of the *ATM* gene. One possible explanation would be that a part of the opposite strand of the *ATM* locus spanning exon 30 was transcribed and processed by the addition of a poly(A) tail, followed by the reverse transcription by the L1 machinery.

To estimate the time of this retrotransposition during primate evolution, we searched for homologous loci in non-human primates. PCR amplification using a primer pair (P1+P2) giving a 3281 bp product in the human (LA07NS01-atmr 1095-4375) gave rise to a similar band in the chimpanzee (*Pan troglodytes*) or the gorilla (*Gorilla gorilla*), but a shorter band

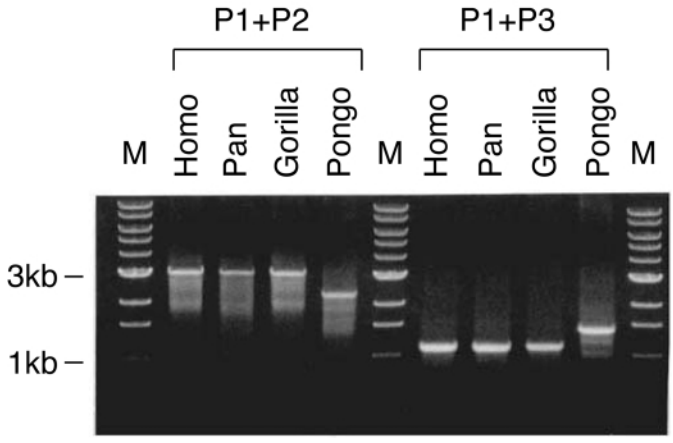

Figure 3. Detection of retrotransposed *ATM* sequences by PCR in non-human primates. The shorter band in the orangutan (P1+P2) represents the ancestral locus without the *ATM* insert. The longer band in the orangutan (P1+P3) reflects the insertion of an *Alu*-Y element between L1PA5 and L1ME ORF2.

in the orangutan (*Pongo pygmaeus*; Fig. 3). Sequence analyses revealed that the bands detected in the chimpanzee and the gorilla represent the loci homologous to the human chromosome-7 locus containing the *ATM* insert, while the shorter band in the orangutan represents the ancestral locus without the *ATM* insert (Fig. 4). The longer PCR band (P1 + P3) seen in the orangutan (Fig. 3) proved to reflect the insertion of a 306 bp *Alu*-Y element between L1PA5 and L1ME ORF2. It is therefore estimated that this retrotransposition event occurred after the divergence of the genus *Pongo* but before the divergence of the genus *Gorilla* from *Hominidae* lineage, that is about 5–10 million years ago. The extent of sequence identity between the two paralogous *ATM* loci (96.2% for the human, 95.7% for the chimpanzee, 95.6% for the gorilla) is consistent with this estimate. We also compared the sequence around exon 30 of the *ATM* locus among primates (Fig. 5). The genomic structure in this region was highly conserved in the six species belonging to the suborder *Anthropoidea*. The genomic structure was somewhat different in the ring-tailed lemur (*Lemur catta*) which belongs to the suborder *Prosimii*, but a full-length L1 element was not present in any species. The results further suggest that the region around *ATM* exon 30 has long been devoid of an active L1 element, which might have driven 3′ transduction, during primate evolution.

Two gene duplication events with similar molecular features have been documented previously. The case of the transposition of exon 9 and flanking introns of the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene is quite similar to ours (25). No L1 element was found around exon 9 of the *CFTR* locus, and the orientation of poly(A) tail is opposite to the *CFTR*'s transcriptional orientation. An inversion has occurred within a partial L1 sequence co-integrated 5′ to the *CFTR* sequence, and the distance between the inversion point and the 3′ end poly(A) tail is 772 bp, which is comparable to ours (729 bp). The other case is the transposition of a part of the melanin-concentrating hormone (*MCH*) gene. This has generated a new chimeric gene, the pro *MCH*-like (*PMCHL*) gene, which contains a partial *MCH* sequence with a poly(A) tail oriented oppositely to *MCH*'s transcriptional orientation (26,27). It is to be noted that, in both cases, the duplicated units include exon-containing genomic DNA of functional genes.

It remains possible, however, that the retrotransposition event we described here resulted from L1-mediated 3′ transduction. Namely, it is possible that there once was a full-length L1 in the opposite orientation of the *ATM* gene, upstream of exon 30, which generated a read-through transcript, and carried the *ATM* sequence to a new genomic location on chromosome 7. The L1 may have been 10 kb or more distant from exon 30, because splicing may have eliminated the sequence between them. Extensive 5′ truncation would leave no evidence of the L1 sequence, resulting in a very similar structure to our case (Fig. 6, upper right). In the *ATM* gene, there are three full-length, currently inactive L1s in intron 18 (*ATM* 48310–54319), intron 63 (*ATM* 144786–150805), as well as in the 3′ untranslated region (*ATM* 160948–167473). However, because they are oriented in the same direction as the *ATM* gene, they could not be the driver of exon-30 transduction. However, recent evidence suggests that there is strong selective pressure against the accumulation of full-length L1s in human DNA (28). It is therefore possible that the once-active L1 involved in this event has now been lost from the population.

An equally plausible explanation we propose is that a partial, antisense *ATM* transcript was retrotransposed in *trans* by the L1 machinery (Fig. 6, lower left). The generation of an aberrant, antisense transcript may have been initiated from a cryptic promoter on the opposite strand of the *ATM* gene. To be captured in *trans* by the L1 machinery, the transcript would have to compete with the L1 RNA for the L1-encoded proteins, either within the nucleus or after being transported to the cytoplasm. By a BLAST search of *ATM* mRNA sequence against dbEST, we identified five human expression sequence tags (ESTs) possibly representing antisense *ATM* transcripts. The ESTs contain the genomic sequence around exon 53 (*ATM* 119076–119232 for GenBank accession no. AA412245, *ATM* 118749–119164 for AA912542 and *ATM* 118749–119154 for AA960806) or exons 43–44 (*ATM* 103532–103912 for GenBank accession no. AA460148 and *ATM* 103488–103957 for AA460746) of the *ATM* locus. Interestingly, three of them (AA412245, AA460148 and AA460746) are expressed in testis. The others may also be of testis-origin because they are derived from a pooled library made from fetal lung, testis and B-cell cDNAs. The antisense *ATM* transcripts may represent some unknown tissue-specific function. Otherwise, their existence may be a consequence of reduced RNA surveillance against antisense or nonsense transcripts in testis (29,30). In view of the elevated L1 activity in the germline, it is conceivable that the L1 machinery has targeted such testis-specific antisense transcripts. The presence of a partial *Alu*-Jb sequence at the 3′ end of the *ATM* transcript may have been another factor. *Alu* elements are thought to amplify their copies by the L1 machinery. *Alu* RNA is a good template for the target-primed reverse transcription by L1 ORF2 protein *in vitro* (31). Two signal recognition particle proteins (SRP9 and SRP14) bind with high affinity to the 7SL RNA domain of *Alu* (32). It is suggested that the SRP9/14-*Alu* complex binds to the

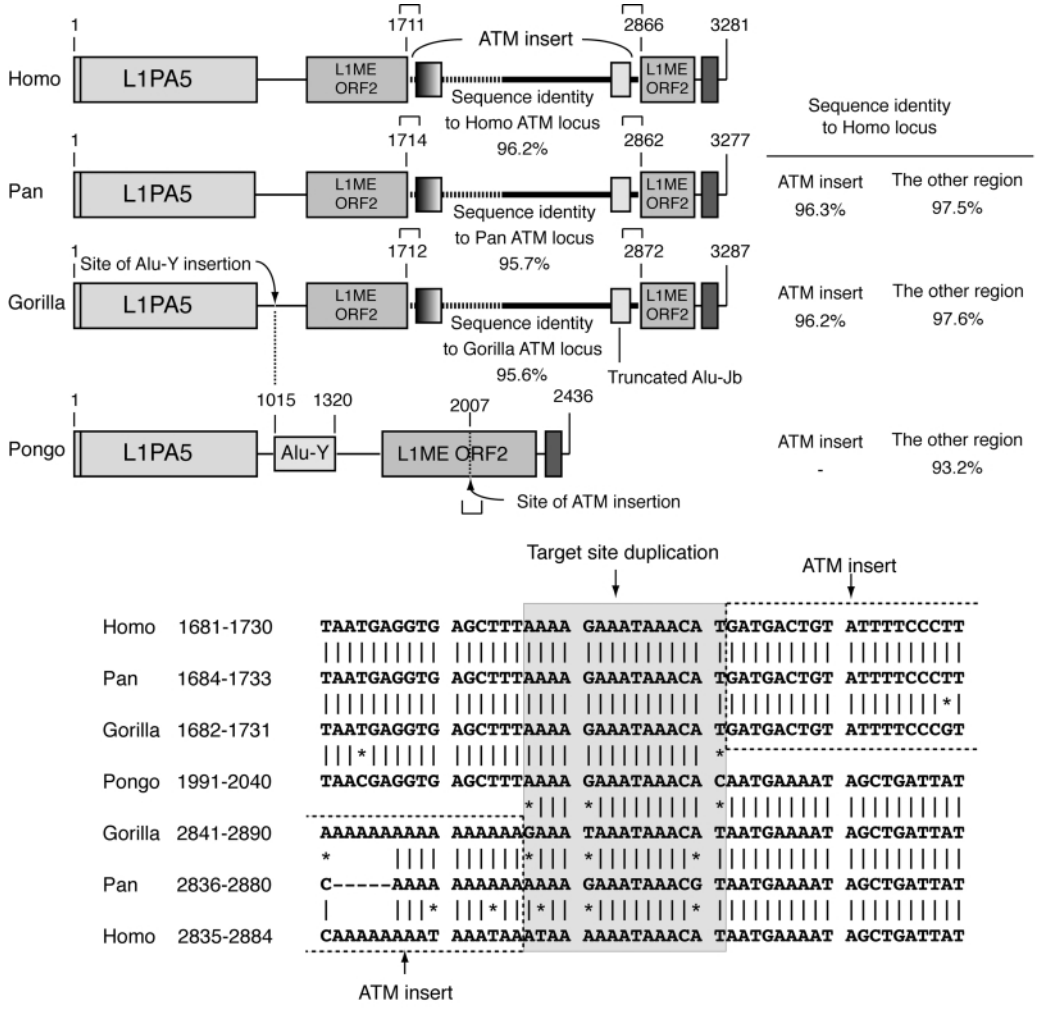

Figure 4. Comparison of genomic structure around the *ATM* insert among primates. Numbers above the sequence give the nucleotide positions within the sequence of PCR products (P1+P2). The sequence identity between the human and the orangutan outside the *ATM* insert (93.2%) was estimated by excluding the *Alu*-Y region. Brackets indicate the regions whose sequences are depicted below. In the sequence data (below), the human, chimpanzee and gorilla sequences around the 5′ and 3′ end of the *ATM* insert are aligned with the orangutan sequence. Target site duplications are shaded. The vertical bar or star represents identical or mismatched nucleotide pair, respectively.

large ribosomal subunit, leaving *Alu* poly(A) suitably positioned to interact with L1 ORF2 protein (17). The *Alu*-like 3′ structure of the *ATM* transcript may have facilitated its access to the L1-encoded proteins.

L1-encoded proteins preferentially mobilize in *cis* the transcript from which they are encoded (16,33). The 3′ transduction (*cis*-mobilization) is therefore an efficient mechanism to mobilize the 3′ flanking exons. On the other hand, the *trans*-mobilization of distant genomic DNA would be another strategy of L1 machinery to shuffle exons. This complements the capability of 3′ transduction because virtually any genomic regions become the potential targets. The length of genomic DNA mobilized by the *trans* action of L1 would, as in other L1-mediated events, usually not exceed 1 kb due to the inherent low processivity of L1 reverse transcriptase (1,10–15,34). Therefore, a certain proportion of short segmental duplications of no more than 1 kb, as well as the resultant exon shuffling events, is likely to be accounted for by this mechanism.

MATERIALS AND METHODS

Cell culture

Non-human primate cell lines were purchased from Coriell Cell Repository: GM03448A for chimpanzee, *Pan troglodytes*; AG05251B for lowland gorilla, *Gorilla gorilla*; GM04272 for orangutan, *Pongo pygmaeus*; GM03446 for cab-eating macaque, *Macaca fascicularis*; AG05352 for black-handed spider monkey *Ateles geoffroyi*; and AG07099A for ring-tailed lemur, *Lemur catta.* Mouse A9 and A9-derived human chromosome-11-containing somatic cell hybrid cell line were gifts from M. Oshimura.

DNA preparation, RNA preparation and nucleotide sequencing

Genomic DNA was prepared using DNA Extraction Kit (Stratagene, CA, USA). Plasmid DNA was prepared using

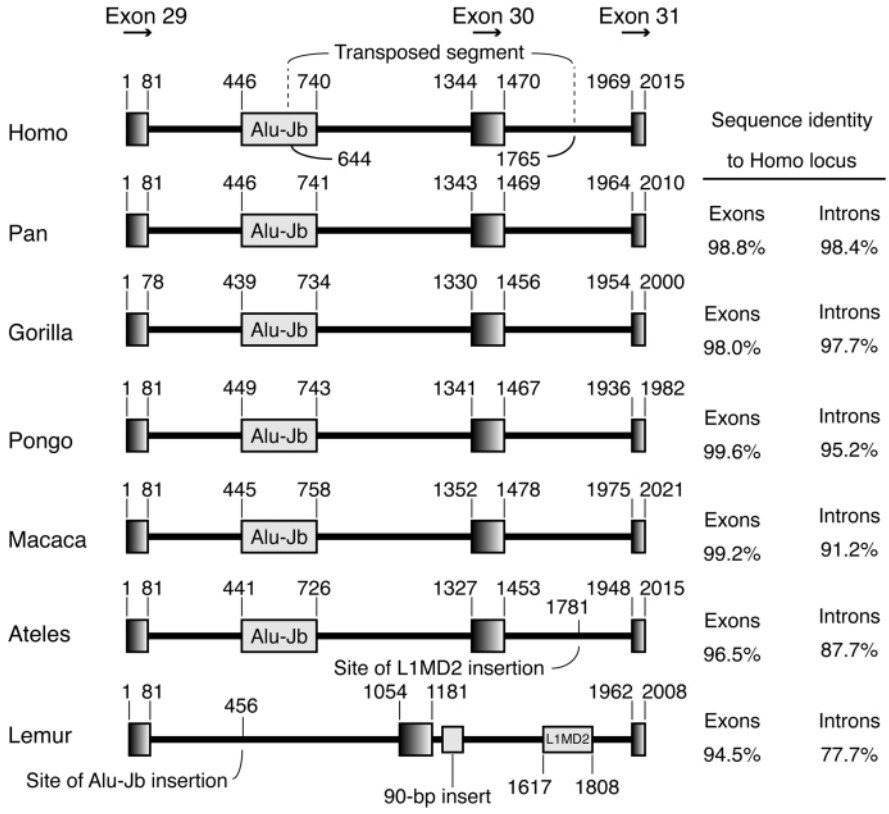

Figure 5. Comparison of the *ATM* loci (exons 29–31) among primates. Numbers give the nucleotide positions within the sequence of PCR products (29F+31R). The sequence identity between the human and the lemur in introns (77.7%) was estimated by excluding the regions of repetitive sequences (*Alu*-Jb, 90 bp insert and L1MD2).

QIAprep spin miniprep column (Qiagen GmbH). Total RNA was prepared using Trizol reagent (Invitrogen, CA, USA). Nucleotide sequencing was performed using BigDye Terminator Cycle Sequencing FS Ready Reaction Kit (Applied Biosystems, CA, USA) and a DNA sequencer (Applied Biosystems, model ABI Prism 310 Genetic Analyzer). Identification of interspersed repeats was performed by the CENSOR program (35).

Southern hybridization

The *ATM* cDNAs were synthesized by the RT–PCR method. Four sense and four antisense primers derived from the *ATM* mRNA sequence were used to synthesize four overlapping cDNA fragments (nucleotides 1–3033, 2414–5433, 5062–8091 and 7453–9171 of *ATM* ORF). Primer sequences and procedures for Southern blot analysis were described previously (36).

Library screening

Human chromosome-7-specific genomic library LA07NS01 constructed by cloning the *EcoR*I-digested DNA from hamster/human MR3.31 6TG6 hybrids into Charon 21A lambda phage vector (37) was purchased from American Type Culture Collection. The library was plated at a density of 10 000–20 000 plaques per plate, and 600 000 clones were screened with a 3.02 kb *ATM* cDNA probe (nucleotides 2414–5433). Secondary and tertiary screening were used to purify positive clones. Lambda DNA was prepared by using Lambda Quick! Spin Kit (BIO 101, CA, USA). Excised inserts were recloned in ZAP Express Vector (Stratagene, CA, USA), and liberated from the plaques with the ExAssist/SOLR *in vivo* excision protocol (Stratagene, CA, USA).

PCR

PCR amplification of reverse-transcribed cDNA, lambda phage insert or genomic DNA from non-human primates was performed by eLONGase amplification system (Invitrogen, CA, USA). For the amplification of phage insert DNA, a pair of Charon 21A vector primers (5′GTTGGCAGGGATATTCTGGC3′; 5′AGGACGTAGCCAGACGGAAC3′) was used. For the amplification of the region around the retrotransposed *ATM* insert, three primers (P1, 5′CTGCAATCTACTCTTCTGAC3′; P2, 5′ACATGAGGTCAAAAGTCTTC3′; and P3, 5′GAAGCACAGTAAAATCAAGC3′) were used. For the amplification of the region around exon 30 of the *ATM* locus, two primers (ATM29F, 5′TGTGGTGGAGTTATTGATGA3′; ATM31R, 5′TCAGCTGCTTGCTCACATAT3′) were used. The PCR products derived from non-human primate DNA were cloned in pGEM-T Easy Vector by TA cloning protocol (Promega, WI, USA).

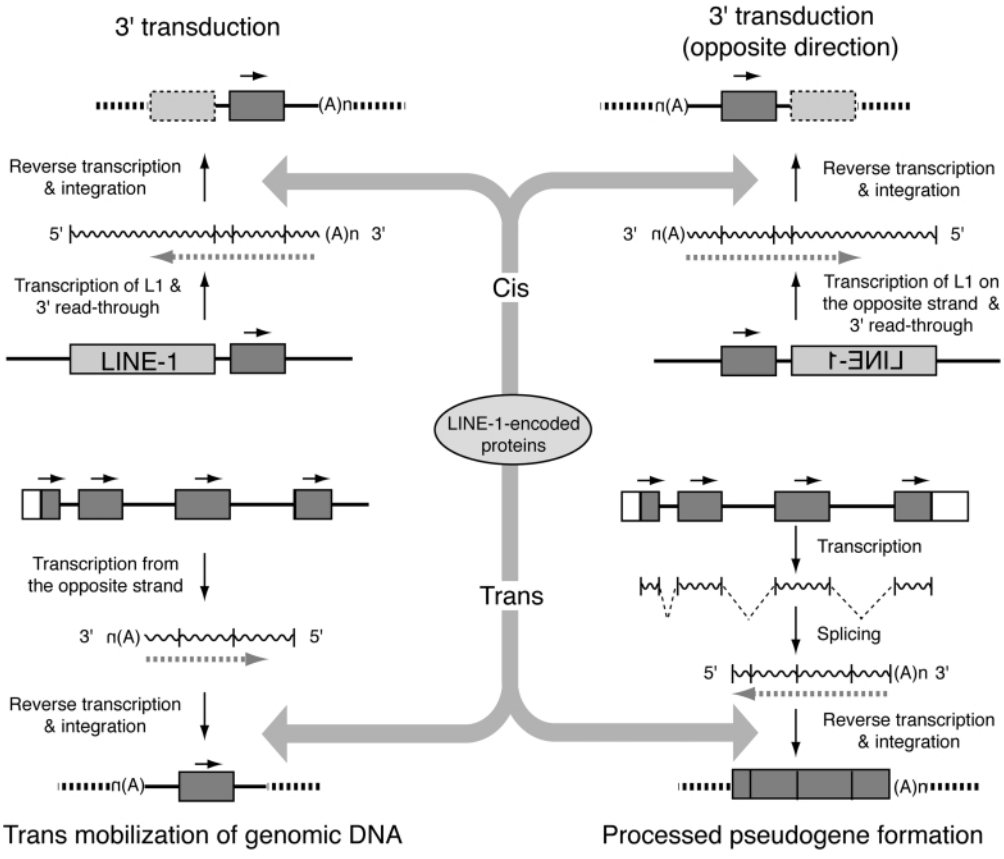

Figure 6. A model for how L1 retrotransposition can mobilize non-L1 sequences. In the *cis* (3′ transduction) pathway, the DNA located 3′ to an active L1 element is mobilized along with L1 retrotransposition. This results in the transduction of downstream exon (upper left). If an active L1 resides in the opposite orientation of the gene, the transduced exon has a poly(A) tail oriented opposite to its own transcriptional orientation (upper right). Retrotransposed L1 sequences are 5′ truncated (dot–lined squares, upper left and upper right) or, occasionally, completely lost. The action of L1 machinery in *trans* on cellular mRNA or antisense transcript results in the formation of intronless processed pseudogene (lower right), or the insertion of intron-containing DNA into a new genomic location (lower left). Arrows on exons indicate the transcriptional orientation of a functional gene. Shaded, thick dotted lines indicate reverse-transcribed cDNAs. Solid, thick dotted lines indicate the recipient genomic DNA where L1-mobilized sequences are integrated. $(A)_n$ denotes poly(A) tail.

ACKNOWLEDGEMENTS

This work was supported by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan to Y.E.

REFERENCES

1. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
2. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
3. Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J. and Eichler, E.E. (2001) Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.*, **11**, 1005–1017.
4. Eichler, E.E. (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.*, **17**, 661–669.
5. Bailey, J.A., Yavor, A.M., Viggiano, L., Misceo, D., Horvath, J.E., Archidiacono, M., Schwartz, S., Rocchi, M. and Eichler, E.E. (2002) Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *Am. J. Hum. Genet.*, **70**, 83–100.
6. Samonte, R.V. and Eichler, E.E. (2002) Segmental duplications and the evolution of the primate genome. *Nat. Rev. Genet.*, **3**, 65–72.
7. Gilbert, W. (1978) Why genes in pieces? *Nature*, **271**, 501.
8. Patthy, L. (1999) Genome evolution and the evolution of exon-shuffling—a review. *Gene*, **238**, 103–114.
9. Jones, J.M., Huang, J.-D., Mermall, V., Hamilton, B.A., Mooseker, M.S., Escayg, A., Copeland, N.G., Jenkins, N.A. and Meisler, M.H. (2000) The mouse neurological mutant flailer expresses a novel hybrid gene derived by exon shuffling between *Gnb5* and *Myo5a*. *Hum. Mol. Genet.*, **9**, 821–828.
10. Moran, J.V., DeBerardinis, R.J. and Kazazian, H.H. Jr. (1999) Exon shuffling by L1 retrotransposition. *Science*, **283**, 1530–1534.
11. Goodier, J.L., Ostertag, E.M. and Kazazian, H.H. Jr. (2000) Transduction of 3′-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.*, **9**, 653–657.
12. Pickeral, O.K., Makalowski, W., Boguski, M.S. and Boeke, J.D. (2000) Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.*, **10**, 411–415.
13. Kazazian, H.H. Jr. and Moran, J.V. (1998) The impact of L1 retrotransposons on the human genome. *Nat. Genet.*, **19**, 19–24.
14. Smit, A.F.A. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.*, **9**, 657–663.
15. Kazazian, H.H. Jr. (2000) L1 retrotransposons shape the mammalian genome. *Science*, **289**, 1152–1153.
16. Esnault, C. Maestre, J. and Heidmann, T. (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.*, **24**, 363–367.
17. Boeke, J.D. (1997) LINEs and *Alus*—the polyA connection. *Nat. Genet.*, **16**, 6–7.
18. Long, M. and Langley, C.H. (1993) Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila. Science*, **260**, 91–95.

19. Long, M. (2000) A new function evolved from gene fusion. *Genome Res.*, **10**, 1655–1657.
20. Sedgwick, R.P. and Boder, E. (1991) Ataxia–telangiectasia. In Vinken, P.J., Bruyn, G.W. and Klawans, H.L. (eds), *Handbook of Clinical Neurology*, Vol. 16. Elsevier Science, Oxford, pp. 347–423.
21. Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D.A., Smith, S., Uziel, T., Sfez, S. *et al.* (1995) A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science*, **268**, 1749–1753.
22. Platzer, M., Rotman, G., Bauer, D., Uziel, T., Savitsky, K., Bar-Shira, A., Gilad, S., Shiloh, Y. and Rosenthal, A. (1997) Ataxia–telangiectasia locus: sequence analysis of 184 kb of human genomic DNA containing the entire *ATM* gene. *Genome Res.*, **7**, 592–605.
23. Cost, G.J. and Boeke, J.D. (1998) Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry*, **37**, 18081–18093.
24. Ostertag, E.M. and Kazazian, H.H. Jr. (2001) Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.*, **11**, 2059–2065.
25. Rozmahel, R., Heng, H.H.Q., Duncan, A.M.V., Shi, X.-M., Rommens, J.M. and Tsui, L.-C. (1997) Amplification of *CFTR* exon 9 sequences to multiple locations in the human genome. *Genomics*, **45**, 554–561.
26. Viale, A., Ortola, C., Richard, F., Vernier, P., Presse, F., Schilling, S., Dutrillaux, B. and Nahon, J.-L. (1998) Emergence of a brain-expressed variant melanin-concentrating hormone gene during higher primate evolution: a gene 'in search of a function'. *Mol. Biol. Evol.*, **15**, 196–214.
27. Courseaux, A. and Nahon, J.-L. (2001) Birth of two chimeric genes in the *Hominidae* lineage. *Science*, **291**, 1293–1297.
28. Boissinot, S., Entezam, A. and Furano, A.V. (2001) Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.*, **18**, 926–935.
29. Van den Eynde, B.J., Gaugler, B., Probst-Kepper, M., Michaux, L., Devuyst, O., Lorge, F., Weynants, P. and Boon, T. (1999) A new antigen recognized by cytolytic T lymphocytes on a human kidney tumor results from reverse strand transcription. *J. Exp. Med.*, **190**, 1793–1799.
30. Sun, X., Li, X., Moriarty, P.M., Henics, T., LaDuca, J.P. and Maquat, L.E. (2001) Nonsense-mediated decay of mRNA for the selenoprotein phospholipid hydroperoxide glutathione peroxidase is detectable in cultured cells but masked or inhibited in rat tissues. *Mol. Biol. Cell*, **12**, 1009–1017.
31. Cost, G.J., Feng, Q, Jacquier, A. and Boeke, J.D. (2002) Human L1 element target-primed reverse transcription *in vitro*. *EMBO J.*, **21**, 5899–5910.
32. Chang, D.-Y., Nelson, B., Bilyeu, T., Hsu, K., Darlington, G.J. and Maraia, R.J. (1994) A human *Alu* RNA-binding protein whose expression is associated with accumulation of small cytoplasmic *Alu* RNA. *Mol. Cell. Biol.*, **14**, 3949–3959.
33. Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Jr., Boeke, J.D. and Moran, J.V. (2001) Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Mol. Cell. Biol.*, **21**, 1429–1439.
34. Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D. and Kazazian, H.H. Jr. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.
35. Jurka, J., Klonowski, P., Dagman, V. and Pelton, P. (1996) CENSOR—a program for identification and elimination of repetitive elements from DNA sequences. *Comput. Chem.*, **20**, 119–121.
36. Ejima, Y. and Sasaki, M.S. (1998) Mutations of the *ATM* gene detected in Japanese ataxia–telangiectasia patients: possible preponderance of the two founder mutations 4612del165 and 7883del5. *Hum. Genet.*, **102**, 403–408.
37. Rommens, J.M., Zengerling, S., Burns, J., Melmer, G., Kerem, B.S., Plavsic, N., Zsiga, M., Kennedy, D., Markiewicz, D., Rozmahel, R. *et al.* (1988) Identification and regional localization of DNA markers on chromosome 7 for the cloning of the cystic fibrosis gene. *Am. J. Hum. Genet.*, **43**, 645–663.