

Transduction of the Human Gene *FAM8A1* by Endogenous Retrovirus During Primate Evolution

Stéphane Jamain,¹ Marc Girondot,² Pascale Leroy,³ Michel Clergue,¹ Hélène Quach,¹ Marc Fellous,¹ and Thomas Bourgeron^{1,*}

¹Laboratoire d'Immunogénétique Humaine, INSERM E021, Institut Pasteur, 25, rue du Docteur Roux, 75724 Paris Cedex 15, France

²Laboratoire d'Ecologie, Systématique et Evolution, Université Paris-Sud, CNRS UPRESA 8079, 91405 Orsay Cedex, France

³INSERM 364 Faculté de Médecine, Avenue de Valombrose, 06107 Nice Cedex 2, France

*To whom correspondence and reprint requests should be addressed. Fax: (33) 1 40 61 31 53. E-mail: thomasb@pasteur.fr.

Capture of cellular mRNA by mobile elements has been an evolutionary catalyst for the spread of genes and a cause of cancer development. Here we present evidence that an orphan gene, *FAM8A1* (family with sequence similarity 8), was captured by a retrovirus, followed by multiple retrotransposition events, during primate evolution between 45 and 58 million years ago. This represents the first record of cellular mRNA transduction in humans. The human gene is localized on chromosome 6p23 with five related pseudogenes (*FAM8A2P–A6P*), each inserted within a human endogenous retrovirus (HERV). Only the functional *FAM8A1* gene is expressed and displays a ubiquitous mRNA and a testis-specific transcript present in the haploid phase of spermatogenesis. The structural features of the *FAM8A1* pseudogenes include two short sequences of similarity between the *FAM8A1* mRNA and the HERV sequences at both the 5' and 3' integration sites. These hallmarks suggest an alternative model to account for the capture of *FAM8A1* cellular mRNA by HERV-K, involving illegitimate recombination events at the two sites of sequence similarity during reverse transcription. Unlike previous models, which assume at least one step of retroviral integration in the genome, our model is consistent with *in vitro* observations showing that multiple template switches occur among packaged viral transcripts. This leads to the speculation that, in some cases, cellular mRNAs may have been captured through similar processes involved in the retroviral life cycle.

Key Words: retrovirus, HERV, AHCP, transduction

INTRODUCTION

During evolution, 10–40% of the mammalian genome may have arisen by retrotransposition events or been derived from mobile elements [1–3]. This retrotransposition process can be achieved by different mechanisms involving retroviruses or other repetitive elements present in the genome and through capture of cellular mRNAs by retroviruses, a process termed transduction [4]. Among the repetitive elements, human endogenous retrovirus (HERV) constitutes about 8% of human DNA [5,6]. It shows the same basic genome organization as sequence infectious retroviruses, possessing regions with sequence similarity to the long terminal repeats (LTRs) and major open reading frames (ORFs) of retroviruses—namely, *gag* (encoding viral core proteins), *pol* (encoding reverse transcriptase), and *env* (encoding envelope proteins). ERVs are encoded within the genomes of all higher

eukaryotes and apparently arose from the infection of germ cells by exogenous retrovirus [7]. These infections generally occurred in the distant past and most ERVs have undergone inactivating mutations or deletions over time. However, some HERV-K (K denotes a lysine tRNA primer binding site) encode functional forms of several viral proteins [5,8]. The HERV-K family is represented in the human genome by 8000 more or less full-length proviruses and solitary LTRs [6,7,9]. Their biological relevance remains largely unknown but the capacity of HERV to undergo the cycle of retrotransposition may represent a source of mutation in humans.

Here, we present evidence indicating that an orphan human gene, *FAM8A1* (family with sequence similarity 8; originally called AHCP for autosomal highly conserved protein), was captured by a retrovirus during primate evolution through a process that resembles oncogene transduction [4]. This gene, localized on chromosome 6p23, produced multiple, closely related, processed pseudogenes (*FAM8A2P–A6P*),

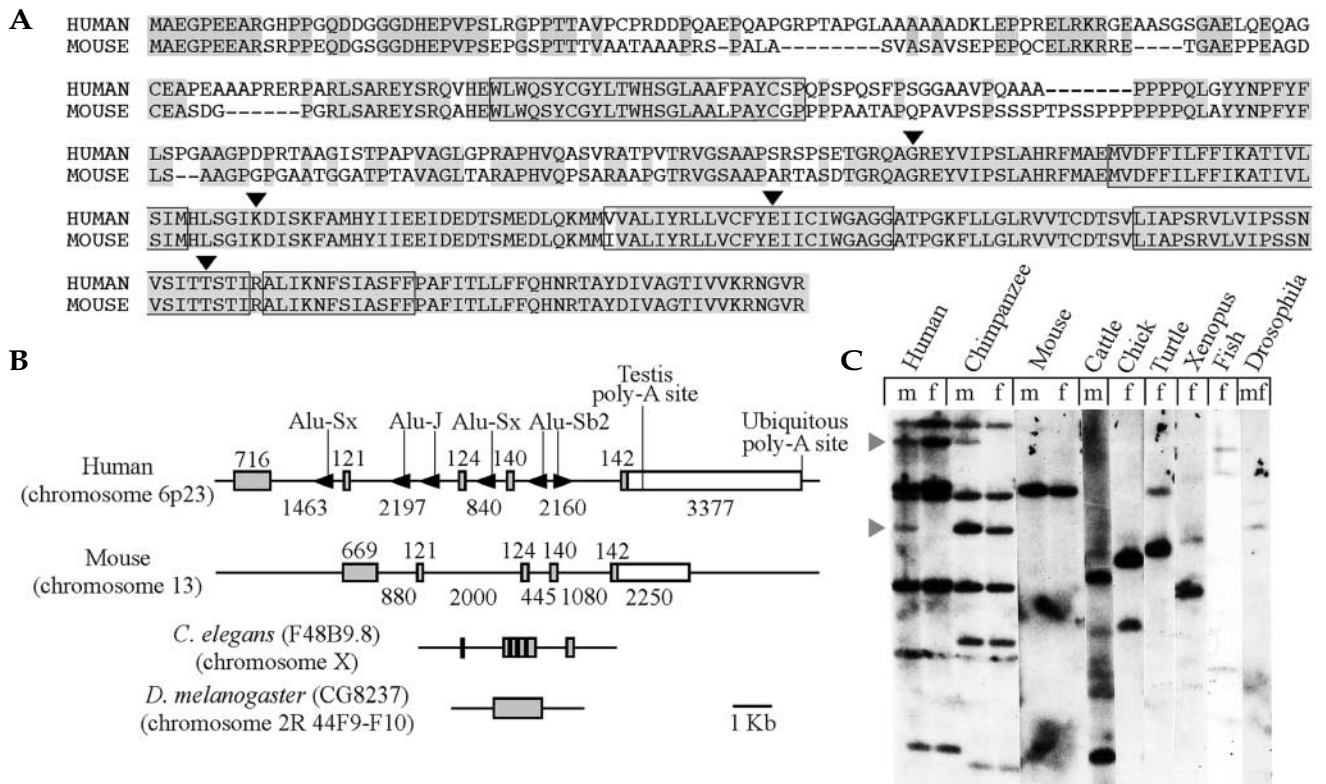


FIG. 1. Sequence, organization, and conservation of *FAM8A1*. (A) Alignment of human and mouse *FAM8A1* protein. Sequences in boxes are putative internal membrane domains. Exon-intron junctions are indicated by arrowheads. (B) Genomic organization of *FAM8A1* in different species. (C) Zoo blot. For each sample, 10 μ g of *Eco*RI-digested DNA was used and the membrane was hybridized with human probe *FAM8A1* F1R1. m, Male; f, female.

each inserted in HERV-K sequences. Considering the hallmarks of *FAM8A1* pseudogene insertion within the HERV, we suggest an alternative model of cellular mRNA transduction that involves two illegitimate recombination events during reverse transcription. In addition, the presence of different *FAM8A1* pseudogenes suggests that this retroelement (ERV + *FAM8A1* pseudogenes) has been copied and inserted in different places in the genome by retrotranscription. This is the first record of cellular mRNA transduction in humans, but other genes may be captured by this process during the HERV life cycle, representing a potential cause of mutation and disease.

RESULTS

Isolating the *FAM8A1* Gene

We cloned *FAM8A1* with the p12f probe of the human Y chromosome [10]. In this probe, we identified *FAM8A4P*, a retrotranscribed pseudogene of *FAM8A1*. To characterize the functional gene, we sequenced the human expressed sequence tag (EST) clone 647960 from the IMAGE consortium presenting similarity with the pseudogene. We obtained the full-length cDNA by 5' rapid amplification of cDNA ends (RACE)

polymerase chain reaction (PCR) on human testis RNA. This 4.7-kb *FAM8A1* cDNA contained a single 1242-bp ORF predicted to encode a 413-amino-acid protein with a molecular mass of 44 kDa (Fig. 1A). We identified the genomic structure of *FAM8A1* by long-range PCR. The size of the human gene is about 12 kb, containing five exons (Fig. 1B).

Evolutionary Conservation of the *FAM8A1* Genes

We hybridized a zoo blot with the cDNA probe and observed positive signals in all animals (Fig. 1C). However, we observed male-specific bands, corresponding to *FAM8A4P*, only in humans and chimpanzees. The complete sequence of the mouse protein has revealed a highly conserved carboxy-terminal domain of the protein encoded by the last four exons (162/163 conserved amino acids) and a less conserved amino-terminal domain encoded by the first exon (57% identity and several gaps). The N-terminal part is a proline-rich polypeptide (about 17% prolines) without a consensus signal peptide. The C-terminal part contains four putative internal membrane domains conserved in both species. Comparison of the *FAM8A1* sequence with other protein sequences in the database shows homology with putative proteins from *Caenorhabditis elegans* (F48B9.8) and *Drosophila melanogaster* (CG8237) and the gene is present in

the genome of the actinopterygian *Tetraodon nigroviridis* (033K23). Homologous cDNA is also present in an EST library from zebrafish olfactory rosettes (AW233323) and fertilized egg from the ascidian *Halocynthia roretzi* (AV382432). However, there is no homology to ESTs from plants and unicellular organisms. Finally, using the Genebridge 4 radiation hybrid panel [11], we localized human *FAM8A1* on chromosome 6p23 in a 2-cM region between *D6S274* and *D6S285* (logarithm of odds > 3.0). In the mouse, the orthologous gene corresponds to the DNA segment D13Pas1 previously localized on chromosome 13 (26 cM) with a polymorphic *SacI* site between *Mus spretus* and C57BL/6 mice [12]. The *C. elegans* F48B9.8 gene is localized on the X chromosome and the *D. melanogaster* CG8237 gene is localized on chromosome 2R 44F9-F10.

Expression of the *FAM8A1* Gene

By northern blot analysis, we detected a transcript of about 5.5 kb in all human tissues tested; an additional mRNA of about 1.6 kb was present in a high amount only in testis (Fig. 2A). We observed the same profile with mouse tissues (Fig. 2B). *In situ* hybridization of mouse testis shows that *FAM8A1* mRNA is strongly expressed in round spermatids (Fig. 2C).

Sequence analysis of both human and mouse cDNAs has shown that the shorter testis-specific transcript of *FAM8A1* results from a different polyadenylation site at the 3' end of the mRNA (Fig. 2D). The poly(A) site used for the testis-specific transcript is not the canonical site AAUAAA, but AUUAAA conserved in the mouse mRNA. In male germ cells, many mRNAs do not have AAUAAA but are efficiently polyadenylated [13].

Characterization of Human *FAM8A1* Pseudogenes

The *FAM8A1* family consists of a minimum of five pseudogenes localized on chromosomes 2 (*FAM8A3P*), 6p21 (*FAM8A5P*), 6q23 (*FAM8A6P*), 11 (*FAM8A2P*), and Y (*FAM8A4P*). These pseudogenes are inserted in a retrovirus with similarity to the consensus sequence of HERVK14 flanked by two LTR14A of 344 bp (Fig. 3A). HERVK14 and LTR14A correspond to the nomenclature used by the repbase database [14]. HERVK14 codes for gag, pro, and pol proteins and shares on average 65% nucleotide identity with the reported members of HERVK-superfamily (HERVK10, HERVKC4). The site of *FAM8A1* pseudogene insertion within the HERV sequence is common to all *FAM8A1* pseudogenes along the genome. Using the consensus sequence of HERVK14 as a standard, *FAM8A1* pseudogenes are inserted at nucleotide 363, creating a 1466-bp deletion of the gag gene. At both 5' (S1) and 3' (S2) insertion sites, there are sequence similarities between the consensus HERVK14 and the *FAM8A1* mRNA (Fig. 3A). All retroposons contain different deletions of the pol gene. The 3' ends of *FAM8A1* pseudogenes are downstream from the testis-specific poly(A) site, which indicates that pseudogenes originate from the ubiquitous transcript instead of from the testis-specific mRNA.

None of the *FAM8A1* pseudogenes contain a complete *FAM8A1* ORF. To determine whether the pseudogenes were transcribed, we used a panel of cDNAs from different human tissues to amplify putative *FAM8A1* pseudogene transcripts. We performed several PCRs with different sets of primers; an example of two PCR experiments is presented in Fig. 3B. Primers used for PCR1 are located in the *FAM8A1* pseudogene sequence and should amplify *FAM8A1* pseudogene transcripts as well as the functional *FAM8A1* mRNA.

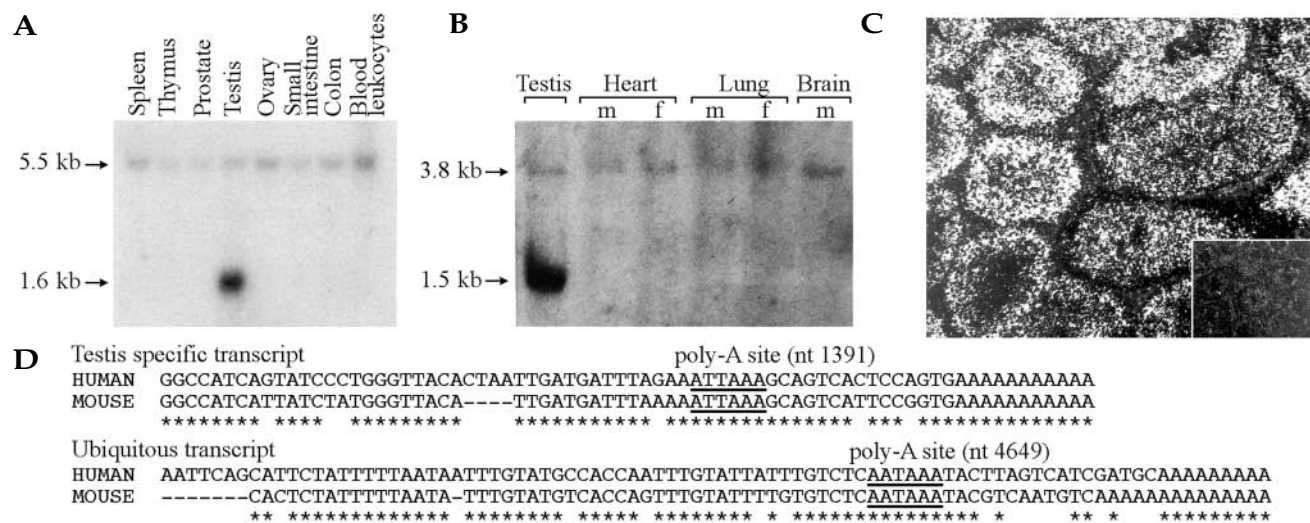
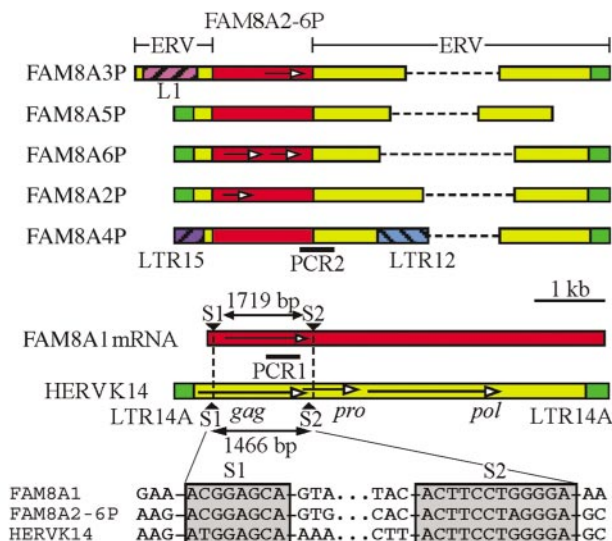
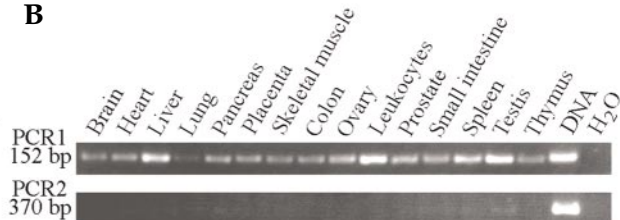


FIG. 2. Characteristics and expression of human and mouse *FAM8A1* mRNA. (A) Northern blot analysis of human *FAM8A1* mRNA. (B) Northern blot analysis of mouse *FAM8A1* mRNA. m, Male; f, female. (C) *In situ* analysis of *FAM8A1* in mouse testis. Slides were hybridized with the mouse antisense (C) or sense (inset) *FAM8A1* FIR1 probe. (D) Sequences of the 3' end of the testis and the ubiquitous transcripts in human and mouse.

A



B



C

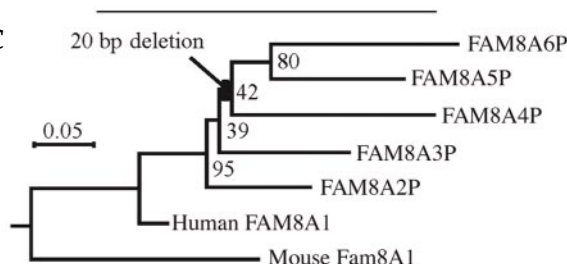


FIG. 3. Characteristics, expression, and phylogeny of *FAM8A1* pseudogenes. (A) Genomic organization of *FAM8A1* pseudogenes. Each pseudogene was inserted in a retrovirus with similarity to HERVK14 flanked by two LTR14A. S1 and S2 correspond to sites of similarity between *FAM8A1* mRNA and HERV sequences. Open arrows correspond to the putative ORF. PCR1 and PCR2 indicate positions of primers used for expression analysis. (B) RT-PCR analysis of the functional gene and related pseudogenes. PCR1 should amplify *FAM8A1* pseudogenes and *FAM8A1* mRNAs. PCR2 is specific to the *FAM8A1* pseudogenes. (C) Bio-NJ tree of exon 1 sequences by using the Tajima and Nei substitution rate [38]. Values at the node are percent bootstraps for 1000 samplings. Black bar shows the presence of a strong synapomorphic signature. *FAM8A3P*, *FAM8A5P*, *FAM8A6P*, *FAM8A2P*, and *FAM8A4P* are located in human genomic clones under acc. nos. AC007394, AC026010, AL121970, AC022882, and AC002992, respectively.

For PCR2, the reverse primer is located in the flanking HERV sequences and so is specific for the potential *FAM8A1* pseudogene transcripts. As expected, with genomic DNA, the positive amplification of both PCRs indicates the presence of the retrotranscribed pseudogenes in the genome. However, direct sequence and restriction enzyme analysis of all the reverse transcription (RT) PCR1 products revealed a homogenous sequence identical to the functional gene. In addition, we could not amplify the PCR2 product by using the cDNA as template. These results indicate that only the gene localized on chromosome 6p23 is expressed in normal human tissues.

Phylogenetic Relationship Among the *FAM8A1* Pseudogene Sequences

A BLAST search reveals that the flanking retroviral sequences are related to HML (human endogenous MMTV (mouse mammary tumor virus)-like) retrovirus. To define the family of the HERV flanking the *FAM8A1* pseudogenes, we included in the HML phylogeny a 265-bp sequence corresponding to the *pol* gene [15]. HERVK14 retrovirus and the retroviral part of *FAM8A1* pseudogenes clustered in 99% of the replicates with HML-1 sequences (data not shown). Sequences corresponding to exon 1, showing the highest substitution rate, have been used for *FAM8A1* pseudogene phylogeny. In the *FAM8A1* tree (Fig. 3C), two nodes are well supported. First, the monophyly of *FAM8A1* is observed in 95% of the replicates; second, *FAM8A5P* and *FAM8A6P* clustered together in

80% of the replicates. The grouping of *FAM8A4P*, *FAM8A5P*, and *FAM8A6P* sequences is not well supported (42%) but these three sequences shared a strong synapomorphic signature (20 bases deleted), which makes their monophyly very likely. However, the relative positions of *FAM8A3P* and *FAM8A2P* are not well supported and remain unresolved. In further analyses, both solutions will be systematically tested. The maximum-likelihood tree gives rise to the same conclusion and the relative positions of *FAM8A3P* and *FAM8A2P* do not significantly change the likelihood of the tree (analysis using PAML 1.3) [16].

Date of *FAM8A1* Pseudogene Spread

Using different sets of primers, we were able to amplify *FAM8A1* pseudogenes in the Old World monkey family (genera *Macaca*, *Cercopithecus*, and *Papio*). However, we could not amplify *FAM8A1* pseudogenes by using DNA from capuchin, a New World monkey (four independent samples: two males and two females). To trace the phylogeny of retroviruses, Johnson and Coffin provide a method based on the divergence of their LTRs [17]. However, this method is unsuitable in our case because the 3' LTR is subject to recombination. The log-likelihood of a tree using only primate autosomal *FAM8A1* pseudogene sequences is -1850.4 with and -1843.7 without a molecular clock; the difference is not statistically significant (LRT test, $\chi^2 = 13.34$, 9 degrees of freedom, $P > 0.1$). The presence of a molecular clock cannot be rejected. The divergence date between

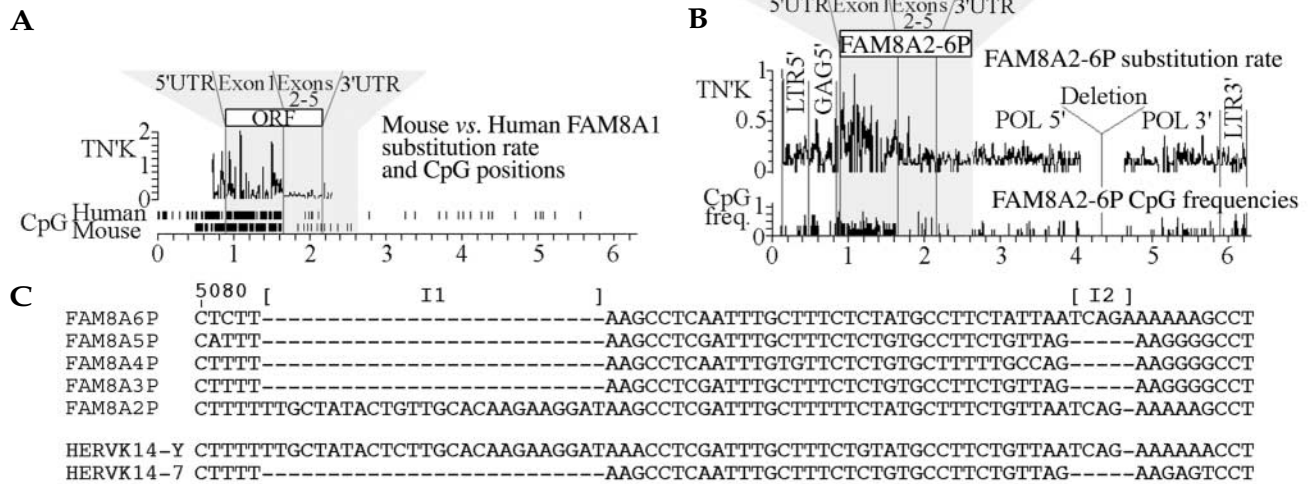


FIG. 4. Sequence analysis of *FAM8A1* and *FAM8A1* pseudogenes. (A) Substitution rate for human versus mouse functional genes. (B) Substitution rate for the different *FAM8A1* pseudogenes. Substitution rates based on the Tajima and Nei method (TNK) are calculated by using a 20-base window. The presence of CpG along the sequences is indicated by lines. (C) Flanking retroviral sequences of the five *FAM8A1* pseudogenes and HERVK14-Y and HERVK14-7 that exhibit two indels [I1] and [I2], showing evidence of recombination.

Hominoidea (genus *Homo*) and Cercopithecoidea is 31 million years [18,19]; consequently, the spread of *FAM8A1* pseudogenes is estimated to have occurred 44.86 million years ago (MYA). This date is coherent with the observation that *FAM8A1* pseudogenes do not exist in Platyrrhini (New World monkeys), which diverged from Catarrhini (Old World monkeys) 57.5 MYA [19]. Thus, the capture and amplification of the cellular *FAM8A1* mRNA probably occurred during the major waves of HERV-K expansion, which happened after the New World monkey and before the Old World monkey branching [7].

Substitution Pattern and Evidence for Gene Conversion in *FAM8A1* Pseudogene Sequences

When mouse and human functional genes were compared, exons 2-5 seem to be more conserved than exon 1, based on substitution pattern (Fig. 4A). Two different phenomena could account for this observation: (1) the first exon is less functionally constrained; or (2) the first exon is more susceptible to mutations. To discern between these possibilities, we calculated two series of substitution rates by using a 20-base sliding window, one with the mouse and human functional *FAM8A1* sequences (Fig. 4A) and another with the nonfunctional *FAM8A1* sequences (Fig. 4B). Correlation of the substitution rate for each homologous position in the alignment between functional and nonfunctional *FAM8A1* sequences was highly significant ($r = 0.365$, $P < 0.0001$). This correlation indicates that a difference in mutation rate is sufficient to explain the differences in substitution rate observed between exons 1 and 2-5 of the functional genes. These differences are probably due to the occurrence of a

CpG island located within exon 1 generating mutational instability.

A phylogeny of the retroposons can be obtained with their internal *FAM8A1* pseudogene sequences (Fig. 3C), but the phylogeny becomes unresolved when flanking ERV sequences are added. The lack of resolution of a tree with complete sequences (*FAM8A1* pseudogene + ERV) is due to the very low substitution rate in most sequences (Fig. 4B), but it could also result from recombination events between the flanking retroviral sequences and other HERVK14 sequences among the genome. To detect such events, we tested whether the base partition along the *FAM8A1* pseudogene sequences is compatible with the phylogeny inferred from exon 1 (Fig. 3C), which is considered to be the true phylogeny. Two regions were detected as being incompatible with the exon 1 phylogeny ($P < 0.05$). These sequences have been checked and evidence of gene conversion has been detected. For example, the insertion (I1) observed in *FAM8A2P* is absent in all other *FAM8A1* pseudogene sequences and matches exactly an insertion found in a HERVK14 sequence located on the Y chromosome (Fig. 4C). The absence of the insertion is found in several other HERVK14s—for example, a sequence located on chromosome 7. In this case, the ancestral state cannot be determined because the position of the *FAM8A2P* sequence is ambiguous (Fig. 3C), but, whatever the solution, at least one recombination event is necessary. The second insertion (I2) is also observed in *FAM8A2P* and HERVK14-Y as well as in *FAM8A6P* and a second event of recombination is necessary to explain such a repartition.

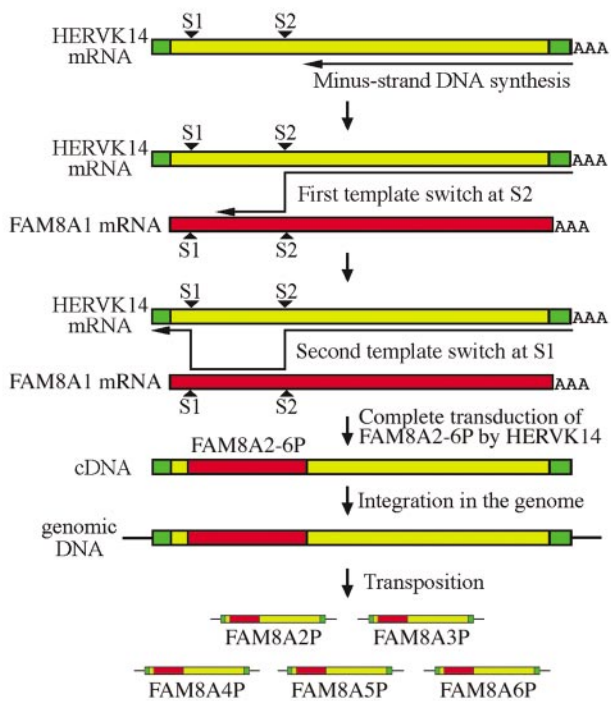


FIG. 5. Model for *FAM8A1* transduction showing postulated events in the creation of *FAM8A1* pseudogenes. *FAM8A1* mRNA and HERVK14 undergo illegitimate recombination during reverse transcription. The first strand switch of the reverse transcription growing point from the retrovirus to the cellular *FAM8A1* mRNA is located at the site with nucleotide identity S2. The second strand switch occurs at the site of similarity S1. After complete *FAM8A1* transduction, the ancestral retroposon (ERV + *FAM8A1* pseudogene) is inserted in the genome. Therefore, cycles of transposition account for multiplication of the *FAM8A1* pseudogenes.

DISCUSSION

Two different mechanisms are involved in the formation of retrotranscribed pseudogenes. First, it has been shown that retroviral proteins can encapsidate RNAs that do not contain retroviral *cis*-acting sequences [20,21]. Such RNAs are reversed transcribed and inserted into the genomes of infected target cells to form cDNA genes. Investigations of such cDNA genes have shown that they are truncated at both the 3' and the 5' ends, do not contain a poly(A) tract, and are not flanked by direct repeats [21]. However, most processed pseudogenes contain the full-length mRNA and a poly(A) sequence and are flanked by direct repeats. For these pseudogenes, LINE elements rather than ERVs are implicated [22].

The formation of *FAM8A1* pseudogenes cannot be accounted for by either of these previous mechanisms but can be referred to the well-known phenomenon of transduction. Capture of cellular mRNA by retroposons has been characterized in different species such as rodents and birds but, to our knowledge, has never been observed in humans [4]. In the

human genome, only two examples of retroviruses have been found to contain an extra ORF in addition to *gag*, *pol*, and *env* genes. First, a doubly spliced HERV-K has been shown to encode K-Rev, a functional homologue of the HIV-1 Rev proteins [23]. However, this protein is of viral origin. Second, some HERV-K also encode a dUTPase located within the *gag* gene and expressed through a ribosomal frameshift. This gene is also of viral origin [24].

Two different hypotheses, referred to as the DNA and RNA models, have been proposed to illustrate the capture of cellular mRNA in retrovirus [25–27]. Both models assume that the 5' virus/cellular gene recombination occurs at the DNA level and results from random integration of a DNA provirus at the 5' region or within an intron of the cellular gene. Considering the 3' recombination, the models diverge. The DNA model predicts that 3' recombination occurs between viral and cellular DNA [27]. Thus, transduction results from a second genomic integration of a complete retrovirus 3' of the coding region of the cellular gene. By contrast, the RNA model requires the formation of a chimeric viral/cellular RNA. This hybrid RNA results either from transcription of DNA after a deletion that fuses 5' viral sequences to cellular sequences or from readthrough transcription. Then, 3' recombination results from a strand switch of the reverse transcription growing point from an infectious retrovirus to the chimeric RNA [25]. This single template switch usually occurs at a site with a small stretch of sequence homology [28].

Interestingly, two stretches of similarity between the HERVK14 and the *FAM8A1* mRNA are present 5' and 3' of *FAM8A1* pseudogene insertion sites (Fig. 3A). This suggests a simple alternative hypothesis to account for *FAM8A1* transduction (Fig. 5). Our model does not require the necessary step shared by the DNA and RNA models, which consists of retrovirus integration at the 5' end or within an intron of the target gene. On the contrary, transduction would take place at the RNA level only (Fig. 5). *FAM8A1* mRNA, expressed from its own promoter, might undergo illegitimate recombination with a wild-type HERVK14 mRNA during reverse transcription. The first strand switch of the reverse transcription growing point from the retrovirus to the cellular *FAM8A1* mRNA should be located at the 3' end of *FAM8A1* at the site with nucleotide similarity (S2) as in the RNA model. A second strand switch is predicted to occur at the 5' site of similarity (S1). Extension of this DNA would produce the capture of cellular mRNA surrounded by retroviral sequences on both sides. Finally, after integration, retroposition events account for amplification of the *FAM8A1* pseudogenes within the genome.

This copackaging model was established exclusively on the basis of sequence and phylogenetic analyses of the *FAM8A1* pseudogenes. However, unlike the RNA model, this model requires only one additional strand switch. Multiple template-switching events have been shown to occur during retroviral recombination at a frequency higher than expected by chance [29]. Furthermore, we tested a similar model *in vitro* and recombination between the two templates occurred with an apparent frequency of 10^{-4} to 10^{-5} per replication cycle in the absence

of homology between the two recombining partners. This frequency increased at least 100-fold if homology was provided at the site of recombination [30]. Thus, we suggest that some cellular mRNA transduction events may now be reinterpreted under this new hypothesis by systematically checking potential short stretches of 5' and 3' sequence similarity between the target mRNA and the retrovirus. Furthermore, other cellular mRNAs may have been captured by this process and the complete genome sequence of different organisms will be useful for detecting such events.

The monophyly of *FAM8A1* pseudogenes indicates that duplication of the pseudogenes arose from one ancestral retroposon (*FAM8A1* pseudogene + ERV)—that is, they did not result from novel retrotranscription events involving the functional *FAM8A1* gene. Moreover, the different *FAM8A1* pseudogenes were produced by retroposition rather than by genomic duplications. This is indicated by the fact that the genomic sequences flanking the LTRs are clearly unrelated and by the existence of 5-bp repeats, still present on both sides of the complete retroposon *FAM8A6P* on chromosome 6q23. These repeats are the universal signature of a transposition [4].

Sequence analysis of the different retroelements (ERV + *FAM8A1* pseudogene) compared with other HERV-K along the genome revealed different recombination events within the flanking retroviral sequences. Gene conversion of retroviral sequences with endogenous sequences has been observed *in vitro* with retrovirus vectors deleted for the U3 region of the 3' LTR [31]. This mechanism, resulting in a sequence similarity of the flanking retroviral sequences that is greater than the internal *FAM8A1* pseudogene sequences, illustrates the concerted evolution of repetitive DNA in eukaryotes [32,33]. Homogenization of repetitive sequence arrays within a species is thought to take place via the mechanism of molecular drive and probably involves unequal chromatid (or chromosome) exchange, gene conversion, or transposition. In this study, with *FAM8A1* pseudogenes and their flanking retroviral sequences, one can observe the concerted evolution of the repetitive sequences (ERV) compared with independent evolution of the low-copy sequences (*FAM8A2P-A6P*).

Transduction is known to give rise to oncogene formation [4]. However, there is no evidence that *FAM8A1* is a cellular oncogene and that captured *FAM8A1* pseudogenes are oncogenic. Northern blot and RT-PCR analyses have shown that none of the identified *FAM8A1* pseudogenes appear to be transcribed in normal human tissues. In particular, *FAM8A1* pseudogene transcripts were not detectable in liver and lung, in contrast to the high expression of some retroviruses of the HML-1 family in those organs [15]. However, the possibility that individual *FAM8A1* pseudogenes could be reactivated during cancerogenesis cannot be excluded.

MATERIALS AND METHODS

DNA and RNA analysis. DNA and total RNA were prepared from the different animal tissues by standard procedures. Southern and northern blots were

done as described [34]. Total RNAs were extracted and reverse transcribed with the Gene Amp RNA PCR kit (Perkin-Elmer Corp). The last four exons of *FAM8A1* cDNA were amplified with primers *FAM8A1F1* (5'-AGGCAGGCCA-GAGAATATGTTATTC-3') and *FAM8A1R1* (5'-GGATACGTATGGC-CITGATAC-3'). After 35 cycles of PCR (95°C, 30 s; 50°C, 30 s; 72°C, 1 min), amplification products were used as a probe. The probe was labeled with [³²P]dCTP using a random priming kit (Boehringer Mannheim). Sequencing was done with fluorescently labeled *Taq* dideoxy terminator reaction mix and determined with a 373A automated DNA sequencer (Applied Biosystems). *In situ* hybridization was done as described [35].

Localization of *FAM8A1*. *FAM8A1* was mapped with the Genebridge 4 radiation hybrid panel [11]. Two PCR products were amplified by using primers in the 3' UTR of the *FAM8A1* cDNA (*FAM8A1THO29*, 5'-CTGCAGTTC-CTCTCGTAATG-3'; *FAM8A1THO28*, 5'-TGTGTGCACAGTTTAAACGT-3') or in intron 1 of *FAM8A1* (*DL2THO1*, 5'-TCTATATGCCATGTATTGAGC-3'; *DL2THO4*, 5'-CAGGCTTGATCTTGACTCTG-3'). The primers amplify human *FAM8A1*-specific products of 954 and 186 bp, respectively. The conditions of the PCR were 35 cycles of 30 s at 94°C, 30 s at 60°C, and 2 min at 72°C with 50 ng of DNA.

Sequence analysis. All sequences were analyzed with the Repeat masker program (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>). Sequences were aligned by using ClustalX 1.64 [36] and further checked by hand. Phylogenetic relationships between the sequences were analyzed by two methods: BIO-NJ algorithm of tree reconstruction [37] using Tajima and Nei genetic distance [38] and maximum-likelihood using DNAML from Phylip package 3.573c [39].

To date the major spread of *FAM8A1* pseudogenes in primate lineage, pseudogenes were amplified from different primate DNA with primers *FAM8A1F1* and *FAM8A1R1* (acc. nos. AF315790, AF315791, AF315792, AF315793, AF315794, AF315795, and AF315796).

A maximum-likelihood tree with molecular clock was produced and tested with the likelihood ratio test against a tree without a clock [40]. For this test, only sequences with nearly equal evolutionary rates are required. Therefore, functional *FAM8A1* sequences have been removed, as their evolutionary rate is lower than that of *FAM8A1* pseudogenes. *FAM8A4P* has also been discarded as its evolutionary rate is higher because of its presence on the Y chromosome [41].

Gene conversion among the *FAM8A1* pseudogenes was detected with the PLATO package [42]. A sliding window is moved along the alignment, and for each position the likelihood of the base partition is established and tested against the true phylogeny as a null hypothesis. If the partition is significantly different, that part of the alignment is checked manually to detect recombination events. For this purpose, six complete HERVK14 sequences have been extracted from the human genome database (acc. nos. K02774.1, AC003100, AC025593.3, AC021560, HS249C1, and AC068828). All sequence alignments performed in this study are available on request.

ACKNOWLEDGMENTS

We thank Chris Ottolenghi (Institut Pasteur, Paris, France) and Thierry Heidmann (Institut Gustave Roussy, Villejuif, France) for helpful discussions and comments during the writing of the manuscript.

RECEIVED FOR PUBLICATION MAY 8; ACCEPTED JULY 26, 2001.

REFERENCES

- Weiner, A. M., Deininger, P. L., and Efstratiadis, A. (1986). Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55: 631-661.
- Brosius, J., and Tiedge, H. (1995). Reverse transcriptase: mediator of genomic plasticity. *Virus Genes* 11: 163-179.
- Smit, A. F. (1996). The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* 6: 743-748.
- Weiss, R., Teich, N., Varmus, H., and Coffin, J. (1985). *Molecular Biology of RNA Tumor Viruses*. Cold Spring Harbor Laboratory Press, Plainview, NY.
- Lower, R. (1999). The pathogenic potential of endogenous retroviruses: facts and fantasies. *Trends Microbiol.* 7: 350-356.
- Lander, E. S., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
- Sverdlov, E. D. (2000). Retroviruses and primate evolution. *Bioessays* 22: 161-171.

8. Patience, C., Wilkinson, D. A., and Weiss, R. A. (1997). Our retroviral heritage. *Trends Genet.* **13**: 116–120.
9. Tonjes, R. R., Czauderna, F., and Kurth, R. (1999). Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K. *J. Virol.* **73**: 9187–9195.
10. Bishop, C. E., et al. (1983). Single-copy DNA sequences specific for the human Y chromosome. *Nature* **303**: 831–832.
11. Gyapay, G., et al. (1996). A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5**: 339–346.
12. Leroy, P., Seboun, E., Mattei, M. G., Fellous, M., and Bishop, C. E. (1987). Testis-specific transcripts detected by a human Y-DNA-derived probe. *Development* **101** (Suppl.): 177–183.
13. Wallace, A. M., et al. (1999). Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells. *Proc. Natl. Acad. Sci. USA* **96**: 6763–6768.
14. Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**: 418–420.
15. Medstrand, P., and Blomberg, J. (1993). Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues. *J. Virol.* **67**: 6778–6787.
16. Yang, Z. (1997). PAML 1.3 package distributed by the author. Department of Integrative Biology, University of California, Berkeley.
17. Johnson, W. E., and Coffin, J. M. (1999). Constructing primate phylogenies from ancient retrovirus sequences. *Proc. Natl. Acad. Sci. USA* **96**: 10254–10260.
18. Purvis, A., Nee, S., and Harvey, P. H. (1995). Macroevolutionary inferences from primate phylogeny. *Proc. R. Soc. London Ser. B* **260**: 329–333.
19. Takahata, N., and Satta, Y. (1997). Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proc. Natl. Acad. Sci. USA* **94**: 4811–4815.
20. Linial, M. (1987). Creation of a processed pseudogene by retroviral infection. *Cell* **49**: 93–102.
21. Dornburg, R., and Temin, H. M. (1990). cDNA genes formed after infection with retroviral vector particles lack the hallmarks of natural processed pseudogenes. *Mol. Cell. Biol.* **10**: 68–74.
22. Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.* **24**: 363–367.
23. Yang, J., et al. (1999). An ancient family of human endogenous retroviruses encodes a functional homolog of the HIV-1 Rev protein. *Proc. Natl. Acad. Sci. USA* **96**: 13404–13408.
24. Harris, J. M., Haynes, R. H., and McIntosh, E. M. (1997). A consensus sequence for a functional human endogenous retrovirus K (HERV-K) dUTPase. *Biochem. Cell Biol.* **75**: 143–151.
25. Swain, A., and Coffin, J. M. (1992). Mechanism of transduction by retroviruses. *Science* **255**: 841–845.
26. Sugden, B. (1993). How some retroviruses got their oncogenes. *Trends Biochem. Sci.* **18**: 233–235.
27. Schwartz, J. R., Duesberg, S., and Duesberg, P. H. (1995). DNA recombination is sufficient for retroviral transduction. *Proc. Natl. Acad. Sci. USA* **92**: 2460–2464.
28. Zhang, J., and Temin, H. M. (1993). Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. *Science* **259**: 234–238.
29. Anderson, J. A., Teufel, R. J. II, Yin, P. D., and Hu, W. S. (1998). Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination. *J. Virol.* **72**: 1186–1194.
30. Stuhlmann, H., Dieckmann, M., and Berg, P. (1990). Transduction of cellular neo mRNA by retrovirus-mediated recombination. *J. Virol.* **64**: 5783–5796.
31. Olson, P., Temin, H. M., and Dornburg, R. (1992). Unusually high frequency of reconstitution of long terminal repeats in U3-minus retrovirus vectors by DNA recombination or gene conversion. *J. Virol.* **66**: 1336–1343.
32. Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* **299**: 111–117.
33. Elder, J. F., Jr., and Turner, B. J. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* **70**: 297–320.
34. Bourgeron, T., Chretien, D., Rotig, A., Munnich, A., and Rustin, P. (1993). Fate and expression of the deleted mitochondrial DNA differ between human heteroplasmic skin fibroblast and Epstein-Barr virus-transformed lymphocyte cultures. *J. Biol. Chem.* **268**: 19369–19376.
35. Sassoon, D. A., Garner, I., and Buckingham, M. (1988). Transcripts of α -cardiac and α -skeletal actins are early markers for myogenesis in the mouse embryo. *Development* **104**: 155–164.
36. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
37. Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* **14**: 685–695.
38. Tajima, F. (1993). Unbiased estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* **10**: 677–688.
39. Felsenstein, J. (1993). Phylip package 3.573c distributed by the author. Department of Genetics, University of Washington, Seattle.
40. Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**: 368–376.
41. Agulnik, A. I., et al. (1998). Evolution of the DAZ gene family suggests that Y-linked DAZ plays little, or a limited, role in spermatogenesis but underlines a recent African origin for human populations. *Hum. Mol. Genet.* **7**: 1371–1377.
42. Grassly, N., and Rambaut, A. (1999). PLATO package distributed by the authors. University of Oxford, U.K.

Sequence data from this article have been deposited with DDBJ/EMBL/GenBank Data Libraries under accession numbers AF097027 (human FAM8A1 mRNA), AF097026 (human FAM8A1 gene), AF07850 (human FAM8A1 protein), AF315797 (human FAM8A3P), AF315798 (human FAM8A5P), AF315799 (human FAM8A6P), AF315800 (human FAM8A2P), AF315801 (human FAM8A4P), AF315790 (*Cercopithecus hamlyni* FAM8A1 clone 16), AF315791 (*C. hamlyni* FAM8A1 clone 11), AF315792 (*C. hamlyni* FAM8A1 clone 2), AF315793 (*Macaca mulatta* FAM8A1 clone 1), AF315794 (*M. mulatta* FAM8A1 clone 20), AF315795 (*Papio hamadryas* FAM8A1 clone 4), and AF315796 (*P. hamadryas* FAM8A1 clone 11).