Evolution of Endogenous Retrovirus-like Elements of the Woolly Mammoth (*Mammuthus primigenius*) and its Relatives

Alex D. Greenwood, Fred Lee,† Cristian Capelli,‡ Robert DeSalle,§ Alexei Tikhonov,‖ Preston A. Marx,†¶ and Ross D. E. MacPhee**

*Division of Vertebrate Zoology, American Museum of Natural History, New York, New York; †Aaron Diamond AIDS Research Center, New York, New York; ‡Istituto di Medicina Legale, Universitá Cattolica del Sacro Cuore, Rome, Italy; §Division of Invertebrate Zoology, American Museum of Natural History, New York, New York; ‖Zoological Institute, Russian Academy of Science, St. Petersburg, Russia; and ¶Aaron Diamond AIDS Research Center at Tulane Regional Primate Research Center, Covington, Louisiana

Endogenous retrovirus-like elements characterizable by a leucine tRNA primer (ERV-Ls) are reiterated genomic sequences known to be widespread in mammals, including humans. They may have arisen from an ancestral *foamy* virus-like element by successful germ line infection followed by copy number expansion. However, among mammals, only primates and rodents have thus far exhibited high copy number amplification and sequence diversification. Conventionally, empirical studies of proviral amplification and diversification have been limited to extant species, but taxa having good Quaternary fossil records could potentially be investigated using the techniques of ''ancient'' DNA research. To examine evolutionary parameters of ERV-Ls across both time and taxa, we characterized this proviral class in the extinct woolly mammoth (*Mammuthus primigenius*) and living elephants, as well as extant members of the larger clade to which they belong (Uranotheria, a group containing proboscideans, sirenians, hyraxes, and their extinct relatives). Ungulates and carnivores previously analyzed demonstrated low copy numbers of ERV-L sequences, and thus it was expected that uranotheres should as well. Here, we show that all uranothere taxa exhibit unexpectedly numerous and diverse ERV-L sequence complements, indicating active expansion within this group of lineages. Selection is the most parsimonious explanation for observed differences in ERV-L distribution and frequency, with relative success being reflected in the persistence of certain elements over a variety of sampled time depths (as can be observed by comparing sequences from fossil and extant elephantid samples).

Introduction

Endogenous retroviruses (ERVs) are a significant group of viral entities with several unusual features. Many ERV classes (including endogenous retrovirus-like elements characterizable by a leucine tRNA primer [ERV-Ls], the subject of this paper) have lost the envelope (*env*) gene sequence and are therefore unlikely to be infectious (Bénit et al. 1999). However, some ERVs are known to retain function, including the capacity to produce infectious viral particles, a potentially serious contraindication for xenotransplantation (Patience, Takeuchi, and Weiss 1997). More beneficially, ERVs have gained novel regulatory functions in the mammalian genome that are now indispensable (Mi et al. 2000). Human immunodeficiency virus (HIV) shares specific functionally homologous sequences with ancient endogenous retroviruses, raising the possibility that recombination with ERVs may change the properties of exogenous retroviruses (Yang et al. 1999). ERVs may also serve as a sequence pool from which exogenous viruses rapidly diversify and could even be the progenitors of exogenous retroviruses (Temin 1980).

In the absence of a ''virological fossil record'' for any mammalian taxon, it is unknown how ERVs have managed to establish their distributions (at the individual, population, and species levels) or how ERV complements have varied across time and host ranges. With respect to ERV-Ls, among major mammalian clades studied to date, only primates and rodents exhibit marked expansion of genomic ERV-L content and diversification of sequences (Bénit et al. 1999). In both primates and rodents, ERV-L copy number and number of unique sequences are high compared with those of carnivores, lagomorphs, and ungulates. For example, Southern blot data indicate that most placental mammals exhibit 10–30 ERV-L copies per genome, whereas primates and rodents have at least 100–200 copies (Bénit et al. 1999).

Not all mammalian orders have been examined for ERV-L incidence, but it is already clear that there are some significant interordinal differences. As noted, most primates and rodents possess a large number of distinct ERV-L elements in their genomes, but artiodactyls have very few (e.g., the cow has two reported unique sequences). Marsupials and monotremes appear to lack this element class altogether (Bénit et al. 1999). Why such differences should exist at all is not obvious, and to make any headway in understanding the evolution of ERV-Ls, it will be necessary to collect empirical data—especially data that permit comparisons between individuals within species and across appreciable lengths of time.

In undertaking this study, we attempted to collect both kinds of information simultaneously to capitalize on the fact that, despite the technical problems that attend their use, fossils are potentially the best empirical source of temporally distributed data on ERV-L preva-

Abbreviation: ERV-L, endogenous retrovirus-like element with tRNA leucine primer.

Key words: endogenous retrovirus, mammoth, ancient DNA, evolution, elephant, microsatellite.

Address for correspondence and reprints: Alex D. Greenwood, Division of Vertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024-5192. E-mail: alexgr@amnh.org.

Mol. Biol. Evol. 18(5):840–847. 2001

Table 1
Mammoth Sites, Dates, Specimens and DNA Extraction Information (see also fig. 1)

Locality/^{14}C Lab[a]	^{14}C Age (in radiocarbon years before present)[b]	Specimen[c]	Among-Clones Variation (cytochrome *b*, microsatellite [substitutions], microsatellite [indels])
Engineer Creek (Alaska)/Ångström Lab, Sweden	13,775 ± 145	Molar (dentine)	1.20, 2.29, 2.76
Chekurovka (Siberia)/Institute of Permafrost Studies, Yakutsk, Russia	26,000 ± 1,600	Rib	1.00, 0.20, 0.71
Wrangel Island (Siberia)/Beta Analytic 134781	4,590 ± 50	Tibia (marrow)	0.50, 0.67, 0.99

[a] Labs cooperating in ancient DNA investigations: Istituto di Medicina Legale (IML) of the Universitá Cattolica del Sacro Cuore, and Aaron Diamond AIDS Research Center (ADARC).

[b] Engineer Creek and Wrangel Island radiocarbon ages are conventional AMS dates (not calibrated); the Chekurovka age is a ''whole bone'' date (accuracy uncertain).

[c] Sampled specimens are from the following institutions: Engineer Creek, American Museum of Natural History (Department of Vertebrate Paleontology 8460); Chekurovka, Mammoth Museum, Yakutsk, Republic of Yakutia, Russia; and Wrangel Island, collection of Wrangel Island State Reserve.

lence—assuming, of course, that the relevant information can be reliably collected. Among mammals, the ideal group for exploring ERV-L paleovirology is Elephantidae (Proboscidea). In addition to living *Loxodonta africana* (African elephant) and *Elephas maximus* (Asian elephant), this family includes *Mammuthus primigenius*, the extinct woolly mammoth of late Quaternary Europe, northern Asia, and North America. Given the record of low ERV-L incidence in investigated ungulates, it might be expected that proboscideans would exhibit low copy numbers and few unique sequences in their genomes. This assumption implies that it should be comparatively easy to fully characterize the incidence and relative diversity of ERV-L sequences among individuals of proboscidean species over various time depths. Fortunately, fossils of woolly mammoths are abundant in many parts of this species' former range. Previous studies have shown that genetic material from high-latitude mammoth sites is often exceptionally well preserved (Johnson, Olson, and Goodman 1985; Yang, Golenberg, and Shoshani 1996; Greenwood et al. 1999), indicating that temporal examination of ERV-L evolution should be possible in principle. To provide a relevant context for interpreting the results of these experiments, equivalent data should be collected for members of the larger taxonomic group of which Proboscidea is a member—Uranotheria, the taxon which includes Asian and African elephants, sirenians (manatees and dugongs), and hyraxes (McKenna and Bell 1997).

Materials and Methods

DNA Extraction

Most ancient DNA studies are concerned only with mitochondrial DNA (mtDNA). Nuclear DNA is usually thought to be harder to retrieve than mtDNA from fossil material, but the task has been eased by several basic advances in the characterization of nuclear DNA from mammoths and other Late Pleistocene mammals (Greenwood et al. 1999). This study is concerned with both mtDNA and nuclear DNA.

DNA extraction from samples of living elephantids presented no special problems. However, investigation of mammoth material required the use of appropriate ancient DNA techniques (table 1) (Pääbo 1989; Janczewski et al. 1992; Greenwood et al. 1999). Ancient DNA extractions were performed in an American Museum of Natural History (AMNH) facility dedicated solely to ancient DNA extraction and PCR setup, in which elephant investigations had previously never been undertaken. Our protocols for avoidance of contamination are described in Greenwood et al. (1999). In addition to the precautions mentioned therein, bacterial transformation and PCR using bacterial colonies as template ''colony PCR'' for all products were carried out (by F.L.) at an institute separate from the ancient DNA laboratory (ADARC) (see Greenwood et al. [1999] for procedures). Modern DNA PCR and cloning was carried out in a separate AMNH laboratory to avoid contamination of the facility in which ancient extractions and PCR setup were carried out. A portion of each sample was sent to the Istituto di Medicina Legale, Universita Cattolica di Sacro Cuore, where extraction and PCR of a nuclear DNA locus were performed independently.

Modern DNA extraction from other uranotheres was performed (again in a laboratory separate from that used for ancient DNA extractions) as described in Greenwood et al. (1999), from blood in the case of the Florida manatee (*Trichechus manatus*) and the Asian elephant (*E. maximus*), and from muscle tissue in the case of the rock hyrax (*Procavia capensis*). African elephant (*L. africana*) DNA was provided by N. Georgiadis, M'Pala Research Centre, Kenya.

PCR, Cloning, and Sequencing

PCR conditions used for ancient samples were performed as described in Greenwood et al. (1999). PCRs

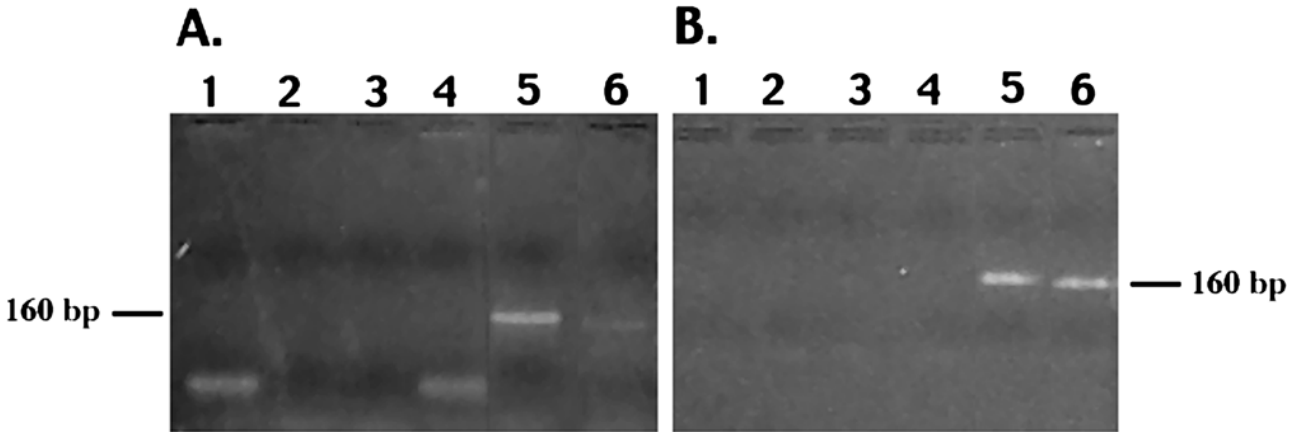

FIG. 1.—Example agarose gels of a mammoth microsatellite PCR amplification and reamplification. Lanes 1 and 4 are negative controls using water as template. Lanes 2 and 3 are negative controls using mock extractions as PCR templates. Lanes 5 and 6 are Engineer Creek and Chekurovka mammoth PCR amplifications, respectively. *A,* A 3% low-melting-point agarose gel in a standard Tris acetate (without EDTA) buffer with the entire 30-μl initial PCR reaction loaded in each lane. The bands were excised, and reamplifications were performed as described in Greenwood et al. (1999). *B,* A 2% agarose gel in a standard Tris, boric acid, EDTA buffer; 17% of each reamplified PCR reaction was loaded in each lane.

of extant taxa included 30 cycles performed in a 50-μl volume with 1 μg DNA using *Taq* polymerase and the standard buffer supplied by Boehringer Mannheim. PCR primers and the annealing temperature used for the elephant-specific microsatellite are described in Nyakaana and Arctander (1999). Primers and PCR conditions for cytochrome *b* and 28S rDNA are described in Greenwood et al. (1999). ERV-L primers and the annealing temperature for the larger *pol* fragment are described in Bénit et al. (1999). Primers for the shorter product applied to mammoth extracts combined the 3′ primer described in Bénit et al. (1999) with a primer designed based on initial elephant ERV-L sequences (5′-CAGCAATACACCTTCACTTG-3′) at a 60°C annealing temperature. PCR product cloning and colony sequencing was done as in Greenwood et al. (1999) except that (1) colony PCR products were purified with QIAquick columns (Qiagen), and (2) sequencing with plasmid-specific primers T7 and SP6 was executed with an ABI 377 sequencer and the manufacturer's protocol. Both strands were sequenced for all ERV-L clones. Prior to attempting proviral sequence amplification in the mammoth specimens, we established that each sample contained endogenous mammoth DNA by demonstrating the presence of specific mitochondrial, multicopy nuclear DNA sequences and a single-copy microsatellite sequence reported here for the first time for mammoth material (fig. 1) (Nyakaana and Arctander 1999). Facilities, extraction of nucleic acids, and PCR amplification are described in detail elsewhere (Greenwood et al. 1999).

Results and Discussion

Verification of Ability to Retrieve Sequences of Mammoth Origin

For each mammoth, a 199-bp portion of the mitochondrial cytochrome *b* gene was successfully amplified, the PCR products were cloned, and four or five clones were sequenced for each mammoth. Among-clones differences (ACDs) are presented in table 1 (GenBank accession numbers AF154864, AF309073, and AF309074 and other data not shown). All negative controls including mock extractions were always devoid of PCR product. Consensus sequences of clones for Engineer Creek (Alaska) and Chekurovka (Siberia) mammoths were identical to each other and to those of four other mammoths in the existing database. Sequence retrieved for the Engineer Creek specimen was identical to a sequence previously retrieved from the same sample (Greenwood et al. 1999). The Wrangel Island (Siberia) mammoth differed by one position from the other mammoths in all clones. African and Asian elephants differed from the mammoths by 3–5 and 6–10 positions, respectively, by 33–34 positions from the manatee (the closest living relative of elephants), and by 36–37 positions from humans. This pattern strongly indicated that the recovered cytochrome *b* sequences were mammoth in origin.

A 28S rDNA fragment with a characteristic amplification product size for elephants and mammoths (179 bp in elephantids versus 150 bp in humans; Greenwood et al. 1999) was successfully amplified for all three mammoths, thereby demonstrating the presence of intact multicopy nuclear DNA (data not shown).

To demonstrate the presence of intact endogenous single-copy sequences, a microsatellite sequence described for elephants was amplified (Nyakaana and Arctander 1999). The expected product, approximately 160 bp, was recovered for all three mammoths, although products were generally much weaker than observed for mitochondrial or multicopy nuclear DNA amplifications (fig. 1 and other data not shown). Again, negative controls were always devoid of PCR product. A portion of each sample was sent to another institute (Istituto di Medicina Legale [IML]) where the DNA was independently extracted and the same microsatellite was amplified (by C.C.). Clones from the two products for each mammoth are shown in figure 2 (for ACDs, see table 1). A greater ACD for the Engineer Creek mammoth is

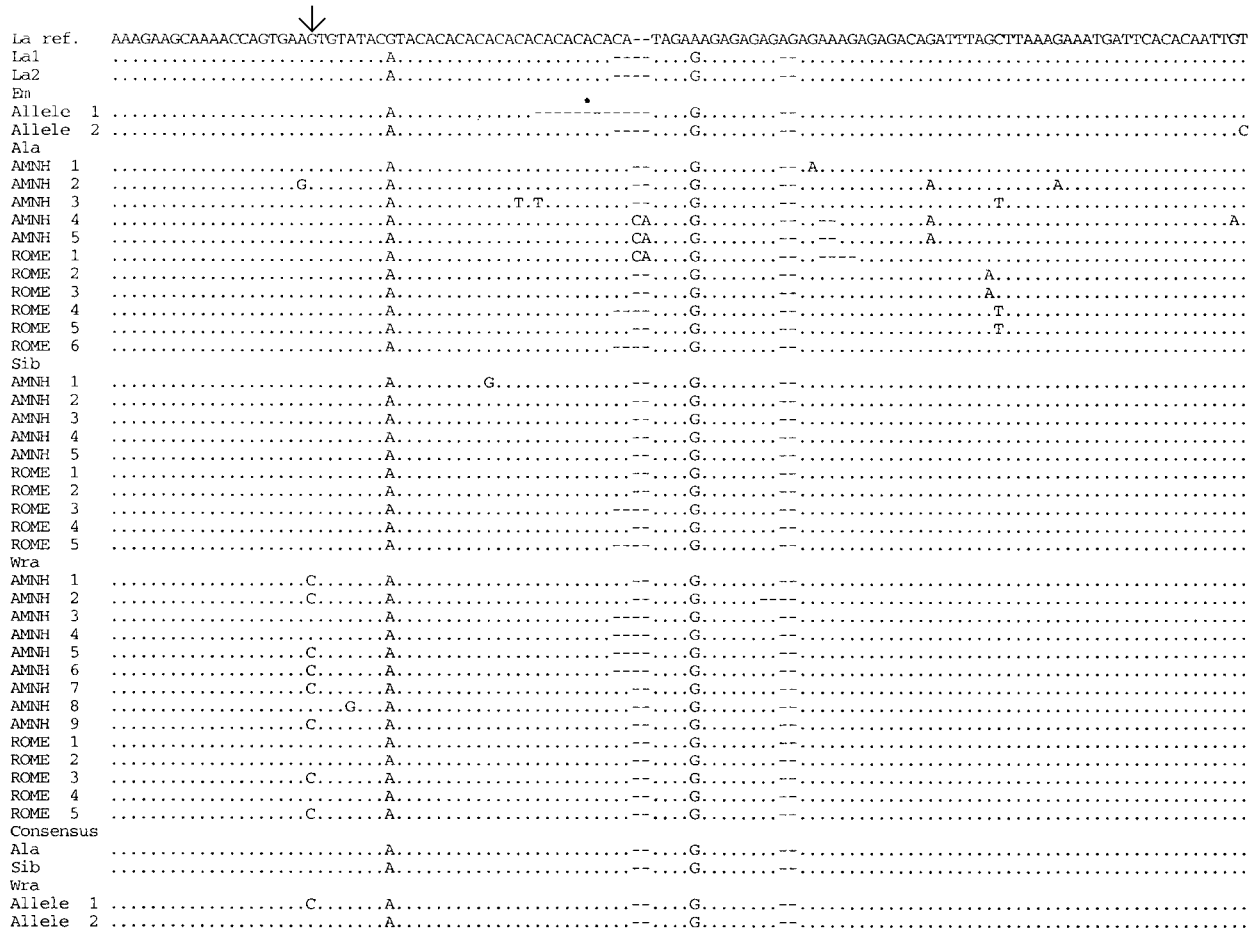

FIG. 2.—Proboscidean microsatellite sequences. Key: La, African elephant; Em, Asian elephant; Ala, Engineer Creek (Alaska) mammoth; Sib, Chekurovka (Siberia) mammoth; Wra, Wrangel Island (Siberia) mammoth. The top sequence is the published African elephant sequence (Nyakaana and Arctander 1999). Column dots indicate identity to bases registered in the top sequence; dashes represent gaps. Individual clones are shown for mammoths. ''AMNH'' designates clones from PCR products obtained from extracts made at the American Museum of Natural History (by A.D.G) and in Rome (by C.C.) for clones from extractions done independently at IML, Rome. Consensus sequences for each mammoth and different microsatellite alleles are also shown. La1 and La2 are African elephants from Botswana (Savute population) and Tanzania (Serengeti population) respectively. Sequences have been deposited in GenBank (accession numbers AF317793–AF317800).

consistent with previous DNA analysis from this sample and may reflect greater postmortem DNA damage and modification (Greenwood et al. 1999). The Wrangel Island mammoth at one position exhibited G in three clones of the first amplification and C at six (↓ in fig. 2). Similar results were obtained for clones from the IML extract (G in three clones, C in two). This difference, which has been observed before in mammoth nuclear DNA sequences, is most likely due to allelic variation. The mammoths differed from the published African elephant sequence by two to three substitutions and by a 2-bp deletion (Nyakaana and Arctander 1999). The mammoths differed from the two Asian elephant alleles by zero to two substitutions. Both Asian elephant alleles shared the same 2-bp deletion observed in the mammoth sequences. Subsequent sequencing of additional African elephants by our group has shown that the African elephant microsatellite-flanking regions also differ from mammoths by zero to two differences and do not exhibit the 2-bp insertion observed in the published sequence (fig. 2, La1 and La2). The same sequences were retrieved for each mammoth at both the Rome and the New York labs, thus providing confirmation of the presence of endogenous mammoth DNA in these samples.

ERV-L Sequences

PCR primers amplifying two different lengths of the ERV-L polymerase (*pol*) gene were used. The first amplified a sequence of approximately 360 bp (Bénit et al. 1999). The second, based on elephant ERV-L sequences, amplified a sequence of approximately 110 bp (excluding primers) and was internal to the 360-bp fragment. To place these results in a useful comparative context, the longer fragment was characterized for extant representatives of Uranotheria (data not shown). Results demonstrated that, contrary to expectation, all living uranotheres possess highly diverse ERV-L complements, indicating that active element bursts have occurred (and possibly still occur in extant lineages). Until more mammalian orders and individuals and longer sequences are retrieved, the phylogenetic implications of these sequences will remain unclear. Nevertheless, it is definite

that uranotheres have an unexpectedly diverse ERV-L complement.

For the mammoth specimens, data had to be acquired from the shorter ERV-L fragment because experience has shown that long nuclear DNA fragments extracted from mammoth material will not amplify with existing techniques (Greenwood et al. 1999). The shorter primer combination excludes humans and cows, two potential contamination sources (Taylor [1996] and other data not shown). As noted above, all elephant amplifications, cloning, and sequencing were done in facilities removed from the ancient DNA laboratory that have never held uranothere collections in order to exclude the possibility of contamination of mammoth ERV-L sequences.

Using the shorter ERV primer combination, products of expected size were obtained from the DNA of all three mammoth specimens, while all negative controls were always devoid of product. Each mammoth sample was independently amplified three times by PCR, and 10 individual clones per PCR product were sequenced to evaluate ERV-L diversity. Thus, 30 clones were generated from three independent PCR reactions for each mammoth. For the Engineer Creek mammoth, the third amplification was derived from an extract that was different from the first two. Subsequent to the mammoth work, the same PCR sampling protocol was applied to DNA extracted from an Asian elephant and an African elephant. For comparative purposes, the same fragment was amplified from DNA derived from the manatee and the hyrax; 10 clones were sequenced for each PCR product.

Again, contrary to expectations, PCR reactions with both mammoth and elephant DNA yielded multiple distinct ERV-Ls (alignment available from GenBank, accession numbers AF312038–AF312207). The average among-clones absolute difference, including substitutions and insertions/deletions, for the entire elephantid data set was 20.42 (table 2). Each PCR revealed approximately the same average level of ACDs, with the mammoth from Engineer Creek having the highest average (table 2). For this latter specimen, first- and second-extract clones did not differ in diversity (table 2).

For the Chekurovka specimen from mainland Siberia, all amplifications had relatively low ACDs, as did the first amplification for the Asian elephant. In the latter, lower values most likely reflected stochastic variation. In both cases, the results led to statistically significant differences among the means ($P < 0.001$), albeit the sample size was low. The consistently lower relative average diversity of sequences in the Chekurovka specimen could possibly be explained by the comparatively weak amplification signal for all PCR products recovered from this individual in other experimental applications (fig. 1 and unpublished data not shown). Weak signal was most likely due to low retrievable DNA concentration in this specimen, which was in turn presumably due to diagenesis. Given initial low concentration in the Chekurovka specimen, the number of template DNA molecules available for initial cycles of any given PCR reaction will be consistently lower than in, for example, the other mammoth specimens or modern samples used in this study. Consequently, ERV-L sequences amplified in the earlier rounds of PCR would tend to dominate in the final PCR product. Thus, in any one PCR reaction, the same sequence may dominate among clones. However, the sequences dominating among different replicates can be very divergent, and thus the overall mean ACD diversity is greater than that within a given replicate. Under this interpretation, we conclude that the lower within-replicate and overall sequence diversity for the Chekorovka specimen reflects lower DNA concentration, not reduced ERV-L complement.

At the beginning of this study, it was not known how many ERV-L sequences might exist in different uranothere lineages, even to an order of magnitude (in some vertebrate species, ERV sequences may comprise up to 1% of the genome) (Coffin 1996). The results of Bénit et al. (1999) in sequencing clones from ERV-L amplifications from several ungulate groups suggested that proboscideans would probably display only a few different elements. Thus, it is of great interest to report that novel ERV-L sequences were found in each replicate in our study, suggesting that many more unique sequences could be obtained if sampling were to continue. Uranotheres evidently genomically maintain a large pool of ERV-L sequences, unlike virtually all other mammalian orders sampled to date with the exception of primates and rodents (Bénit et al. 1999). In order to compare results across species and major taxa, we as-

Table 2
ERV-L Among-Clones Variability for Mammoths and Elephants

Provenience[a]	Replication	Average[b]	Coding Potential[c]
Ala	1	20.87 (0–28)	
	2	21.31 (1–29)	
	3	19.36 (0–28)	
	Total	20.80 (0–31)	10
Sib[d]	1	11.67 (0–22)	
	2	11.20 (0–19)	
	3	12.80 (0–20)	
	Total	16.28 (0–26)	20
Wra	1	19.24 (0–29)	
	2	21.13 (0–45)	
	3	20.56 (0–28)	
	Total	21.74 (0–45)	16
Em[d]	1	13.27 (0–25)	
	2	18.24 (11–28)	
	3	22.76 (14–32)	
	Total	18.72 (0–34)	22
La	1	1789 (1–23)	
	2	24.34 (0–37)	
	3	20.91 (12–27)	
	Total	21.01 (0–38)	13
Elephantidae	Total	20.42 (0–45)	81

[a] Species designations are the same as those used in figure 2.

[b] Average absolute differences (substitutions and insertions/deletions, minimum and maximum values in parentheses) are given for clones from each PCR replication, as well as the total for each individual tested.

[c] Overall average and number of potential coding sequences shown for each individual.

[d] ANOVA and *t*-tests (assuming unequal variance) revealed that the total means for Sib and Em were significantly from those of the other elephantids ($P < 0.001$).

Table 3
Observed Frequencies of the Same or Similar Sequences Among Clones

	Ala	Sib	Wra	Em	La	Tm	ACD[a]
1[b]		3.1, 3.2, 3.6, 3.10					0
2		3.4, 3.5, 3.7					0
3[b]			1.5, 1.7				0
4			1.9, 1.10				0
5			2.1, 2.2, 2.3, 3.9				0–1
6[b]				1.6, 2.9, 3.8			0
7	1.1, 2.5, 2.8, 2.9, 3.6, 3.10	2.4, 2.7	3.7				0–5
8	1.4	1.3, 1.5, 1.6, 1.8, 1.10					0–1
9[b]	1.6		3.8				2
10	2.7		1.2, 1.3, 1.6				0–1
11[b]	2.4		3.4				5
12[b]	2.6	2.1, 2.2, 2.3, 2.6, 2.9, 2.10	2.5, 2.6, 3.6, 3.10	3.3			0–2
13[b]	1.2, 3.4		2.8, 3.1	3.9			0–4
14[b]		1.1, 1.2, 1.4, 1.7, 1.9		3.2			0–3
15[b]			1.1, 2.9, 3.2	3.6			0–8[c]
16[b]	3.2, 3.9		3.3		2.5		0—3
17	1.8, 2.1, 2.2, 3.1, 3.3, 3.5, 3.7				1.4, 1.7, 3.10		0–6
18[b]	1.3, 1.5, 1.10	3.3, 3.8, 3.9			2.9		0
19	2.3				2.8		2
20	1.7			2.7	2.10, 3.8		0–6
21			2.4, 2.10	1.4	1.10, 2.3, 2.6		0–8
22[b]	3.8			1.3	1.3		0–1
23[b]				1.2, 1.5, 1.7, 1.8, 1.10	1.5, 1.9, 3.1		0–2
24				2.1	1.1, 2.4, 3.9		0–1
25[b]				1.9	1.2		0
26[b]				2.2		6, 7	8–11[d]
27[b]				2.1, 2.5	3.6	4, 5, 10	7–12[d]

NOTE.—Species designations are as in figure 2. The first number designates the PCR replication. The number following the period designates the clone number.

[a] Among-clones differences.

[b] Groups of sequences with coding potential.

[c] Although among-clones variation was observed in the microsatellite sequences (0–8 for the Engineer Creek mammoth, fig. 1), mammoth clones that differed by ≤8 positions were still considered orthologous (although it is possible that some are paralogs). However, most sequences differed by 0–2 positions.

[d] For the manatee (Tm) and elephant comparisons, sequences were chosen that differed from each other by less than 10 substitutions. Longer fragments are needed to determine phylogenetic relationships among specific elements and to identify other sequences shared by descent in more distantly related taxa.

sumed that detection was related to abundance, such that the sequences exhibiting the highest copy numbers in any individual's genome were the ones most likely to be detected by PCR. This assumption was justified by the fact that our primers, although doubtlessly excluding some sequences, were nevertheless able to amplify a great diversity of sequences across all selected uranothere taxa.

Although diverse, not all recovered ERV-L sequences were unique (table 3). The Wrangel Island and Chekurovka mammoths and the Asian elephant each exhibited sequences that were individually unique but appeared more than once among clones, with the Wrangel Island mammoth showing four groups of multiply occurring sequences (table 3). Among the mammoths, five sequences were shared (table 3). The Engineer Creek and Wrangel Island mammoths exhibited more sequences in common than either did with the Chekurovka mammoth. Thus, it is arguable that the same elements were present with little or no change in mammoth populations ranging temporally from 26,000 to 4,500 years ago and geographically from north-central Siberia to central Alaska. Interestingly, the mammoths shared four sequences with Elephas and four different ones with Loxodonta, while the Asian elephant and the African elephant exclusively shared three sequences. Additionally, three sequences were shared by mammoths and both elephant species. Since these three lineages probably diverged from one another during the Early Pliocene 4–5 MYA (Todd and Roth 1996), it is evident that some elephantid ERV-Ls have been maintained in descendant taxa over a considerable time span.

A common feature among the groups now known to have experienced active ERV-L expansion is the presence of potential coding capacity, i.e., sequences that are in frame with the mouse ERV-L open reading frame (MuERVL, GenBank accession number Y12713) and contain no stop codons. Fifty-four percent of all elephantid sequences retained coding potential (table 2). Among shared proboscidean sequences, 14 of 25 groups had coding potential, while two groups had individual members with coding potential (table 3). Primates and rodents (both of which experienced expansion events) are the only known mammalian groups showing ERV-L coding potential (Bénit et al. 1999). Thus, long-term maintenance of ERV-Ls may depend on some species in a group retaining such potential. Indeed, functionality may be diagnostic of proviral activity bursts. While 110 bp is insufficient for absolute determination of the existence of open reading frames, results from sequencing of larger ERV-L PCR fragments derived from modern uranotheres are similar (data not shown).

Five manatee sequences had 7–11 differences from three of the modern elephant sequences (table 3). All of

these sequences were rare in extant elephants; interestingly, none occurred in the mammoths. It is likely that the sequences were shared by the last common ancestor of manatees and elephants. This is consistent with the view that sirenians and proboscideans are related, albeit distantly (their last common ancestor probably lived in Middle to Late Paleocene, 56–60 MYA; Fischer 1996). No comparable sequences were observed in the hyrax, indicating that the possibility of persistence of a given ERV-L is not unlimited.

In cases where sequences are identical or nearly so (i.e., among mammoths and elephants), the simplest explanation is that the shared sequences represent orthologs. However, it is formally possible that more divergent sequences (such as those of the manatee) are paralogous. Yet, even if descriptively identical loci are not identical by virtue of actual descent, paralogy is nevertheless consistent with the proposition that both taxa shared a common progenitor element. Lack of a high degree of sequence sharing is also consistent with observations utilizing the longer *pol* gene fragment. Burst sequences within monophyletic groups showing significant ERV-L activity, such as primates, tend to be much more similar within a group than between unrelated groups (Bénit et al. [1999] and other data not shown).

The 110-bp product retrieved is insufficient to draw any strong phylogenetic conclusions regarding uranothere interrelationships. However, preliminary phylogenetic analysis of a subset of uranothere sequences and those of other mammalian orders revealed that the primate and rodent sequences tend to cluster as monophyletic groups, while ERV-Ls from ungulates, lagomorphs, and carnivores are randomly distributed, a finding which is consistent with previous analysis (Bénit et al. 1999). In addition, with few exceptions, the uranothere sequences also formed a monophyletic group. This is consistent with observations and interpretations regarding rodent and primate ERV-L expansion events (Bénit et al. 1999). However, as tree support was not robust, wider sampling of mammalian orders is necessary, and accumulation of longer sequences will be required to draw any further conclusions.

Explanation for ERV Expansion

Horizontal transfer and repeated novel infection are unlikely sources of overall diversity of ERV-Ls, as these entities generally lack an *env* gene, tend to cluster phylogenetically by vertebrate class, and are noninfectious (Herniou et al. 1998). Lack of an abundance of shared sequences within higher-level taxa (e.g., uranotheres) also argues against horizontal transmission as the main force generating diversity (Bénit et al. 1999). More generally, ERVs activate during fertilization and embryogenesis (Löwer 1999). During gametogenesis and development, the normal suppression of ERVs is released. The effects can be dramatic, with the appearance of RNA transcripts from multiple ERVs and ERV-induced chromosomal rearrangements in offspring of interspecific hybrids (Waugh O'Neill, O'Neill, and Marshall Graves 1998; Löwer 1999). In the present case, the sequences that appear most often among individual mammoths are a combination of elements that have expanded successfully in this species. Some elements, many with apparent coding potential, have been retained within specific elephantid lineages for millions of years, presumably having generated enough copies of themselves to avoid removal by selection, recombination, or genetic drift (Bénit et al. 1999).

The fact that primates, rodents, and proboscideans each display unique expanded elements suggests that ERV-Ls, which are generally nonfunctional, may have acquired functionality to expand. An alternative explanation is that lineages were infected with functional exogenous viruses that subsequently expanded endogenously. It is also possible that ERVs could recombine in the genome or during reverse transcription to form novel elements, as do HERV-K elements (Berkhout, Jebbink, and Zsíros 1999). These combinations of events could lead to remarkable lineage-specific bursts of transposition, formation of novel elements, and expansion of ERV-Ls. Although more evidence is required, if such events do occur, the significance of ERVs for phylogenetic investigations could be great indeed.

In summary, the existence of high-quality mammoth remains from Late Quaternary paleontological sites provides a unique opportunity for examining different levels of sequence evolution within an unquestionably monophyletic group of mammals. In particular, our data demonstrate that a high proportion of elephantid ERV-Ls have been able to successfully persist despite the effects of time, geographical distance, and speciation. Finally, recovery of numerous ERV-L sequences from mammoth material suggests that it should be possible to use molecular probes to search for evidence of exogenous viruses in well-preserved fossil remains. Eventually, this may lead to important insights into the epizootiological history of now-extinct populations and species (cf. MacPhee and Marx 1997).

Acknowledgments

We thank S. Pääbo, W. Schartau, and the University of Munich for support during the early phases of this project. We are indebted to R. Connor for improving the manuscript. For various forms of support and encouragement, we also acknowledge V. L. Pascali, Universitá Cattolica di Sacro Cuore, Rome; G. Destro-Bisol, University La Sapienza, Rome; and Y. Guo and D. Ho, Aaron Diamond AIDS Research Center, New York. For assistance and providing samples, we thank P. Lazarev and G. Boeskorov, Mammoth Museum, Yakutia; N. Georgiadis, M'Pala Research Centre, Kenya; J. Hektor, Tierpark Hellabrunn, Munich; G. Amato, Wildlife Conservation Society, New York; D. Murphy, Lowry State Park Zoo, Florida; S. Vartanyan, Wrangel Island State Reserve, Russia; Jeffrey Saunders, Illinois State Museum, Springfield; and R. Tedford, C. Flemming, and R. Liberman, American Museum of Natural History. Dr. Jennifer White (Augustana College) gave valuable advice on statistics. This work was supported by grants

from the National Science Foundation, the Irene Diamond Fund, and the Adler Fund.

LITERATURE CITED

BÉNIT, L., J.-P. LALLEMAND, J.-F. CASELLA, H. PHILIPPE, and T. HEIDMANN. 1999. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. J. Virol. **73**:3300–3308.

BERKHOUT, B., M. JEBBINK, and J. ZSÍROS. 1999. Identification of an active reverse transcriptase enzyme encoded by a human endogenous HERV-K retrovirus. J. Virol. **73**:2365–2375.

COFFIN, J. M. 1996. Retroviridae: the viruses and their replication. Pp. 1767–1848 *in* B. N. FIELDS, D. M. KNIPE, and P. M. HOWLEY, eds. Fields virology. 3rd edition. Lippincott-Raven, Philadelphia.

FISCHER, M. S. 1996. On the position of Proboscidea in the phylogenetic system of Eutheria: a systematic review. Pp. 35–38 *in* J. SHOSHANI and P. TASSY, eds. The Proboscidea: evolution and palaeoecology of elephants and their relatives. Oxford University Press, London.

GREENWOOD, A. D., C. CAPELLI, G. POSSNERT, and S. PÄÄBO. 1999. Nuclear DNA sequences from Late Pleistocene megafauna. Mol. Biol. Evol. **16**:1466–1473.

HERNIOU, E., J. MARTIN, K. MILLER, J. COOK, M. WILKINSON, and M. TRISTEM. 1998. Retroviral diversity and distribution in vertebrates. J. Virol. **72**:5955–5966.

JANCZEWSKI, D. N., N. YUHKI, D. A. GILBERT, G. T. JEFFERSON, and S. J. O'BRIEN. 1992. Molecular phylogenetic inference from saber-tooth cat fossils of Rancho La Brea. Proc. Natl. Acad. Sci. USA **89**:9769–9773.

JOHNSON, P. H., C. B. OLSON, and M. GOODMAN. 1985. Isolation and characterization of deoxyribonucleic acid from tissue of the woolly mammoth, *Mammuthus primigenius.* Comp. Biochem. Physiol. B **81**:1045–1051.

LÖWER, R. 1999.The pathogenic potential of endogenous retroviruses: facts and fantasies. Trends Microbiol. **7**:350–356.

MCKENNA, M., and S. K. BELL. 1997. Classification of mammals above the species level. Columbia University Press, New York.

MACPHEE, R. D. E., and P. A. MARX. 1997. The 40,000-year plague: humans, hyperdisease, and first-contact extinctions. Pp. 169–217 *in* S. M. GOODMAN and B. D. PATTERSON, eds. Natural change and human impact in Madagascar. Smithsonian Institution Press, Washington, D.C.

MI, S., X. LEE, X. LI et al. (12 co-authors). 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature **403**:785–789.

NYAKAANA, S., and P. ARCTANDER. 1999. Isolation and characterization of microsatellite loci in the African elephant, *Loxodonta africana.* Mol. Ecol. **7**:1436–1437.

PÄÄBO, S. 1989. Ancient DNA: extraction, characterization, molecular cloning and enzymatic amplification. Proc. Natl. Acad. Sci. USA **86**:1939–1943.

PATIENCE, C., Y. TAKEUCHI, and R. A. WEISS. 1997. Infection of human cells by an endogenous retrovirus of pigs. Nat. Med. **3**:282–286.

TAYLOR, P. 1996. Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. Mol. Biol. Evol. **13**:2839.

TEMIN, H. M. 1980. Origin of retroviruses from cellular moveable genetic elements. Cell **21**:599–600.

TODD, N., and V. L. ROTH. 1996. Origin and radiation of the Elephantidae. Pp. 193–202 *in* J. SHOSHANI and P. TASSY, eds. The Proboscidea: evolution and palaeoecology of elephants and their relatives. Oxford University Press, London.

WAUGH O'NEILL, R. J., M. J. O'NEILL, and J. A. MARSHALL GRAVES. 1998. Undermethylation associated with retroelement activation and chromosome remodeling in an interspecific mammalian hybrid. Nature **393**:68–72.

YANG, H., E. M. GOLENBERG, and J. SHOSHANI. 1996. Phylogenetic resolution within the Elephantidae using fossil DNA sequences from the American mastodon (*Mammut americanum*). Proc. Natl. Acad. Sci. USA **93**:1190–1194.

YANG, J., H. P. BOGERD, S. PENG, H. WIEGAND, R. TRUANT, and B. R. CULLEN. 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.

RODNEY HONEYCUTT, reviewing editor

Accepted January 15, 2001