Evolutionary Dynamics and Evolutionary History in the RTE Clade of Non-LTR Retrotransposons

Vera Župunski, Franc Gubenšek,*† and Dušan Kordiš**

*Department of Biochemistry and Molecular Biology, Jožef Stefan Institute, Ljubljana, Slovenia; and †Department of Chemistry and Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Ljubljana, Slovenia

This study examined the evolutionary dynamics of Bov-B LINEs in vertebrates and the evolution of the RTE clade of non-LTR retrotransposons. The first full-length reptilian Bov-B LINE element is described; it is 3.2 kb in length, with a structural organization typical of the RTE clade of non-LTR retrotransposons. The long-term evolution of Bov-B LINEs was studied in 10 species of Squamata by analysis of a PCR-amplified 1.8-kb fragment encoding part of apurinic/apyrimidinic endonuclease, the intervening domain, and the palm/fingers subdomain of reverse transcriptase. A very high level of conservation in Squamata Bov-B long interspersed nuclear elements has been found, reaching 86% identity in the nearly 600 amino acids of ORF2. The same level of conservation exists between the ancestral snake lineage and Ruminantia. Such a high level is exceptional when compared with the level of conservation observed in nuclear and mitochondrial proteins and in other transposable elements. The RTE clade has been found to be much more widely distributed than previously thought, and novel representatives have been discovered in plants, brown algae, annelids, crustaceans, mollusks, echinoderms, and teleost fishes. Evolutionary relationships in the RTE clade were deduced at the amino acid level from three separate regions of ORF2. By using different independent methods, including the divergence-versus-age analysis, several examples of horizontal transfer in the RTE clade were recognized, with important implications for the existence of HT in non-LTR retrotransposons.

Introduction

Transposable elements (TEs) are widely distributed in prokaryotes and eukaryotes, where they can move within the host genome or between unrelated genomes. They play important roles in the structural organization and evolution of the genomes they inhabit (Kidwell and Lisch 1997, 2000; Kumar and Bennetzen 1999; Shapiro 1999; Bennetzen 2000; Fedoroff 2000; Silva and Kidwell 2000). Eukaryotic TEs are divided into two main classes according to their structural organization and their mechanism of transposition. Class I elements use an RNA-mediated mode of transposition and encode a reverse transcriptase (RT), while class II elements, the transposons, use a DNA-based mode of transposition (Kumar and Bennetzen 1999).

Non–long terminal repeat (LTR) retrotransposons, one of the oldest groups of retrotransposons, have been classified into 12 clades (Malik, Burke, and Eickbush 1999; Malik and Eickbush 2000). Vertebrate genomes contain representatives from the L1, CR1, and RTE clades. Representatives of the remaining nine clades are distributed in invertebrates, fungi, and plants (Malik, Burke, and Eickbush 1999).

The evolutionary dynamics of TEs in their hosts differ, but in most cases they involve vertical transmission and occasionally also horizontal transmission. The life cycles of different groups of TEs consist of vertical inactivation of active elements, stochastic loss of TEs from populations, and eventual reintroduction by horizontal transfer (Lohe et al. 1995; Hartl, Lohe, and Lozovskaya 1997).

The phenomenon of horizontal transfer (HT) has been observed and well documented in most TE classes (Kidwell 1993; Capy, Anxolabehere, and Langin 1994; Hartl, Lohe, and Lozovskaya 1997). The most-studied and best-understood examples are those from DNA transposons represented by *mariner* (Robertson 1993; Gueiros-Filho and Beverley 1997; Hartl, Lohe, and Lozovskaya 1997) and the *P* element (Houck et al. 1991; Kidwell 1993; Clark and Kidwell 1997; Silva and Kidwell 2000). In the most primitive RT-containing elements, the mobile introns, a very large number of HTs have been discovered in higher plant groups (Cho et al. 1998). In the LTR retrotransposons, several examples of HT have been documented between closely related species (mostly among *Drosophila*), such as *copia* (Jordan and McDonald 1998; Jordan, Matyunina, and McDonald 1999) and *gypsy* (Terzian et al. 2000), and between more distantly related species, represented by SURL elements from sea urchins (Gonzalez and Lessios 1999). In contrast to the general acceptance of HT of DNA transposons, mobile introns, and LTR-retrotransposons, several proposed cases of HT in non-LTR retrotransposons have met with strong criticism (Malik, Burke, and Eickbush 1999). This issue was further investigated, and new evidence for HT of non-LTR retrotransposons is presented.

Members of the RTE clade are the shortest of the non-LTR retrotransposons and have a highly conserved structural organization (Malik and Eickbush 1998; Malik, Burke, and Eickbush 1999). They have been studied in nematodes and insects (Malik and Eickbush 1998), mammals (Szemraj et al. 1995; Okada and Hamada 1997; Malik and Eickbush 1998), trematodes (Drew et al. 1999), and teleost fishes (Volff et al. 1999). On the basis of discontinuous distribution, extreme nucleotide

Abbreviations: AP-EN, apurinic/apyrimidinic endonuclease; HT, horizontal transfer; LCA, last common ancestor; LINE, long interspersed nuclear element; LTR, long terminal repeat; NJ, neighbor joining; ORF, open reading frame; RT, reverse transcriptase; TE, transposable element.

Key words: non-LTR retrotransposon, RTE clade, Bov-B LINE, horizontal transfer, evolutionary dynamics.

Address for correspondence and reprints: Dušan Kordiš, Department of Biochemistry and Molecular Biology, Jožef Stefan Institute, Jamova 39, 1001 Ljubljana, Slovenia. E-mail: dusan.kordis@ijs.si.

Mol. Biol. Evol. 18(10):1849–1863. 2001

sequence conservation, genetic distances, and evolutionary relationships, the HT of Bov-B long interspersed nuclear elements (LINEs) 40–50 MYA from the ancestral snake lineage (Boidae) to the ancestor of ruminants has been demonstrated (Kordiš and Gubenšek 1998, 1999*a*, 1999*b*). At present, only one full-length bovine Bov-B LINE element is known (Szemraj et al. 1995; Okada and Hamada 1997; Malik and Eickbush 1998), and no reptilian species have so far been investigated. Here we present the first full-length reptilian Bov-B LINE element and examine the evolutionary dynamics of Bov-B LINEs in vertebrates and the distribution and evolutionary relationships of the RTE clade in eukaryotes.

Materials and Methods

Species Analyzed

To analyze the distribution of Bov-B LINEs by PCR, the genomic DNA of the same species tested previously (Kordiš and Gubenšek 1998) was used. We tested some additional mammalian species, such as *Tragulus* (chevrotain), *Hippopotamus* (hippopotamus), and *Tursiops* (dolphin), with the tissues being kindly provided by Prof. J. A. Lenstra from Utrecht University (the Netherlands). A blood sample of *Macropus* (kangaroo, Marsupialia) was obtained from Ljubljana Zoo. Genomic DNA extractions were performed as previously described (Kordiš and Gubenšek 1998).

PCR Amplification of the 1.8-kb Fragment of Bov-B LINEs

All experiments were performed in parallel with negative and positive controls. Different rooms, reagents, equipment, and positive displacement pipettes were used, according to the general precautions for PCR performance. The sense (ME1: 5′-CACRGTRATY-CAAGYCTAYRCYCCAAC-3′) and antisense (ME2: 5′-CWGCAWATCTGAGGTTMKTKAKATTTCT-3′) degenerate oligonucleotide primers were based on conserved regions in the endonuclease (TVIQVYAPT—ME1) and RT (RNINNLRYA—ME2) domains of *Bos taurus* Bov-B LINE and *Caenorhabditis elegans* RTE-1 element (Malik and Eickbush 1998). PCR amplification of the 1.8-kb fragment of Bov-B LINEs was performed in a 100-μl volume with 1 μg of genomic DNA, each dNTP at 200 μM, 50 pmol of each primer, 2.5 U of AmpliTaq DNA polymerase, and buffer provided by the supplier of the enzyme (Perkin-Elmer). After an initial denaturation step of 5 min at 95°C, the PCR reactions were subjected to 30 cycles of amplification consisting of 2 min denaturation at 95°C, 2 min annealing at 60°C, and 2 min extension at 72°C, with a 5-min final extension at 72°C. The resulting PCR products were directly ligated into a pGEM vector using a pGEM-T-easy cloning kit (Promega) for sequence determination. The inserts were sequenced on both strands with an ABI fluorescent sequencing kit on an ABI 310 sequencer.

Isolation of a Full-Length Bov-B LINE Element from a *Vipera ammodytes* Genomic Library

The λ GEM-12 (Kordiš and Gubenšek 1996) genomic library was screened with the ^{35}S-labeled 1.8-kb fragment of the Bov-B LINE using the plaque hybridization method. Hybridization was carried out at 42°C for 20 h in a mixture of 6 × standard saline citrate (SSC), 5 × Denhardt's solution, 0.5% SDS, and denatured herring sperm DNA at 100 μg/ml in 50% formamide. The filters were washed successively with 6 × SSC and 2 × SSC at 35°C for 20 min each. The positive clones were rescreened by the same procedure. Phage DNA was prepared from plate lysates (Sambrook, Fritsch, and Maniatis 1989, p. 2.64) and digested with *Bam*HI, *Eco*RI, *Sac*I, and *Xho*I restriction enzymes. The resulting fragments were separated by gel electrophoresis on 0.7% agarose, transferred to Hybond-N membranes (Amersham), and hybridized at 42°C with the same probe as described above. Positive genomic fragments were subcloned into pUC19.

DNA- and Protein-Based Sequence Analyses

Computer-based nucleotide and protein searches of the GenBank databases with reptilian Bov-B LINEs and several representatives from the RTE clade were performed with the different BLAST (Altschul et al. 1990) search programs of the NCBI. Specialized EBI databases were searched with different Fasta programs (Pearson 1990). The following sequence databases were searched at NCBI (www.ncbi.nlm.nih.gov): nonredundant (NR), dbEST (expressed sequence tags database), dbSTS (sequence tagged sites database), dbGSS (genome survey sequences database), and HTGS (unfinished high-throughput genomic sequences—phases 0, 1, and 2); while at EBI (www.ebi.ac.uk), we searched parasite genomes (www.ebi.ac.uk/parasite-genome.html) and eukaryotic genomes (www.ebi.ac.uk/fasta33/genomes.html). Evolutionary rates were estimated by standard methods (Nei and Kumar 2000, p. 20). Poisson correction distances (d) were estimated by the equation $d = -\ln(1 - p)$, where p represents the proportion of different amino acids. The rate of amino acid substitution (r) was estimated by the standard equation $r = d/2T$, where T is the divergence time of the last common ancestor (LCA) of the species compared. Amino acid distances used in divergence-versus-age analysis were calculated from sequences of the complete RT domain using the MEGA2 program (Kumar et al. 2000).

Phylogenetic Analyses

The amino acid sequences of Bov-B LINE and other RTE representatives were aligned using CLUSTAL W (Thompson, Higgins, and Gibson 1994) with some manual refinements. Phylogenetic trees were inferred using the neighbor-joining (NJ) method (Saitou and Nei 1987) as implemented in TREECON and MEGA2 (Van de Peer and De Wachter 1994; Kumar et al. 2000). The significance of the various phylogenetic lineages was assessed by bootstrap analysis. Phylogenetic analyses

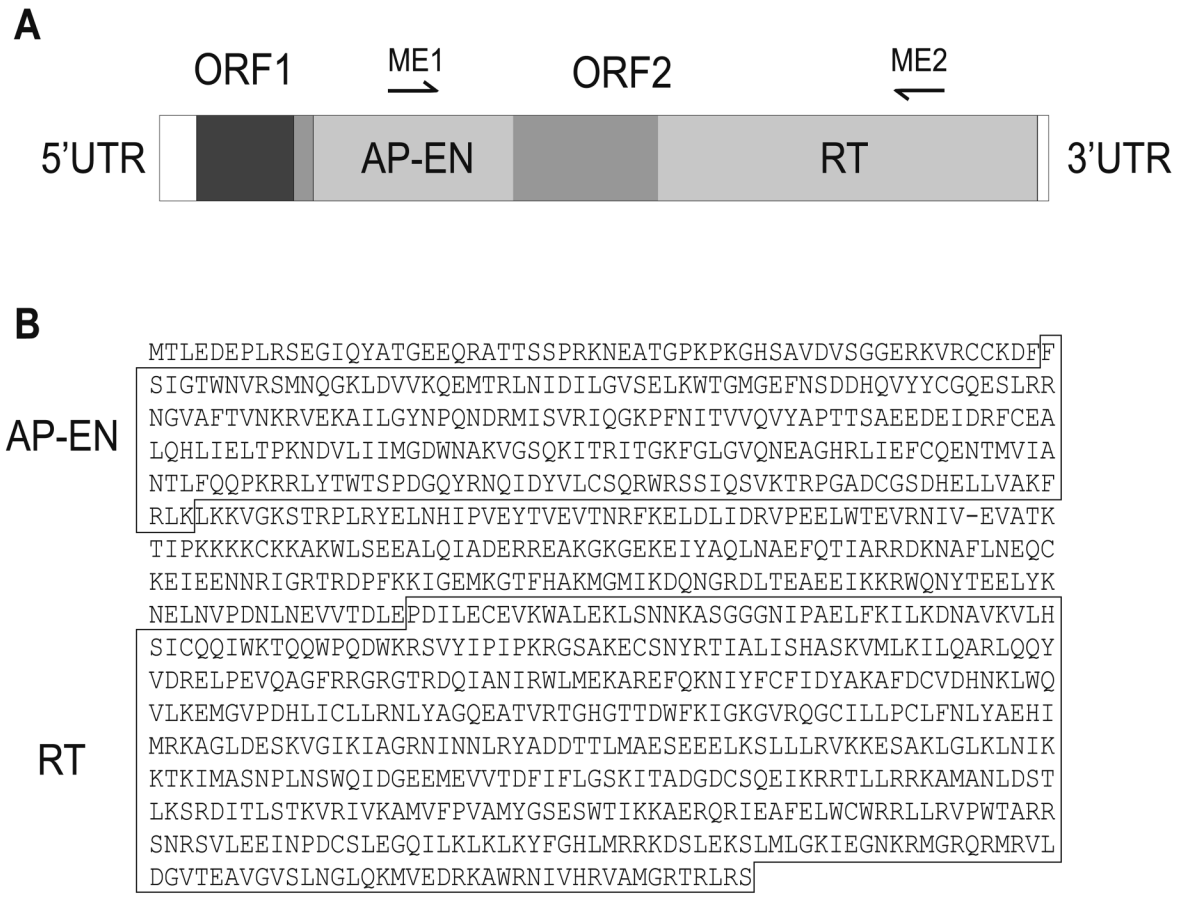

FIG. 1.—Structure and amino acid sequence of the full-length *Vipera ammodytes* Bov-B LINE element. *A,* The structure of the full-length *V. ammodytes* Bov-B LINE element. The locations of the primers (ME1, ME2) used for PCR amplification of the 1.8-kb fragment are indicated above. *B,* The deduced amino acid sequence of ORF2. AP-EN = apurinic/apyrimidinic endonuclease; RT = reverse transcriptase.

were performed on separate regions of the RTE elements, such as apurinic/apyrimidinic endonuclease (AP-EN), the palm/fingers subdomain of RT, and the thumb subdomain of RT. Several potential outgroups were tested, and group II intron RT (from Neurospora, accession number S07649), Schistosoma SR2 element AP-EN, and the thumb subdomain of RT (Drew et al. 1999) were selected, since the phylogenetic trees were in accord with the species phylogeny and taxonomy. Gaps in aligned sequences were removed for the purpose of analysis.

Results and Discussion

Structural Organization of the Full-Length Reptilian Bov-B LINE Element

After screening the *V. ammodytes* genomic library with the 1.8-kb PCR-amplified fragment of the *V. ammodytes* (Vam) Bov-B LINE, several genomic clones with strong hybridization signals were selected. One of them was subjected to further analysis. The full-length Bov-B LINE element from *V. ammodytes* is 3,229 bp long, and its structure, with amino acid translation, is shown in figure 1.

The Vam Bov-B LINE encodes a 1,027-amino-acid-long ORF2. Conceptual translation of the Vam Bov-B LINE sequence upstream of its major ORF revealed a 38-amino-acid-long ORF1 that does not overlap with ORF2. This very short putative ORF1 is typical for the RTE class of non-LTR retrotransposons and shows no similarity to any protein sequence in databases (Malik and Eickbush 1998; Drew et al. 1999). The 5′ untranslated region (UTR) is only 64 bp long, similar to the ruminant Bov-B LINE (*B. taurus*) and *C. elegans* RTE-1 elements (Malik and Eickbush 1998). The extremely short 3′ UTR of the Vam Bov-B LINE consists of four CAA repeats only, followed by TTCTA tandem repeats. The stop codon that terminates ORF2 is located immediately before the first CAA repeat. On both sides of the full-length reptilian Bov-B LINE element, different numbers of pentanucleotide repeats (TTCTA) exist, indicating that the Bov-B LINE might specifically insert in the microsatellite sequences. The Vam Bov-B LINE encodes an RT domain with 11 conserved segments, as found in all non-LTR retrotransposons (Malik and Eickbush 1998; Malik, Burke, and Eickbush 1999). The size of the entire RT domain (palm/fingers and thumb) is 503 amino acids, while the AP-EN domain contains 244 amino acids, and the intervening domain, located between these two domains, contains 194 amino acids.

On the basis of the nucleotide and amino acid sequences of the *V. ammodytes* Bov-B LINE, it is possible to reconstruct the amino acid sequence of the single pre-

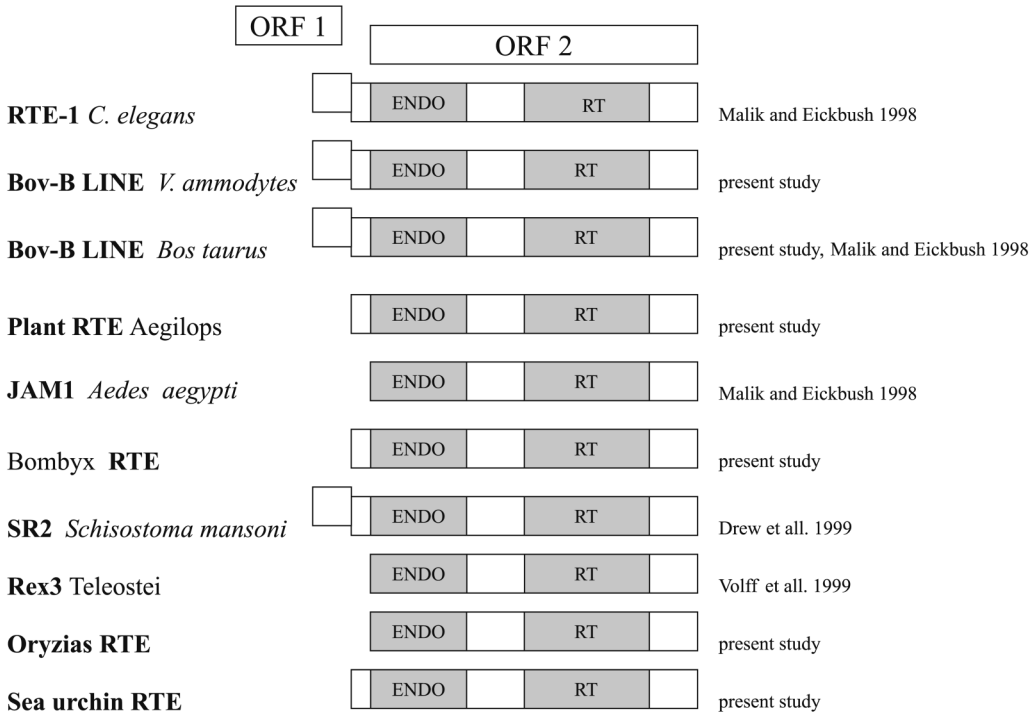

FIG. 2.—Structural organization of the full-length elements in the RTE clade. All currently known full-length elements from the RTE clade are presented. The endonuclease (ENDO) and RT domains within ORF2 are indicated by darker shading, while the putative ORF1 is shown as an unshaded box.

viously available full-length ruminant Bov-B LINE element (Szemraj et al. 1995; Okada and Hamada 1997). This element was previously reconstructed on the basis of the *C. elegans* RTE-1 element (Malik and Eickbush 1998) but contained some errors, such as the part of AP-EN region from *C. elegans* RTE-2 element. Nematode RTE-1 and bovine Bov-B LINE show only 25%–30% identity at the amino acid level, in contrast to the 75% identity between the full-length Vam and *B. taurus* Bov-B LINE elements.

Elements from the RTE clade are the shortest among the non-LTR retrotransposons and have a highly conserved structural organization (Malik and Eickbush 1998; Malik, Burke, and Eickbush 1999). Several representatives from the RTE clade have been analyzed, with some of them being full-length, such as the *C. elegans* RTE-1 element (Malik and Eickbush 1998), the *B. taurus* Bov-B LINE (BDDF) element (Szemraj et al. 1995; Okada and Hamada 1997; Malik and Eickbush 1998), the *Schistosoma mansoni* SR2 element (Drew et al. 1999), the *Rex3* element from teleost fishes (Volff et al. 1999), the JAM1 element from Aedes (Malik and Eickbush 1998), and now the full-length reptilian *V. ammodytes* Bov-B LINE (fig. 2). We discovered several novel nearly full length elements in GenBank databases, such as *Bombyx mori* Bov-B/RTE, *Strongylocentrotus* Bov-B/RTE, Oryzias RTE, and plant RTEs (fig. 2), and these will be described below in more detail.

Distribution of Bov-B LINEs

The discovery of the widespread distribution of Bov-B LINEs in Squamata through the use of the extreme C-terminal part of the RT domain (thumb region) (Kordiš and Gubenšek 1998) prompted an investigation of the distribution among vertebrates of a much larger fragment of Bov-B LINEs encoding both AP-EN and the palm/fingers subdomain of RT. A 1.8-kb fragment of Bov-B LINEs was amplified by PCR using primers complementary to conserved regions in ORF2 (fig. 1). Under normal-stringency reaction conditions, the primers yielded a product in the same Squamata species studied previously (Kordiš and Gubenšek 1998), in ruminants and in a marsupial, but not in any other vertebrate species tested.

We cloned 1.8-kb fragments and sequenced several (two to four) independent clones from each of 10 selected Squamata species. Intraspecies variability among several clones was very low and the same has been found for interspecies variability. Genetic distances were even smaller than those previously obtained for the extreme C-terminal part of the Bov-B LINEs (Kordiš and Gubenšek 1998). In some sequences, we found a few stop codons and short deletions; these result in frameshift mutations or in frame stop mutations.

It has been reported that marsupials, but not monotremes, also contain Bov-B LINEs (Gilbert and Labuda 1999, 2000), but no experimental data were provided. We amplified the 1.8-kb fragment and a shorter 0.5-kb (thumb subdomain) fragment from a marsupial (*Macropus*) and from the ruminants tested. Since the monotremes and marsupials form a sister group, Marsupionta (Janke, Xu, and Arnason 1997), the acquisition of Bov-B LINEs in marsupials or their loss in monotremes needs additional study with sequence and phylogenetic

```
Bos       FLGSKSLQDGDCSHEIKRHLLLGRKVMTNLDSILKSRDITLPTRSRLVKAMVFPVVLYGC
Python    --------------....RF.....A.AS.........V....KI.I.........M...
Macropus  --...ITA.N.YN.RM..R......AVA.............SSKLLIA.......AI..W

Bos       ESWTMKKAECRRIDAFELWCWRRLLRVPWTARRSNQSILKEISPGCSLEGLMLKLKLQYF
Python    ....I....R.................................N.D.....SIM....R..
Macropus  ...SIR...HD.....K..Y......I...E.T..HL.....NSEY...CQI...-----

Bos       SHLMRKYDLIGKTLMLGGIFFGGRRGR-RMRWLDGITHSMALFDGELWELVMDRRAWKAV
Python    G....REESLK......K.EGQK....Q......S..EA.SMKLSK.R.A.K..K..CNM
Macropus  -MRGQ--.SVE...F..K.KGKRK..QQ....I.SVMEQQKVSLDR.Q.I.V.....CTI

Bos       IHGVTKSQTRLGD
Python    V......R...R-
Macropus  FNR....R...R-
```

FIG. 3.—Alignment of the thumb subdomain of reverse transcriptase of Bov-B LINE elements from the cow, the ancestral snake lineage, and the marsupial. The alignment was constructed using the program CLUSTAL W (Thompson, Higgins, and Gibson 1994). The dots represent the amino acids conserved between all sequences.

analyses in order to clarify their origin in mammals. Bov-B LINEs may have originally been present in an ancestor of the mammals, or, alternatively, they could have been introduced on two separate occasions by HT into marsupial and ruminant genomes.

It is important to note that all other mammalian species tested—representatives of different mammalian orders: *Cetartiodactyla* (*Tragulus* [PCR was negative, although the AP-EN domain is already known], *Hippopotamus*, *Tursiops*, *Sus*), Carnivora (*Canis*), Rodentia (*Mus*), and Primates (*Homo*)—were negative by our PCR analyses. Additional support for the restricted and discontinuous presence of Bov-B LINEs in mammals provided an examination of mammalian sequences in GenBank databases. The currently known distribution of Bov-B LINEs in mammals is highly discontinuous; they are absent in most of the mammalian orders. Hybridization and PCR amplification data (Buntjer, Hoff, and Lenstra 1997; Kordiš and Gubenšek 1997, 1998; Shimamura et al. 1999; Gilbert and Labuda 2000) also do not support their widespread distribution in mammals, or even panvertebrate distribution, as suggested recently (Malik and Eickbush 1998; Gilbert and Labuda 1999). In order to determine their evolutionary origins in mammalian genomes, a much wider study in the context of the latest mammalian phylogeny (Madsen et al. 2001; Murphy et al. 2001) is needed. Marsupial Bov-B LINEs are not inconsistent with the hypothesis of their horizontal origin in Ruminantia.

Our preliminary sequence data of the marsupial Bov-B LINE encoding the thumb domain of RT (fig. 3) shows up to 78% identity at the DNA level and 65% at amino acid level with the ancestral snake lineage, a level of conservation similar to that previously observed for ruminants (Kordiš and Gubenšek 1998). This may indicate an additional horizontal transfer of Bov-B LINEs into the ancestor of marsupials at least 120 MYA. However, more extensive analyses of the monotremes, marsupials, and placental mammals will be required to conclusively answer the question of the origin of Bov-B LINEs and their current distribution in mammals.

Extreme Conservation of Bov-B LINEs Between Vertebrate Classes

Comparison of the newly available reptilian Bov-B LINEs (partial and full-length elements) with the elements from ruminants shows a very high level of conservation, both at the nucleotide and at the amino acid levels, throughout the entire length of the Bov-B LINE elements. It is not limited to the C-terminal part of RT (thumb) which was examined previously (Kordiš and Gubenšek 1998). The highest level of conservation between reptiles and ruminants was once again observed (as in Kordiš and Gubenšek 1998) between *Python* and *B. taurus*, reaching 86% identity at the amino acid level in the nearly 600 amino acids compared (fig. 4). Since the LCA of reptiles and mammals existed 310 MYA (Kumar and Hedges 1998; Hedges and Poling 1999), it is very difficult to explain the extreme nucleotide and amino acid conservation, together with the discontinuous distribution, without invoking the horizontal transfer of these elements as previously proposed (Kordiš and Gubenšek 1998). On the other hand, the only additional representatives of the RTE clade in vertebrates, the teleost-specific *Rex3* elements, show lower levels of amino acid similarity with vertebrate Bov-B LINEs than those between vertebrate Bov-B LINEs and invertebrate Bov-B/RTE elements (from *B. mori* and the sea urchin *Strongylocentrotus*).

We also compared the conservation of other available vertebrate non-LTR retrotransposons (table 1) for which no HT has been previously observed and a strict vertical mode of transmission has been recognized. Surprisingly, we found that the divergence between *Python* and bovine Bov-B LINEs (LCA 310 MYA) is the same as that between *Mus* and *Rattus* L1 elements (LCA 15 [40] MYA), providing another argument for HT.

```
Boa       TVIQVYAPTTDAEEAEVDRFYEDLQHLLELTPKKDVLFIIGDWNAKVGSQEITGITGKFG
Python    ..................................................I.......
Podarcis ..........G...T.I.Q........I.V.......L..............K.T.....
Bos       ....A.T..SN.K....EQ...N..R.........................TP.V....S

Boa       LGVQNEAGQRLIEFCQENTLAIGNTLFQQHKRRLYTWTSPDGQYRNQIDYVLCSQRWRSS
Python    ..........................................................
Podarcis ..D...................T.....................H......I.........
Bos       ..I.......S.......A.......................H...V..I.........

Boa       IQSVKTRPGADCGSDHELLIAKFRLKLKNVGKTTRPFRYDLNHIPYEYTVEVRNRFKGLD
Python    ..A................................L....D...C....K.M...R...
Podarcis ................Q......K....K.....G.V..N.SQ..............D..
Bos       ........E.........R...V.....K.............QT.CD.....T......Y

Boa       LVDRVPEELWTEVCNIIQKTATQTIPKKKKCKKAKWLSEEALQIAEERRRAKGKGERERY
Python    ...................A..................D......V...N..........
Podarcis ..........M.AR.....A......M.R.........N..........Q..CE.DS...
Bos       .L.K..D.......D.....GI....M-..............-..VK..EV.....K...

Boa       KQLNADFQRIARRDKKAFLNEQCKEIEENNRMGKTRDLFKKIGDIKGRFCAKMGMIKDKK
Python    ..........................................................
Podarcis RK.......................Q........R..........M....RT.ITI....S
Bos       .H...E.............SN...D..K..........L...R.T....H....L...RN

Boa       GRDLTEAEEIKKRWQEYTEELYKKDLNVPDQHKDVVTDLEPDILESEVKWALGSITNNKA
Python    ...............................V.A..........................
Podarcis .K......D.............Q..MD.SYTPGS..A..............E.L.KY..
Bos       .M......D.................HNQ.N.D..I.H.-.....C...R.VE...M...

Boa       SGDDSIPAELFKILKEDAVKVLHSICQQIWKTQQWPQDWKRSVYVPIPKKGNAKECSNYR
Python    ....................................R.....................
Podarcis .ES.G..................................E..R....I......S........
Bos       ..G.G..V...Q......L.........V..............FKL..E......S...H

Boa       TIALISHASKVMLKILQARLQQYVNRELPEVQAGFRRGRGTRDQIANLRWIMEKAREFQK
Python    ........................Q...D.............................
Podarcis ..................G..K...D.....................M............
Bos       ........................S...D......KS..........C.......D...

Boa       NIYFCFIDYAKAFDCVDHNKLWQILQEMGVPDHLTCLLRNLYAGQEATVRTGHGTTDWFQ
Python    ..........................................................
Podarcis D..................S....V..........I.......V.........Y.......
Bos       ......................K....V.I.E....S.....T................E

Boa       IGKGVRQGCILSPFLFNLYAEYIMRNARLDESQAGVKIAGRNLTNLRFA
Python    ............................E...................
Podarcis ........................K.G...................Y.
Bos       .....C...................R.G.E.T..........IN...Y.
```

FIG. 4.—Alignment of Bov-B LINE elements from the ancestral snake lineage, the lizard, and the cow. The alignment was constructed using the program CLUSTAL W (Thompson, Higgins, and Gibson 1994). The dots represent the amino acids conserved between all sequences.

The Bov-B LINEs are considerably more conserved than the typical cellular and mitochondrial proteins. This is demonstrated by analysis of the conservation of cellular and mitochondrial proteins between mammals (Ruminantia) and reptiles (Squamata). In most cellular proteins analyzed, we found 50%–70% divergence at the amino acid level (unpublished results). In a few exceptional cases, we found a very low level of amino acid divergence of proteins throughout all kingdoms, such as in TFIID, histones, glutamate receptors, and lactate dehydrogenases. However, in these cases, the level of conservation and the widespread distribution are correlated, with a very slow rise in divergence with time. Among vertebrates, the amino acid divergence is, in the extreme cases, 0%–20%, while in some rare cases such a low level of amino acid divergence can be observed through several kingdoms. Surprisingly, the comparison of some highly conserved mitochondrial genes, such as cytochrome *b*, between *Python* and *B. taurus* shows only 59% identity, contrasted with the 86% identity between Bov-B LINEs in the same species. In the reptiles we compared, the divergence of cytochrome *b*

Table 1
Conservation of the ORF2s in Vertebrate Non-LTR Retrotransposons: Amino Acid Divergence Versus time

Non-LTR Retrotransposon	LCA (divergence time in Myr)	Amino Acid Divergence (%)
LINE1		
Homo versus *Gorilla*	7–9	3
Homo versus *Nycticebus* (prosimian)	63	36
Homo versus *Canis*	65 (100)	38
Homo versus *Mus*	100	38
Homo versus *Didelphis* (marsupial)	150	53
Mus versus *Rattus*	15 (40)	18
LINE2		
Homo versus *Trimeresurus* (snake)	310	53
Homo versus *Fugu* (fish)	450	57
Trimeresurus versus *Fugu*	450	57
CR1 LINE		
Platemys (turtle) versus *Gallus* (chicken)	220	42
Gallus versus shark	450	70
Bov-B LINE		
Natrix (Colub.) versus *Walterinnesia* (Elap.)	45	10
Walterinnesia (Elap.) versus *Python* (Boidae)	150	18
Podarcis (lizard) versus *Natrix* (Colub.)	>150	18
Python (Reptilia) versus *Bos* (Mammalia)	310	17 (14.8 in RT)

NOTE.—ORF = open reading frame; LTR = long terminal repeat; LCA = last common ancestor; LINE = long interspersed nuclear element; RT = reverse transcriptase.

between the ancestral snake lineage, represented by *Python*, and evolutionary younger snake lineages shows 14%–25% amino acid divergence and an exceptional 39% with blind snakes; similar values were observed for lizards (36%–40% amino acid divergence). We may exclude any strong functional constraints acting on Bov-B LINEs for such a long period of time and any important biological role. In conclusion, our analyses show the very unusual conservation of Bov-B LINEs in relation to the evolutionary relationships of their hosts.

Wide Distribution of the RTE Clade

The distribution of the RTE clade of non-LTR retrotransposons was reanalyzed and found to be much wider than previously thought (table 2). By searching specific databases (NR, GSS, HTGS, EST, STS, parasite genomes, and completed eukaryote genomes) and by searching within specific taxonomic groups using the full-length sequence of the Vam Bov-B LINE and other

Table 2
Phylogenetic Distribution of the RTE Clade of Non-LTR Retrotransposons

Taxonomic Group	Phylum	Class	Order/Genus	Element Name	Reference
Stramenopiles	Phaeophyceae (brown algae)		*Laminaria*	—	Present study
Viridiplantae	Embryophyta	''Dicots''	*Solanum, Nicotiana, Platanus, Glycine*	Plant RTE	Present study
		''Monocots''	*Aegilops, Hordeum, Zea*	Plant RTE	Present study
Metazoa	Platyhelminthes	Trematoda	*Schistosoma, Paragonium*	SR 2 element	Drew et al. (1999)
	Nematoda		*Caenorhabditis,* other species	RTE-1, RTE-2	Malik and Eickbush (1998)
	Annelida	Oligochaeta	*Lumbricus*	—	Present study
	Arthropoda	Insecta	*Diptera* (*Aedes, Anopheles*)	JAM 1	Malik and Eickbush (1998)
			Lepidoptera (*Bombyx*)	Bombyx Bov-B	Malik and Eickbush (1998); present study
	Mollusca	Gastropoda	*Helix*	Mollusca Bov-B	Present study
		Cephalopoda	*Ommastrephes*	Mollusca RTE	Malik and Eickbush (1998)
	Echinodermata	Echinoidea	*Strongylocentrotus*	Sea urchin Bov-B	Present study
	Chordata	Actinopterygii	*Teleostei*	Rex 3	Volff et al. (1999)
			Oryzias, Mola	Oryzias RTE	Present study
		''Reptilia''	Squamata	Bov-B LINE	Kordiš and Gubenšek (1997, 1998, 1999*a*, 1999*b*); present study
		Mammalia	Ruminantia	Bov-B LINE	Szemraj et al. (1995); Okada and Hamada (1997)
			Marsupialia	Bov-B LINE	Gilbert and Labuda (1999, 2000); present study

Table 3
Novel Representatives of the RTE Clade

Taxonomic Group	Species	Element name	Accession No.
Stramenopiles	*Laminaria digitata*		AW400610, AW401018
Embryophyta			
Monoctos	*Aegilops tauschii*	Plant RTE	AF091802 (complete element)
	Hordeum vulgare		AF064563, AJ001317
	Zea mays		BE475977, U90128, AF100768, M81603, AF123535
	Triticum aestivum		BE402326
	Saccharum (sugarcane)		AJ293564
Dicots	*Glycine max*		AW350284, BF595877
	Medicago truncatula		AJ132891, BF644841, AC087771
	Solanum tuberosum		BF054569, U20345, S66866
	Solanum lycopersicum		U32444
	Lycopersicum esculentum		AI1896609, AQ367282, AF273333
	Nicotiana sylvestris		AB012638
	Nicotiana tabacum		AF261032
	Petunia		AF130352
	Pisum sativum		X90996
	Spinacia oleracea		X17031
	Malus domestica		AF053126
	Platanus racemosa		AF106842
Metazoa			
Trematoda	*Paragonimus heterotremus*	SR2	AZ254640
Annelida	*Lumbricus terrestris*		J05161
Arthropoda	*Anopheles gambiae*	JAM1	AL141965, AL155508
	Aedes albopictus		AF065437
	Bombyx mori	Bov-B/RTE	AP-EN: AV405248, AV404913, AV403564, AV403523; RT: AV406121, AV405248, AV404804, AU004927; thumb: AV406078, AF130337, D86212, U49854
Crustacea	*Artemia franciscana*		X81967
	Penaeus monodon		AF077590
Mollusca	*Helix pomatia*		AF109924
Echinodermata	Strongylocentrotus	Cel RTE-1 like	AF228874
		BovB/RTE	AP-EN: AZ202731, AZ158776, AZ175290, AZ177815; RT: AZ201020, AZ155116, AZ186300, AZ209815; thumb: AZ149051, AZ198614, AZ190360, AZ145938
Vertebrata	*Oryzias latipes*	Oryzias RTE	AB021490 (complete element)
	Mola mola		AF134630
	Tetraodon nigroviridis		AL189200, AL214774, AL311143

novel representatives of the RTE clade, we discovered several new representatives (table 3).

In insects, a nearly full length Bov-B-like element exists in *B. mori*, while in echinoderms, a nearly full length Bov-B-like element exists in the sea urchin *Strongylocentrotus*. In medaka fish (*Oryzias latipes*, accession number ABO21490) we found a nearly full length RTE element which was only 30% identical to the *Rex3* element from the same species or from other teleost fishes. Surprisingly, by searching public databases with the *Oryzias* RTE element, we found the closest relatives to this element in the plant kingdom. The plant RTE lineage is particularly interesting, since in a few species we found nearly full length elements (*Aegilops* and *Hordeum*). By TBLASTN searching of different databases with AP-EN, RT, and the C-terminal part of RT, we identified their widespread presence in a number of different monocots and dicots. We found that in closely related Gramineae species, such as *Aegilops* and *Hordeum*, they are highly conserved, but less so than vertebrate RTE lineages such as Bov-B LINEs or *Rex3* at similar divergence times. When they were compared between monocots and dicots (LCA around 200 MYA), the level of similarity dropped sharply.

We also detected short fragments of RT in brown algae (*Laminaria*); the size of the translated sequence was relatively short, encoding only 130 amino acids. Shorter fragments with significant similarity were also detected in annelids (*Lumbricus*), mollusks (*Helix pomatia*), and crustaceans (*Artemia*, *Pennaeus*). Additionally, there was a relatively high level of conservation (40%) between *Laminaria* and sea urchin RTEs, in contrast to the lower level between echinoderm RTEs and teleost *Rex3* elements, which showed only 30% amino acid identity.

In previous analyses of the RTE clade (Malik and Eickbush 1998; Drew et al. 1999; Volff et al. 1999), representatives from nematodes, arthropods, mollusks,

trematodes, and vertebrates have been described. These data were extended in the present study to cover a much wider taxonomic distribution. The RTE clade is seen to originate very early in the evolution of eukaryotes, since they are present in stramenopiles, in plants, and apparently in most metazoan phyla. Most of the novel representatives recognized by TBLASTN searching were their C-terminal domains only. Available sequence data of the RTE clade show that the level of conservation on a global scale is quite low, as was found for other clades of non-LTR retrotransposons (Malik and Eickbush 1998; Malik, Burke, and Eickbush 1999). The availability of these new basal eukaryotic RTE representatives will enable the design of novel oligonucleotide primers and consequently much denser sampling that will clarify their long-term evolution in eukaryotes. The RTE and L1 clades are the most widespread of the known eukaryotic non-LTR retrotransposons.

Evolutionary Relationships Among Bov-B LINEs and Other Representatives of the RTE Clade

To determine the evolutionary relationships of the Bov-B LINE elements and other representatives of the RTE clade, the amino acid sequences of AP-EN, RT, and the C-terminal part of RT (thumb subdomain) were subjected to phylogenetic analysis using the NJ algorithm (Saitou and Nei 1987). Because the level of sequence divergence within each species was insignificant compared with that between species, either a consensus or a single Bov-B LINE sequence was used to represent the elements of each species. Most Squamata Bov-B LINE elements are seen to be grouped in phylogenetic trees according to their species of origin (fig. 5).

In order to study the evolutionary relationships among reptilian and mammalian Bov-B LINEs, two mammalian DNA contaminants were included with all known reptilian (*Squamata*) and several ruminant Bov-B LINEs. One was from the brown rat (accession number M28630), and its origin is bovine, since calf thymus DNA was used as a carrier for transfected SV40 DNA. It has been suggested (Malik and Eickbush 1998) that rats contain Bov-B LINEs, but all other rat sequences in the GenBank database lack the Bov-B LINE. PCR and hybridization experiments were similarly negative. The other contaminant was taken from the human EST database, but once again, the PCR and hybridization experiments for humans were always negative. In NJ phylogenetic trees, the ruminant Bov-B LINEs always group together with Boidae snakes (boas and pythons), but lizards are in some cases (fig. 5*b*) branched out separately from snakes (and ruminants). Also, the brown rat and human contaminants always group together with ruminants, indicating the origin of the contamination.

NJ phylogenetic trees, based on amino acid sequence alignments of AP-EN, the palm/fingers subdomain of RT, and the thumb subdomain of RT domains, produce the same general pattern. In all of these trees, the Ruminantia Bov-B LINEs group together with the ancestral snake lineage (Boidae), as if they were the closest relatives (fig. 5). The same feature has previously been recognized at the DNA level, from the thumb subdomain of RT (Kordiš and Gubenšek 1998;, 1999*a*, 1999*b*). The unusual position of mammalian (ruminant) Bov-B LINEs inside the reptilian cluster is thus confirmed and is a clear indication of HT. These evolutionary relationships show that potential HT of Bov-B LINEs or any other TE can easily be recognized from phylogenetic trees based on either DNA or amino acid sequences of the short parts of RT or AP-EN, or even from analyses of the full-length elements.

In the phylogenetic analyses of the RTE clade, we included mostly elements without a large number of stop codons or difficulties in translation. Frameshift mutations and stop codons were frequently observed during TBLASTN searching, but in most species and RTE lineages the elements are relatively well conserved and can be easily translated. As an outgroup, we used the very ancient group II introns (*Neurospora*) for the palm/fingers subdomain of RT, while for AP-EN and the thumb subdomain of RT this ancient sequence was too divergent and not useful for alignment or phylogenetic analyses. *Schistosoma* SR2 elements AP-EN and the thumb subdomain were therefore used as outgroups in such trees. Phylogenetic reconstruction of evolutionary relationships among other RTE representatives showed several well-separated groups, such as *B. mori* Bov-B/RTE, sea urchin (*Strongylocentrotus*) Bov-B/RTE, nematode RTE, Rex3 elements from teleost fishes, and plant RTEs, together with *Oryzias* (medaka fish) RTE. The positions of trematoda SR2 and mosquito JAM1 are not well resolved, since they show only about 30% amino acid identity with all other representatives of the RTE clade. These phylogenetic analyses of the RTE clade used larger numbers of RTE representatives, from plants to vertebrates, than previous phylogenetic analyses (Malik and Eickbush 1998; Drew et al. 1999; Volff et al. 1999).

Divergence-Versus-Age Analysis of the RTE Clade

Amino acid distances between the RT domains of the RTE clade representatives are plotted against estimates of host divergence time in figure 6. The RTE lineages show a regular pattern of increased divergence with time that approaches saturation at the time of the split of Tetrapoda and Teleostei at 400 MYA (Kumar and Hedges 1998). Comparisons of species separated by more than 600 Myr have little resolution, as has been observed for R1 and R2 elements (Malik, Burke, and Eickbush 1999). Comparisons of amino acid distances versus host divergence time were made within and between plant and metazoan RTE lineages. One of the comparisons between monocot and dicot plants fell above the curve and probably represented a paralogous comparison. Three examples in which points fell markedly below the RTE curve were Squamata versus Ruminantia Bov-B LINEs (at 310 Myr), *B. mori* Bov-B/RTE versus vertebrate Bov-B LINEs (at 670 Myr), and plant RTE versus *Oryzias* (medaka fish) RTE (at 1,200 Myr). It has been proposed (Malik, Burke, and Eickbush 1999) that such points can be interpreted as an indication of

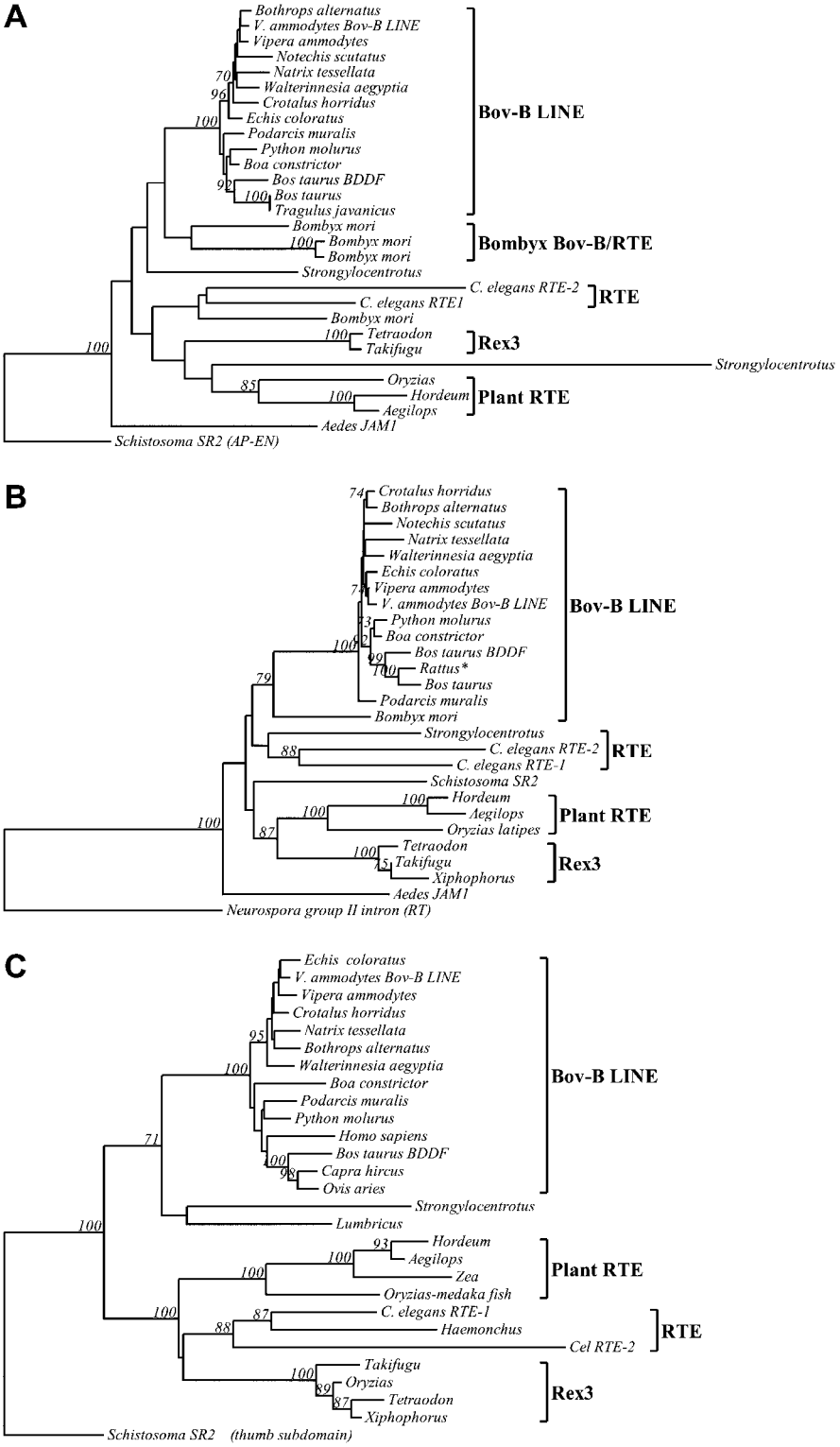

FIG. 5.—Neighbor-joining (NJ) phylogenetic trees of the RTE clade. The rooted NJ phylogenetic tree using the Poisson correction model and (*A*) the AP-EN domain of the *Schistosoma* SR2 element (Drew et al. 1999), (*B*) reverse transcriptase (RT) of the *Neurospora* group II intron, and (*C*) the thumb subdomain of RT of the *Schistosoma* SR2 element (Drew et al. 1999) as outgroups. All NJ trees were drawn by the TREECON program (Van de Peer and De Wachter 1994). They represent the bootstrap consensus following 1,000 replicates; nodes with confidence values greater than 70% are indicated.

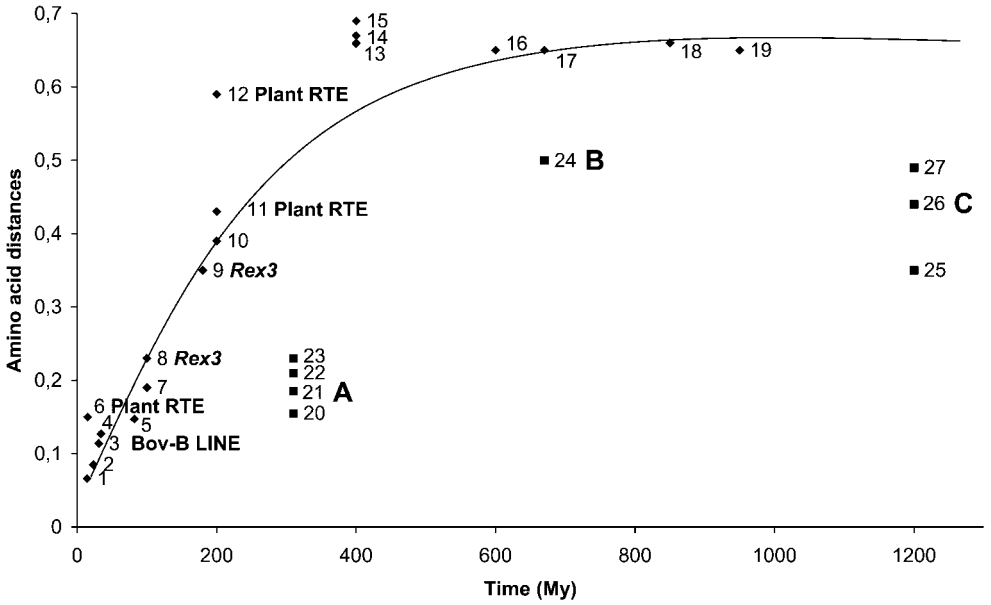

FIG. 6.—Divergence-versus-age analysis of the RTE clade. Amino acid distances were calculated from sequences of the complete reverse transcriptase (RT) domain. The RTE lineage from which each comparison was obtained is shown adjacent to data points; comparison points are in parentheses. For each host divergence time estimate, the elements used are as follows: Bov-B LINEs compared at 14 MYA: *Vipera ammodytes* (Vam) versus *Echis* (1), at 23 MYA; Vam versus *Bothrops* (2), at 31 MYA; *Crotalus* versus *Natrix* (Nte) (3), at 34 MYA; Vam versus Nte (4), at 82 MYA; *Python* (Pmo) versus *Podarcis* (Pmu) (5). *Rex3* elements compared at 100 MYA: *Xiphophorus* (Xma) versus *Esox* (7), Xma versus *Danio* (8), at 180 MYA; Xma versus *Anguilla* (9). Plant RTEs compared at 15 MYA: *Aegilops* (Ata) versus *Hordeum* (Hvu) (6), at 200 MYA; Ata versus *Nicotiana* (Nta) (10), Ata versus *Glycine* (Gma) (11), Ata versus *Platanus* (12). Comparisons between different RTE lineages: Xma *Rex3* versus Vam Bov-B LINEs (13), compared at 400 MYA, *Oryzias* RTE versus Vam Bov-B LINE (14), at 400 MYA, *Oryzias* RTE versus cow (Bta) Bov-B LINE (15), at 400 MYA; Vam Bov-B LINEs versus sea urchin (Spu) Bov-B/RTE (16), at 600 MYA; Xma *Rex3* versus *Bombyx* (Bmo) Bov-B/RTE (17), at 670 MYA; Spu versus Bmo Bov-B/RTEs (17), at 670 MYA; Vam Bov-B LINE versus *Caenorhabditis elegans* (Cel) RTE-1 (18), at 850 MYA; and *Schistosoma mansoni* SR2 versus Cel RTE-1 (19), at 950 MYA. Three examples of HT are shown by (*A*) Bov-B LINEs compared at 310 MYA: Bta versus Pmo (20), Bta versus Vam (21), Bta versus Pmu (22), and Bta versus Nte (23); (*B*) Vam Bov-B LINE versus Bmo Bov-B/RTE compared at 670 MYA (24); and (*C*) *Oryzias* RTE compared with plant RTE lineage at 1,200 MYA: *Oryzias* RTE versus Gma (25), *Oryzias* RTE versus Hvu (26), and *Oryzias* RTE versus Ata (27). Species divergence times are based on estimates by Kumar and Hedges (1998) for comparisons within vertebrates, by Hedges and Poling (1999) (and our unpublished data) for comparisons within reptiles, by Volff et al. (1999) for comparisons within teleost fishes, and by Feng, Cho, and Doolittle (1997) and Wang, Kumar, and Hedges (1999) for comparisons within plants and eukaryotes.

HT occurrence. This finding has important implications for the existence of HT in non-LTR retrotransposons.

An explanation for the low amino acid divergence and sequence homogeneity of the Bov-B LINEs in Squamata is the strong selective constraint in the ORF2 sequence coupled with strict vertical transmission. On the other hand, the very low amino acid divergence between Ruminantia and Squamata seems unlikely to indicate selection pressure acting on Bov-B LINEs separated from a common ancestor for such a length of time. Thus, the anomalous position of Bov-B LINEs (LCA 310 MYA) in figure 6 is a clear indication of an HT event.

The level of amino acid divergence between vertebrate Bov-B LINEs and invertebrate *B. mori* Bov-B-like elements seems relatively low with regard to their taxonomic position and their LCA with vertebrates (around 670 MYA). How is it possible that invertebrate Bov-B-like elements are much more similar to vertebrate Bov-B LINEs than they are to the teleost *Rex3* elements? The most parsimonious explanation, as supported by the anomalous position in figure 6, would be separate horizontal transfers of Bov-B LINEs of invertebrate origin into the ancestor of Squamata, the ancestor of Ruminantia and the ancestor of marsupials.

Regarding the amino acid divergence observed between medaka fish (*Oryzias*) and the plant RTE lineage with the age of the LCA between the plant and metazoan kingdoms being >1,200 Myr (Wang, Kumar, and Hedges 1999), their evolutionary relationships are explainable only by HT. Medaka fish, however, could not be the donor of the RTE element in the potential HT for the plants, since plant RTE elements are distributed throughout the Embryophyta.

Evolutionary Rates in the RTE Clade of Non-LTR Retrotransposons

Evolutionary rates for reptilian and mammalian Bov-B LINEs and the other representatives of the RTE clade have been estimated (table 4), leading to several interesting observations. First, the evolutionary rates in Bov-B LINEs are very low, as in *Rex3* elements from bony fishes, another representative of the RTE clade in vertebrates. Second, evolutionary rates of representatives of the RTE clade in vertebrates are slower than those in invertebrates. Third, evolutionary rates between endothermic (Ruminantia) and ectothermic (Squamata, Teleostei) vertebrates are different; slower rates of evolution were found in ectotherms. Fourth, long-term evo-

Table 4
Evolutionary Rates in the RTE Clade

Element Name	Species Compared	LCA (in Myr)	Evolutionary Rate (10^{-9})	Vertical Transfer	Horizontal Transfer
Squamata Bov-B LINE	*Boa* versus *Python*	48	0.42	+	−
	Python versus *Podarcis*	81.6	0.67	+	−
Ruminantia Bov-B LINE	*Bos* versus *Capra*	20	4.1	+	−
	Bos versus *Tragulus*	50	3.4		
Squamata versus *Ruminantia* Bov-B LINE	*Python* versus *Bos*	310	0.27	−	+
	Vipera ammodytes versus *Bos*	310	0.35	−	+
Bov-B LINE versus *Rex* 3	*V. ammodytes* versus *Xiphophorus*	400	1.46	+	−
Bov-B LINE versus *Oryzias* RTE	*V. ammodytes* versus *Oryzias*	400	1.9	+	−
Bov-B LINE versus sea urchin RTE	*V. ammodytes* versus *Strongylocentrotus*	600	0.72	+	−
Rex 3	*Xiphophorus* versus *Anguilla*	180	1.18	+	−
Plant RTE versus *Oryzias* RTE	*Hordeum* versus *Oryzias*	1,200	0.27	−	+
	Glycine versus *Oryzias*	1,200	0.18	−	+
	Aegilops versus *Oryzias*	1,200	0.31	−	+
Plant RTE	*Aegilops* versus *Platanus*	200	2.1	+	−
	Aegilops versus *Glycine*	200	1.31	+	−
Bombyx RTE versus *Rex* 3	*Bombyx* versus *Tetraodon*	670	0.73	+	−

lutionary rates in the RTE clade correlate with divergence times, with a significantly lower evolutionary rate suggesting the occurrence of HT.

The finding that HT events may cause a big slowdown in evolutionary rate is novel for the TE field. The observed slowdown in evolutionary rate between evolutionarily distantly related taxa (Squamata and Ruminantia) is comparable with that of the evolutionary rate of histones, which is highly unlikely, since both groups show, among evolutionarily closely related taxa, higher (Squamata), or even much higher (Ruminantia), evolutionary rates. Evolutionary rates of Bov-B LINEs in ruminants are much higher than in Squamata, with the same level of amino acid divergence observed between *Python* and the cow (LCA 310 MYA) being reached in ruminants in 20 Myr. This anomalous slowdown in evolutionary rate is typical of potential HT events in both classes of TEs (unpublished data).

Multiple Lineages of RTE Elements Are Present in Some Genomes

Previous studies have confirmed that multiple RTE lineages coexist in nematode genomes, such as the RTE-1 and RTE-2 lineages in *C. elegans* (Malik and Eickbush 1998). In a search of the sea urchin genome database, one group of elements was found to be more closely related to the *C. elegans* RTE-1 elements, while another lineage was much more similar to the Bov-B LINEs (table 3). Another example of multiple lineages is demonstrated in the genome of medaka fish (*Oryzias*), where the *Rex3* (Volff et al. 1999) and *Oryzias* RTE elements coexist, showing only 30%–35% identity, but where the *Oryzias* RTE element surprisingly shows the strongest similarity to the plant RTE lineage.

Evolutionary History of Bov-B LINEs in Deuterostomia

By searching the GSS database, we found that Bov-B-like elements are present in the most ancestral lineage of Deuterostomia, Echinodermata, represented by the sea urchin *Strongylocentrotus*. These elements exhibit 40%–50% amino acid identity with the Squamata and Ruminantia Bov-B LINEs. However, they have not been retained in Urochordata (e.g., *Ciona* [GSS database] and *Halocynthia* [EST database]), Cephalochordata (*Branchiostoma* [HTGS database]), hagfishes (*Myxine*), lampreys (Petromyzontiformes), bony fishes (Teleostei—they contain another lineage, the Rex3 elements), cartilaginous fishes (Chondrichthyes), or Amphibia (PCR data and *Xenopus* [EST database]). They appear only in the most ancestral reptiles, in Squamata (Janke et al. 2001), but are absent from the turtles and Archosauria (crocodiles and birds). In mammals, they appear only in marsupials and ruminants. Why, then, are Bov-B LINEs not present in all vertebrates?

Discontinuous distribution of Bov-B LINEs in Deuterostomia can be explained by two alternative hypotheses. The first proposes that Bov-B LINEs were present in the LCA of Deuterostomia and have been transmitted vertically. A complex evolutionary scenario involving an unreasonably high number of evolutionary losses is necessary to explain the very sporadic distribution of Bov-B LINEs. Even the Bov-B LINE lineage is not retained in any of the vertebrate lineages leading to reptiles and mammals. Once lost from a lineage, it would be absent from all descendant taxa. The second hypothesis involves the possibility of HT. Three events of HT in the ancestors of Squamata, Ruminantia, and Marsupialia can quite simply explain their current distribution, providing the most parsimonious explanation for the discontinuous distribution.

Horizontal Transfer Versus Alternative Hypotheses

Evolutionary studies of TEs often yield element-generated phylogenies that are incongruent with the host species phylogenies, such as those we report, but these inconsistencies are not always indicative of HT (Capy, Anxolabehere, and Langin 1994). Many factors can obscure phylogenetic reconstruction of multicopy TEs. Comparison of paralogous copies of elements and varying rates of sequence evolution of TE copies within and

between species are factors which can yield incongruent phylogenies even under conditions of strict vertical transmission. Ancestral polymorphism, coupled with independent assortment of copies into the descendant species, inequality of substitution rates in TE sequences in different species, and the stochastic loss of TEs from a few taxa, could obscure phylogenetic reconstruction of TEs and lead to incongruence in phylogenetic reconstruction. Alternative explanations are often hard to dismiss conclusively, especially when TEs from closely related taxa are compared.

HT has been traditionally inferred when a high degree of similarity exists between TEs, coupled with a long divergence time of their respective host species. Incongruence between TE and host phylogenies, or the absence of the TE in question from taxa closely related to that into which the TE was supposedly transferred horizontally, can help corroborate that inference.

By comparing the divergence of TE amino acid sequences with those observed for host genes evolving under similar or stronger selective constraints, an HT event can be inferred whenever the divergence among TE sequences is significantly lower than that observed for the host proteins. We found that the Bov-B LINEs compared between Ruminantia and Squamata (LCA 310 MYA) show much lower divergence than the typical cellular and mitochondrial proteins from the same taxa, strongly indicating HT. No other vertebrate non-LTR retrotransposons, evolving by strict vertical transmission, show such an anomaly.

Selection and recent HT have very different consequences both for the level of congruence expected between TE and host phylogenies and for the proportionality of TE divergence relative to host gene divergence when sister groups of different ages are compared (Silva and Kidwell 2000). If the low TE divergence resulted from selection and the elements were transmitted vertically, TE and host phylogenies would be expected to be congruent. TE and host gene divergence should be correlated (Silva and Kidwell 2000). We observed this pattern in the Squamata Bov-B LINEs, as shown by very low sequence divergence and an element phylogeny that is in accord with the host phylogeny.

If recent HT caused the low TE divergence, then the TE phylogeny is not necessarily expected to be congruent with that of the host species. Also, TE and host gene divergence should not be correlated (Silva and Kidwell 2000). The erroneous phylogenetic position of Ruminantia and the comparison with typical cellular and mitochondrial proteins shows this pattern and strongly supports the occurrence of HT.

Evolutionary analysis of non-LTR retrotransposons (Malik, Burke, and Eickbush 1999) suggests that vertical transmission is the most common mode of inheritance of these elements, but, according to the present study, one should not exclude the possibility of occasional HT events. Only five HT events, between Squamata/Ruminantia, Marsupialia/Ruminantia, invertebrates/vertebrates, *Oryzias* fishes/plants, and *Laminaria*/sea urchin, are needed to explain all distributional and phylogenetic discrepancies in the RTE clade. It is obvious that Bov-B LINEs and some other RTE clade representatives are evolving at a very slow rate, at least in reptilian genomes (Bov-B LINEs) and teleost fishes (*Rex3*). Taken together, the data from the present study suggest several explanations which are not necessarily mutually exclusive to explain the apparent discrepancies between RTE clade and species evolution.

Mechanism of HT

One of the problems in dealing with the HT of TEs is that very little is known about the actual mechanisms through which this transmission could occur. Viruses, bacteria, fungi, parasites, or symbionts could act as potential vectors. All that the vector would need is a broad host range and the ability to somehow gain access to the germ line (Kidwell 1993). The potential HT events in the RTE clade have occurred between different classes, phyla, or kingdoms. For all of these cases, a possible ecological connection exists between them (parasitizing or feeding). The ability of any element from the RTE clade to function in such diverse hosts indicates that they are not dependent on specific host factors for their activity.

Even though we cannot directly show experimentally how the HT events may have occurred, we propose that the evidence for HT in our data indicates that the non-LTR retrotransposons (at least among several RTE lineages) do occasionally transfer horizontally. Previous conclusions to the contrary (Malik, Burke, and Eickbush 1999) are based on the relative rarity of HT events, which makes it necessary that many species be sampled in order to detect them.

Conclusions

The results presented here show that reptiles contain full-length Bov-B LINE elements, which are highly conserved in Squamata. A very high level of sequence conservation exists between the ancestral snake lineage and Ruminantia. Bov-B LINEs are considerably more conserved than the typical cellular and mitochondrial proteins. The RTE clade, together with the L1 clade, has been shown to be one of the most widespread non-LTR retrotransposons, originating very early in the evolution of eukaryotes. Evolutionary rates in some RTE lineages are very low. Using different independent methods, including the divergence-versus-age analysis, several examples of HT in the RTE clade have been recognized, with important implications for the existence of HT in non-LTR retrotransposons.

Supplementary Material

The new sequences used in this paper have been deposited in the GenBank database (accession numbers AF332663–AF332697).

Acknowledgments

We thank Prof. R. H. Pain for critical reading of the manuscript, Prof. J. A. Lenstra for tissue samples,

and anonymous reviewers for critical comments on an earlier version of this manuscript. This work was supported by the Ministry of Science and Technology of Slovenia by research program PO-050-0106.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

BENNETZEN, J. L. 2000. Transposable element contributions to plant gene and genome evolution. Plant Mol. Biol. **42**:251–269.

BUNTJER, J. B., I. A. HOFF, and J. A. LENSTRA. 1997. Artiodactyl interspersed DNA repeats in cetacean genomes. J. Mol. Evol. **45**:66–69.

CAPY, P., D. ANXOLABEHERE, and T. LANGIN. 1994. The strange phylogenies of transposable elements: are horizontal transfers the only explanation? Trends Genet. **10**:7–12.

CHO, Y., Y. L. QIU, P. KUHLMAN, and J. D. PALMER. 1998. Explosive invasion of plant mitochondria by a group I intron. Proc. Natl. Acad. Sci. USA **95**:14244–14249.

CLARK, J. B., and M. G. KIDWELL. 1997. A phylogenetic perspective on P transposable element evolution in Drosophila. Proc. Natl. Acad. Sci. USA **94**:11428–11433.

DREW, A. C., D. J. MINCHELLA, L. T. KING, D. ROLLINSON, and P. J. BRINDLEY. 1999. SR2 elements, non-long terminal repeat retrotransposons of the RTE-1 lineage from the human blood fluke *Schistosoma mansoni.* Mol. Biol. Evol. **16**: 1256–1269.

FEDOROFF, N. 2000. Transposons and genome evolution in plants. Proc. Natl. Acad. Sci. USA **97**:7002–7007.

FENG, D., G. CHO, and R. F. DOOLITTLE. 1997. Determining divergence times with a protein clock: update and reevaluation. Proc. Natl. Acad. Sci. USA **94**:13028–13033.

GILBERT, N., and D. LABUDA. 1999. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. Proc. Natl. Acad. Sci. USA **96**:2869–2874.

———. 2000. Evolutionary inventions and continuity of CORE-SINEs in mammals. J. Mol. Biol. **298**:365–377.

GONZALEZ, P., and H. A. LESSIOS. 1999. Evolution of sea urchin retroviral-like (SURL) elements: evidence from 40 echinoid species. Mol. Biol. Evol. **16**:938–952.

GUEIROS-FILHO, F. J., and S. M. BEVERLEY. 1997. Trans-kingdom transposition of the Drosophila element *mariner* within the protozoan Leishmania. Science **276**:1716–1719.

HARTL, D. L., A. R. LOHE, and E. R. LOZOVSKAYA. 1997. Modern thoughts on an ancyent marinere: function, evolution, regulation. Annu. Rev. Genet. **31**:337–358.

HEDGES, S. B., and L. L. POLING. 1999. A molecular phylogeny of reptiles. Science **283**:998–1001.

HOUCK, M. A., J. B. CLARK, K. R. PETERSON, and M. G. KIDWELL. 1991. Possible horizontal transfer of Drosophila genes by the mite *Proctolaelaps regalis.* Science **253**:1125–1128.

JANKE, A., D. ERPENBECK, M. NILSSON, and U. ARNASON. 2001. The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. Proc. R. Soc. Lond. B Biol. Sci. **268**:623–631.

JANKE, A., X. XU, and U. ARNASON. 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationships among Monotremata, Marsupialia, and Eutheria. Proc. Natl. Acad. Sci. USA **94**: 1276–1281.

JORDAN, I. K., and J. F. MCDONALD. 1998. Evolution of the *copia* retrotransposon in the *Drosophila melanogaster* species subgroup. Mol. Biol. Evol. **15**:1160–1171.

JORDAN, I. K., L. V. MATYUNINA, and J. F. MCDONALD. 1999. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. Proc. Natl. Acad. Sci. USA **96**: 12621–12625.

KIDWELL, M. G. 1993. Lateral transfer in natural populations of eukaryotes. Annu. Rev. Genet. **27**:235–256.

KIDWELL, M. G., and D. LISCH. 1997. Transposable elements as sources of variation in animals and plants. Proc. Natl. Acad. Sci. USA **94**:7704–7711.

———. 2000. Transposable elements and host genome evolution. Trends Ecol. Evol. **15**:95–99.

KORDIŠ, D., and F. GUBENŠEK. 1996. Ammodytoxin C gene helps to elucidate the irregular structure of Crotalinae group II phospholipase A_2 genes. Eur. J. Biochem. **240**:83–90.

———. 1997. Bov-B long interspersed repeated DNA (LINE) sequences are present in *Vipera ammodytes* phospholipase A_2 genes and in genomes of Viperidae snakes. Eur. J. Biochem. **246**:772–779.

———. 1998. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. Proc. Natl. Acad. Sci. USA **95**:10704–10709.

———. 1999*a*. Horizontal transfer of non-LTR retrotransposons in vertebrates. Genetica **107**:121–128.

———. 1999*b*. Molecular evolution of Bov-B LINEs in vertebrates. Gene **238**:171–178.

KUMAR, A., and J. L. BENNETZEN. 1999. Plant retrotransposons. Annu. Rev. Genet. **33**:479–532.

KUMAR, S., and S. B. HEDGES. 1998. A molecular timescale for vertebrate evolution. Nature **392**:917–920.

KUMAR, S., K. TAMURA, I. JAKOBSEN, and M. NEI. 2000. MEGA 2: molecular evolutionary genetics analysis program. Version 2.0. Pennsylvania State University, University Park, Penn.; and Arizona State University, Tempe, Ariz.

LOHE, A. R., E. N. MORIYAMA, D. A. LIDHOLM, and D. L. HARTL. 1995. Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. Mol. Biol. Evol. **12**:62–72.

MADSEN, O., M. SCALLY, C. J. DOUADY, D. J. KAO, R. W. DEBRY, R. ADKINS, H. M. AMRINE, M. J. STANHOPE, W. W. DE JONG, and M. S. SPRINGER. 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature **409**:610–614.

MALIK, H. S., W. D. BURKE, and T. H. EICKBUSH. 1999. The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. **16**:793–805.

MALIK, H. S., and T. H. EICKBUSH. 1998. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. Mol. Biol. Evol. **15**:1123–1134.

———. 2000. NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from *Caenorhabditis elegans.* Genetics **154**:193–203.

MURPHY, W. J., E. EIZIRIK, W. E. JOHNSON, Y. P. ZHANG, O. A. RYDER, and S. J. O'BRIEN. 2001. Molecular phylogenetics and the origin of placental mammals. Nature **409**: 614–618.

NEI, M., and S. KUMAR. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

OKADA, N., and M. HAMADA. 1997. The 3′ ends of tRNA-derived SINEs originated from the 3′ ends of LINEs: a new example from the bovine genome. J. Mol. Evol. **44**(Suppl. 1):S52–S56.

PEARSON, W. R. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. Meth. Enzymol. **183**:63–98.

ROBERTSON, H. M. 1993. The *mariner* transposable element is widespread in insects. Nature **362**:241–245.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SHAPIRO, J. A. 1999. Transposable elements as the key to a 21st century view of evolution. Genetica **107**:171–179.

SHIMAMURA, M., H. ABE, M. NIKAIDO, K. OHSHIMA, and N. OKADA. 1999. Genealogy of families of SINEs in cetaceans and artiodactyls: the presence of a huge superfamily of $tRNA^{Glu}$-derived families of SINEs. Mol. Biol. Evol. **16**: 1046–1060.

SILVA, J. C., and M. G. KIDWELL. 2000. Horizontal transfer and selection in the evolution of *P* elements. Mol. Biol. Evol. **17**:1542–1557.

SZEMRAJ, J., G. PLUCIENNICZAK, J. JAWORSKI, and A. PLUCIENNICZAK. 1995. Bovine *Alu*-like sequences mediate transposition of a new site-specific retroelement. Gene **152**: 261–264.

TERZIAN, C., C. FERRAZ, J. DEMAILLE, and A. BUCHETON. 2000. Evolution of the *gypsy* endogenous retrovirus in the *Drosophila melanogaster* subgroup. Mol. Biol. Evol. **17**: 908–914.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 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.

VAN DE PEER, Y., and R. DE WACHTER. 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput. Appl. Biosci. **10**:569–570.

VOLFF, J. N., C. KORTING, K. SWEENEY, and M. SCHARTL. 1999. The non-LTR retrotransposon *Rex3* from the fish Xiphophorus is widespread among teleosts. Mol. Biol. Evol. **16**:1427–1438.

WANG, D. Y., S. KUMAR, and S. B. HEDGES. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proc. R. Soc. Lond. B Biol. Sci. **266**:163–171.

PIERRE CAPY, reviewing editor

Accepted May 28, 2001