Automated Scanning for Phylogenetically Informative Transposed Elements in Rodents

ASTRID FARWICK,¹ URSULA JORDAN,¹ GEORG FUELLEN,²,³,⁴ DOROTHÉE HUCHON,⁵ FRANÇOIS CATZEFILS,⁶ JÜRGEN BROSIUS,¹ AND JÜRGEN SCHMITZ¹

¹Institute of Experimental Pathology, ZMBE, University of Münster, Von-Esmarch-Str. 56, 48149 Münster, Germany; E-mail: jueschm@uni-muenster.de (J.S.)
²Department of Medicine, AG Bioinformatics, University of Münster, Domagkstr. 3, 48149 Münster, Germany
³Department of Biology, AG Bioinformatics, University of Münster, Schloßplatz 4, 48149 Münster, Germany
⁴Institute for Mathematics and Computer Science, University Greifswald, Jahnstr. 15a, 17487 Greifswald, Germany
⁵Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel
⁶Institut des Sciences de l’Evolution, Université Montpellier 2, 34095 Montpellier, France

Abstract.—Transposed elements constitute an attractive, useful source of phylogenetic markers to elucidate the evolutionary history of their hosts. Frequent and successive amplifications over evolutionary time are important requirements for utilizing their presence or absence as landmarks of evolution. Although transposed elements are well distributed in rodent taxa, the generally high degree of genomic sequence divergence among species complicates our access to presence/absence data. With this in mind we developed a novel, high-throughput computational strategy, called CPAL (Conserved Presence/Absence Locus-finder), to identify genome-wide distributed, phylogenetically informative transposed elements flanked by highly conserved regions. From a total of 232 extracted chromosomal mouse loci we randomly selected 14 of these plus 2 others from previous test screens and attempted to amplify them via PCR in representative rodent species. All loci were amplifiable and ultimately contributed 31 phylogenetically informative markers distributed throughout the major groups of Rodentia.

Rodentia is the most diverse order among placental mammals. Extant rodent species represent half of all placental diversity (2021 species divided into 28 families; Wilson and Reeder, 1993). The phylogenetic relationships among rodents have been strongly debated in the last decades. Based on morphological and paleontological data, rodents have a single origin and were traditionally divided into either two (Hystricognathi and Sciurognathi; Tullberg, 1899) or three (Myomorpha, Sciuromorpha and Hystricomorpha; Brandt, 1855) suborders. However, the characters underlying these classifications have been shown to be homoplastic and the taxonomic divisions do not represent phylogenetic clades (e.g., Vianey-Liaud, 1985; Hartenberger, 1985; Nedbal et al., 1996; Huchon et al., 2000, 2002; Adkins et al., 2001). Debates concerning the relationships within the suborder Sciurognathi and their phylogenetic relationships with Hystricognathi have been the subject of numerous morphological papers (reviewed in Luckett and Hartenberger, 1985; Jaeger, 1988). By contrast, the monophyly of most of the rodent families is usually accepted (Hartenberger, 1985; Wilson and Reeder, 1993; McKenna and Bell, 1997).

Although early molecular studies complicated our understanding of rodent evolution by suggesting a paraphyletic grouping of rodents (e.g., Graur et al., 1991; D’Erchia et al., 1996; Reyes et al., 1998), rodent paraphyly has been highly contested. The current prevailing view is that rodent paraphyly was erroneously inferred due to limited taxon sampling, long-branch attraction artifacts, and/or the use of oversimplified models (Luckett and Hartenberger, 1993; Philippe, 1997; Sullivan and Swoford, 1997). Recent molecular analyses suggest that Rodentia are monophyletic and divided into seven, well-supported groups: (1) Myodonta (rats, mice, jerboas); (2) Anomaluroomorpha (scaly-tailed flying squirrels, springhares); (3) Castoridae (beavers); (4) Geomyoidea (pocket gophers, pocket mice); (5) Ctenohystrica (gundi, porcupines, guinea-pigs); (6) Sciuroidea (mountain beavers, squirrels, woodchucks); and (7) Gliridae (dormice). The phylogenetic relationships among these lineages have not been reliably inferred from these studies, but a general consensus based on independent sequence data indicates that rodents are formed of three major clades, a “mouse-related clade” (1 to 4), Ctenohystrica (5), and a “squirrel-related clade” (6, 7) (Adkins et al., 2001, 2003; Murphy et al., 2001a; Huchon et al., 2002; DeBry, 2003). Although the second and third of these major clades are also supported by morphological and paleontological data, the categorization of the first clade was surprising because it had never been previously suggested. Moreover, a recent analysis using the nuclear-encoded gene IRBP (DeBry and Sagel, 2001) supports the classical hypothesis that groups Castoridae with Sciuroidea and not the grouping of Castoridae with Geomyoidea, Anomaluroidea, and Myodonta, as suggested by other molecular data. Thus, the relationships among rodents, and in particular the validity of this latter grouping, is far from conclusive; additional information is necessary to determine whether the “mouse-related clade” is valid.

In the period following their divergence from a common ancestor with the primate lineage, about 85 million years ago (Mya) (Yang et al., 2004), rodents have experienced intense mobile element activity of 7SL RNA- and tRNA-derived SINEs (Short Interspersed Nucleotide Elements).
SINEs in Rodents

**7SL RNA**-derived SINEs

- FAM
- FLAM-A-like

**tRNA**-derived SINEs

- MIR

**Rodent specific SINEs**

- PB1 (FLAM-A)
- BC1
- PB1D7, D9, D10
- ID-family
- B1, 4.5S
- B1-dID
- B4 (RSINE2 or B1-ID)
- MEN

**4.5S**

**I** and **B3** are **B2**-derived small RNAs (Serdobova and Kramerov, 1998; Lee et al., 1998). The origin of **DIP** elements from a **B2**-like ancestor is uncertain (dashed arrow). **B1** and **ID** combined to form two dimeric SINE families, **B1-dID** (Kramerov and Vassetzky, 2001) and **B4** (RSINE2 or B1-ID) (Lee et al., 1998; Waterston et al., 2002). In the squirrel genome, an additional dimer is represented by **MEN**, an **ID-like-B1** chimera (Serdobova and Kramerov, 1998). The curved lines connect the original monomeric parts of the dimers.
high rates of evolution (e.g., Wu and Li, 1985; Catzeflis et al., 1999; Kramerov and V assetzky, 2001); and B2 and B4, SINE elements are present in Dipodidae; g/s-B1dID elements are shared by Sciuroids, Gliridae, Hystricognathi, and Castoridae (Kramerov et al., 1999; Kramerov and V assetzky, 2001); and B2 and 4.5S, SINE elements are restricted to Muroids (Serdobova and Kramerov, 1998; Gogolevskaia and Kramerov, 2002). Thus, there exists a strong case for the use of presence/absence patterns of retrotransposed SINEs as phylogenetic markers in Rodentia, a strategy that has successfully provided virtually ambiguity-free reconstructions of various other evolutionary histories (e.g., Nikaido et al., 1999; Salem et al., 2003; Roos et al., 2004; Schmitz et al., 2001; Tatout et al., 1999; Kriegs et al. 2006). However, rodent sequences frequently show a very high sequence divergence making comparative presence/absence analyses of SINEs difficult and, thus far, not very successful. Experimentally screening for phylogenetic informative SINEs by establishing genomic libraries from selected taxa, isolating clones that contain SINE loci, determining the sequences, and designing PCR primers in flanking regions, followed by Zoo-PCR of representative species, offer a powerful strategy for certain taxonomic groups. However, the predominantly intergenic locations of randomly selected SINEs are too diverged in rodents for successful comparative PCR in rodents. On the other hand, selecting conserved loci with embedded SINEs from available genomic sequences can eliminate labor-intensive, time-consuming work.

Thus, to improve the efficacy, and hopefully the success, of this approach, we have developed a fully automatic, computational strategy to efficiently identify retrotransposed SINE presence/absence with highly conserved flanking regions. Application of this strategy yielded hundreds of possible phylogenetically informative presence/absence patterns at orthologous loci throughout the various rodent species that promise to fully elucidate the evolutionary history of rodent species. Our initial extensive characterization of selected loci demonstrates the validity of the screen and provides the first useful rodent SINE markers.

**Materials and Methods**

**Computational Strategies to Find Informative Presence/Absence Loci**

Sequence data and DNA-DNA hybridization have indicated that some rodent taxa (particularly Muridae, Caviomorpha, and Geomyoidea) are characterized by high rates of evolution (e.g., Wu and Li, 1985; Catzeflis et al., 1987; Philippe, 1997; Huchon et al., 1999, 2000; Adkins et al., 2001). The high rate of substitution in specific rodent genomes presents a challenge in the development of special strategies to identify phylogenetically informative presence/absence patterns. Highly conserved sequence regions flanking insertions are essential for successful PCR amplification of a spectrum of orthologous loci in diverged species. Manually searching for suitable loci is time consuming and not amenable to high throughput techniques. This prompted us to create CPAL (Conserved Presence/Absence Locus-finder), a fully automated computational search program utilizing Bioperl (Stajich et al., 2002). The underlying objective of this program is to locate mouse SINEs inserted in introns and to identify conserved regions in the flanking exons in order to design universal rodent PCR primers. As shown in Figure 2, CPAL is organized into four main processes: (1) An NCBI GenBank search for annotated exon-intron-exon structures in the species of choice, in our case *Mus musculus* (Bioperl module: Bio::DB::Query::GenBank). Introns no larger than 1 kb with their flanking exons are extracted for further analyses. Restricting the structures to introns of <1 kb facilitates their amplification by PCR. (2) A local RepeatMasker screen of the selected sequences for interspersed repeats, in our case all SINEs known in rodents (Fig. 1) (Bioperl module: Bio::Tools::RepeatMasker). (3) A StandAloneBlast alignment of the mouse exons against sequences of a reference species of choice, in our case all human chromosomes, to reveal loci conserved in both mouse and human (Bioperl module: Bio::Tools::Run::StandAloneBlast). (4) A Clustal W (Thompson et al., 1997) alignment of all mouse exon-flanked repeat loci against the human reference sequences from point 3 (Bioperl module: Bio::Tools::Run::Alignment::Clustalw). The program is generally applicable to various organisms with annotated sequences and any available sequences of reference species. Intron lengths and specific repeat searches are optional parameters.

An exhaustive search for mouse SINE loci with conserved flanks corresponding to human sequences was performed on a Dual 1.8 GHz PowerPC G5 in about 20 h and recovered 374 mouse/human alignments. Based on the intron-flanking conserved sequences of 16 different loci, we generated sets of PCR primers to exemplarily amplify the loci in a representative taxonomic sampling of rodents. The Bioperl script is available at http://zmbe2.uni-muenster.de/expath/addmat/CPAL.

**Amplification of Informative Presence/Absence Loci**

DNA was processed by standard protocols (Sambrook et al., 1989) from tissues of the following rodent species: Cricetidae: *Cricetus cricetus* (black-bellied hamster), *Ondatra zibethicus* (muskrat); Muridae: *Meriones unguiculatus* (Mongolian gerbil); Spalacidae: *Nannospalax ehrenbergi* (Ehrenberg’s mole-rat); Dipodidae: *Jaculus jaculus* (lesser Egyptian jerboa); Pedetidae: *Pedetes capensis* (springhare); Anomaluridae: *Anomalurus* sp. (scaly-tailed squirrel); Castoridae: *Castor fiber* (Eurasian beaver); Heteromyidae: *Heteromys gaumeri* (Gaumer’s spiny...
**Figure 2.** Flow chart of the Conserved Presence/Absence Locus-finder computational strategy. The CPAL Bioperl script begins with a GenBank search using the query “Mus musculus and exon” and excludes mRNAs and cDNAs. All hits are scanned for introns not longer than 1 kb, flanked by exonic sequences. With a local version of the RepeatMasker, CPAL selects introns with rodent-specific SINEs (-rod) and excludes sequences of low complexity (-nolow). A StandAloneBlast search of the exonic flanks against all human chromosomes reveals loci conserved in mouse and human. A CLUSTAL W alignment produces a FASTA output file with potential presence/absence loci in rodent species flanked by highly conserved regions facilitating the generation of conserved primers for Zoo-PCR in rodents.

**Presence/Absence Analyses**

We applied the RepeatMasker option (-rod) of the CPAL script to mouse intronic sequences, thereby focusing on rodent-specific and mammalian-wide repeats. Thus, all selected loci represented the “presence” state of a given SINE in mouse, in which the intronic transposed elements were visible as larger fragments after PCR amplification (for an example see Fig. 3). Conversely, the “absence” states of the transposed element, denoting the plesiomorph condition prior to integration, were visible as smaller PCR products. All species carrying a given transposed element at the orthologous position were inferred to have derived from a common ancestor.

For some loci, the PCR patterns were more difficult to interpret because of additional independent SINE insertions in different lineages. The sequencing of all PCR products was thus an important step to fully characterize specific presence/absence patterns. This is exemplified by locus 11 (Fig. 3), in which sequence analysis revealed a PB1 SINE in the large fragments of all Muroidea that was not present in the small fragments of other rodents. Additionally, some taxa (Cricetus cricetus and Ondatra zibethicus) contained larger bands than the mouse because of the independent integration of an additional B1 SINE in the ancestor of these two species. Furthermore, an independent insertion of a B2 SINE occurred in Ondatra zibethicus. Hence, markers in this locus support both the monophyly of Muroidea and the monophyly of Cricetidae. Additional, small size variations are due to random
indels. Careful sequence analysis revealed the presence of several of these additional retroposed elements in some species that were not present in mouse. These fortuitous findings proved to be very useful in analyzing phylogenetic lineages not leading to mouse.

**Phylogenetic Analyses**

There are two different strategies for using presence/absence data for phylogenetic inferences. The first strategy includes a phylogenetic reconstruction based on a presence/absence matrix. For this type of data the ancestral state is safely defined by the clear character polarity. Using the 31 presence/absence markers for phylogenetic inference we search for the most parsimonious tree using the irrev.up option of character transformation implemented in PAUP* 4.0b10 (Swofford, 2000) in a heuristic search using the lago-

**RESULTS**

**CPAL-Derived SINE-Loci**

A run of CPAL on mouse annotated sequences extracted 374 mouse/human alignments with potential phylogenetically informative presence/absence loci of transposed elements flanked by highly conserved exonic sequence regions. The deletion of duplicated sequences and ambiguous BLAST hits (BLAST hits < 30 basepairs) from our data set reduced this to 232 loci. From these we randomly selected 14 loci for Zoo-PCR in rodents. In addition, during the early test phases of CPAL we also scanned annotated sequences of the hamster Cricetulus griseus against human reference sequences, which yielded marker 12 (for chromosomal location see Supplementary Table 2, online at http://systematicbiology.org). Furthermore, while inversely testing CPAL for human phylogenetic informative loci against mouse reference sequences, we discovered the rodent-specific marker 16a. This rodent specific ID element was found as an additional insertion after aligning human and mouse sequences. Because its flanking exons are not annotated in mouse, we would not detect this marker when scanning mouse sequences with CPAL, and have therefore included it in this analysis. Owing perhaps to the stringent conditions and criteria of the CPAL screen, most loci were PCR-amplifiable in a critical species sampling, allowing to draw our phylogenetic conclusions. In some cases multiple-fragment PCR products were amplified, all of which were sequenced to exclude false amplification and potential pseudogenes. However,
considering the high degree of genomic sequence divergence among rodent species, a large number of the investigated taxa were not amplifiable for one or the other markers.

As the search strategy was based on transposed elements in mouse, many of the markers were informative only for evolutionary splits associated with the lineage leading to this species. Fortunately, apart from the transposed elements detected in mouse, some of the loci also contained additional integrations in other rodent taxa that were not necessarily present in mouse. The screen revealed 92 such fortuitous integrations, 17 of which could be used as additional phylogenetic markers (Fig. 4). The remaining 75 additional insertions appear to be species-specific, at least with respect to the reduced taxonomic sampling used in our analysis (see Supplementary Table 3, online at http://systematicbiology.org). Thus, from the 16 loci examined, a total of 31 phylogenetically informative, virtually homoplasy-free presence/absence markers were identified. All flanking regions of the 31 markers could be well aligned and insertions could be clearly assigned to specific SINE consensus sequences (for examples see Fig. 5).

**Retroposed SINEs in the “Mouse-Related Clade”**

As summarized in Figures 4 and 6, the 16 loci could be amplified in various species distributed throughout the entire rodent order and yielded a variety of retroposed elements. Two B1 elements (markers 1 and 2a) group *Mus musculus*, *Rattus norvegicus*, and *Meriones unguiculatus* and are not found in other rodents. Six transposed elements (markers 3 to 7, 8a) were present in all Muroidea except *Nannospalax ehrenbergi* (Spalacidae). Taking the clade Muridae as a clear prior hypothesis, these six markers significantly confirm its monophyly (*P < 0.004; [6 0 0]; Waddell et al., 2001). Furthermore, all members of the Muroidea that we examined contained an orthologous ID4 element (marker 9) that is absent in *Jaculus jaculus* and other rodents. Thus, these markers support the classification of Muroidea as Spalacidae plus Eumuroidea, here represented by *Mus* and *Rattus* (Murinae), *Meriones* (Gerbillinae), *Cricetus* (Cricetinae), and *Ondatra* (Arvicolinae) (Steppan et al., 2004). Three SINES, ID_B1, PB1, and ID4 (markers 10, 11a, and 12, respectively), were amplified in all Muroidea species plus *Jaculus jaculus* but not in other rodents. Together, these three markers statistically confirm the monophyly of all Myodontia (*P < 0.04; [3 0 0]). A B1 SINE element (marker 13a) was present in all Myodontia plus the two Anomaluridae species as well as in the Castoridae and Geomyoidea species, but was absent in all other rodents, providing positive evidence for the monophyly of the “mouse-related” rodent clade.

In addition to the 14 transposed elements extracted from the mouse genome by the CPAL screen, three additional integrations in some of the same orthologous loci were revealed during the Zoo-PCR that did not occur in the lineage leading to mouse but do shed light on evolutionary relationships within the mouse-related clade. One of these, a B1 SINE element (marker 11b), integrated in the common ancestor of *Cricetus cricetus* and *Ondatra zibethicus* (Cricetidae). Likewise, PB1 and ID4 SINEs (markers 13b, c) were found in both the *Pedetes capensis* and *Anomalurus sp.* and were absent in all other rodents, thus supporting the monophyly of the Anomaluridae.

**Retroposed SINEs in Ctenohystrica**

Because the CPAL search was a screen of the mouse genome, all the Ctenohystrica-specific markers belonged necessarily to the group of fortuitous, additional integrations absent in *Mus musculus* and the mouse-related clade. A PB1 element (marker 13d) was detected in all the Ctenohystrica and was absent in all other rodents. Two B1 elements (markers 5b, 13e) were detected in all investigated Caviomorpha but were absent in *Ctenodactylus gundi* and other rodents. Three SINE elements, two B1 and an ID2 (markers 2b–d) integrated at various positions in the same locus of *Proechimys cunvieri*, *Myocastor coypus*, and *Octodon degus*, thus providing significant support for a common ancestor of these three (*P < 0.04; [3 0 0]). Two further integrations, both ID4 SINEs, (markers 14a, b) cluster *Proechimys cunvieri* and *Myocastor coypus*, and ID4, two ID2, and B1 elements (markers 15, 16b, c) were found at orthologous positions in both *Cavia porcellus* and *Dolichotis patagonum* (significant support *P < 0.04; [3 0 0]); none of which were found in other members of the Ctenohystrica or in other rodents. It should be mentioned that loci 14 and 15 were not amplifiable in some species, making it impossible to identify in which lineage the SINE insertion occurred. However, in these two loci we did find other insertions (markers 14a, b and 15) with conclusive presence/absence patterns that were not present in mouse (see Fig. 4).

**Retroposed SINEs in the “Squirrel-Related Clade”**

A PB1 and ID4 SINE (markers 8b, c) were revealed in both *Sciurus vulgaris* and *Marmota marmota* that were not present in either *Glis glis* or representative members of the mouse-related and Ctenohystrica clades.

**Rodent Monophyly**

In early preliminary testing of the CPAL program we found an ID4 element in rodents (marker 16a). Identification of the corresponding empty target site in lagomorphs and humans supports the monophyly of all rodents.

**Distribution of SINEs at the 16 Analyzed Loci**

In deep rodent splits we found predominantly ID4 and PB1-related SINEs (Fig. 6), both of which are also present in terminal branches of the tree (Supplementary Table 3, online at http://systematicbiology.org). This indicates a long period of activity for these elements and emphasizes their suitability for phylogenetic analyses. Note that all
FIGURE 4. Presence/absence patterns of analyzed SINE insertion loci. (+) denotes the presence and (−) absence of a given diagnostic transposed element for all investigated loci (1 to 16) in representative rodent species. Question marks indicate that PCR amplification of the given transposed element was not successful in the corresponding species. The triangles indicate large deletions including the orthologous regions of the insertions. The lagomorphs (Lepus europaeus or Oryctolagus cuniculus) were used as the outgroup to rodents.
FIGURE 5. Representative presence/absence alignments. The 5′ and 3′ parts of the inserted SINE elements are shown together with their immediate flanking regions. Possible direct repeats (DR) are boxed. Marker names correspond to those in Figures 4 and 6. Mm = Mus musculus; Mu = Meriones unguiculatus; Cc = Cricetus cricetus; Ne = Nannospalax ehrenbergii; Pca = Pedetes capensis; Gg = Glis glis; Jj = Jaculus jaculus; Sv = Sciurus vulgaris; Pcu = Proechimys cuvieri; Mc = Myocastor coypus; Dp = Dolichotis patagonum; Cop = Coendou prehensilis; Mma = Marmota marmota; Cgu = Ctenodactylus gundi; Le = Lepus europaeus. (A)n = A-rich region.
analyzed transposed elements were identified and assigned by the RepeatMasker program.

**DISCUSSION**

**SINEs as Phylogenetic Markers**

SINEs as presence/absence markers provide an excellent, virtually homoplasy-free source to substantiate phylogenetic scenarios and have, therefore, gained widespread application in evolutionary biology. To take advantage of orthologous SINEs as phylogenetic markers in a given group, there must be a sufficient number of SINEs that actively transposed before the divergence points of speciation. However, even though the large number of transposed SINEs in rodents (Kramerov and Vassetzky, 2005) satisfies this criteria and promises an enormous resource for evolutionary studies, the high sequence divergence inherent in Rodentia hampers a comparative study of SINE presence/absence data at orthologous loci in the various species. This is especially evident using classical methods based on experimentally tracing SINES in predominantly intergenic sequence regions. Consequently, even though the distribution of rodent SINE families and subfamilies (i.e.,
B2 SINEs, Serdobova and Kramerov, 1998; 4.5S RNA, Gogolevskaya et al., 2005; BC1 RNA, Martignetti and Brosius, 1993) has contributed to our understanding of rodent phylogeny (see points I to III in Fig. 6), no significant phylogenetic information was available based on the analysis of SINE insertions.

Schmitz et al. (2001) utilized an alternative strategy to identify mono-locus SINE markers representative of primate splits that occurred more than 55 Mya by selecting SINEs in short intronic sequences flanked by highly conserved exons. The conserved exon sequences facilitated the design of conserved PCR primers for Zoo-PCR. This approach demands an extensive, manual GenBank search to identify a few suitable loci. In the present work, we automated this process by generating and applying a Bioperl script, called CPAL, to perform an exhaustive screen of all suitable loci in rodent species. The two principal prerequisites for such a search are a well-annotated genome, like that of the mouse, and extensive sequence information from a distinct reference species, like human. To demonstrate its effectiveness, we applied CPAL to the mouse genome and extracted 232 phylogenetically informative SINE loci flanked by highly conserved exons. Despite a high rate of evolution in the mouse genome, the stringent search profile of CPAL enabled us, in most cases, to recognize direct repeats flanking the insertions as well as the unoccupied, empty target sites in species lacking the insertion (e.g., Fig. 5). A comparative Zoo-PCR analysis of 16 such sites in orthologous loci in 22 rodent species provides the first 31 phylogenetically informative SINE presence/absence markers in rodents. Based on the volume of significant data from these first analyses, we expect to find hundreds of reliable markers when all of the initial 232 loci are analyzed, presumably enough to cover all major splits of the rodent tree.

However, such phylogenetic SINE data are not without their caveats; the new sequences from such a screen must be meticulously examined to verify the orthology of both the SINE elements themselves and their insertion sites. Two possible sources of homoplasy in SINE presence/absence data have been reported: precise deletion (reversion) and independent insertion at the same genomic locus.

The search for SINE presence/absence markers based on mouse sequence information revealed many phylogenetically informative integrations in the mouse-related...
clade. Fortunately, the high retropositional activity in rodents also promises the solution of branchings in the other major rodent clades as well. Moreover, in addition to the retroposons detected in mouse, 17 additional integrations, not present in mouse, were found in other rodent species that could be used to corroborate evidence one way or the other in open questions of rodent phylogeny. A definite problem using SINEs as landmarks of evolution, in particular for highly diverged sequences such as those in rodents, is to PCR amplify loci from sufficient representatives of all taxonomic groups. Our strategy optimized the PCR efficiency by amplifying presence/absence loci from highly conserved exonic flanks. However, the preconditions were still not sufficient to achieve a complete species sampling due largely to the presence of large indels or high sequence divergences (see Fig. 4).

Rodent Phylogeny
The SINE presence/absence information retrieved from our comparative Zoo–PCR analyses of loci uncovered by the CPAL script are summarized in Figures 4 and 6. The monophyly of rodents is supported by the integration of an ID4 element at orthologous loci in all analyzed rodents. It is worth noting that the monophyly of rodents is also supported by the presence of BC1 small cytoplasmic RNA in all rodents and its absence in other mammals (Martignetti and Brosius 1993). Among the three major infraordinal clades identified by molecular studies (DeBry 2003; Huchon et al., 2002; Adkins et al., 2001), our results support the monophyly of the mouse-related clade (marker 13a) and of the Ctenohystrica clade (marker 13d), but we have yet to find support for the squirrel-related clade.

SINEs in the Mouse-Related Clade
This clade, unanticipated by morphological studies, groups various rodents from diverse biogeographical origins. For example, it has been suggested that Myodonta originated from Asian hystricomorph rodents; that Anomaluromorpha might have originated from African hystricomorphs and that Geomyidae and Castoridae are sciurid rodents from North America (e.g., Vianey-Liaud 1985; Hartenberger 1998; Korth 1994). Molecular data only marginally support the monophyly of the mouse-related clade (BP < 80%) (Adkins et al., 2001, 2003; Murphy et al., 2001a; Huchon et al., 2002; DeBry 2003). Even though the mouse-related clade, as such, was first identified by molecular studies, the very recent analysis of dental characters in extant and fossil rodents by Marivaux et al. (2004) provides the first morphological support for this clade. Although more data are necessary for full verification, our recovery of a common B1 insertion (marker 13a) strengthens support for this unexpected rodent clade.

The mouse-related clade has been divided into three lineages based on molecular data: Myodonta, Anomalauromorpha and Castoridae + Heteromyidae (Adkins et al., 2003; Murphy et al., 2001a; Huchon et al., 2002; DeBry, 2003). These three lineages are not corroborated by morphological studies (e.g., Marivaux et al., 2004; Hartenberger, 1985). In particular, the monophyly of Anomalauromorpha (Anomaluridae and Pedetidae) is still an open question. Although middle ear structure (Lavocat and Parent, 1985) and carotidial arterial pattern (Bugge, 1985) support the grouping of Pedetidae and Anomaluromorpha, incisor enamel microstructure supports the grouping of Pedetidae with Ctenohystrica because these lineages share a derived multiserial enamel (Martin, 1995). Two SINE markers (13b, c) corroborate the monophyly of Anomaluromorpha, thus supporting the molecular results of Montgelard et al. (2002) and Huchon et al. (2002). Concerning the Myodonta, three of the SINE markers (10, 11a, and 12) presented here support the monophyly of this clade. Several nodes among this clade previously established by both molecular and morphological data, including, for example, the monophyly of Muroidea (e.g., Michaux et al., 2001; Steppan et al., 2004; Musser and Carleton 1993), are also supported by SINE insertion sites (marker 9) and thus provide validation of the methodology presented here.

SINEs in Ctenohystrica and the Squirrel-Related Clade
There are also several phylogenetic relationships within the Ctenohystrica that have been supported by sequence data but contested by morphological data, which are now corroborated by our data. These include the monophyly of Ctenohystrica (Ctenodactylyidae plus Hystricognathi) (review in Huchon et al., 2000), the monophyly of Caviomorpha, and the monophyly of the group of Proechimys, Myocastor, and Octodon (review in Huchon and Douzery 2001). Furthermore, we found two markers (8b, c) present exclusively in Sciuroidea.

Conclusion and Outlook
The present work contributes an initial step to facilitate the application of orthologous SINE data to the solution of rodent phylogeny. Although it has been previously recognized that SINE insertions might provide a reliable source of phylogenetic data, the extremely laborious search methods available rendered these valuable pieces of information beyond our grasp. Here we have demonstrated the successful application of the exhaustive and automated CPAL search strategy by the analysis of a small, random selection of orthologous loci and have already contributed to a better understanding of the evolution of rodent taxa. The remaining, more than 200, as yet unanalyzed, potential phylogenetically informative presence/absence loci located in rodents are currently being processed in a high-throughput experimental approach and are expected to shed significant light on the hitherto highly controversial problems of rodent evolution. The unresolved interrelationships among the mouse-related, the Ctenohystrica, and the squirrel-related clades is only one of several issues that requires further investigation. However, insertions of ID and B1 related elements, as shown in the present study, were active in the critical time window and therefore are
promising candidates to provide abundant information concerning the various questionable branching orders. We are also in the process of converting the underlying Bioperl algorithm of CPAL to a user-friendly, Java-based application that can flexibly perform exhaustive searches in other organisms as well.

ACKNOWLEDGMENTS

We thank Sandra Honigs for providing tissue of *Ctenodactylus gundi*, Michael Hermes for *Castor fiber*, Peter Vogel for *Glis glis*, Jörg Molten for *Dolichotis patagonum*, and Georg Hüging for *Ondatra zibethicus* and *Myocastor coypus*. Dominique Allaine and Aurélie Cohas for *Marmota marmota*, and Jan Ole Krieger for organizing the tissue transfer. We appreciate the assistance of Kathrin Weber, Huyen Pham, Hermann Kräling, Anna Shaik, and Carolin Wesseling in PCR amplification and sequencing. Many thanks go to Marsha Bundman for editorial assistance. This work was supported by the Deutsche Forschungsgemeinschaft (SCHM 1469 to I.S. and J.B.) and the United States–Israel Binational Science Foundation (BSF) (2004–07 to D.H.).

REFERENCES

Brosius, J. 1999. RNAs from all categories generate retrosequences that may be exalted as novel genes or regulatory elements. Gene 238:115–134.


First Submitted 4 November 2005; reviews returned 31 December 2005; final acceptance 10 May 2006

Associate Editor: Andrew Shedlock