Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

Syst. Biol. 55(6):936–948, 2006
Copyright © Society of Systematic Biologists
ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150601064806

Automated Scanning for Phylogenetically Informative Transposed Elements in Rodents

ASTRID FARWICK,[1] URSULA JORDAN,[1] GEORG FUELLEN,[2,3,4] DOROTHÉE HUCHON,[5] FRANÇOIS CATZEFLIS,[6] JÜRGEN BROSIUS,[1] AND JÜRGEN SCHMITZ[1]

[1]*Institute of Experimental Pathology, ZMBE, University of Münster, Von-Esmarch-Str. 56, 48149 Münster, Germany; E-mail: jueschm@uni-muenster.de (J.S.)*
[2]*Department of Medicine, AG Bioinformatics, University of Münster, Domagkstr. 3, 48149 Münster, Germany*
[3]*Department of Biology, AG Bioinformatics, University of Münster, Schlossplatz 4, 48149 Münster, Germany*
[4]*Institute for Mathematics and Computer Science, University Greifswald, Jahnstr. 15a, 17487 Greifswald, Germany*
[5]*Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel*
[6]*Institut 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. [Automated scanning; CPAL; mouse; phylogeny; rodents; SINE.]

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 Swofford, 1997). Recent molecular analyses suggest that Rodentia are monophyletic and divided into seven, well-supported groups: (1) Myodonta (rats, mice, jerboas); (2) Anomaluromorpha (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

Astrid Farwick and Georg Fuellen contributed equally to this publication.

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

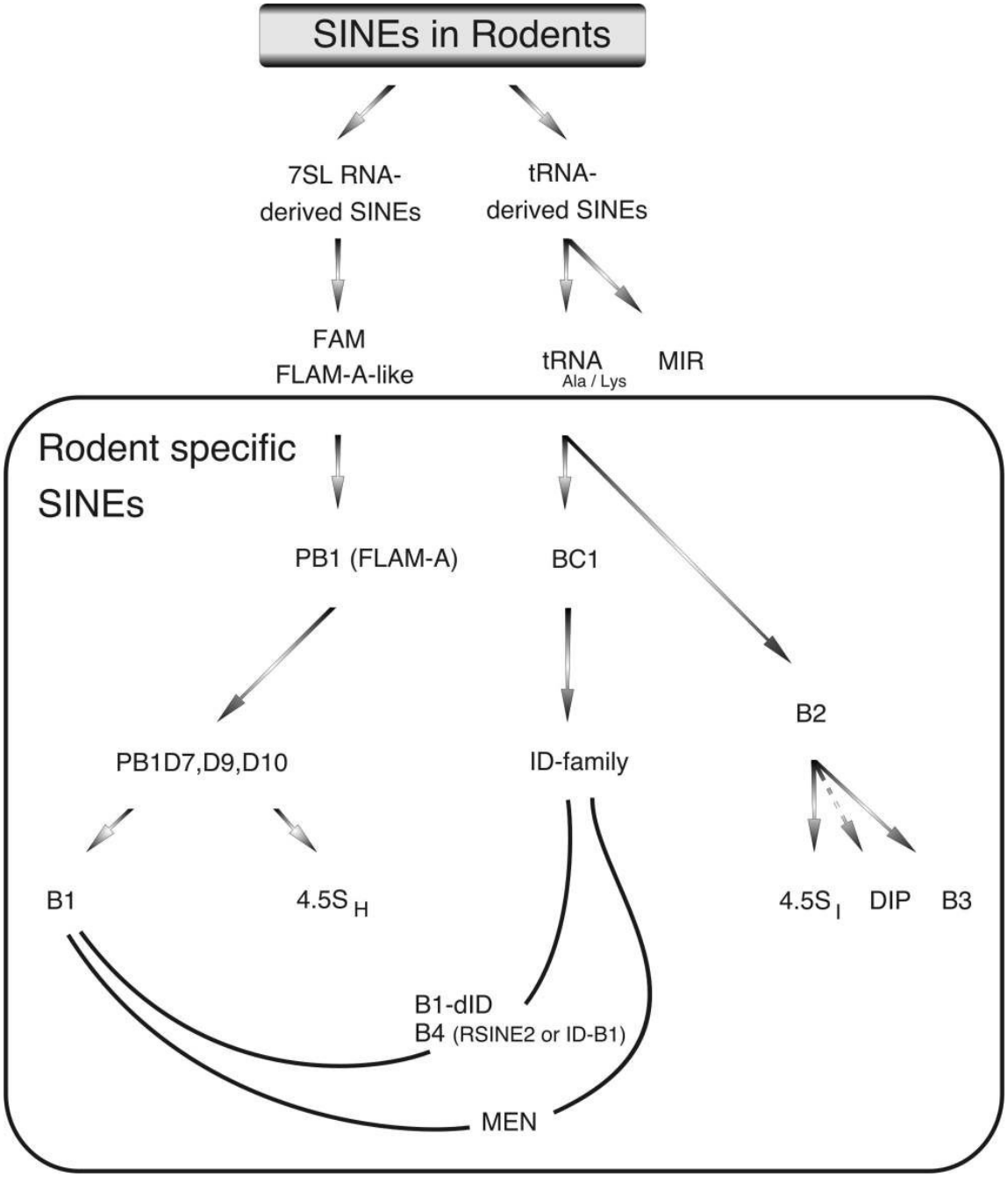

FIGURE 1. SINEs in rodents and their possible ancestral relationships. 7SL RNA-derived SINEs chiefly occur in primates and rodents and evolved from an SRP RNA (7SL) via a "fossil" *Alu* monomer (FAM) and a free left *Alu* monomer (FLAM-A, Quentin, 1992). Proto B1 (PB1) is restricted to rodents. Its consensus sequence differs by only three base exchanges compared to the FLAM–A-like SINE that might have been present in the common ancestor of rodents and primates. In the rodent lineage, various PB1 subfamily members (PB1D7, PB1D9, and PB1D10) appeared prior to the first modern B1 elements (Quentin, 1994; Zietkiewicz and Labuda, 1996). $4.5S_H$ is probably derived from a PB1D10 progenitor (Gogolevskaya et al., 2005). Transfer RNA (tRNA)-derived SINEs are prominent in many vertebrate and plant lineages. One of the ancient tRNA-derived SINEs is represented by MIR elements that were mainly active prior to the mammalian radiation. In rodents, tRNA-derived SINEs are predominantly derived from $tRNA^{Ala}$ (ID elements) or $tRNA^{Lys}$ (B2 elements) (Serdobova and Kramerov, 1998). The master gene for ID elements is the small, neuronally expressed (and at lower levels also in testes) BC1 RNA (Martignetti and Brosius, 1993; Brosius, 1999). Currently, there is some ambiguity as to the sub-classification of ID elements. In the rat, Kim et al. (1994) designated elements that are directly derived from BC1 RNA as ID1 and the ones that were generated from one or several transcribed ID elements as ID2 to ID4. Lee et al. (1998) classified ID elements "in all rodents" from ID1 to ID6 according to their presumed age, whereby ID1 is supposedly the youngest and ID6 the oldest subfamily. $4.5S_I$ 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.

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

Elements) (Fig. 1). Approximately 32.4% of the mouse genome consists of rodent-specific repeats (Waterston et al., 2002) with continuous activity over their evolution. The initial sequencing and comparative analysis of the mouse genome revealed 564,000 B1 elements, 348,000 B2, 391,000 B4, and 79,000 ID-related elements (Waterston et al., 2002). Moreover, different SINE families were distributed in various rodent lineages at different periods. For example, the B1 and ID elements are common to all rodent families (D.H., unpublished data and results in the present work); DIP elements are present in Dipodidae; g/s-B1dID elements are shared by Sciuroids, Gliridae, Hystricognathi, and Castoridae (Kramerov et al., 1999; Kramerov and Vassetzky, 2001); and B2 and $4.5S_I$ SINE elements are restricted to Muroids (Serdobova and Kramerov, 1998; Gogolevskaya and Kramerov, 2002). Thus, there exists a strong case for the use of presence/absence patterns of retroposed 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 retroposed SINE presence/absence loci 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). Intron sequences 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

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

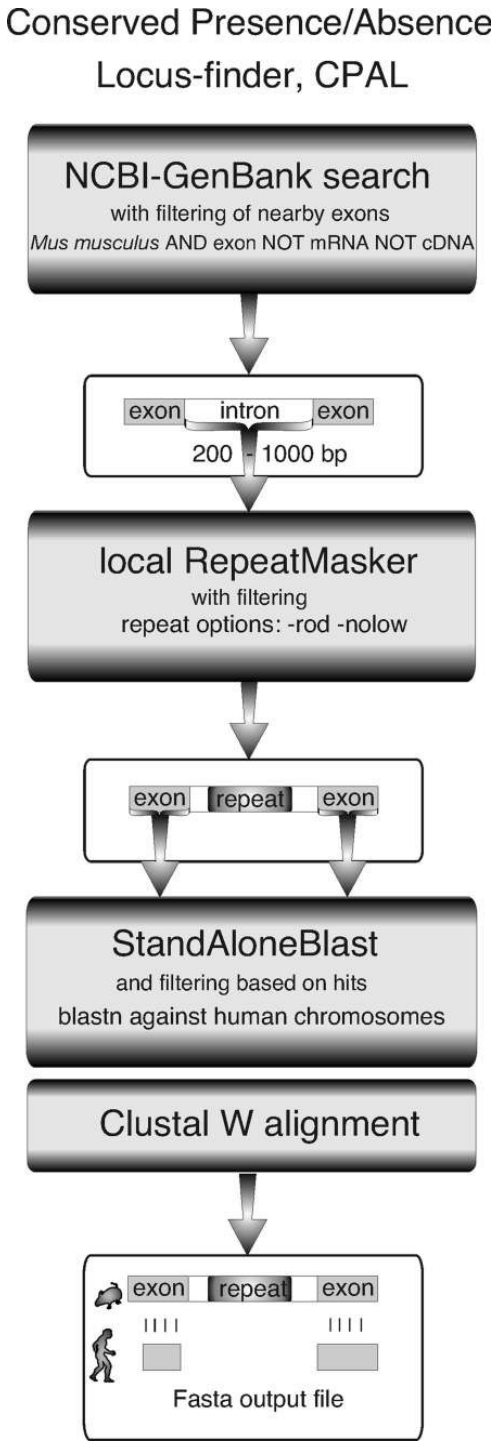

FIGURE 2. Flow chart of the Conserved Presence/Absence Locus-finder computational strategy. The CPAL Bioperl script begins with a GenBank search using the query "*Mus musculus* and exon" and excludes mRNAs and cDNAs. All hits are scanned for introns not longer than 1 kb, flanked by exonic sequences. 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.

pocket mouse); Ctenodactylidae: *Ctenodactylus gundi* (northern gundi); Thryonomyidae: *Thryonomys swinderianus* (greater cane rat); Echimyidae: *Proechimys cuvieri* (Cuvier's spiny rat); Myocastoridae: *Myocastor coypus* (nutria); Octodontidae: *Octodon degus* (degu); Caviidae: *Cavia porcellus* (domestic guinea pig), *Dolichotis patagonum* (Patagonian mara); Erethizontidae: *Coendou prehensilis* (Brazilian porcupine); Sciuridae: *Sciurus vulgaris* (Eurasian red squirrel), *Marmota marmota* (European marmot); Gliridae: *Glis glis* (fat dormouse); Lagomorpha: (*Lepus europaeus*). Additional mouse, rat, guinea pig, and rabbit sequences were obtained from the NCBI GenBank.

PCRs were performed for 4 min at 94°C followed by 35 cycles of 30 s at 94°C, 30 s at the primer-specific annealing temperature and 60 s at 72°C. The PCRs were finished with 5 min at 72°C. The PCR fragments were purified on agarose gels, ligated into the pDrive cloning vector (Qiagen, Hilden) and electroporated into TOP10 cells (Invitrogen, Groningen). Inserts from each of three individual clones were sequenced using the Ampli Taq FS Big Dye Terminator Kit (PE Biosystems, Foster City) and standard M13 forward and reverse primers (Supplementary Table 1, online at http://systematicbiology.org). The sequence data from this study have been submitted to NCBI GenBank under accession numbers DQ321375 to DQ321487, DQ451018 to DQ451089, and DQ521259 to DQ521268. The TreeBASE study and matrix accession numbers are S1568 and M2820, respectively (www.treebase.org).

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. Additonally, 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

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

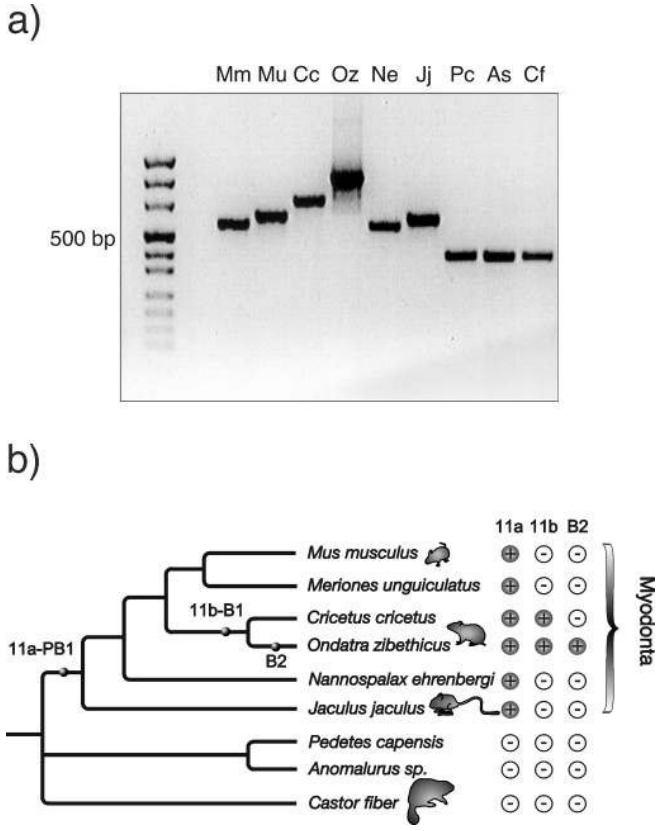

FIGURE 3. Presence/absence pattern of a PB1 retroposon integrated at locus 11. (a), The larger PCR fragments in Muroidea include an intronic PB1 element inherited from a common ancestor. Amplification in all other investigated rodent species produced smaller PCR fragments without PB1 at the orthologous position. (b), Phylogenetic conclusions drawn from locus 11. The dot (11a) denotes the integration event of the retroposon in the common ancestor of Muroidea. An additional integration of a B1 element (11b) indicates a close relationship between the genera *Cricetus* and *Ondatra*. Presence is symbolized by (+), absence by (−). The *Ondatra zibethicus*-specific B2 element is an autapomorphic integration (see Supplementary Table 3, online at www.systematicbiology.org). Mm = *Mus musculus;* Mu = *Meriones unguiculatus;* Cc = *Cricetus cricetus;* Oz = *Ondatra zibethicus;* Ne = *Nannospalax ehrenbergi;* Jj = *Jaculus jaculus;* Pc = *Pedetes capensis;* As = *Anomalurus* sp.; Cf = *Castor fiber.*

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 lagomorphs as an outgroup. Gaps were treated as missing data and the searches were performed using 1000 random sequence addition and TBR branch swapping. The data matrix for phylogenetic reconstruction based on the presence/absence of all 31 markers was used to generate a strict consensus tree (Fig. S-1, online at http://systematicbiology.org). We got 18 equally parsimonious trees, including that shown in Figure 6, all of length 31 and with a consistency index (CI) = 1 indicating the homoplasy-free character of the underlying data.

A second strategy tests certain branching points (e.g., of sequence-based tree reconstructions; Schmitz et al., 2001; Roos et al., 2004; Kriegs et al., 2006) and hence is not dependent on a complete data matrix. Inasmuch as the primary aim of the present study was to first test the efficacy of the computational CPAL screen to return phylogenetically informative rodent SINE markers when no other study had successfully done so, we analyzed only a limited number of loci. Hence, the small number of markers and partial gaps in the data matrix favored this latter strategy and we used the most accepted sequence-based rodent tree as a framework (Fig. 6). Clear prior hypotheses concerning the validity of given branch points were statistically confirmed using the method of Waddell et al. (2001). It is important to stress that no incongruence was found between the phylogenetic tree reconstructed with PAUP and the most accepted rodent tree.

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,

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

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 Eumuroida, 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 Myodonta ($P < 0.04$; [3 0 0]). A B1 SINE element (marker 13a) was present in all Myodonta plus the two Anomaluromorpha 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 Anomaluromorpha.

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 clades. 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 cuvieri*, *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 cuvieri* and *Myocastor coypus*, and ID4, two ID2, and B1 elements (markers 13f, 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

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

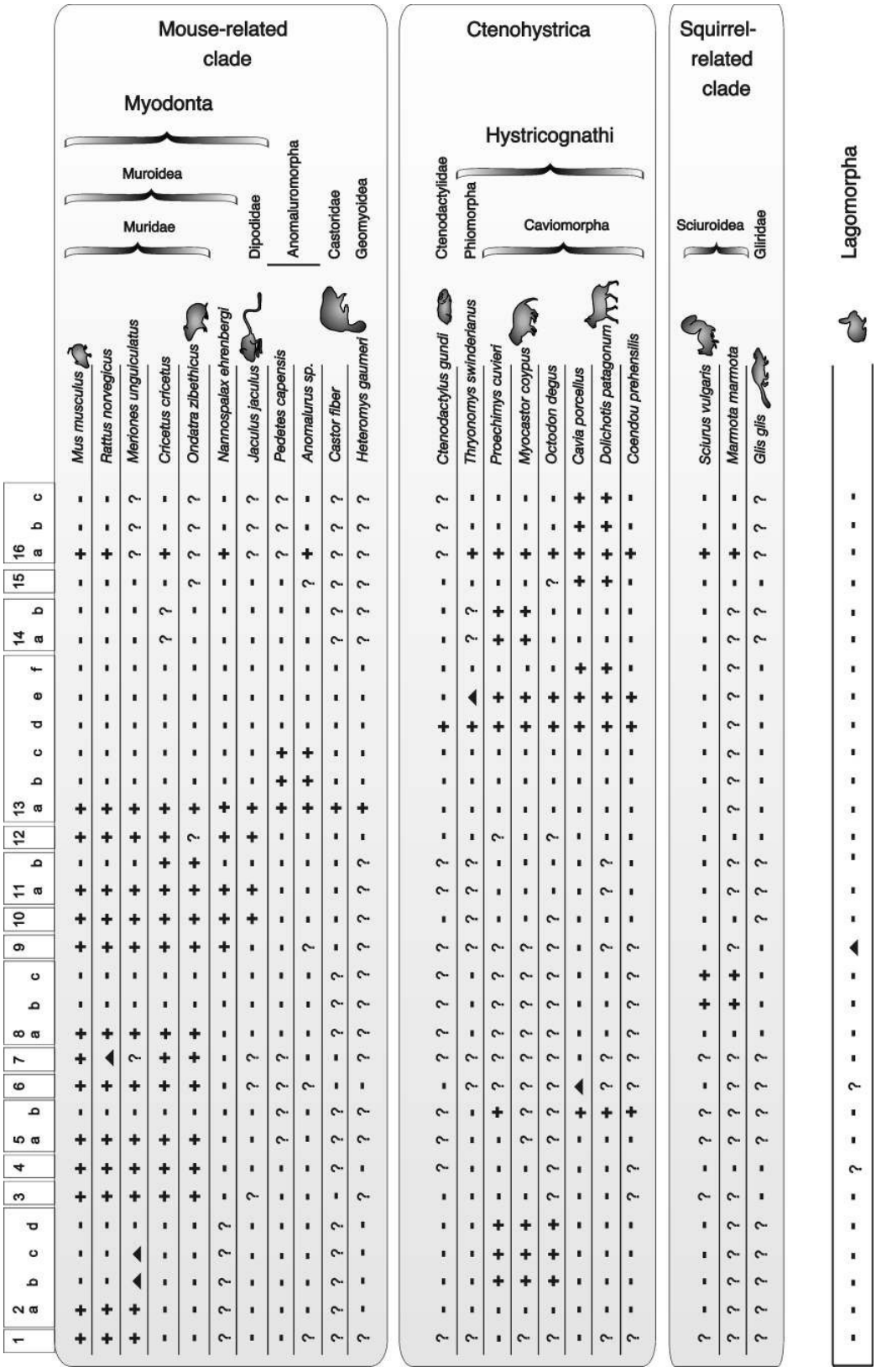

FIGURE 4. Presence/absence patterns of analyzed SINE insertion loci. (+) denotes the presence and (−) absence of a given diagnostic transposed element for all investigated loci (1 to 16) in representative rodent species. Question marks indicate that PCR amplification of the given transposed element was 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).

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

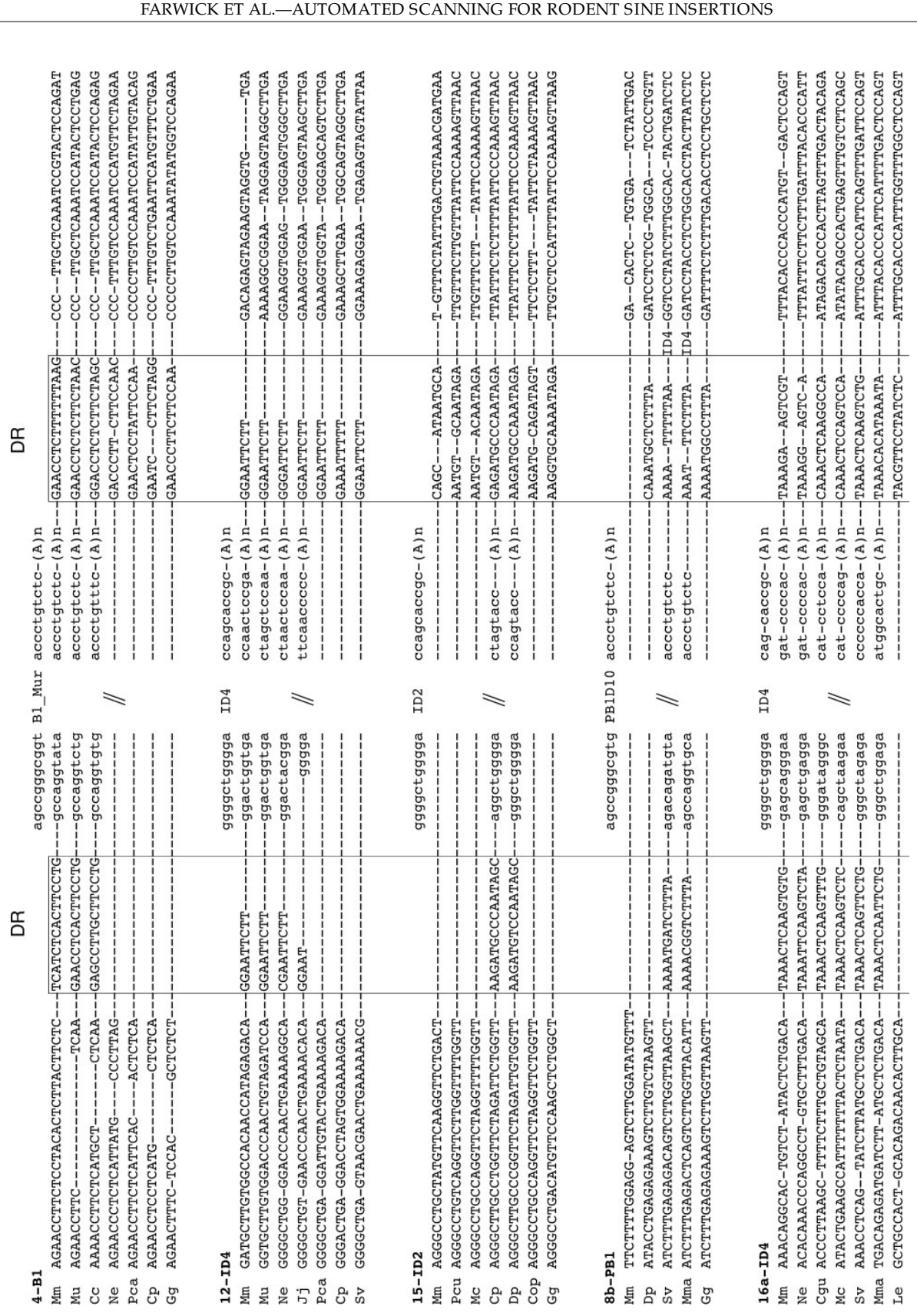

FIGURE 5. Representative presence/absence alignments. The 5′ and 3′ parts of the inserted SINE elements are shown together with their immediate flanking regions. Possible direct repeats (DR) are boxed. Marker names correspond to those in Figures 4 and 6. Mm = *Mus musculus;* Mu = *Meriones unguiculatus;* Cc = *Cricetus cricetus;* Ne = *Nannospalax ehrenbergi;* Pca = *Pedetes capensis;* Cp = *Cavia porcellus;* 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.

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

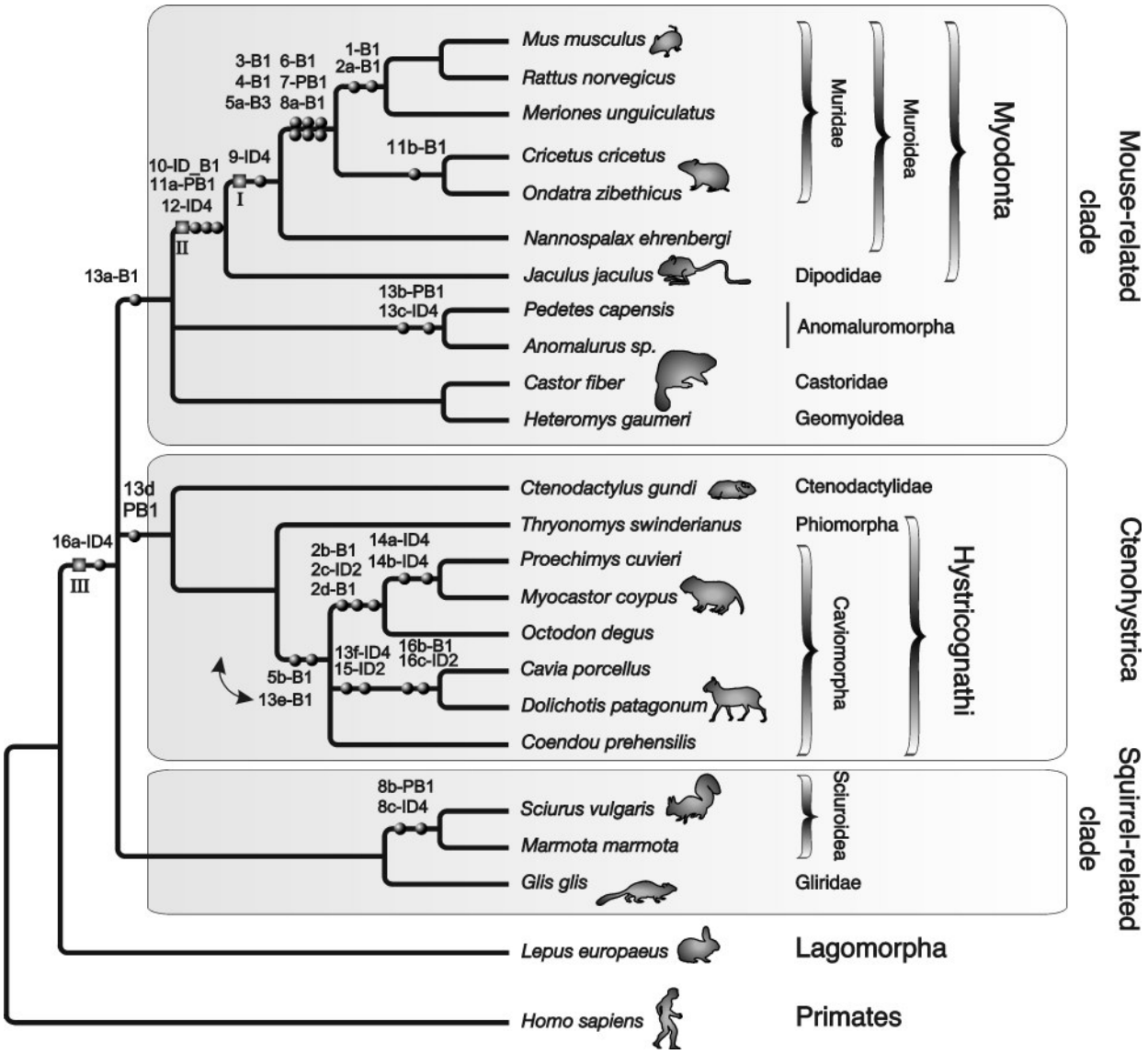

FIGURE 6. SINEs as molecular cladistic markers in rodent phylogeny. The evolutionary tree of representative rodent species used as a framework is based on molecular data (Michaux et al., 2001; Huchon et al., 2002; Adkins et al., 2003; DeBry et al., 2003; Steppan et al., 2004). Arabic numerals refer to markers found in this study (see also Fig. 3). Because of a large deletion in the 13e-B1 locus of *Thryonomys swinderianus*, it was not possible to determine in which branch of the tree the SINE insertions occurred. An arrow indicates the ambiguous placement of this SINE. Roman numerals refer to other, previously discovered multilocus markers uniting (I) Muroidea (B2 SINEs; Serdobova and Kramerov, 1998), (II) Myodonta (4.5S_H RNA; Gogolevskaya et al., 2005), (III) Rodents (BC1 RNA; Martignetti and Brosius, 1993). The lagomorphs (*Lepus europaeus* or *Oryctolagus cuniculus*) were used as the outgroup to rodents.

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.,

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

B2 SINEs, Serdobova and Kramerov, 1998; $4.5S_H$ 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 position (convergence). Van de Lagemaat et al. (2005) recently described cases of precise deletions of transposed elements in primates that rendered the deleted loci indistinguishable from pre-integration sites. An important advantage of using transposed elements as phylogenetic markers is the clear, unambiguous separation of the two presence/absence patterns; "presence" describing the derived condition and "absence," the plesiomorph state. A prevalence of precisely deleted transposed elements would necessarily cloud the use of transposed elements in phylogeny. However, precise deletion of transposed elements was detected only in rare cases and is thought to be due to recombination between 10-20 bp of perfect direct repeats (Van de Lagemaat et al., 2005). In evolutionary time frames, direct repeats diverge extensively and are, therefore, no longer compatible for precise deletion. Furthermore, precise deletion that took place in ancient lineages would, today, be indistinguishable from incomplete lineage sorting, an obvious problem of all phylogenetic marker systems. The second potential source of SINE homoplasy is the independent integration into orthologous loci of different species (Rothenburg et al., 2002; Ludwig et al., 2005). Cantrell et al. (2001) described hot spots of integration as a rare but potential source of homoplasy. Parallel integration in orthologous loci is explicable by a preference of SINE integration into A-rich genomic regions (Jurka, 1997; Tatout, et al., 1998). However, misinterpretation of independent insertions at the same genomic locus is only possible if identical SINEs are involved in parallel integrations or in the case of gene conversion, though gene conversions are also rare events. For example, among 133 hominoid *Alu*-insertion loci, Salem et al. (2003) found only three cases of gene conversion, and only one of them involved full gene conversion. Integrations of truncated retroposons as well as indels that occur in an ancestral lineage before species divergence are additional indicators of true orthologous integrations (Schmitz et al., 2001). Cases of homoplasy with SINE insertion are very rare in the literature and when they occurred they are best explained by incomplete lineage sorting of ancestral polymorphism (Shedlock et al., 2000, 2004). For example, from among 99 published phylogenetic informative SINE insertions in primates and other mammals (Roos et al., 2004; Schmitz et al., 2005; Singer et al., 2003; Kriegs et al., 2006), none were found to demonstrate any homoplasy. Furthermore, among his 133 amplified hominid *Alu*-insertion loci, Salem et al. (2003) found only a single case of homoplasy, which was probably a result of incomplete lineage sorting of an ancestral polymorphism.

Despite their rarity, misinterpretations caused by homoplastic SINE insertions cannot be fully excluded. That SINEs may integrate independently into the same integration site could be shown for a species-specific B1 element in *Ondatra zibethicus* and a species-specific ID_B1 element in *Ctenodactylus gundi* (locus 14, data not shown). Nevertheless, because both elements are from different sources, they are not homoplastic and will not cause erroneous interpretation of relationships. On the other hand, one cannot fully rule out that, in rare cases, the same could happen with identical elements. But this should result in conflicting presence/absence patterns that were not evidenced in the current data set. A similar situation was observed for locus 8 where a *Dolichotis patagonum* specific ID element inserted into the same integration site, but in opposite orientation, as a *Glis glis*-specific ID element (data not shown).

The most reliable strategy to validate the conclusions drawn from SINE insertions is to identify several independent markers to support the monophyly of each phylogenetic group and Waddell et al. (2001) has provided a means of statistically showing this. Due to the restricted number of loci investigated and their highly distributed nature, the present report includes some singular integration events that await further support.

The search for SINE presence/absence markers based on mouse sequence information revealed many phylogenetically informative integrations in the mouse-related

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

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 sciuromorph 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, Anomaluromorpha 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 Anomaluromorpha (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 Anomaluridae, 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 (Ctenodactylidae 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

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

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 Kriegs for organizing the tissue transfer. We appreciate the assistance of students Kathrin Weber, Huyen Pham, Hermann Krähling, 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 J.S. and J.B.) and the United States–Israel Binational Science Foundation (BSF) (2004-407 to D.H.).

REFERENCES

Adkins, R. M., E. L. Gelke, D. Rowe, and R. L. Honeycutt. 2001. Molecular phylogeny and divergence time estimates for major rodent groups: Evidence from multiple genes. Mol. Biol. Evol. 18:777–791.

Adkins, R. M., A. H. Walton, and R. L. Honeycutt. 2003. Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. Mol. Phylogenet. Evol. 26:409–420.

Brandt, J. F. 1855. Beiträge zur nahern Kenntniss der Säugethiere Russlands. Mem. Acad. Imp. St. Petersbourg Ser. 69:1–375.

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

Bugge, J. 1985. Systematic value of the carotid arterial pattern in rodents. Pages 381–402 *in* Evolutionary relationships among rodents: A multidisciplinary analysis (W. P. Luckett and J.-L. Hartenberger, eds.). Plenum Press, New York and London.

Cantrell, M. A., B. J. Filanoski, A. R. Ingermann, K. Olsson, N. DiLuglio, Z. Lister, and H. A. Wichman. 2001. An ancient retrovirus-like element contains hot spots for SINE insertion. Genetics 158:769–777.

Catzeflis, F. M., F. H. Sheldon, J. E. Ahlquist, and C. G. Sibley. 1987. DNA-DNA hybridization evidence of the rapid rate of muroid rodent DNA evolution. Mol. Biol. Evol. 4:242–253.

D'Erchia, A. M., C. Gissi, G. Pesole, C. Saccone, and U. Arnason. 1996. The guinea-pig is not a rodent. Nature 381:597–600.

DeBry, R. W. 2003. Identifying conflicting signal in a multigene analysis reveals a highly resolved tree: The phylogeny of Rodentia (Mammalia). Syst. Biol. 52:604–617.

DeBry, R. W., and R. M. Sagel. 2001. Phylogeny of Rodentia (Mammalia) inferred from the nuclear-encoded gene IRBP. Mol. Phylogenet. Evol. 19:290–301.

Gogolevskaya, I. K., A. P. Koval, D.A. Kramerov. 2005. Evolutionary history of 4.5SH RNA. Mol Biol Evol. 22:1546–1554.

Gogolevskaya, I. K., and D. A. Kramerov. 2002. Evolutionary history of 4.5SI RNA and indication that it is functional. J. Mol. Evol. 54:354–364.

Graur, D., W. A. Hide, and W.-H. Li. 1991. Is the guinea-pig a rodent? Nature 351:649–652.

Hartenberger, J.-L. 1985. The order Rodentia: Major questions on their evolutionary origin, relationships and suprafamilial systematics. Pages 1–33 *in* Evolutionary relationships among rodents: A multidisciplinary analysis (W. P. Luckett and J.-L. Hartenberger, eds.). Plenum Press, New York and London.

Hartenberger, J.-L. 1998. Description of the radiation of the Rodentia (Mammalia) from the Late Paleocene to the Miocene; phylogenetic consequences. Comptes Rendus de l'Academie des Sciences Serie II 326:439–444.

Huchon, D., F. M. Catzeflis, and E. J. P. Douzery. 1999. Molecular evolution of the nuclear von Willebrand factor gene in mammals and the phylogeny of rodents. Mol. Biol. Evol. 16:577–589.

Huchon, D., F. M. Catzeflis, and E. J. P. Douzery. 2000. Variance of molecular datings, evolution of rodents, and the phylogenetic affinities between Ctenodactylidae and Hystricognathi. Proc. R. Soc. Lond. B 267:393–402.

Huchon, D., and E. J. P. Douzery. 2001. From the Old-World to the New-World: A molecular chronicle of the phylogeny and biogeography of hystricognath rodents. Mol. Phylogenet. Evol. 20:238–251.

Huchon, D., O. Madsen, M. J. J. B. Sibbald, K. Ament, M. Stanhope, F. Catzeflis, W. W. De Jong, and E. J. P. Douzery. 2002. Rodent phylogeny and a timescale for the evolution of glires: Evidence from an extensive taxon sampling using three nuclear genes. Mol. Biol. Evol. 19:1053–1065.

Jaeger, J.-J. 1988. Rodent phylogeny: New data and old problems. Pages 177–199 *in* The phylogeny and classification of the tetrapods (M. J. Benton, ed.). Clarendon Press, Oxford.

Jurka, J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc. Natl. Acad. Sci. USA 94:1872–1877.

Kim, J., J.A. Martignetti, M.R. Shen, J. Brosius, and P. Deininger. 1994. Rodent BC1 RNA gene as a master gene for ID element amplification. Proc. Natl. Acad. Sci. USA 91:3607–3611.

Korth, W. W. 1994. The Tertiary record of rodents in North America. Plenum Press, New York.

Kramerov, D. A., and N. S. Vassetzky. 2001. Structure and origin of a novel dimeric retroposon B1-dID. J. Mol. Evol. 52:137–143.

Kramerov, D. A., and N. S. Vassetzky. 2005. Short retroposons in eukaryotic genomes. Int. Rev. Cytol. 247:165–221.

Kramerov, D., N. Vassetzky, and I. Serdobova. 1999. The evolutionary position of dormice (Gliridae) in Rodentia determined by a novel short retroposon. Mol. Biol. Evol. 16:715–717.

Kriegs, J.O., G. Churakov, M. Kiefmann, U. Jordan, J. Brosius, J. Schmitz. 2006. Retroposed elements as archives for the evolutionary history of placental mammals. PLoS Biol. 4:e91.

Lavocat, R., and J.-P. Parent. 1985. Phylogenetic analyses of middle ear features in fossil and living rodents. Pages 333–354 *in* Evolutionary relationships among rodents: A multidisciplinary analysis (W. P. Luckett and J.-L. Hartenberger, eds.). Plenum Press, New York and London.

Lee, I. Y., D. Westaway, A. F. A. Smit, K. Wang, J. Seto, L. Chen, C. Acharya, M. Ankener, D. Baskin, C. Cooper, H. Yao, S. B. Prusiner and L. E. Hood. 1998. Complete genomic sequence and analysis of the prion protein gene region from three mammalian species. Genome Res. 8:1022–1037.

Luckett, W. P., and J.-L. Hartenberger. 1985. Evolutionary relationships among rodents a multidisciplinary analysis. Plenum Press, New York and London.

Luckett, W. P., and J.-L. Hartenberger. 1993. Monophyly or polyphyly of the order Rodentia: possible conflict between morphological and molecular interpretations. J. Mammal. Evol. 1:127–147.

Ludwig, A., T. S. Rozhdestvensky, V. Yu. Kuryshev, J. Schmitz and J. Brosius. 2005. An unusual primate locus that attracted two independent *Alu* insertions and facilitates their transcription. J. Mol. Biol. 350:200–214.

Marivaux, L., M. Vianey-Liaud, and J. J. Jaeger. 2004. High-level phylogeny of early Tertiary rodents: Dental evidence. Zool. J. Linn. Soc. 142:105–134.

Martignetti, J. A., and J. Brosius. 1993. Neutral BC1 RNA as an evolutionary marker: Guinea pig remains a rodent. Proc. Natl. Acad. Sci. USA 90:9698–9702.

Martin, T. 1995. Incisor enamel microstructure and phylogenetic interrelationships of Pedetidae and Ctenodactyloidea (Rodentia). Berliner Geowiss. Abh. E 16:693–707.

McKenna, M. C., and S. K. Bell. 1997. Classification of mammals above the species level. Columbia University Press, New York.

Michaux, J., A. Reyes, and F. Catzeflis. 2001. Evolutionary history of the most speciose mammals: Molecular phylogeny of muroid rodents. Mol. Biol. Evol. 18:2017–2031.

Montgelard, C., S. Bentz, C. Tirard, O. Verneau, and F. M. Catzeflis. 2002. Molecular systematics of Sciurognathi (Rodentia): The mitochondrial cytochrome *b* and 12S rRNA genes support the Anomaluroidea (Pedetidae and Anomaluridae). Mol. Phylogenet. Evol. 22:220–233.

Downloaded By: [EBSCOHost Content Distribution] At: 17:55 2 April 2007

Murphy, W. J., E. Eizirik, W. E. Johnson, Y. P. Zhang, O. A. Ryder, and S. J. O'Brien. 2001a. Molecular phylogenetics and the origins of placental mammals. Nature 409:614–618.

Murphy, W. J., E. Eizirik, S. J. O'Brien, O. Madsen, M. Scally, C. J. Douady, E. Teeling, O. A. Ryder, M. J. Stanhope, W. W. de Jong, and M. S. Springer. 2001b. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294:2348–2351.

Musser, G. M., and M. D. Carleton. 1993. Family Muridae. Pages 501–756 *in* Mammal species of the world: A taxonomic and geographic reference (D. E. Wilson, and D. M. Reeder, eds.). Smithsonian Institution Press, Washington and London.

Nedbal, M. A., R. L. Honeycutt, and D. A. Schlitter. 1996. Higher-level systematics of rodents (Mammalia, Rodentia): Evidence from the mitochondrial 12S rRNA gene. J. Mammal. Evol. 3:201–237.

Nikaido, M., A. P. Rooney, and N. Okada. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. USA 96:10261–10266.

Philippe, H. 1997. Rodent monophyly: Pitfalls of molecular phylogenies. J. Mol. Evol. 45:712–715.

Quentin, Y. 1992. Origin of the *Alu* family: A family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. Nucleic Acids Res. 20:3397–3401.

Quentin, Y. 1994. A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes. Nucleic Acids Res. 22:2222–2227.

Reyes, A., G. Pesole, and C. Saccone. 1998. Complete mitochondrial DNA sequence of the fat dormouse, *Glis glis*: Further evidence of rodent paraphyly. Mol. Biol. Evol. 15:499–505.

Roos, C., J. Schmitz, and H. Zischler. 2004. Primate jumping genes elucidate strepsirrhine phylogeny. Proc. Natl. Acad. Sci. USA 101:10650–10654.

Rothenburg, S., M. Eiben, F. Koch-Nolte, and F. Haag. 2002. Independent integration of rodent identifier (ID) elements into orthologous sites of some RT6 alleles of *Rattus norvegicus* and *Rattus rattus*. J. Mol. Evol. 55:251–259.

Salem, A. H., D. A. Ray, J. C. Xing, P. A. Callinan, J. S. Myers, D. J. Hedges, R. K. Garber, D. J. Witherspoon, L. B. Jorde, and M. A. Batzer. 2003. Alu elements and hominid phylogenetics. Proc. Natl. Acad. Sci. USA 100:12787–12791.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A laboratory manual, second edition. Cold Spring Harbor Laboratory Press, New York.

Schmitz, J., M. Ohme, and H. Zischler. 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. Genetics 157:777–784.

Schmitz, J., C. Roos, and H. Zischler. 2005. Primate phylogeny: Molecular evidence from retroposons. Cytogenet Genome Res. 108:26–37.

Serdobova, I. M., and D. A. Kramerov. 1998. Short retroposons of the B2 superfamily: Evolution and application for the study of rodent phylogeny. J. Mol. Evol. 46:202–214.

Shedlock , A. M., and N. Okada. 2000. SINE insertions: Powerful tools for molecular systematics. Bioessays 22:148–160.

Shedlock, A. M., K. Takahashi, and N. Okada. 2004. SINEs of speciation: Tracking lineages with retroposons. Trends Ecol. Evol. 19:545–553.

Singer, S. S., J. Schmitz, C. Schwiegk, and H. Zischler. 2003. Molecular cladistic markers in New World monkey phylogeny (Platyrrhini, Primates). Mol. Phylogenet. Evol. 26:490–501.

Stajich, J. E., D. Block, K. Boulez, S. E. Brenner, S. A. Chervitz, C. Dagdigian, G. Fuellen, J. G. R. Gilbert, I. Korf, H. Lapp, H. Leväslaiho, C. Matsalla, C. J. Mungall, B. I. Osborne, M. R. Pocok, P. Schattner, M. Senger, L. D. Stein, E. Stupka, M. D. Wilkinson, and E. Birney. 2002. The Bioperl Toolkit: Perl modules for the life sciences. Genome Res. 12: 1611–1618.

Steppan, S. J., R. M. Adkins, and J. Anderson. 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. Syst. Biol. 53:533–553.

Sullivan, J., and D. L. Swofford. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mammal. Evol. 4:77–86.

Swofford, D.L., 2000. PAUP: Phylogenetic analysis using parsimony. Sinauer Associates, Sunderland, Massachusetts.

Tatout, C., L. Lavie, and J. M. Deragon. 1998. Similar target site selection occurs in integration of plant and mammalian retroposons. J. Mol. Evol. 47:463–470.

Tatout, C., S. Warwick, A. Lenoir, and J.-M. Deragon. 1999. SINE insertions as clade markers for wild crucifer species. Mol. Biol. Evol. 16:1614–1621.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.

Tullberg, T. 1899. Ueber das System der Nagetiere: Eine phylogenetische studie. Nova Acta Reg. Soc. Sci. Upsala Ser. 3 18:1–514.

van de Lagemaat, L. N., L. Gagnier, P. Medstrand, and D. L. Mager. 2005. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. Genome Res. 15:1243–1249.

Vianey-Liaud, M. 1985. Possible evolutionary relationships among Eocene and Lower Oligocene rodents of Asia, Europe and North America. Pages 277–309 *in* Evolutionary relationships among rodents: A multidisciplinary analysis (W. P. Luckett, and J.-L. Hartenberger, eds.). Plenum Press, New York and London.

Waddell, P. J., H. Kishino, and R. Ota. 2001. A phylogenetic foundation for comparative mammalian genomics. Genome Inform. Ser. Workshop Genome Inform. 12:141–154.

Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.

Wilson, D. E., and D. M. Reeder. 1993. Mammal species of the world: A taxonomic and geographic reference, edition 2. Smithsonian Institution Press, Washington and London.

Wu, C.-I., and W.-H. Li. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA 82:1741–1745.

Yang, S., A. F. Smit, S. Schwartz, F. Chiaromonte, K. M. Roskin, D. Haussler, W. Miller, and R. C. Hardison. 2004. Patterns of insertions and their covariation with substitutions in the rat, mouse, and human genomes. Genome Res. 14:517–527.

Zietkiewicz, E., and D. Labuda. 1996. Mosaic evolution of rodent B1 elements. J. Mol. Evol. 42:66–72.

First Submitted 4 November 2005; reviews returned 31 December 2005; final acceptance 10 May 2006
Associate Editor: Andrew Shedlock