

Rodent Evolution: Back to the Root

Gennady Churakov,^{†1} Manoj K. Sadasivuni,^{†1} Kate R. Rosenbloom,² Dorothée Huchon,³ Jürgen Brosius,¹ and Jürgen Schmitz^{*1}

¹Institute of Experimental Pathology (ZMBE), University of Münster, Münster, Germany

²Department of Biomolecular Engineering, University of California

³Department of Zoology, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv, Israel

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: jueschm@uni-muenster.de.

Associate editor: Dan Graur

Abstract

Some 70 Ma, rodents arose along a branch of our own mammalian lineage. Today, about 40% of all mammalian species are rodents and are found in vast numbers on almost every continent. Not only is their proliferation extensive but also the rates of DNA evolution vary significantly among lineages, which has hindered attempts to reconstruct, especially the root of, their evolutionary history. The presence or absence of rare genomic changes, such as short interspersed elements (SINEs), are, however, independent of high molecular substitution rates and provide a powerful, virtually homoplasmy-free source for solving such phylogenetic problems. We screened 12 Gb of rodent genomic information using whole-genome three-way alignments, multiple lineage-specific sequences, high-throughput polymerase chain reaction amplifications, and sequencing to reveal 65 phylogenetically informative SINE insertions dispersed over 23 rodent phylogenetic nodes. Eight SINEs and six indels provide significant support for an early association of the Mouse-related and Ctenohystrica (guinea pig and relatives) clades, the Squirrel-related clade being the sister group. This early speciation scenario was also evident in the genomewide distribution pattern of B1-related retroposons, as mouse and guinea pig genomes share six such retroposon subfamilies, containing hundreds of thousands of elements that are clearly absent in the ground squirrel genome. Interestingly, however, two SINE insertions and one diagnostic indel support an association of Ctenohystrica with the Squirrel-related clade. Lineage sorting or a more complex evolutionary scenario that includes an early divergence of the Squirrel-related ancestor and a subsequent hybridization of the latter and the Ctenohystrica lineage best explains such apparently contradictory insertions.

Key words: SINE, retroposon, rodent evolution.

Introduction

Although about 195 Ma, *Hadrocoidium wui*, the potential fossil ancestor of all mammals, looked very much like a small mouse (Luo et al. 2001), rodents evolved only some 130 My later, 62–100 Ma, from a common ancestor with lagomorphs, forming the clade Glires (Benton and Donoghue 2007). Glires share a common ancestry with primates, tree shrews, and the flying lemurs (Murphy et al. 2001; Huchon et al. 2002; Kriegs, Churakov, et al. 2007). Probably favored by their small size, short breeding cycles, and wide variety of foods eaten, rodents rose fast to become one of the most successful mammalian groups, occupying nearly all continents and comprising close to half of all living mammalian species. Many rodents, murines (e.g., mice and rats) in particular, are typical r-strategists (favoring quantity over quality in offspring) and evolved extremely short generation times and extraordinary explorative and adaptive abilities, exerting significant impacts on population structures and their evolutionary rates (Spradling et al. 2001).

Although the significantly high rate of nucleotide substitution in murines, when compared with primates, is the classical example of rate heterogeneity in mammals

(Wu and Li 1985; Gibbs et al. 2004), the reason for this phenomenon is hotly debated (Martin and Palumbi 1993; Kumar and Subramanian 2002; Bininda-Emonds et al. 2007). However, not all rodents share the high nucleotide substitution rates of mouse and rat; the ground squirrel genome in particular appears to have evolved more slowly than most other rodents (Gissi et al. 2000; Douzery et al. 2003) and has a more conserved organization (Stanyon et al. 2003). In phylogenetic reconstruction analyses, rapidly evolving lineages often lead to a phenomenon called long-branch attraction (LBA). Such lineages are robustly grouped, regardless of their true evolutionary relationships. Because the closest outgroup is usually distantly related and is a long branch per se, LBA often leads to artifactual trees in which fast-evolving lineages emerge at the base of the tree attracted by the outgroup branch (Philippe and Laurent 1998). One now classical example of LBA led to the assumption that “The guinea pig is not a rodent” (D’Erchia et al. 1996), based on the placement of the fast-evolving murine mitochondrial sequences at the root of the placental tree, leaving the guinea pig behind in a closer relationship to primates. It was shown that an increase in species and gene sampling and the use of probabilistic models of sequence evolution provide strong

support in favor of rodent monophyly (Huchon et al. 2002, 2007; Blanga-Kanfi et al. 2009). However, these approaches have still not enabled the resolution of the most debated question of rodent phylogeny, the first divergence after their emergence. This part of the tree is still completely unresolved. Morphological data have never been able to robustly solve this issue (Marivaux et al. 2004), and trees based on sequence substitution provide conflicting results (Montgelard et al. 2008; Blanga-Kanfi et al. 2009). Molecular analyses do support the division of rodents into three clades: the Mouse-related clade, Ctenohystrica (guinea pig and relatives), and the Squirrel-related clade (Huchon et al. 2002; DeBry 2003), but the Mouse- and Squirrel-related clades have both been proposed as the most divergent rodent lineages (Montgelard et al. 2008; Blanga-Kanfi et al. 2009); thus, the interrelationships among these three lineages remains nebulous. One possible way of resolving this controversial issue is to use a phylogenetic marker system that is insensitive to the effects of LBAs, nucleotide composition biases, etc.

Retroposed elements insert into genomes at random genomic locations and provide powerful, virtually noise-free cladistic landmarks of relatedness (Shedlock and Okada 2000). Because of their insertion complexity, involving target site duplications and random integrations, they offer an extremely large number of possible unique character states (corresponding to insertion sites) such that maximum parsimony analyses converge to maximum likelihood estimators (Steel and Penny 2000). The informative character of such markers lies in their random genomic insertion in an ancestral germline, such that specific insertions in the ancestor of two species reliably document their common ancestry after fixation and speciation. The evolutionary power of retroposon presence–absence data in primate phylogeny was first recognized by Ryan and Dugaiczky (1989). In the following subsequent years, the Okada group pioneered the usage of presence–absence markers to resolve phylogenetic questions (e.g., Murata et al. 1993; Shimamura et al. 1997).

Although retroposed elements occur frequently in rodent genomes (Gibbs et al. 2004), their use as phylogenetic markers in these species can be challenging. Due to the accelerated sequence evolution of murine genomes, the elements tend to have highly diverged sequences. Therefore, only the most conserved sequence regions provide reliable information for reconstructing orthologous retroposon insertion and absence sites in reference species. We designed bioinformatic tools to select phylogenetically informative retroposons inserted in conserved loci in the lineage leading to mouse (the only rodent genome available at the time; Farwick et al. 2006). This screen identified 35 phylogenetically informative, diagnostic markers, including one for rodent monophyly and one grouping the entire Mouse-related clade, but several branch points of the rodent tree were unresolved due to insufficient information (Farwick et al. 2006; Huchon et al. 2007). In the intervening time, new data sources have become available that promise to help elucidate the remaining uncertain branches of the

rodent evolutionary tree. Genomic sequences of the mouse, rat, kangaroo rat, guinea pig, and the 13-lined ground squirrel now enable us to screen for retroposon markers in species of the three main rodent branches. In addition, a novel program was designed to conduct exhaustive genomewide searches using three-way alignments (Warren et al. 2008) of the major rodent lineages to specifically screen for retroposed elements to resolve the root of the rodent tree. In the present study, a high-throughput computational and experimental screening recovered 65 phylogenetically informative markers that provide valuable information about the root of the rodent tree and resolve several internal rodent branches. Moreover, we also have discovered four novel short interspersed element (SINE) subfamilies.

Materials and Methods

Genomic Sources

Mus musculus (house mouse): The July 2007 mouse genome data (http://www.sanger.ac.uk/Projects/M_musculus/) were obtained from the Build 37 assembly and included approximately 2.6 Gb of sequences. These sequences are considered “essentially complete” genomic information. *Rattus norvegicus* (Brown rat): The November 2004 rat genome assembly is based on version 3.4 (<http://www.hgsc.bcm.tmc.edu/project-species-m-Rat.hgsc?pageLocation=Rat>) and covers more than 90% of the estimated 2.8 Gb of rat genomic data. *Dipodomys ordii* (Ord’s kangaroo rat): About 7,350 million available traces representing a 2.5× genomic coverage (<http://www.hgsc.bcm.tmc.edu/project-species-m-Kangaroo%20rat.hgsc?pageLocation=Kangaroo%20rat>). All selected *D. ordii* trace sequences were crosschecked for contamination by blasting (BlastN) against available genomic sequences (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). *Cavia porcellus* (guinea pig): The guinea pig February 2008 draft assembly (<http://www.broadinstitute.org/science/projects/mammals-models/guinea-pig/guinea-pig>) includes 3,143 scaffolds and contains 2.7 Gb. *Spermophilus tridecemlineatus* (13-lined ground squirrel): About 8,250 million available traces representing nearly a 2.5× genomic coverage (<ftp://ftp.ncbi.nih.gov/TraceDB/>). Selected trace sequences were crosschecked for contamination (see above). About 1–2% of the 13-lined ground squirrel genomic traces are contaminations of *M. musculus*. The calculations of the genomic coverage based on trace sequences were conducted in relation to the comparable genomic proportion of mammalian interspersed elements (MIR) that were established before the divergence of rodents and are comparable in all rodent genomes and in relation to the coverage of completely assembled genomes.

Retroposon Screening from Mouse Genomic Sequence Sources

The CPAL (Conserved Presence/Absence Loci) finder was previously generated to explore the annotated genome of mouse (Farwick et al. 2006). CPAL automatically searches the NCBI database for all short mouse introns (<1 kb) with conserved mouse/human flanks, including potentially

informative retroposed elements. In a first screening, we recovered 35 such retroposons predominantly located on the branch leading to mouse (Farwick et al. 2006; Huchon et al. 2007). In the present work, we doubled the species sampling for those loci and included two additional informative loci also detected by CPAL.

Retroposon Screening from *Dipodomys* Trace Sequence Sources

To solve the Mouse-related branch from the mouse distant end of the group, we blasted (BlastN) all “empty” short mouse introns (<http://genome.ucsc.edu/cgi-bin/hgTables>; 25,056 short introns, 100–800 nt, devoid of retroposons) against trace data of *D. ordii*. Corresponding *Dipodomys* hits were screened for lineage-specific retroposon insertions (RepeatMasker; Smit, AFA, Hubley R, and Green, P, <http://www.repeatmasker.org>). Conserved exon-based polymerase chain reaction (PCR) primers were derived for Zoo-PCR.

Retroposon Screening from Guinea Pig Genomic and Ground Squirrel Trace Sequences

The same strategy as described above for the *Dipodomys* traces was used to screen for lineage-specific markers based on *C. porcellus* and *S. tridecemlineatus* sequence information.

All experimental procedures of PCR amplification and sequencing were performed as described previously in Farwick et al. (2006) (for oligos see [supplementary table S1](#), Supplementary Material online). The orthology of derived sequences was confirmed by analyzing the open reading frames (ORFs) and splice sites of all analyzed loci. The 305 newly obtained sequences so derived are deposited in GenBank under the accession numbers GQ506666–GQ506970.

Three-Way Alignments to Analyze the Root of Rodents

To facilitate an exhaustive and unbiased search for rare phylogenetic markers at the root of the rodent tree, we prepared specific three-way alignments with the MULTIZ program (Blanchette et al. 2004) using two pairwise BlastZ alignments (Chiaromonte et al. 2002; Kent et al. 2003; Schwartz et al. 2003) of sequences from mouse/guinea pig and 13-lined ground squirrel/mouse.

The alignments were linked into chains using a dynamic programming algorithm (“axtChain”) that finds maximally scoring chains of gapless subsections of the alignments organized in a kd tree. The parameters for axtChain were selected based on phylogenetic distance from the reference (chainMinScore = 3,000, chainLinearGap = medium). High-scoring chains were then filled in with lower-scoring chains to construct an alignment net (using the program “chainNet”), and the resulting pairwise net alignments used as the basis of the multiple alignment.

The first three-way alignment (mouse/guinea pig/ground squirrel) comprised 3.08 Gb divided into 5,847,869 blocks in multiple alignment format (MAF). From

this alignment, we analyzed 20,847 mouse/guinea pig sequences with corresponding gaps in the ground squirrel and 18,214 sequence regions with gaps in the guinea pig. Of these, 1,132 and 304 regions, respectively, contained retroposons in the remaining two species. The first three-way alignment was based on mouse genomic information as the “leading sequence.” This alignment was suitable to search for mouse + guinea pig and mouse + ground squirrel but not for guinea pig + ground squirrel shared retroposons. To investigate the latter phylogenetic direction, we used the second three-way alignment based on the guinea pig as the leading sequence (guinea pig/mouse/ground squirrel) that comprised 3.94 Gb and 8,113,940 MAF blocks. For guinea pig and ground squirrel, we found 19,831 sequences with gaps in mouse, 890 of which contained retroposed elements. All retroposon-containing, potential phylogenetic informative loci were selected for orthologous insertions (i.e., retroposon with similar orientation, retroposon flanking direct repeats, internal diagnostic indels, and clear absence in the outgroup). Clear orthologs were examined further and supplemented by screening in additional species. Among the 2,326 selected loci, only 70 were found to be phylogenetic informative under these criteria.

We also used the three-way alignments to search for diagnostic random insertions and deletions (indels conserving the ORF, that is, divisible by 3 nt) in protein-coding regions. For this, we located 191,078 mouse protein-coding exons (<http://genome.ucsc.edu/cgi-bin/hgTables>) in the corresponding regions of the two three-way alignments; 2,402 potential informative exons were selected and inspected by mouse BLAT for available information in additional mammalian species. Overlapping indels, indels located in repeated regions, as well as all homoplastic indels (e.g., indels present in primate and some rodent species but not in rabbit), were excluded.

Genomewide Retroposon Statistics

For a comprehensive survey of the spectrum of retroposed elements in rodents, we analyzed the genomic sequences of mouse, rat, and guinea pig and trace data from the ground squirrel. The RepeatMasker was used to identify retroposed elements using the “-species Rodentia” option and the RepeatMasker library version 14.06 (09-07-2009). All full-length rodent-specific SINEs were extracted and aligned against the RepeatMasker consensus sequence. To make all analyses comprehensible, for known elements, we strictly used the standard RepeatMasker or the Genetic Information Research Institute (giri, <http://www.girinst.org>) nomenclature. For obvious misannotations of the RepeatMasker program (e.g., detection of apparent jerboa-specific DIP elements in other rodent species), we extracted and characterized the misannotated elements from genomic data, derived consensus sequences in comparison to known and novel elements (e.g., ID-Spe, GPIDL), and built new user-defined RepeatMasker libraries for a final species-specific screening. These modified species-specific RepeatMasker libraries (available upon request) were then used to derive the

complete landscape of retroposed elements for mouse, rat, guinea pig, and ground squirrel (supplementary table S2, Supplementary Material online). Elements were assigned to specific internal branches of the phylogenetic tree according to their presence/absence in specific species. If no further information of distribution was available, certain elements were placed at terminal branches of species where they were first described. This does not necessarily mean that they “are” restricted to those species.

TinT Analysis for Detecting the Activity Period of Rodent Retroposed Elements

We recently developed a likelihood-based strategy to retrospectively explore ancestral fixation activity periods of retroposon insertions by analyzing nested transpositions (transpositions in transpositions, TinT; Kriegs, Matzke, et al. 2007). The TinT method was used here to establish the relative time periods of retroposon activity profiles of SP-D-Geo and IDL-Geo.

Results

Retroposons as Phylogenetic Markers Providing Evidence for Lineage-Specific Splits

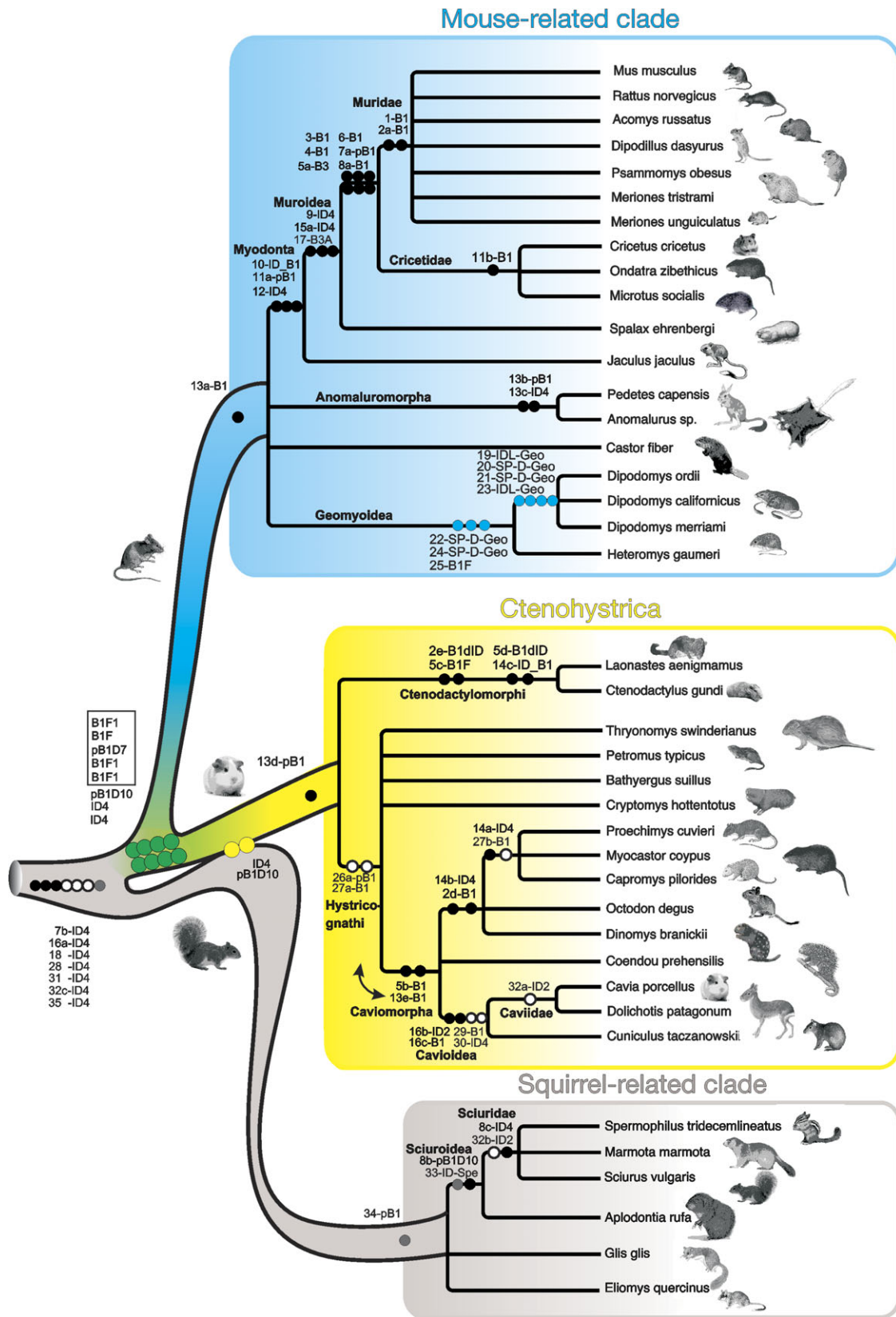
We used two approaches to identify novel informative SINE insertion loci and elements. First, we examined the previously investigated loci, originally derived from a screen of the mouse genome (Farwick et al. 2006; Huchon et al. 2007) in 17 additional rodent species representing all three major lineages. The increased taxonomic sampling generated support for five additional branches of the rodent tree: In the Ctenohystrica, marker 14a-ID4 supports a common ancestry of *Proechimys*, *Myocastor*, and *Capromys*; marker 14b-ID4 supports a common ancestry of Octodontoidea and Chinchilloidea; markers 16b-ID2 and 16c-B1 support the monophyly of Caviioidea. In the Squirrel-related clade, marker 8c-ID4 supports the monophyly of Sciuridae, and marker 8b-pB1D10 supports the monophyly of Sciuroidea (fig. 1; supplementary table S3, Supplementary Material online). The addition of material from *Cuniculus taczanowskii* resulted in significant support (four markers; Waddell et al. 2001) for the branch leading to Caviioidea. Moreover, two additional loci were identified using the CPAL finder, each containing one retroposon marker (supplementary table S3, Supplementary Material online). The new marker 17 consolidates the Muroidea clade, and marker 18 confirms the monophyly of rodents (fig. 1).

Second, we screened the predicted intronic sequences (mouse orthologs) derived from trace data (see Materials and Methods) of *D. ordii*, *C. porcellus*, and *S. tridecemlineatus* for additional lineage-specific conserved SINE loci. The screen of *D. ordii* intronic sequences yielded seven novel informative loci, each containing one retroposon marker (supplementary table S3, Supplementary Material online; markers 19–25). All are located within the Geomyoidea lineage, and four of them comprise a novel SINE subfamily, SP-D-Geo (see below). The screen of *C. porcellus* revealed 10 novel informative markers from 7 intronic loci. Three of them (markers 28, 31, and 32c) support the monophyly of rodents, two the monophyly of Hystricognathi (markers 26a, 27a), one (marker 27b) the common ancestry of *Proechimys*, *Myocastor*, and *Capromys*, two (markers 29 and 30) support the monophyly of Caviioidea, one (marker 32a) provides evidence for the monophyly of Caviidae, and the 10th (marker 32b) supports a common ancestry for all Sciuridae. The screen of *S. tridecemlineatus* revealed three additional loci with retroposon markers confirming the monophyly of rodents (marker 32c), the common origin of all Squirrel-related species (marker 34), and the monophyly of Sciuroidea (marker 33).

Retroposons as Phylogenetic Markers Solving the Root of Rodents

As none of the above markers helped us to resolve the root of the rodent tree, we computationally analyzed specifically generated three-way genome alignments, in MAF blocks (Materials and Methods; Warren et al. 2008), of mouse, guinea pig, and ground squirrel for gaps in one of them that were not present in the other two. The sequence regions in the other two corresponding to the region of the gaps were then analyzed for retroposons. We initially found 60 potential retroposon-containing loci in mouse and guinea pig that were absent in the ground squirrel and 10 in the guinea pig and ground squirrel that were not in mouse. No such sequences were found in mouse and ground squirrel that were absent in guinea pig. These 70 loci were aligned with additional available sequences, including the outgroups human and rabbit as well as rat and kangaroo rat. Unambiguously clear, recognizable, orthologous insertion sites were identified in the outgroup for only 10 of these loci: eight retroposon insertions with orthologs in mouse and guinea pig and two elements in guinea pig and ground squirrel (Supplementary Material online). Thus, although the preponderance of evidence provides statistically

Fig. 1. Phylogenetic tree of rodents derived from retroposon presence–absence data. Black circles represent markers detected by screening *Mus musculus* introns; blue circles show markers derived from *Dipodomys ordii* trace data, white circles from *Cavia porcellus* genomic sequences and gray circles from *Spermophilus tridecemlineatus* genomic sequences. Enlarged green/yellow circles represent retroposons detected from the genomic three-way alignments (eight markers supporting the grouping of the Mouse-related clade with Ctenohystrica, five of which are from specific B1-related subfamilies [framed] that are not present in the Squirrel-related clade and only two markers supporting the grouping of Ctenohystrica with the Squirrel-related clade). Labeled markers are listed in supplementary table S3 (Supplementary Material online) and include the retroposon type specification. Note that the ID subtypes ID2 or ID4 differ by only two diagnostic nucleotide substitutions. The two melting branches and their shared ID4 and pB1D10 elements indicate the introgressive hybridization or incomplete lineage sorting of the Ctenohystrica and the Squirrel-related clades.



significant support for the early separation of the Squirrel-related clade from those of the Mouse-related and Ctenohystrica clades (Waddell et al. 2001), there is also some evidence for the Mouse-related clade being the sis-

ter group of both the Ctenohystrica and Squirrel-related clades.

As additional independent evidence to resolve the basal evolutionary relationships of rodents we selected all

annotated mouse exons from the three-way alignments and searched for diagnostic random indels in protein-coding regions. After the first screen, we added all available sequence information from rat and kangaroo rat as well as from human, rabbit, or pica and dog as outgroups. Only seven indels survived our criteria (nonoverlapping, nonhomoplastic; see Materials and Methods). Six of them were shared by mouse and guinea pig but not ground squirrel, further supporting the Squirrel-related clade as the root of the rodent tree. But, there was also one indel shared by guinea pig and ground squirrel but not mouse. Once again, mouse and ground squirrel shared no indels.

Genomewide Retroposon Distribution in Major Rodent Lineages

Thus, due to the apparently conflicting evidence in the above two analyses, we sought further evidence to resolve the root of the rodent tree by examining the occurrence of rodent-specific retroposon subfamilies in the available assembled genomic information of *M. musculus*, *R. norvegicus*, *C. porcellus*, and *S. tridecemlineatus*. The mouse and rat retroposon copy numbers obtained in this work were carefully crosschecked with the estimates published in the corresponding genome papers (Waterston et al. 2002; Gibbs et al. 2004). To these data we also added information from previously published work on SINE subfamilies in other rodent species and from the novel retroposons discovered in the *D. ordii*, *C. porcellus*, and *S. tridecemlineatus* genomes during the course of the current study (see below). This information was then mapped on the phylogenetic tree of rodents to derive a comprehensive picture of the distribution of rodent retroposons (fig. 2). It is worth noting that no apparent conflicts were detected in this distribution pattern. Instead, the distribution analysis clearly supported a grouping of the Mouse-related clade and Ctenohystrica. Mouse (137,584 total copies) and guinea pig (380,365 total copies) share five B1-related SINE subfamilies (pB1D7, pB1D9, B1F, B1F1, and B1F2) as well as a specific dimeric ID_B1 element (111,246 and 28,330 copies, respectively; fig. 2; supplementary table S2, Supplementary Material online). These elements are absent in the squirrel genome. The B1-elements of mouse and guinea pig (B1F, B1F1, and B1F2) share a specific 29-nt duplication (B1F₂₉). This duplication is absent in their putative progenitor the pB1D7 element (Veniaminova et al. 2007). The ground squirrel genome contains B1-related elements with a 20-nt duplication located at the same position as the 29-nt duplication of the B1 elements. These B1-like elements are specific to members of the squirrel-related clade and most likely arose independently from a pB1D10 progenitor (Veniaminova et al. 2007). They are called B1L₂₀ for B1-like. All B1L₂₀-derived elements (e.g., the dimeric B1L-dID and MEN elements of the Squirrel-related clade) share the characteristics of the B1L₂₀ progenitor.

Novel Retroposons

SP-D-Geo. In screening *D. ordii* trace sequences for phylogenetically informative retroposons, we detected an unknown dimeric SINE with some distal similarity to the previously

reported IDL-Geo element (Gogolevsky and Kramerov 2006). The new element, called SP-D-Geo, is composed of a pB1D10 (SP-D; first defined in Kriegs, Churakov, et al. 2007) and a Geo monomer. ID4 elements show a high similarity to Geo (~80%) and are probably their progenitor. Both of the ID4 promoter A and B boxes are diverged but still recognizable in Geo. Thus IDL-Geo probably resulted from the dimerization/recombination of two ID elements (fig. 3a). In the genome of *D. ordii*, we estimated about 40,700 copies of IDL-Geo and 61,000 copies of SP-D-Geo (fig. 3b). The transpositions in transpositions analysis (TinT; Kriegs, Matzke, et al. 2007; data not shown) clearly identified SP-D-Geo as being older than the IDL-Geo elements. Both elements are restricted to Geomyoidea.

ID-Spe. A novel SINE, ID-Spe, was initially detected in *S. tridecemlineatus*. This element is composed of a 5' ID4-like monomer with an A-rich region followed by a 23-nt sequence (probably the remains of a partially deleted second ID monomer; see twinID-Spe) that ends in an A-tail (fig. 3c). A phylogenetically informative ID-Spe element was detected in *S. tridecemlineatus* and at an orthologous position in *Aplodontia rufa*, indicating that it is not Spermophilus specific but has a more ancestral history. We calculated about 185,000 copies of the ID-Spe element in *S. tridecemlineatus*.

tri-Spe. The analysis of *S. tridecemlineatus* genomic sequences also revealed a novel trimeric element built from two B1L monomers and one ID element that we call tri-Spe. The structure is comparable with trimeric elements in tree shrews (Tu type II) (Nishihara et al. 2002) and the colugo (CYN-III) (Schmitz and Zischler 2003) except that a 5' B1-like element provides the A and B-boxes for transcription as opposed to the 5' tRNA-derived monomers in tree shrew and colugo. We calculated about 15,000 copies of tri-Spe in the *S. tridecemlineatus* genome (fig. 3d).

twinID-Spe. A novel dimeric ID element called twinID-Spe was detected in genomic data of *S. tridecemlineatus*. There are about 18,000 calculated copies of the element in the assembled *S. tridecemlineatus* genome. Based on sequence comparison (data not shown), we hypothesize that twinID-Spe (fig. 3e) is the progenitor of the ID-Spe element.

Discussion

The random genomic insertion of a retroposed element in a common ancestor of two species can serve a 100 My later as a virtually homoplasy-free marker of their shared ancestry. Thus, retroposon insertions are powerful indicators of relatedness, and as long as they are frequent enough and master copies were active over long evolutionary periods, they enable us to resolve complete phylogenetic tree topologies (Shedlock and Okada 2000; Kriegs et al. 2006; Nishihara et al. 2006). In rodents, SINEs are frequent (~8% of the mouse genome; Waterston et al. 2002) and were active over the entire evolution of rodents (see figs. 1 and 2). However, due to extremely high substitution rates in many species, in particular in the Mouse-

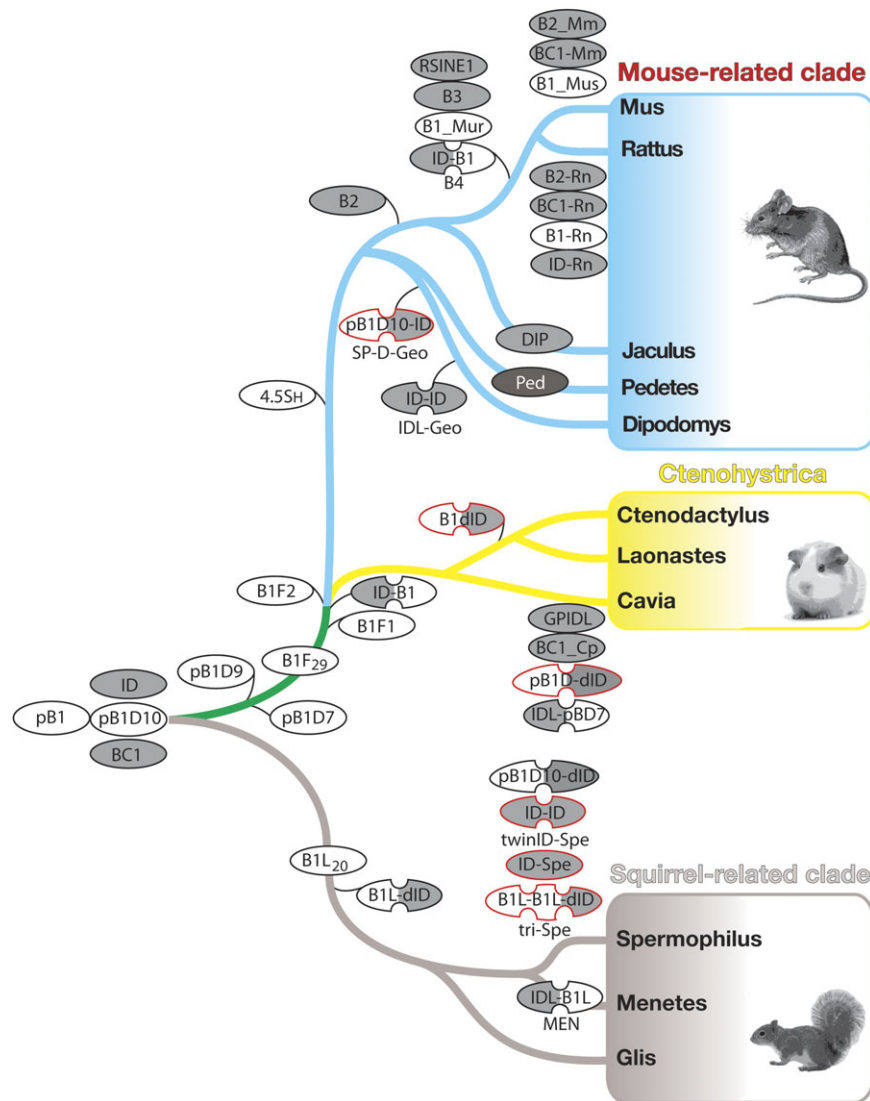


Fig. 2. Distribution and activity of SINE element families in rodents. The common ancestral genome of rodents already contained active tRNA-(BC1 and ID) and 7SL-derived elements (pB1 and pB1D10). pB1 and pB1D10 (pB1D10 is a pB1 derivate with a typical 10-nt deletion) were progenitors for all B1-related SINEs in rodents. ID elements continued their activity along the main rodent branches with an increased activity in rat and contributed to diverse dimeric elements (ID subtypes are not shown). BC1 elements experienced a boost of retropositional activity in rat. pB1D7, pB1D9, B1F₂₉, B1F1, B1F2, and ID-B1 evolved on the common branch of the Mouse-related and Ctenohystrica clades. 4.5S_H elements in the Mouse-related clade are possibly derived from a B1-like ancestor. DIP elements are probably tRNA-derived elements specific for jerboas. Ped elements are specific for *Pedetes capensis* and propagate via a retroposon-like transposable element (RTE)-related distribution machinery. On the branch leading to the Squirrel-related clade a specific 20-nt duplication took place in B1L₂₀, the origin of all subsequently evolving B1-related monomers and dimers (e.g., B1L-dID) in the Squirrel-related species. tRNA-derived elements are gray (including B2 and B3 elements), B1-related elements are white, and the RTE-mobilized Ped element is dark gray. Dimeric and trimeric elements are indicated by notched ovals. Trivial names of such dimers are indicated below the ovals. The six newly discovered SINE subfamilies (SP-D-Geo, twinID-Spe, ID-Spe, tri-Spe, and the yet uncharacterized pB1D10-dID (in *Cavia*) and B1dID (in *Ctenodactylus* and *Laonastes*) are outlined in red. All references are compiled in [supplementary table S4](#), Supplementary Materials online.

related clade, finding reliable diagnostic retroposons is extremely laborious compared with other mammalian orders (e.g., Xenarthra [Moeller-Krull et al. 2007] and primates [Schmitz et al. 2005]). The diverged sequences hamper the use of retroposons as phylogenetic markers in two ways: First, PCR amplification throughout the full spectrum of rodents is problematic; second, intronic alignments are difficult to build. Thus, we were forced to exclude most of the preselected, potentially informative markers. From every 100 loci identified as containing retroposons,

only about 10 can be correctly aligned and amplified in other rodent lineages. To acquire a large enough number of informative markers, bioinformatic screening, including whole genomes and high-throughput amplification procedures, is indispensable in rodents.

In the present case, hundreds of investigated genomic loci yielded only 35 intronic sequences containing 55 reliable phylogenetically informative markers spread over nearly all internal branches of the rodent phylogenetic tree. These provided “significant” support (three or more

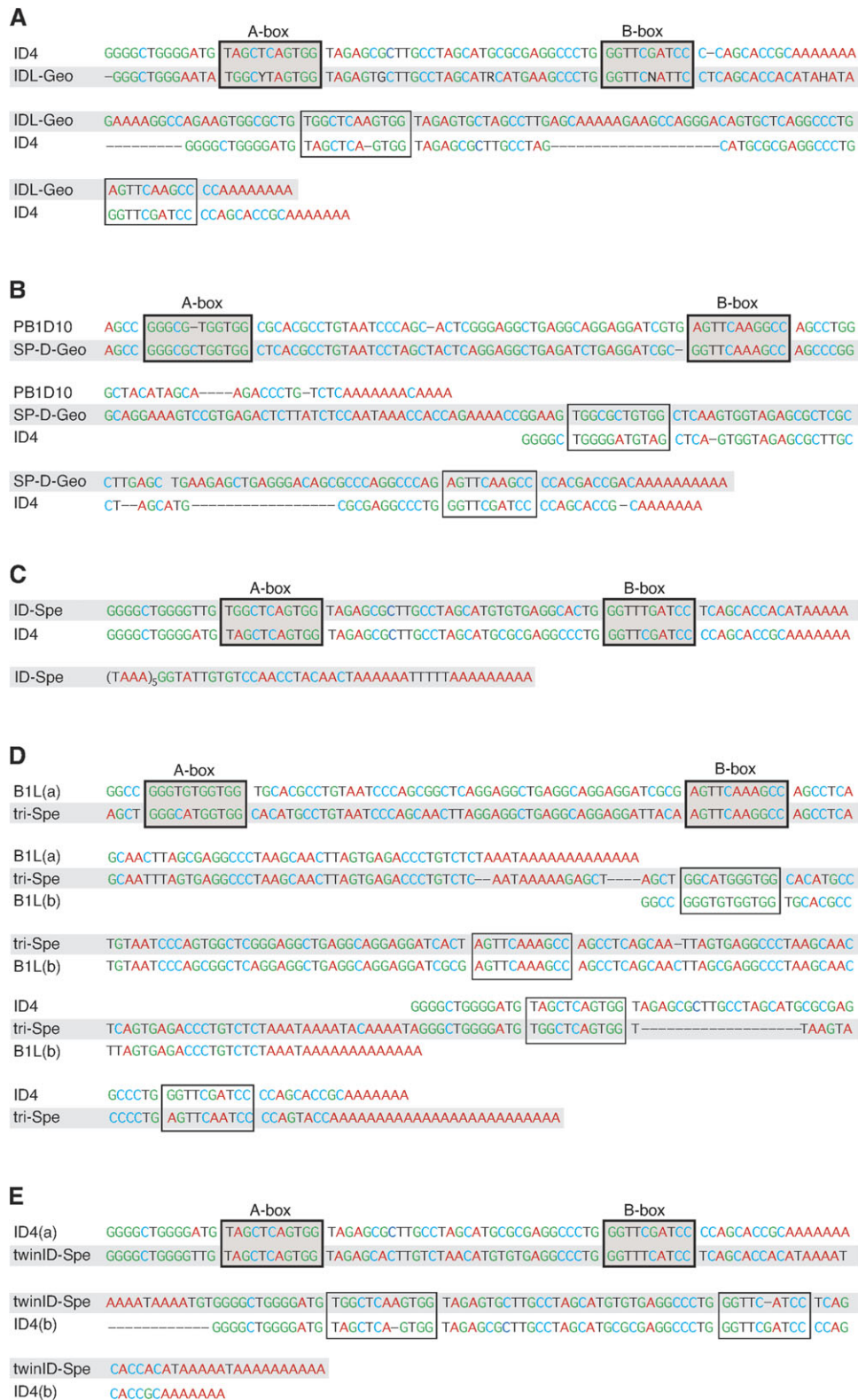


Fig. 3. Novel SINE elements detected in *Dipodomys ordii* and *Spermophilus tridecemlineatus* genomic sequences. Promoter boxes necessary for transcription are labeled as A and B boxes (boxed and shaded in gray; diverged promoter sequences are simply boxed). Novel found/analyzed retroposons are shaded in gray. Alignment gaps are shown as dashes. (A) Newly detected relationship of both IDL-Geo element monomers (Gogolevsky and Kramerov 2006) to ID4. (B) Newly detected SP-D-Geo element and its relationship to pB1D10 (first monomer) and ID4 (second monomer). (C) New SINE ID-Spe element in *S. tridecemlineatus* derived from an ID4 element. (D) Novel tri-Spe element in *S. tridecemlineatus* composed of two B1L monomers and one ID element. (E) New dimeric ID element (twinID-Spe) detected in *S. tridecemlineatus*.

markers for each branch (sensu Waddell et al. 2001) for eight rodent branches, “strong” evidence for eight branches (two markers each), and “preliminary” information for five additional branches (one marker each) (fig. 1). For example, the monophyly of rodents, including guinea pigs, once strongly questioned (Graur et al. 1991; D’Erchia et al. 1996) (but see also Martignetti and Brosius 1993), is now solidly confirmed by seven ID element insertions that are clearly present in all major rodent lineages but absent in the outgroup human/rabbit. Strong support for the three major clades (Mouse-related clade, Ctenohystrica, and Squirrel-related clade) awaits further SINE investigation, as our results produced only one marker for each branch. However, these nodes are strongly supported by sequence data (Huchon et al. 2007; Montgelard et al. 2008; Blanga-Kanfi et al. 2009). Our results “do” provide a much clearer picture of the evolutionary events at the root of rodents. Grouping the Mouse-related clade with Ctenohystrica is significantly supported (Waddell et al. 2001) by eight retroposons and six indels. However, we also found two retroposons and one indel present in both the Ctenohystrica and the Squirrel-related clades. Such apparent conflict among retroposon insertion patterns is extremely rare in deep mammalian branches but was also described for the early divergence of placentals (Churakov et al. 2009; Nishihara et al. 2009). The most parsimonious interpretation of our data, similar to that shown recently for the base of the placentals (Churakov et al. 2009) is that the ancestral rodent populations were probably affected by events of introgressive hybridization or incomplete lineage sorting. In addition to incomplete lineage sorting (Shedlock et al. 2004), a likely scenario describing the first rodent divergence involves an early separation of the pre-Squirrel-related clade from the common ancestor of rodents followed by a later separation of the pre-Mouse-related and pre-Ctenohystrica clades, and possibly an introgression of squirrel-related genes into the pre-Ctenohystrica genome or vice versa.

Biogeographical locations of fossils have previously enlightened molecular results (Teeling et al. 2005). Although numerous rodent fossil families have been described, their relationships with extant species are often debated; consequently, a clear picture of the past distribution of extant lineages is still not available. Similar to many placental lineages, the current paleontological data support the probable origin of rodents in Asia, from where they colonized all other continents (except South America) at the Paleocene–Eocene boundary (~56.3 Ma) (Beard 2002). The Paleocene fossil record attests to an early division of Asian rodents into two main lineages the Ischyromyidae and the “ctenodactyloids.” The Ischyromyidae was the first rodent family to colonize Europe and North America and the only one known from all three continents (Asia, Europe, and North America), whereas ctenodactyloids were endemic Asian rodents through the Eocene (Dawson 2003). Members of the Squirrel-related clade are thought to be closely related to North American and European ischyromyid (Korth 1994; Hartenberger

1998). Ctenohystrica, instead, are considered to be ctenodactyloid descendents (Marivaux et al. 2004). This early taxonomic and geographical dichotomy of rodent fossils would fit our SINE-based inference if members of the Mouse-related clade were known to have originated from ctenodactyloid ancestors. Although such relationships have been suggested in the past (e.g., Flynn et al. 1985), the current paleontological view considers, instead, that the family Sciuravidae is the ancestral stock for Geomyoidea, Muroidea, and Dipodoidea (Korth 1994; Marivaux et al. 2004; Emry 2007). The Sciuravidae is an Early-Eocene family endemic of North America that was traditionally related to the Ischyromyidae (Wilson 1949). Our SINE inference of the rodent root suggests that the morphological characteristics of Sciuravidae and of the oldest members of the Mouse-related clade should be reconsidered in the light of a ctenodactyloid origin of the Mouse-related clade.

In contrast to statistical analyses of nucleotide substitutions, the interpretation of retroposed sequences, due to their clear character polarity (presence of an element as the derived state) and the low probability of orthologous, exact deletions or parallel insertions, is straightforward and thus is ideal for conserving such ancestral signals of population dynamics and to identify rapid, successive speciation events that deviate from clear dichotomic patterns. According to our SINE-based scenario, we expect that analyses of sequence substitutions might lead, depending on the gene sampled, to support for a basal position of either the Squirrel-related clade or the Mouse-related clade or might lead to an absence of support for any hypothesis, which is exactly what has been observed. Montgelard et al. (2008) analyzed two concatenated mitochondrial genes and six nuclear ones (two exons and four introns; ~7,600 nt) in 30 rodent species and found support for a basal position of the Mouse-related clade when fast-evolving characters were excluded. In contrast, Blanga-Kanfi et al. (2009) analyzed six nuclear exons from 41 rodent species (6,255 nt) and did not solve the basal rodent trifurcation but did favor a basal position of the Squirrel-related clade using rate-shift models. They explained the difficulty in resolving the rodent root by the rapid rodent radiation rather than by conflicting phylogenetic signals. It is expected that with the increase in sequence data, analyses of sequence substitution should eventually converge toward a basal position of the Squirrel-related clade (e.g., Hallström and Janke 2008), although branch support at the base of the rodent tree will probably always be low when considering a large taxon sampling. By contrast, our retroposon insertion data overwhelmingly support the basal position of the Squirrel-related clade.

Newly Discovered SINE Elements

Broad genomic screening for phylogenetically informative presence–absence markers is also well suited for detecting previously unknown retroposed sequences and is a starting point for gaining an *in silico*–based image of the genomic distribution of such elements (which has

phylogenetic value too; see below). Comparing the presence of an element in one species and its absence in another provides reliable information about the retroposon boundaries and their mechanisms of integration. With the aid of newly available, large-scale genome information, we identified a new SP-D-Geo SINE that bears similarity to the IDL-Geo element previously described by Gogolevsky and Kramerov (2006). The latter element represents the first dimeric ID–ID-derived element discovered in rodents. Di- and trimerizations of SINE elements have been described as progressive advantages for increasing retropositional efficiency (Borodulina and Kramerov 2006), but the reason for this is unclear. It is possible that remnants of the second part of dimeric elements, such as the one found in the newly discovered Spermophilus ID-Spe SINE, provide structural elements for such higher retropositional efficiency. In *Spermophilus*, we also discovered tri-Spe, the first trimeric element in rodents, and twinID-Spe, a dimeric ID element with a similar composition to IDL-Geo in Geomyoidea.

The Activity Spectrum of SINEs in Rodents

In accordance with the observed SINE and indel presence–absence patterns, the spectrum of SINE activity in rodents obtained from the whole genomewide analyses also provides strong support for the early separation of the Squirrel-related clade. This is best exemplified by the B1-elements (pB1D7-derived) with the characteristic 29-nt duplication present in the Mouse-related clade (represented by mouse, rat, and kangaroo rat) and in Ctenohystrica (represented by the guinea pig) but absent in the Squirrel-related clade (except for a few hits in trace data that were determined to be mouse contaminations) and therefore thought to be derived in a common ancestor of mouse and guinea pig. It is worth noting that because the squirrel genome, compared with the mouse genome, is slowly evolving, it is unlikely that B1 elements were misidentified by our transposon search. A close relationship of the Mouse-related clade and Ctenohystrica was first proposed by Veniaminova et al. (2007) based on the distribution pattern of experimentally found B1-related retroposons. This first indication is now borne out at the genomewide level. In the Squirrel-related clade, the predominant B1-“like” elements (pB1D10-derived) are similar to B1 elements in mouse and guinea pig but differ by an independent duplication of a 20-nt sequence region. The duplication shows strong similarity to the flanking, squirrel-specific B1 region and is, therefore, considered an independent duplication event rather than a 9-nt deletion from a potential B1D29 progenitor. The screening for individual orthologous elements shared by the Mouse-related clade and Ctenohystrica but absent in the Squirrel-related clade “via” three-way alignments revealed that five such elements (pB1D7, B1F, and 3 × B1F1) belong to element subfamilies absent in the Squirrel-related clade (fig. 1). The remaining three diagnostic elements (pB1D10, 2 × ID) as well as the two elements

shared between Ctenohystrica and the Squirrel-related clade (ID4, pB1D10) belong to retroposon subfamilies that were active over a long range of rodent evolution (e.g., ID4 and pB1D10 elements were found that resolve both basal and terminal branches of the tree; fig. 1). This demonstrates the congruence of orthologous single loci comparisons and the diagnostic genomewide distribution patterns of retroposons.

Conclusion

Retroposons offer an invaluable window into rodent evolution. A detailed knowledge of the elements, their distributions, and changes over time are key attributes helping to provide a virtually homoplasy-free reconstruction of history. In the present study, we considered different aspects of retroposon evolution, identified new elements, and described a genomewide picture of their distribution. We applied diverse bioinformatical search strategies to exhaustively screen for informative retroposon presence/absence patterns. Our data provide an understanding of early speciation in rodents, a process revealed to be more complicated than can be explained by clear bifurcation events. Such complexity cannot be resolved by classical, sequence-based statistical approaches. Our data support an early divergence of a pre-Squirrel-related clade from a common ancestor of a pre-Mouse-related-Ctenohystrica clade. The time that elapsed between the separation of the Squirrel-related clade and the split at the origin of the Mouse-related and Ctenohystrica clades was sufficient to fix at least eight orthologous retroposon elements and six indels in the common ancestor of the pre-Mouse-related and Ctenohystrica clades. However, the two retroposon markers shared by Ctenohystrica and the Squirrel-related clade indicate that the separation at that time was not strictly complete and that possibly a limited hybridization between the pre-Ctenohystrica and presquirrel ancestor occurred. However, incomplete lineage sorting is another possible explanation for the results observed (Shedlock et al. 2004; Churakov et al. 2009; Nishihara et al. 2009). A close relationship of the Mouse-related clade and Ctenohystrica was also strongly supported by the general distribution pattern of B1-related elements in rodents with pB1D7, pB1D9, B1F₂₉, B1F1, B2F2, and ID-B1 elements that arose in the common ancestor of these two. In total, we found 65 retroposons that support 23 different branches of the rodent evolutionary tree (e.g., in cementing the monophyly of this order by seven orthologous elements shared by all rodents), representing the most extensive retroposon study in rodents to date that offers an ideal starting point to unlock the last remaining secrets in the evolution of rodents.

Supplementary Material

Supplementary tables S1–S4 and Supplementary Material S1 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Marsha Bundman for editorial assistance. We very much appreciate the help of Swathy Sadasivuni and Ursula Jordan for the PCR amplifications. We thank Mary Dawson, Dmitri A. Kramerov, and Rodney L. Honeycutt for helpful discussions, two anonymous reviewers for their comments, and Christen Williams, Chris Bonar, François Catzeflis, David Eilam, Chris Faulkes, Jack Hayes, Reginald Hoyt, John Trupkiewicz, Noga Kronfeld-Schor, and Marian Mensink for providing tissue samples. This work was supported by the Deutsche Forschungsgemeinschaft (SCHM1469/3-1). J.B. notes that an earlier grant proposal of his on rodent phylogeny was rejected by the National Science Foundation.

References

- Beard C. 2002. Paleontology. East of Eden at the Paleocene/Eocene boundary. *Science* 295:2028–2029.
- Benton MJ, Donoghue PCJ. 2007. Paleontological evidence to date the tree of life. *Mol Biol Evol.* 24:26–53.
- Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. 2007. The delayed rise of present-day mammals. *Nature* 446:507–512.
- Blanchette M, Kent WJ, Riemer C, et al. (12 co-authors). 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* 14:708–715.
- Blanga-Kanfi S, Miranda H, Penn O, Pupko T, DeBry RW, Huchon D. 2009. Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evol Biol.* 9:71.
- Borodulina OR, Kramerov DA. 2006. t-SINE or simple SINE? Can be both. *Gene* 375:111.
- Chiaromonte F, Yap VB, Miller W. 2002. Scoring pairwise genomic sequence alignments. *Pac Symp Biocomput.* 7:115–126.
- Churakov G, Kriegs JO, Baertsch R, Zemmann A, Brosius J, Schmitz J. 2009. Mosaic retroposon insertion patterns in placental mammals. *Genome Res.* 19:868–875.
- Dawson MR. 2003. Paleogene rodents of Eurasia. *Deinsea.* 10:97–126.
- DeBry RW. 2003. Identifying conflicting signal in a multigene analysis reveals a highly resolved tree: the phylogeny of Rodentia (Mammalia). *Syst Biol.* 52:604–617.
- D'Erchia AM, Gissi C, Pesole G, Saccone C, Arnason U. 1996. The guinea-pig is not a rodent. *Nature* 381:597–600.
- Douzery EJP, Delsuc F, Stanhope MJ, Huchon D. 2003. Local molecular clocks in three nuclear genes: divergence times for rodents and other mammals and incompatibility among fossil calibrations. *J Mol Evol.* 57(1 Suppl):S201–S213.
- Emry RJ. 2007. The middle Eocene North American myomorph rodent *Elymys*, her Asian sister *Aksyiromys*, and other Eocene myomorphs. *Bull Carnegie Mus Natur Hist.* 39:141–150.
- Farwick A, Jordan U, Fuellen G, Huchon D, Catzeflis F, Brosius J, Schmitz J. 2006. Automated scanning for phylogenetically informative transposed elements in rodents. *Syst Biol.* 55: 936–948.
- Flynn LJ, Jacobs LL, Lindsay EH. 1985. Problems in muroids phylogeny: relationship to other rodents and origin of major groups. In: Luckett WP, Hartenberger J-L, editors. *Evolutionary relationships among rodents*. New York: Plenum Press. p. 589–616.
- Gibbs RA, Weinstock GM, Metzker ML, et al. (230 co-authors). 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428:493–521.
- Gissi C, Reyes A, Pesole G, Saccone C. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. *Mol Biol Evol.* 17:1022–1031.
- Gogolevsky KP, Kramerov DA. 2006. Short interspersed elements (SINEs) of the Geomyoidea superfamily rodents. *Gene* 373:67–74.
- Graur D, Hide WA, Li WH. 1991. Is the guinea-pig a rodent? *Nature* 351:649–652.
- Hallström BM, Janke A. 2008. Resolution among major placental mammal interordinal relationships with genome data imply that speciation influenced their earliest relations. *BMC Evol Biol.* 8:162.
- Hartenberger J-L. 1998. Description de la radiation des Rodentia (Mammalia) du Paléocène supérieur au Miocène; incidences phylogénétiques. *CR Acad Sci.* 326:439–444.
- Huchon D, Chevret P, Jordan U, Kilpatrick CW, Ranwez V, Jenkins PD, Brosius J, Schmitz J. 2007. Multiple molecular evidences for a living mammalian fossil. *Proc Natl Acad Sci U S A.* 104:7495–7499.
- Huchon D, Madsen O, Sibbald MJ, Ament K, Stanhope MJ, Catzeflis F, de Jong WW, Douzery EJ. 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.
- Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A.* 100:11484–11489.
- Korth WW. 1994. *The Tertiary record of rodents in North America*. Plenum Press, New York.
- Kriegl JO, Churakov G, Jurka J, Brosius J, Schmitz J. 2007. Evolutionary history of 7SL RNA-derived SINEs in supraprimates. *Trends Genet.* 23:158–161.
- Kriegl JO, Churakov G, Kieffmann M, Jordan U, Brosius J, Schmitz J. 2006. Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol.* 4:e91.
- Kriegl JO, Matzke A, Churakov G, Kuritzin A, Mayr G, Brosius J, Schmitz J. 2007. Waves of genomic hitchhikers shed light on the evolution of gamebirds (Aves: Galliformes). *BMC Evol Biol.* 7:190.
- Kumar S, Subramanian S. 2002. Mutation rates in mammalian genomes. *Proc Natl Acad Sci U S A.* 99:803–808.
- Luo ZX, Crompton AW, Sun AL. 2001. A new mammaliaform from the early Jurassic and evolution of mammalian characteristics. *Science* 292:1535–1540.
- Marivaux L, Vianey-Liaud M, Jaeger J-J. 2004. High-level phylogeny of early Tertiary rodents: dental evidence. *Zool J Linn Soc.* 142:105–134.
- Martignetti JA, Brosius J. 1993. Neural BC1 RNA as an evolutionary marker: guinea pig remains a rodent. *Proc Natl Acad Sci U S A.* 90:9698–9702.
- Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci U S A.* 90:4087–4091.
- Moeller-Krull M, Delsuc F, Churakov G, Marker C, Superina M, Brosius J, Douzery EJ, Schmitz J. 2007. Retroposed elements and their flanking regions resolve the evolutionary history of xenarthran mammals (armadillos, anteaters, and sloths). *Mol Biol Evol.* 24:2573–2582.
- Montgelard C, Forty E, Arnal V, Matthee CA. 2008. Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. *BMC Evol Biol.* 8:321.
- Murata S, Takasaki N, Saitoh M, Okada N. 1993. Determination of the phylogenetic relationships among Pacific salmonids by using

- short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc Natl Acad Sci U S A*. 90:6995–6999.
- Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* 409:614–618.
- Nishihara H, Hasegawa M, Okada N. 2006. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc Natl Acad Sci U S A*. 103:9929–9934.
- Nishihara H, Maruyama S, Okada N. 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proc Natl Acad Sci U S A*. 106:5235–5240.
- Nishihara H, Terai Y, Okada N. 2002. Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins. *Mol Biol Evol*. 19:1964–1972.
- Philippe H, Laurent J. 1998. How good are deep phylogenetic trees? *Cur Opin Genet Dev*. 8:616–623.
- Ryan SC, Dugaiczak A. 1989. Newly arisen DNA repeats in primate phylogeny. *Proc Natl Acad Sci U S A*. 86:9360–9364.
- Schmitz J, Roos C, Zischler H. 2005. Primate phylogeny: molecular evidence from retroposons. *Cytogenet Genome Res*. 108:26–37.
- Schmitz J, Zischler H. 2003. A novel family of tRNA-derived SINEs in the colugo and two new retrotransposable markers separating dermopterans from primates. *Mol Phylogenet Evol*. 28:341–349.
- Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. 2003. Human–mouse alignments with BLASTZ. *Genome Res*. 13:103–107.
- Shedlock AM, Okada N. 2000. SINE insertions: powerful tools for molecular systematics. *Bioessays* 22:148–160.
- Shedlock AM, Takahashi K, Okada N. 2004. SINEs of speciation: tracking lineages with retroposons. *Trends Ecol Evol*. 19:545–553.
- Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T, Goto M, Munechika I, Okada N. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* 388:666–670.
- Spradling TA, Hafner MS, Demastes JW. 2001. Differences in rate of cytochrome-b evolution among species of rodents. *J Mammalogy* 82:65–80.
- Stanyon R, Stone G, Garcia M, Froenicke L. 2003. Reciprocal chromosome painting shows that squirrels, unlike murid rodents, have a highly conserved genome organization. *Genomics* 82:245–249.
- Steel M, Penny D. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Mol Biol Evol*. 17:839–850.
- Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ, Murphy WJ. 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307:580–584.
- Veniaminova NA, Vassetzky NS, Kramerov DA. 2007. B1 SINEs in different rodent families. *Genomics* 89:678–686.
- Waddell PJ, Kishino H, Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Inform*. 12:141–154.
- Warren WC, Hillier LW, Graves JAM, et al. (104 co-authors). 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453:175–183.
- Waterston RH, Lindblad-Toh K, Birney E, et al. (222 co-authors). 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562.
- Wilson RW. 1949. Tertiary rodents of North America. *Carnegie Inst Washington Publ*. 584:67–164.
- Wu CI, Li WH. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc Natl Acad Sci U S A*. 82:1741–1745.