J Mol Evol (1996) 42:7–14

JOURNAL OF MOLECULAR EVOLUTION

Sporadic Amplification of ID Elements in Rodents

David H. Kass,[1] Joomyeong Kim,[1] Prescott L. Deininger[1,2]

[1] Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, 1901 Perdido St., New Orleans, LA 70112, USA
[2] Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, New Orleans, LA 70121, USA

Received: 14 November 1994 / Accepted: 6 December 1994

Abstract. ID sequences are members of a short interspersed element (SINE) repetitive DNA family within the rodent genome. The copy number of individual ID elements varies by up to three orders of magnitude between species. This amplification has been highly sporadic in the order Rodentia and does not follow any phylogenetic trend. Using library screening and dot-blot analysis, we estimate there are 25,000 copies of ID elements in the deer mouse, 1,500 copies in the gerbil (both cricetid rodents), and 60,000 copies of either ID or ID-like elements in a sciurid rodent (squirrel). By dot-blot analysis, we estimate there are 150,000, 4,000, 1,000, and 200 copies of ID elements in the rat, mouse, hamster, and guinea pig, respectively (which is consistent with previous reports) and 200 copies in the hystricognath rodent, nutria. Therefore, a rapid amplification took place not only after the divergence of rat and mouse but also following the deer mouse (*Peromyscus*) and hamster split, with no evidence of increased amplifications in hystricognath rodents. No notable variations of sequences from the BC1 genes of several myomorphic rodents were observed that would possibly explain the varied levels of ID amplification. We did observe subgenera and species-group-specific variation in the ID core sequence of the BC1 gene within the genus *Peromyscus*. Sequence analysis of cloned ID elements in *Peromyscus* show most ID elements in this genus arose prior to *Peromyscus* subgenus divergence. Correspondence of the consensus sequence of individual ID elements in gerbil and deer mouse further confirms BC1 as a master gene in ID amplification. Several possible mechanisms responsible for the quantitative variations are explored.

Key words: Rodent evolution — ID repeats — SINEs — BC1 — Retroposition

Introduction

ID elements are members of a family of short interspersed elements (SINEs) (Deininger 1989), with copy numbers that vary by several orders of magnitude between rodent species (Sapienza and St-Jacques 1986; Anzai et al. 1987). Previously, it was observed that the copy numbers decreased with increased taxonomic distance from the rat, including over an order-of-magnitude variation with the mouse (Sapienza and St-Jacques 1986; Anzai et al. 1987). Because rat and mouse diverged about 25 million years ago (Brownell 1983; Sarich 1985; O'hUigin and Li 1992), over 100,000 copies must have been produced from that time to the present-day rat.

Characteristic features of SINEs include A box and B box RNA polymerase III promoter sequences, oligo dA-rich 3′ regions, and flanking direct repeats indicative of retroposition as the primary mode of dispersal (Deininger 1989). The subfamily structure of *Alu* repeats, based on diagnostic nucleotide positions (reviewed in Deininger and Batzer 1993), and the divergence of nucleotide sequences between individual elements, imply a sequential amplification from a very limited number of

The nucleotide sequences reported in this paper have been submitted to the GenBank/EMBL Data Bank with accession numbers: U33850, U33851, U33852 (BC1 sequences); and U33853, U33854, U33855, U33856, U33857, U33858, U33859, U33860, U33861, U33862, U33863, U33864, U33865, U33866, U33867 (ID sequences)
Correspondence to: D.H. Kass

master or source loci (Deininger et al. 1992). Most of the 500,000 *Alu* elements among primates belong to the oldest subfamily, suggesting a decrease of amplification rates through evolutionary time (Shen et al. 1991). This scenario provides a striking contrast to ID evolution.

Based on the close relationship between the consensus of ID sequences and the BC1 RNA gene within different rodent species (Kim et al. 1994) and the dominant RNA polymerase III–derived transcription of the BC1 RNA (DeChiara and Brosius 1987), the BC1 RNA gene has been determined to be a master gene for ID SINE amplifications. A likely explanation for the vast copy number increase in the rat is the creation of one or more new, highly active master genes, as suggested by the recent subfamily structure of ID elements (Kim et al. 1994). Possible explanations for the lower copy number in guinea pig include differences in the transcriptional regulatory sequences of the BC1 RNA gene, and the less homogeneous A-tail (Kim et al. 1994), which has been proposed to play a role in priming of reverse transcription (Jagadeeswaran et al. 1981).

The possibility exists that several mechanisms are involved in the phenomenon of copy number variation that may have occurred at different branch points of rodent evolution. We therefore examined several additional rodent species, with varied taxonomic relationships, to see if there had been a gradual trend of increased amplification directed toward evolution of the rat, or if there had been varied levels of copy number in the rodent order indicative of more stochastic events.

Methods

Specimens. Rat (*Rattus norvegicus*), mouse (*Mus musculus*), guinea pig (*Cavia porcellus*), and hamster (*Mesocricetus auratus*) liver specimens were purchased from Harlan Sprague Dawley, Inc. Deer mice (*Peromyscus maniculatus*), oldfield mice (*P. polionotus*), and white-footed mice (*P. leucopus*) were obtained from the *Peromyscus* Stock Center at the University of South Carolina. The California mouse (*P. californicus*) was donated by Dr. Susan Hoffman, Mammal Division, Museum of Zoology, University of Michigan. Pinyon mouse (*P. truei*) and rock mouse (*P. difficilis*) livers were obtained courtesy of Drs. J. Cook and T.L. Yates, Museum of Southwestern Biology, University of New Mexico. Squirrel (*Sciurus carolinensis*), cotton mouse (*P. gossypinus*), and nutria (*Myocaster coypus*) were wild caught. The nutria was obtained courtesy of Dr. John Doucet. Gerbil (*Meriones unguiculatus*) DNA was donated by Dr. M. Richard Shen.

DNA Isolation. Genomic DNA was isolated from homogenized liver or spleen (*P. californicus*) by proteinase K digestion, phenol/chloroform (Amresco) extractions, and ethanol precipitation (Sambrook et al. 1989). RNA was removed by digestion with pancreatic RNase A (Sigma). DNA was quantitated by UV absorbance of 260 nm.

Library Production. Genomic DNA libraries from squirrel, gerbil, and deer mouse were constructed following partial digestion with *Tsp*5091 (NEB) to an average size of 5 kb, and insertion into the λZAP (Stratagene) vector, followed by transformation into *E. coli* XL1-blue M′R′F′ cells (Stratagene). To determine the average insert size, random clones were selected, followed by in vivo excision of the pBluescript plasmid from the λZAP vector according to the manufacturer's protocol. DNA was isolated from overnight cultures using Wizard mini prep columns (Promega), digested with *Bam*HI and *Hin*dIII restriction enzymes (NEB), and analyzed by electrophoresis on a 1.5% agarose gel.

DNA Quantitation. Copy numbers of ID elements were determined by dot-blot analysis and library screening. One microgram of alkaline-denatured DNA from various species was loaded on a nylon membrane (Hybond: Amersham) in a Schleicher and Schuell dot-blot apparatus, along with plasmid standards calculated to correspond with various copy numbers. The filters were then baked in a vacuum oven at 80°C for 10 min and UV-irradiated in a BIO-RAD cross-linker. DNA probes were radioactively labeled by the random primer method (Feinberg and Vogelstein 1983). Filters were prehybridized in 4× SSC, 0.2% SDS, 0.1 M NaH_2PO_4, 10 mM sodium pyrophosphate, 10× Denhardt's solution, 200 mg/ml denatured salmon sperm DNA, for at least 1 h at 65°C, and hybridized overnight in a similar solution (substituting 1× Denhardt's) at 65°C. Unless noted, filters were washed in 2× SSC, 0.2% SDS, at 65°C, 3 times for 40 min each, and examined by autoradiography with exposure to Kodak XAR film. Filters were stripped in a basic solution (0.2 M NaOH, 0.1× SSC, 0.1% SDS) at 65°C for hybridization with additional probes. ID copy number was also quantitated in squirrel, gerbil, and deer mouse by library screening. Nylon (Hybond: Amersham) filters were hybridized to the rat BC1 gene, obtained by PCR (see below), and washed at 4× SSC, 0.2% SDS, 65°C. Values were determined by calculation using the average insert size of at least ten randomly isolated clones, and the proportion of clones that hybridized to the probe.

Cloning of the BC1 RNA Gene and ID Elements from Different Rodents. The BC1 RNA gene from several rodents was isolated by PCR amplification in a 10-μl volume, with primers (5′-0TTTGGAAGGTATCTCTGATC-3′ and 5′-CCTAAAGGGCAGATAATAAAT-3′) that flank the transcribed portion of the gene (sequences kindly provided by Drs. J.A. Martignetti and J. Brosius), using a Hybaid thermal cycler (block control) under the following conditions: 1× buffer (Promega), 3 mM $MgCL_2$, 200 μM dNTPs, 500 nM each primer, and 1 U Taq polymerase (Promega). PCR products were then ligated into the pCRII cloning vector (Invitrogen) and transformed into *E. coli* using the manufacturer's protocol. DNA was isolated from overnight cultures using Wizard mini prep columns (Promega).

The squirrel BC1 gene could not be amplified with the above primers and was isolated by a modified version of the rapid amplification of cDNA ends (RACE) technique (Frohman et al. 1988). Squirrel brain RNA was isolated (Chomczynski and Sacchi 1987) and size-fractionated (120–200 nt) from a 6% polyacrylamide gel, tailed at the 3′ end with CTP using poly A polymerase (Sigma), and converted to cDNA with the primer 5′-GCCTTCGAATTCAGGT-TGGGGGGGGGGG-3′ (DeChiara and Brosius 1987). cDNA was then amplified in a (Perkin-Elmer 9600) thermal cycler with the primers 5′-GGGGTTGGGGATTT-3′ and 5′-GCCTTCGAATTCAGG-3′ under the following conditions: 94°C, 20 s, 42°C, 20 s, 72°C, 30 s, for 35 cycles.

ID elements were picked from libraries following standard screening procedures (Sambrook et al. 1989) with the BC1 (PCR product) probe (see above) and purified by secondary and tertiary screens. The vector was excised from the phage by *in vivo* excision. (See above).

DNA Sequence Analysis. DNA was sequenced by the dideoxy method (Sanger et al. 1977) using Sequenase (USB) and examined by separation on 5% Long Ranger Gels (AT Biochem) and autoradiography. Alignments were initiated with the Clustal program of PC/GENE (Intelligenetics).

Results

Quantitation of ID Elements in Rodents

ID copy numbers in rodents vary by several orders of magnitude, with the guinea pig, a hystricognath rodent,

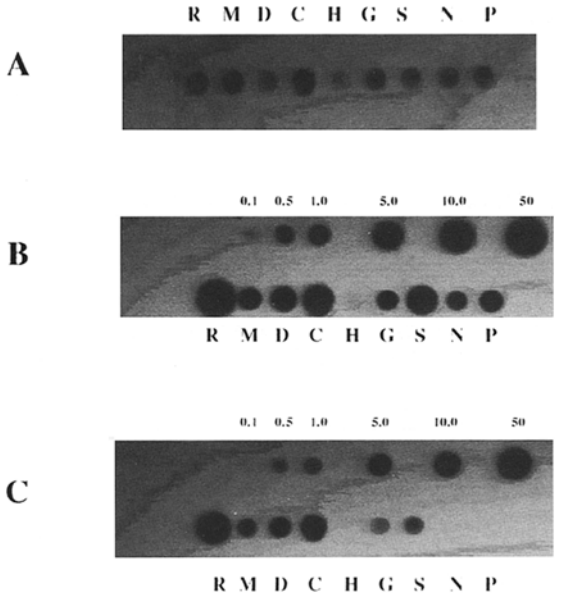

Fig. 1. Copy number determination by dot blot. *A* Hybridization of 1 μg DNA of rodents to the human estrogen receptor gene. *B* Hybridization of the same blot to the BC1 probe (isolated PCR product), washed at 2× SSC, 0.2% SDS, 65°C. **C** Same blot as **B,** washed at 0.5× SSC, 0.2% SDS, 65°C. *R* = rat, *M* = mouse, *H* = hamster, *G* = gerbil, *D* = deer mouse, *C* = California mouse, *S* = squirrel, *N* = nutria, *P* = guinea pig. Standards are based on the number of copies a 75-nt fragment would be observed in 1-μg genomic DNA (haploid equivalent).

representing the lower end at 100–200 copies (Sapienza and St-Jacques 1986). Among myomorphic rodents, the hamster, a cricetid, has 1,000–2,000 copies (Sapienza and St-Jacques 1986), and the murid rodents, mouse and rat, have 10,000–20,000 and 130,000 copies, respectively (Sapienza and St-Jacques 1986). These values are consistent with the relative representations in the rodent GenBank data base (Kim et al. 1994). Lower values were reported using a highly stringent dot blot (Anzai et al. 1987), though rat copy number (40,000) was still 2 orders of magnitude greater than the mouse and the hamster (450 and 150, respectively), and 3 orders of magnitude greater than the guinea pig (50). We therefore examined the copy numbers of other rodents of various phylogenetic relationship. Although the DNA was quantitated before loading the dot blot, as an additional control of relative DNA availability on the blots we hybridized the dot blot to a single-copy gene in the human (estrogen receptor, Oncor) (Fig. 1A). Control hybridization values, standardized by densitometry to the deer mouse, are shown in Table 1. We then hybridized the stripped dot blot to a BC1 probe (Fig. 1B,C). We chose to use a *P. maniculatus* copy number of 25,000 as a standard, because it was confirmed by library screening (see below), and represented an intermediate value. This value, in conjunction with the DNA ratio values (Table 1), was used to determine the relative copy number data in Fig. 1 and Table 1. Estimates of copy number by this means are relatively crude; however, it is obvious that the blot (Fig. 1B,C) shows order-of-magnitude variations

Table 1. Copy number estimates from dot blot (Fig. 1)[a]

Rodent	DNA ratio	2× SSC	0.5× SSC
Guinea pig	0.32	3,000	200
Nutria	0.34	2,000	200
Squirrel	0.80	50,000	10,000
Gerbil	0.56	3,000	3,000
Hamster	2.01	1,000	1,000
California mouse	0.22	27,000	37,000
Deer mouse	1.00	25,000	25,000
Mouse	0.34	4,000	4,000
Rat	0.37	140,000	150,000

[a] DNA ratios are based on the inverse of proportional densitometric values to *P. maniculatus* (1.0) with the estrogen receptor control. Determined values are based on a constant factor that yields 25,000 copies for *P. maniculatus* when multiplied by a subjectively obtained value from the dot blot. Values are rounded off to the thousand or to the hundred (guinea pig and nutria).

in ID copy number. In the myomorphic rodents (rat, mouse, deer mouse, California mouse, hamster, gerbil), the copy number remains fairly consistent irrespective of the stringency of washing. In several trials, our value for the mouse is intermediate to previous reports (Sapienza and St-Jacques 1986; Anzai et al. 1987). Our values for hamster, rat, and guinea pig are fairly consistent with that of Sapienza and St-Jacques (1986) and Kim et al. (1994), even under relatively high stringency (Table 1). The nutria showed levels on the same order of magnitude as the guinea pig, suggesting ID amplification never reached high levels in hystricognath rodents, though other species along this taxonomic branch would need to be examined, as extensive variation is observed between more closely related species in the myomorph branch. This is obvious not only between the rat and the mouse, but also between the hamster and the *Peromyscus* species, as greater than a full order of magnitude of ID amplification has occurred since the 32-million-year split (Sarich 1985) between these two branches. ID copy number in the gerbil, though, is slightly higher than the hamster, but is still approximately an order of magnitude less than *Peromyscus.* The value for the squirrel, the only sciurid rodent assessed in this study, is approximately 60,000, though this value is greatly reduced under more stringent conditions, suggesting either that the upper value represents ID sequences divergent from myomorph sequences or that the squirrel contains another SINE family with sequence similarity to ID repeats. In either case, the squirrel presents a value greater than hystricognaths and is intermediate among the myomorphs. There was also a large drop in copy number of the guinea pig and the nutria with stringent washes, suggestive of divergent ID members or ID-like families.

To aid in corroborating our copy number values, as well as in determining sequences of individual ID elements, we produced DNA libraries from squirrel, gerbil, and *P. maniculatus* genomic DNA, determined the size of random clones, and screened with a BC1 PCR product

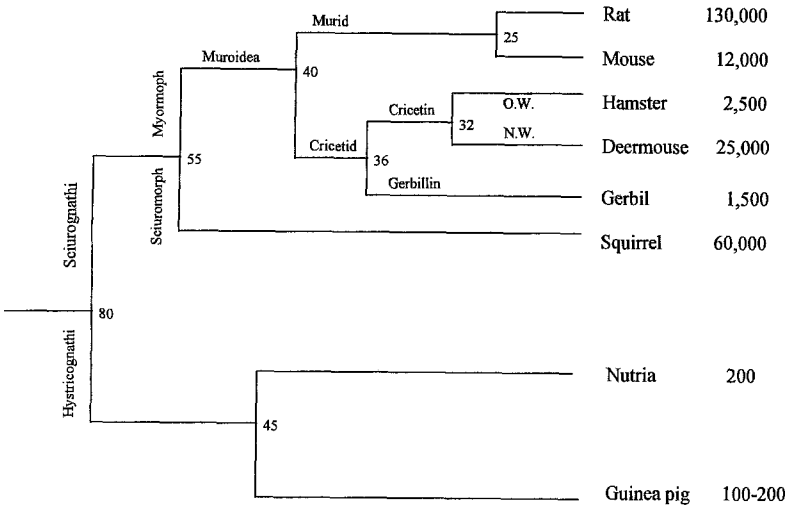

Fig. 2. Dendrogram of rodent phylogeny with ID copy numbers. Estimates for rat, mouse, hamster, and guinea pig are based on Sapienza and St-Jacques (1986). Deer mouse, gerbil, and squirrel estimates are based on library screens from this study, nutria copy number is based on a dot-blot estimate (Fig. 1, Table 1). *Numbers at nodes* refer to times of evolutionary separation in millions of years, based on molecular clock data (Sarich 1985; O'hUigin and Li 1992), and DNA/DNA hybridizations (Brownell 1983). O.W. = old world, N.W. = new world.

as a hybridization probe. Under moderate washing stringency (4× SSC, 65°C) we observed copy number values consistent with the dot blots. These values were determined by estimating the proportion of plaques that hybridized to BC1 and the average insert size of at least ten random clones isolated from this library. For example, 2.7% of the plaques from the *P. maniculatus* library hybridized to BC1, with an average insert size of 3.3 kb, which corresponds to about 25,000 copies per haploid genome. These values are compared with previously determined values, and with the phylogenetic relationships of the rodents (Fig. 2).

BC1 Sequence Analysis

We set out to determine if a relationship exists between the sequence of the BC1 genes of the various rodents and the level of amplification. We isolated the BC1 RNA gene from the gerbil and several species of *Peromyscus* by PCR amplification using primers from the unique flanking sequences. The squirrel BC1 RNA gene could not be amplified by these primers, almost certainly due to sequence divergence around the primer sites, and therefore a cDNA clone was obtained from RNA from the brain by a modified RACE technique. The rat, hamster, and mouse sequences were from Martignetti (1992), and the guinea pig sequences were from Kim et al. (1994). We focused our analysis on regions of the sequence that may have an impact on retroposition events, primarily at the transcriptional level. No variations were observed in the proximal sequence element (PSE), TATA box, A box, and B box regulatory sites in the myomorphic rodents (Fig. 3). Additionally, among the Sciurognathi rodents, there was no correlation of size or maintenance of the A-rich region, which has been proposed to have an effect on reverse transcription (Kim et al. 1994) with copy number. The guinea pig was highly variable in the 5′ region, including the PSE site, and had a highly punctuated A-rich region, which may help to explain the low copy number in this species. The ID core region (the region integrated into new sites by retroposition forming new ID elements) was highly conserved among the myomorphs. Variations among the myomorphs within the ID core sequence were mainly observed in the 3′ end. In relation to a myomorph consensus sequence, the hamster had a variant at position 72, the rat at positions 74 and 76, and *P. maniculatus* at positions 48, 76, and 77. The three *P. maniculatus* variants were observed to be subgeneric-specific variations (Figs. 3, 4), and an additional variant was found within members of the *P. truei* species group (position 69; data not shown). In *P. difficilis,* though, we also observed a BC1 (ID core) sequence identical to the myomorph consensus, which could be due to either allelic forms of BC1 or to a duplicated BC1 gene. The finding that *P. californicus* has two forms (variants at position 72) is suggestive of allelic variation within a species. It is possible that both forms existed prior to the hamster/*Peromyscus* split, since the two forms in *P. californicus* are either a myomorph consensus form or a hamster-like form. If the two allelic forms were maintained throughout evolution of *Peromyscus,* it would be expected that a certain proportion of individual elements in *P. maniculatus* would contain a C at position 72, which was not observed in any element. (See below.) The two guinea pig sequences, which likely arose as a duplication of one from the other (Kim et al. 1994), show several nucleotide variations from the myomorph rodents but are much more closely related to each other.

ID Sequence Analysis

We isolated and sequenced several ID elements from the squirrel, gerbil, and deer mouse libraries to ascertain that these clones contain ID elements and to analyze their relationship to the BC1 RNA genes (Fig. 5). Twelve out of 15 clones in *P. maniculatus* sequenced effectively using primers designed from the ID consensus sequence. This confirms that the vast majority of the clones de-

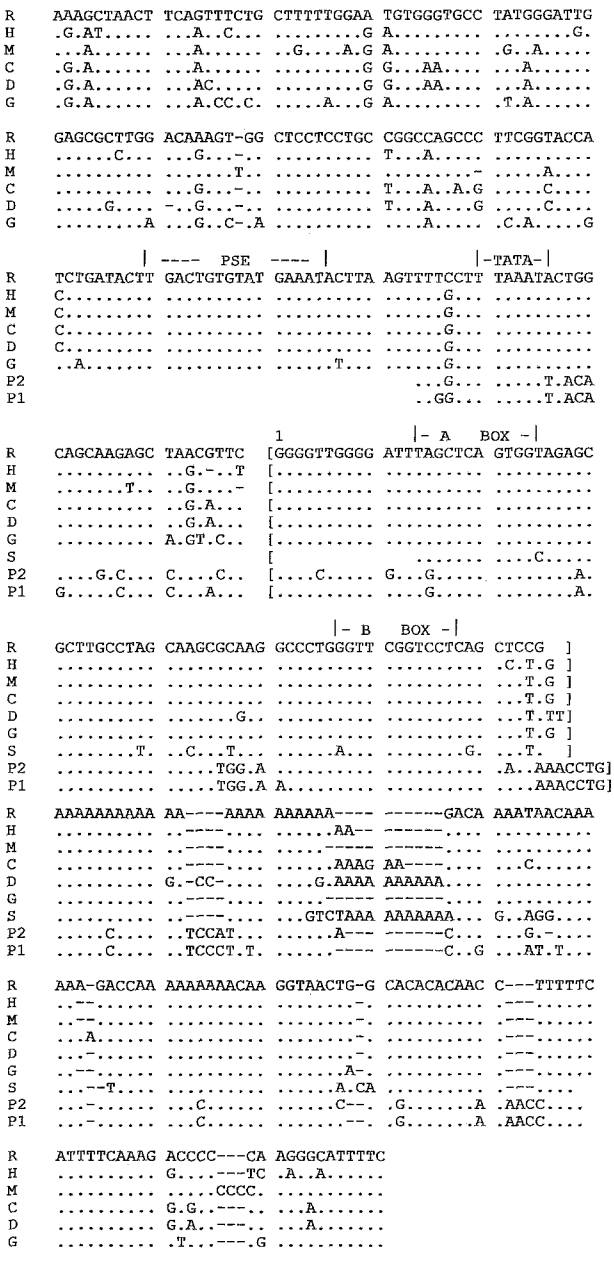

Fig. 3. Alignment of BC1 genes from several rodent species. Squirrel sequences at the 5′ and 3′ ends are not known. Guinea pig sequences at the 5′ and 3′ ends were omitted due to the high level of divergence. The ID core sequence is enclosed in *brackets,* with the first nucleotide labeled *1*. Regulatory regions are indicated above the sequence. *R* = rat, *H* = hamster, *M* = mouse, *C* = California mouse, *D* = deer mouse, *G* = gerbil, *S* = squirrel, *P1* and *P2* = guinea pig. *Dots* refer to sequences identical to the rat; *dashes* indicate gaps. Alignments were based on the clustal program of PC/GENE, with some subjective realigning.

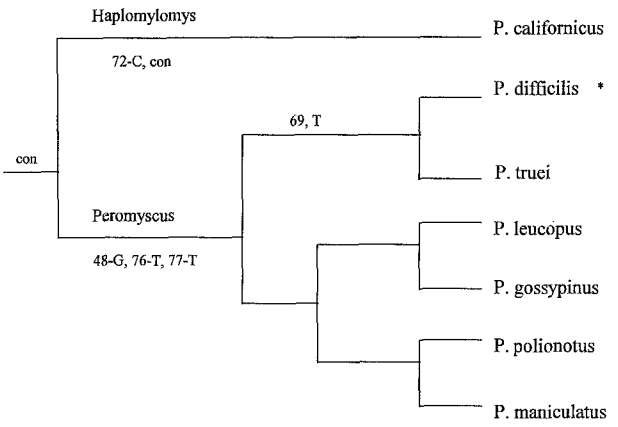

Fig. 4. Evolution of the BC1-ID core sequence in *Peromyscus.* Numbers refer to position in the ID core sequence. (No. 1 is indicated in Fig. 3). *Con* refers to the myomorph ID core consensus sequence (see Fig. 5). The *asterisk* indicates that *P. difficilis* contains the consensus form and the form with the four acquired variants. The nodes furthest to the right represent the taxonomic species-group level. The node furthest to the left represents the break point of the two subgenera. Phylogenetic relationships are based on accumulated data (Carleton 1989).

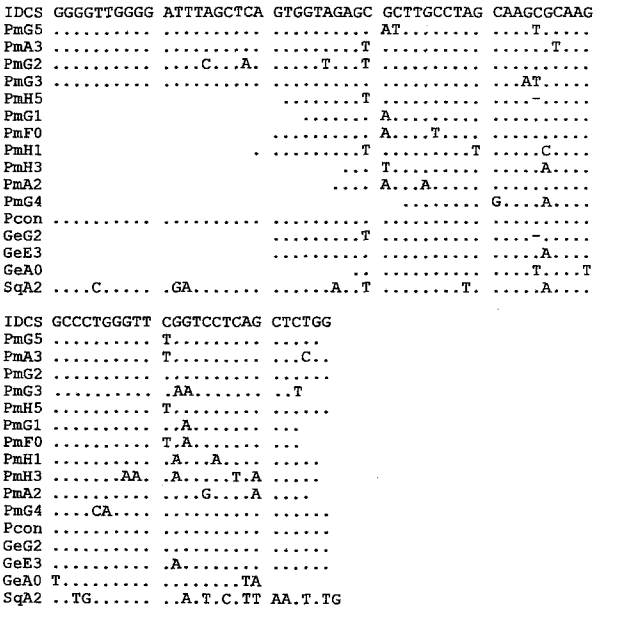

Fig. 5. Alignment of individual ID elements isolated from genomic libraries. *IDCS* refers to the myomorph consensus sequence; *Pcon* refers to the consensus sequence from the *P. maniculatus* clones. All but the squirrel were sequenced with 5′ and 3′ consensus oligonucleotide primers. Sequence utilizing the 3′ primer was only attainable for the first four *Peromyscus* clones. The squirrel sequence was obtained using vector primers. *Pm* = deer mouse, *Ge* = gerbil, *Sq* = squirrel followed by a letter and a number simply to designate a particular clone. Spaces at the 5′ end refer to unattainable sequence; *dots* refer to identity to IDCS; *dashes* refer to gaps. Sequences are shown prior to the beginning of the A-tails.

tected in the library screening represented authentic ID copies. In *P. maniculatus,* we observed that the 12 sequenced clones did not contain any of the most recent changes represented in the current BC1 gene (positions 48, 76, and 77). This is perhaps not surprising in that *P. californicus* and *P. maniculatus* have similar copy numbers and have only recently diverged from one another, within the last 2 million years (Hibbard 1968). Therefore most IDs were probably made prior to the introduction of these diagnostic changes in the BC1 locus. This is further confirmed by the high level of divergence of these elements relative to the consensus. They diverge at about 3.5% of the non-CpG positions and over two-thirds of the CpG positions have mutated. This is significantly higher divergence than for rat ID repeats (J.K. and P.D.,

unpublished), for instance. It would be of interest to find IDs with the subgeneric-specific variants to demonstrate that amplification from the locus has been occurring since the *Peromyscus-Haplomylomys* subgeneric split.

Sequences were obtained from three gerbil ID elements and their consensus agreed with that of the gerbil BC1 gene. The clones from the squirrel library could not be sequenced by the ID core sequence primers, probably due to nucleotide divergence around the primer sites. One fragment was small enough to be sequenced by flanking vector primers and was found to contain several mutations in relation to the squirrel BC1 gene. The squirrel BC1 contains seven variants from the myomorph consensus, and only one of these was present out of the five that were covered in the sequenced region of the ID element.

Discussion

Conservation of the BC1 RNA Gene

The BC1 RNA gene is found only within the genomes of rodents (Martignetti and Brosius 1993) and was suggested to have been ancestrally derived from a $tRNA^{Ala}$ (Daniels and Deininger 1985). This single locus is likely to be the original copy of the ID SINE family (Kim et al. 1994). Previous analysis of the BC1 gene suggested that, unlike the vast majority of other SINE loci which mutate at a neutral rate, the BC1 locus was well conserved. Our sequence analysis of the BC1 locus in other rodents confirms this conservation. We have confirmed that the BC1 RNA is expressed in *P. maniculatus* and *P. californicus* (data not shown), with a slight size difference consistent with the BC1 sequences presented for those species (Fig. 3). The expressed regions of these sequences are highly conserved. There is only one change (excluding the two to three nucleotides preceding the A-tail, which seem subject to change) in the ID-related portion relative to the myomorphic consensus in *P. maniculatus* (position 48) and none in *P. californicus* or in the gerbil. There are a few changes in the A-rich region of the BC1 RNA gene, but overall the conservation is high relative to the flanking sequences. There are a few more changes in the squirrel, showing that it is clearly intermediate in divergence relative to the Hystricognathi. However, all of the changes observed in any of the myomorphic rodents result in conservative changes relative to the proposed BC1 RNA structure (Deininger et al. 1995). Thus, these findings continue to be consistent with the suggestion that this tRNA-derived RNA sequence has now "exapted" (Brosius and Gould 1992) to a new function that places selective pressure on the gene sequence.

The BC1 RNA Gene as a Master Gene

Previous evolutionary analyses of *Alu* elements have demonstrated the sequential amplification of distinct subfamilies from a limited number of master loci (Shen et al. 1991; Deininger et al. 1992; Deininger and Batzer 1993). Analysis of the BC1 RNA gene locus and ID repeats in rat, mouse, and guinea pig demonstrated that the BC1 RNA locus has served as a master gene for ID family amplification and evolution (Kim et al. 1994). The evidence for this was that the ID repeats formed in the various rodent species followed the same pattern of evolutionary change as the BC1 RNA gene locus. Our findings in other rodent species are consistent with this previous analysis. The gerbil sequences continue to match well with the BC1 RNA sequence. The one squirrel ID sequence shows one of the new squirrel BC1 changes, but not the others. However, this is easily explained as a copy that was generated at an intermediate point in the evolution of the squirrel BC1 RNA gene. These ID copies can actually represent fossils of the various evolutionary intermediates that are formed during the evolution of the BC1 RNA gene (Deininger et al. 1995). This is further confirmed by the finding of four woodchuck (sciurid) ID sequences from the data base (data not shown), of which most showed the majority of the squirrel BC1 sequence changes. This suggests that the BC1 RNA gene has continued to be a master gene for at least a significant portion of sciurid ID amplifications.

The Sporadic Nature of ID Amplification

Our data demonstrate that rather than the originally observed trend of increasing copy number in species more closely related to the rat, there is a sporadic pattern of amplification in different rodent lineages. Most striking are the high amplification rates in the deer mouse relative to the hamster (both cricetin, Fig. 2) and in the squirrel. These amplifications are easiest to explain by specific stochastic events occurring in those lineages that increased the amplification rate. Other models, such as the specific removal of elements in some rodent lineages, are inconsistent with studies of orthologous SINE loci in primates, which demonstrate that there is no apparent mechanism for actively removing SINE elements (Sawada et al. 1985; Sawada and Schmid 1986; Shen et al. 1991; Bailey and Shen 1993).

In the rat, the subfamily structure suggests that a major factor in increasing the amplification rate was the formation of one or more highly active master genes (Kim et al. 1994). This may also be the explanation for the higher copy numbers in species such as the deer mouse and the squirrel. The probability of forming a new SINE master gene, possibly either by gene duplication or retroposition into a new sequence environment that is highly effective at amplifying, seems to be quite low, and the occurrence of one or two fortuitous events may be sufficient to initiate a rapid expansion.

One alternative to the formation of new master genes as a cause of the expansions in some species would be

fortuitous changes in the BCI RNA gene which led to a greatly increased amplification capability from that site. This could be the result of mutations that lead to higher expression levels. Although we do not see any significantly higher expression levels of BC1 RNA in the organisms with higher ID copy numbers, it may be that the expression is only higher in some limited germ cell type that we cannot identify. Alternatively, changes in BC1 RNA structure could also influence its interaction with the retroposition process. One good possibility for this would be the ability of BC1 RNA to self-prime its own reverse transcription (Shen et al. unpublished; Deininger et al. 1995). There are no obvious changes in the 3′ unique region of the BC1 RNA that could account for the copy number variation seen in the Sciurognathi. However, in the guinea pig BC1 RNA, the A-rich region is significantly less homogeneous. The A-rich region is an almost universal feature of SINEs and almost certainly contributes to some extent to the amplification mechanism.

The changes in amplification rate could also be the result of stochastic changes occurring unrelated to ID master genes. Examples of this type of event might be changes in availability of reverse transcriptase, either through explosive increases in active LINE amplification during specific periods of rodent evolution (Casavant and Hardies, 1994; Furano et al. 1994) or through specific introduction of exogenous reverse transcriptase through horizontal transmission, a mechanism suggested for the evolution of several transposable elements in *Drosophila* (Capy et al. 1994). Alternatively, adenovirus infection has been shown to stimulate RNA polymerase III transcription of *Alu* elements in HeLa cells (Panning and Smiley 1993), and may suggest a contribution of different types of viral infections in the course of rodent evolution to ID copy number variations.

Is the tremendous variation of SINE copy number seen for the ID repeats in rodent evolution unusual? The primate *Alu* family shows little copy number variation throughout the primates. However, this is largely explained by the amplification of the vast majority of SINEs prior to the major primate radiation. Most *Alu* family members inserted prior to the divergence of the various monkey species, and *Alu* amplification decreased tremendously following those divergences (Shen et al. 1991). Thus, the current *Alu* amplification rate in primates is too low to have a major impact on copy number relative to the bulk of *Alu* made early in primate evolution. Thus, *Alu* also demonstrates tremendous differences in amplification rates at different times, but because *Alu* amplification was more active at an early time, recent differences are less obvious. However, a small subset (HS/PV subfamily) of *Alu* elements, specific to the human/great ape clade of primates, recently amplified to different extents between the gorilla (2 copies; Leeflang et al. 1993b), the chimpanzee (10–100; Leeflang et al. 1993a), and the human (500–2,000; Batzer and Deininger 1991). Many ID repeats have also amplified recently, as evidenced both by the species specificity of the specific ID subfamily variants (Kim et al. 1994) and the relative homogeneity of the ID sequences to each other. It seems likely that functional selection on the BC1 RNA gene helps maintain its activity, which in turn provides a stable master gene for ID amplifications, which may be strongly activated to generate an even more effective master copy on rare occasions.

Ackowledgments. This work was supported by a grant from the Cancer Crusaders and the Stanley S. Scott Cancer Center (D.H.K.) and by Public Health Service grant GM 45668 (P.L.D.).

References

Anzai K, Kobayashi S, Suehiro Y, Goto S (1987) Conservation of the ID sequence and its expression as small RNA in rodent brains: analysis with cDNA from mouse brain-specific small RNA. Mol Brain Res 2:43–49

Bailey AD, Shen C-KJ (1993) Sequential insertion of *Alu* family repeats into specific genomic sites of higher primates. Proc Natl Acad Sci USA 90:7205–7209

Batzer MA, Deininger PL (1991) A human-specific subfamily of *Alu* sequences. Genomics 9:481–487

Brosius J, Gould SJ (1992) On genomenclature: A comprehensive (and respectful) taxonomy for pseudogenes and other 'junk DNA.' Proc Natl Acad Sci USA 89:10706–10710

Brownell E (1983) DNA/DNA hybridization studies of muroid rodents: symmetry and rates of molecular evolution. Evolution 37:1034–1051

Capy P, Anxolabehere D, Langin T (1994) The strange phylogenies of transposable elements: are horizontal transfers the only explanation? Trends Genet 10:7–12

Carleton MD (1989) Systematics and evolution. In: Kirkland GL Jr, Layne JN (eds) Advances in the study of *Peromyscus* (Rodentia). Texas Tech University Press, Lubbock, pp 7–142

Casavant NC, Hardies SC (1994) The dynamics of murine LINE-1 subfamily amplification. J Mol Biol 241:390–397

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid-guanidinium thiocyanate-pheno-chloroform extraction. Anal Biochem 162:156–159

Daniels GR, Deininger PL (1985) Repeat sequence families derived from mammalian tRNA genes. Nature 317:819–822

DeChiara TM, Brosius J (1987) Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. Proc Natl Acad Sci USA 84:2624–2628.

Deininger PL (1989) SINEs: short interspersed repeated DNA elements in higher eucaryotes. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington, DC, pp 619–636

Deininger PL, Batzer MA (1993) Evolution of retroposons. Evol Biol 27:157–196

Deininger PL, Batzer MA, Hutchison III CA, Edgell MH (1992) Master genes in mammalian repetitive DNA amplification. Trends Genet 8:307–311

Deininger PL, Tiedge H, Kim J, Brosius J (1995) The BC1 RNA gene as a master gene for ID amplification: evolution, expression and possible function. In: Cohn WE, Moldave K (eds) Progress in nucleic acid research and molecular biology, Academic Press, New York (in press)

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-

length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA 85: 8998–9002

Furano AV, Hayward BE, Chevret P, Catzeflis F, Usdin K (1994) Amplification of the ancient murine Lx family of long interspersed repeated DNA occurred during the murine radiation. J Mol Evol 38:18–27

Hibbard CW (1968) Palaeontology. In: King JA (ed) Biology of *Peromyscus* (Rodentia). Spec Publ 2, American Society of Mammalogists, pp 6–26

Jagadeeswaran P, Forget BG, Weissman SM (1981) Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts? Cell 26:141–142

Kim J, Martignetti JA, Shen MR, Brosius J, Deininger PL (1994) Rodent BC1 RNA gene as a master gene for ID element amplification. Proc Natl Acad Sci USA 91:3607–3611

Leeflang EP, Chesnokov IN, Schmid CW (1993a) Mobility of short interspersed repeats within the chimpanzee lineage. J Mol Evol 37:566–572

Leeflang EP, Liu W-M, Chesnokov IN, Schmid CW (1993b) Phylogenetic isolation of a human *Alu* founder gene: drift to new subfamily identity. J Mol Evol 37:559–565

Martignetti JA (1992) PhD thesis. BC1 and BC200 RNA: sequence and transcriptional analysis. The Mount Sinai School of Medicine, City University of New York, New York

Martignetti JA, Brosius J (1993) Neural BC1 RNA as an evolutionary marker: Guinea pig remains a rodent. Proc Natl Acad Sci USA 90:9698–9702

O'hUigin C, Li W-H (1992) The molecular clock ticks regularly in muroid rodents and hamsters. J Mol Evol 35:377–384

Panning B, Smiley JR (1993) Activation of RNA polymerase III transcription of human *Alu* repetitive elements by adenovirus type 5: requirement for the E1b 58-kilodalton protein and the products of E4 open reading frames 3 and 6. Mol Cell Biol 13:3231–3244

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sapienza C, St-Jacques B (1986) "Brain-specific" transcription and evolution of the identifier sequence. Nature 319:418–420

Sarich VM (1985) Rodent macromolecular systematics. Evolutionary relationships among rodents. Plenum Press, New York, pp 423–452

Sawada I, Schmid CW (1986) Primate evolution of the a-globin gene cluster and its *Alu*-like repeats. J Mol Biol 192:693–703

Sawada I, Willard C, Shen C-KJ, Chapman B, Wilson AC, Schmid CW (1985) Evolution of *Alu* family repeats since the divergence of human and chimpanzee. J Mol Evol 22:316–322.

Shen MR, Batzer MA, Deininger PL (1991) Evolution of the master *Alu* gene(s). J Mol Evol 33:311–320