Research

Long terminal repeat retrotransposons of *Mus musculus*

Eugene M McCarthy and John F McDonald

Address: Genetics Department, University of Georgia, Athens, GA 30602, USA.

Correspondence: Eugene M McCarthy. E-mail: mccarthy@uga.edu

Published: 13 February 2004

Genome ***Biology*** 2004, **5:**R14

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2004/5/3/R14

Received: 4 September 2003
Revised: 12 November 2003
Accepted: 9 January 2004

Abstract

Background: Long terminal repeat (LTR) retrotransposons make up a large fraction of the typical mammalian genome. They comprise about 8% of the human genome and approximately 10% of the mouse genome. On account of their abundance, LTR retrotransposons are believed to hold major significance for genome structure and function. Recent advances in genome sequencing of a variety of model organisms has provided an unprecedented opportunity to evaluate better the diversity of LTR retrotransposons resident in eukaryotic genomes.

Results: Using a new data-mining program, LTR_STRUC, in conjunction with conventional techniques, we have mined the GenBank mouse (*Mus musculus*) database and the more complete Ensembl mouse dataset for LTR retrotransposons. We report here that the *M. musculus* genome contains at least 21 separate families of LTR retrotransposons; 13 of these families are described here for the first time.

Conclusions: All families of mouse LTR retrotransposons are members of the *gypsy*-like superfamily of retroviral-like elements. Several different families of unrelated non-autonomous elements were identified, suggesting that the evolution of non-autonomy may be a common event. High sequence similarity between several LTR retrotransposons identified in this study and those found in distantly related species suggests that horizontal transfer has been a significant factor in the evolution of mouse LTR retrotransposons.

Background

Retrotransposons are mobile genetic elements that make up a large fraction of most eukaryotic genomes. All retrotransposons are distinguished by a life cycle involving an RNA intermediate. The RNA genome of a retroelement is copied into a double-stranded DNA molecule by reverse transcriptase, which is subsequently integrated into the host's genome. Retrotransposons fall into two main categories: those with long terminal repeats (LTRs), such as retroviruses and LTR retrotransposons, and those that lack such repeats, for example, long interspersed nuclear elements (LINEs).

Retrotransposons are particularly abundant in plants, where they are often a principal component of nuclear DNA. In corn, 50-80%, and in wheat fully 90%, of the genome is made up of retrotransposons [1,2]. This percentage is generally lower in animals than in plants but it can still be significant. For example, about 8% of the human genome is now known to be composed of LTR retrotransposons [3]. In the mouse genome this figure has been estimated at 10% [4].

This article presents the results of a recent survey (December 2002) of the GenBank mouse (*M. musculus*) database

(GBMD) and the 2.9 Gbp Ensembl [5] mouse dataset (EMD) for the presence of LTR retrotransposons. We have employed a new search program, LTR_STRUC (LTR retrotransposon structure program), as the initial data-mining tool in our survey [6]. Identified elements were subjected to sequence analyses to identify open reading frames (ORFs) encoding reverse transcriptase (RT) and other retroviral proteins. LTR_STRUC finds only full-length elements, that is, ones having two LTRs and a pair of target site duplications (TSDs). We therefore augmented our search approach by conducting BLAST searches using reverse transcriptase queries. These queries are of two types: previously known RTs in the public database from mouse and other mammals, and RTs obtained from our initial scan of the EMD with LTR_STRUC. Subsequent RT sequence alignments were carried out, followed by construction of phylogenetic trees.

An LTR retrotransposon 'family' is defined as a group of elements with RTs at least 90% similar at the amino acid level [7]. Experience has shown that when two elements have RTs that are 90% similar, their LTRs are typically about 60% similar. Thus, non-autonomous elements, lacking an RT ORF, are assigned to the same family if their LTRs are at least 60% similar. Many LTR retrotransposons replicate non-autonomously. Four different families of murine LTR retrotransposons have non-autonomous members. (*MalR* elements, *ETn* elements, *VL30* elements and a new type identified in this study, related to *IAP* elements). These non-autonomous elements are discussed below. Non-autonomous elements can reach a high copy number even though they lack an RT ORF [4,8-11].

Currently there is no standard mouse retrotransposon nomenclature. In our system of classification for mouse, LTR retrotransposons are specified by the acronym *Mmr* (*M. musculus* retrotransposon). Distinct families are indicated by number (for example, *Mmr1*, *Mmr2*, *Mmr3*). We have chosen to adopt the *Mmr* nomenclature in this study because it is consistent with the systematic logic ('Mm' indicative of the genus and species of the host organism; 'r' indicates retrotransposon) used in previous articles [8,12]. In each case where we use the *Mmr* acronym in this article to refer to a previously named family, we also include any pre-existing name for the family.

Results and discussion

RTs from elements identified in our survey fall into numerous distinct families. All autonomous LTR retrotransposons identified were of the *gypsy*-like elements (Classes I, II, and III). Autonomous retroviral-like elements in the mouse genome usually have an overall length of between 6,000 and 9,000 bp. Results of our study indicate that the TSDs of mouse LTR retrotransposons are four to six base pairs long and that within each of the three major classes of these elements a single TSD length is characteristic (see below). With the exception of a few mutated copies, mouse LTR retrotransposons seem to have the same canonical dinucleotides terminating the LTRs as are typically found in other species (TG/CA). The LTRs of murine retroviral-like elements are generally 300-600 bp long, with the exception of mouse mammary tumor virus (MMTV) where the LTRs are some 1,300 bp in length. Our survey shows that at least 21 distinct LTR retrotransposon families exist in the mouse genome, 13 of which have not been described previously.

LTR retrotransposon families of the murine genome

Overview

To date, LTR retrotransposon diversity has been rigorously classified into families for only a few organisms (for example, *Oryza sativa* [8], *Drosophila melanogaster* [7] and *Caenorhaditis elegans* [12]). This article represents a first attempt to establish a similar uniform classification and nomenclature for the domestic mouse. Previous studies have classified murine retrotransposons into broad categories only, which ignore the standard definition of 'family' (see above). For example, the term 'intracisternal type A particle' (IAP) has been used to refer to elements that belong to several distinct LTR-retrotransposon phylogenetic groups. The autonomous elements identified in our survey of the GBMD and EMD fall into 20 families on the basis of degree of RT divergence (greater than 10% denotes family). In addition, we have classified *MalR* elements, which are non-autonomous, into a twenty-first family that is closely related to *MuERV-L* elements, because these two types of transposons have similar LTRs. *MusD* and *ETn* elements form a second pair of related autonomous and non-autonomous elements; *MmERV* and *VL30* elements constitute a third. These three paired families are discussed in more detail below.

Our analysis supports previous categorization [4] of mouse LTR retrotransposons into three distinct classes (Figure 1): Class I, containing elements related to retroviral leukemia viruses in mouse (*MuLV*) and other species (for example, gibbon: *GALV* and cat: *FeLV*); Class II which contains the *IAP* elements, mouse mammary tumor virus (*MMTV*) and the *MusD2*/*ETn* family; and Class III which comprises the *MalR* and *MuERV-L* elements. In using these names for the three main categories of murine LTR retrotransposons we follow the usage of the Mouse Genome Sequencing Consortium [4], but the reader is cautioned that the same terminology has been used to designate RNA-based transposons (Class I) and DNA-based transposons (Class II). Here, however, all three classes are RNA-based LTR retrotransposons.

Class I (families 1-4)

Members of this class make up 0.68% of the mouse genome (copy number about 34,000) [4]. They have 4-bp TSDs and are related to murine leukemia virus (*MuLV;* AF033811), a C-type retrovirus that occurs only in mice and is a major cause of cancer in that genus. Class I, to which *MuLV* belongs, contains at least three other families: *Mmr1_MmERV,*

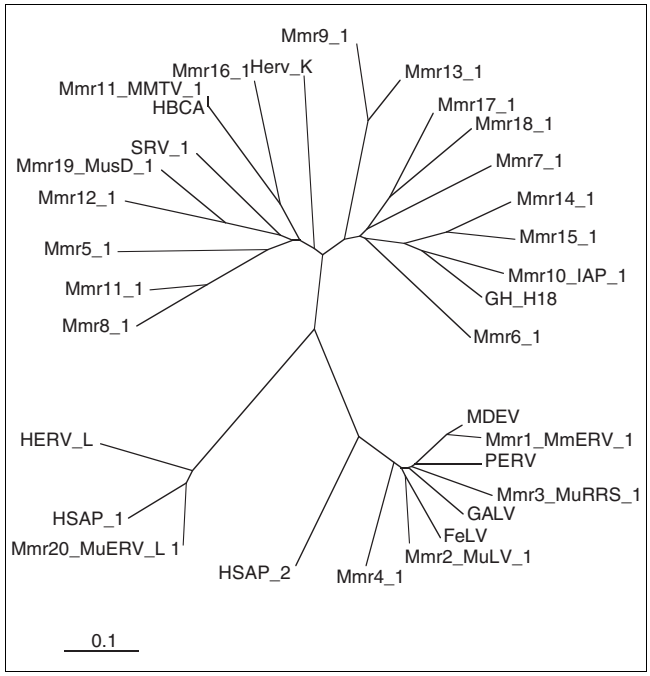

Figure 1
Unrooted RT-based neighbor-joining tree for all three classes of murine retrotransposons. RT sequences from host species other than mouse are included for comparison.

Mmr3_MuRRS, and *Mmr4*. In this article, *MuLV* is referred to as family *Mmr2_MuLV*. Class I endogenous retroviruses are more closely related to elements in other species than to mouse retroelements belonging to Classes II or III. RTs from endogenous retrovirus in pig (*PK15;* AF038601) and koala (*KoRV;* AAF15098), as well as from leukemia viruses in gibbon (*GALV;* AAA466810) and cat (*FeLV;* L06140), group with this class; their RTs are all about 80% similar at the amino acid level to those of murine Class I elements. One member of Class I is found in two different mouse species, *M. musculus* and *M. dunni*, and has previously been referred to as either *MmERV* (in *M. musculus*) or *MDEV* (in *M. dunni*) [13]; here it is referred to as *Mmr1_MmERV*. The identity of this family in these two species is demonstrated by the presence of an element (AAC31805) in the *M. dunni* (Indian pigmy mouse) genome, which is 96% similar (at the amino acid level) to members of *Mmr1_MmERV* resident in *M. musculus* (Figure 2). This finding is consistent either with a recent common origin of these two mouse species or with a horizontal transfer of this retrovirus. This virus may be infectious since an envelope protein sequence is present in the GenBank database (AAC31806) for the *M. dunni* retrovirus and has also been detected in copies of this family during our own survey of *M. musculus*. *Mmr4* is a previously unrecognized Class I family, with members about 80% similar to those of *Mmr2_MuLV*. Family *Mmr3_MuRRS* includes the so-called murine retroviral related sequences (*MuRRS)*. A known human endogenous retrovirus type C oncoviral sequence (AAA73090) is approximately 56% similar at the amino acid level to members of Class I. BLAST searches with RT queries from Class I indicate that at least some elements in the human genome are even more similar (>65%) to Class I elements in mouse (for example, *HSAP-2*; Figure 2 and Table 1).

Class II (families 5-19)

Class II retroviral-like elements make up 3.14% of the mouse genome (copy number approximately 127,000) [4]. This class contains 15 of the 21 murine LTR families. Its members have 6 bp TSDs and are related to MMTV (NC_001503), an oncogenic B-type retrovirus that causes breast cancer in mice. Our survey has revealed only three full-length copies of a member of this family (*Mmr11_MMTV*) in the mouse genome. *MMTV* contains an ORF coding for envelope protein (BAA03768). *Mmr11_MMTV* RTs are also 75% similar to those of a separate endogenous mouse family, *Mmr16*. For the most part, *Mmr16* seems to be represented in the mouse genome by fragmentary elements, but the full-length element *Mmr16-1* described in Table 2 has a full complement of retroviral genes, including an envelope ORF, as is the case with *MMTV*.

Another family in Class II, *Mmr19_MusD*, has been previously described under the name *MusD*. Mager and Freeman [9] who discovered this family, showed that the non-autonomous mouse *ETn* retroelements (early transposons) are deletion derivatives of *Mmr19_MusD*. They are so closely related to *MusD* elements that we have assigned them to the same family. Most copies of the former are around 5,500 bp long, while those of the latter are usually around 7,400 bp in length. *ETn* elements (Y17107; AB033509), first reported by Brulet *et al*. [14], are a moderately repetitive family of murine retrotransposons that lack most of the usual retroviral ORFs. Our survey with LTR_STRUC suggests that full-length copies of *ETn* elements are about half as common again as full-length *MusD* elements. Family *Mmr12* is about 80% similar to *Mmr19_MusD*. Both of these families are 70% similar to Mason-Pfizer Monkey Virus (MPMV; NC_001550). The RTs of *MusD* elements have an unusual active site sequence: FTDDVLM ('T' is not canonical for an active site) [14]. Class II contains an additional clade (See Figure 3), comprising at least eight additional families (*Mmr6, Mmr7, Mmr9, Mmr10_IAP, Mmr14, Mmr15, Mmr17*, and *Mmr18*) with no two families differing from any other by more than 70%. The major constituents of this clade are the IAP retrotransposons, the second most abundant family in the mouse genome, here referred to as family *Mmr10_IAP*. They lack complete *env* genes [15] and thus are considered non-infective. Murine elements identified in GenBank as IAP (for example, GNPSIP and GNMSIA) are restricted to family *Mmr10_IAP*. Nevertheless, members of any of the eight families listed above have been described as IAP by various authors. In addition, a family of retroelements in golden hamster (*GH-G18* }; Figure 3) have been described as 'IAP' but do not actually belong to the *Mmr10_IAP* family (their RT ORFs differ from those of

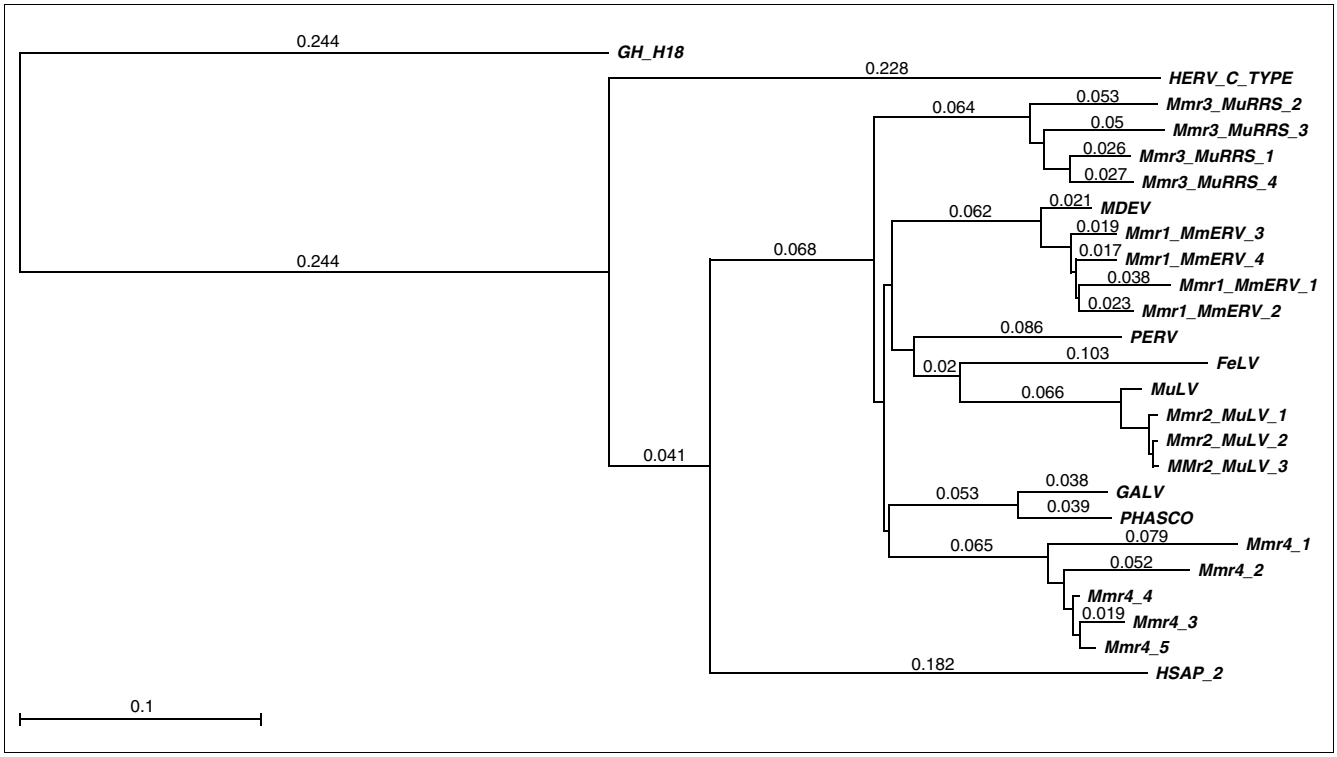

Figure 2
RT-based neighbor-joining tree for Class I murine retrotransposons. The distances (uncorrected 'p') appear next to each of the branches. RT sequences from host species other than mouse are included for comparison. The outgroup is the Class II element *GH-H18* (from golden hamster, *Mesocricetus auratus*; see Table 3 and Figure 3).

Mmr10_IAP by about 18% at the amino acid level). Thus, in mice, the term IAP might best be restricted to *Mmr10_IAP*. Numerous IAP elements share a common, 1,800-bp deletion that includes the upstream end of the RT. Yet these elements were, and perhaps still are, capable of transposing as evidenced by the fact that copies with the same deletion were found on many different chromosomes. Even shorter, internally-deleted elements, with two LTRs and ostensibly capable of transposition, can be assigned to *Mmr10_IAP* on the basis of LTR similarity (down to about 2,700 bp in overall length).

Class III (families 20 and 21)

Members of this class make up 5.40% of the mouse genome (copy number about 442,500) [4]. They have 5 bp TSDs and Class III has two constituents: murine *ERV-L* elements, which have an estimated copy number of 37,000 [4]; and the non-autonomous *MalR*s (mammalian apparent LTR retrotransposons), which are the most common retroviral element in the mouse genome, making up 4.8% of the mouse genome [4]. *MuERV-L* elements are closely related to human endogenous retrovirus L (*HERV-L*). In BLAST searches we have identified a human element (*HSAP-1*; Table 1 and Figure 4) that is 85% similar at the amino acid level to *MuERV-L* RTs. Because alignments show that their LTRs are 51% similar, we conclude that murine *MalR*s and *MuERV-L* elements share a recent common ancestor. However, as they are not quite sufficiently similar to be members of the same family, we have assigned these families the names *Mmr20_ MuERV-L* and *Mmr21_MaLR*.

Like *MalR*s in other species, murine *MalR*s are all internally deleted. The internal region contains only non-coding repetitive DNA. Nevertheless they have typical LTRs, primer binding site and polypurine tract. Members of *Mmr21_MaLR* are of two types: MT *MalR*s - the most common type of LTR retrotransposon in the mouse genome (mean length approximately 1,980 bp); and ORR1 *MalR*s (mean length approximately 2,460 bp). Our survey suggests that in the mouse genome, MT *MalR*s are about ten times as common as their longer relatives, the ORR1 *MalR*s. Non-truncated copies of *Mmr20_ MuERV-L* elements have an overall length of about 6,400 bp.

Length variation in murine LTR retrotransposons

Although all copies of family *Mmr10_IAP* found by LTR_STRUC have two LTRs and recognizable TSDs (as required by the search algorithm employed by the program), the individual members of this abundant family vary widely in overall length (2,700-7,200 bp) due to the presence of internal deletions of varying length. On the other hand, the two abundant types of non-autonomous Class III elements (MT and ORR1 *MalR*s) exhibit a markedly different pattern of

Table 1

Non-murine RTs obtained from translating BLAST

Name	Name of retrotransposon	Accession number	Position of RT in file	Host genus
*HSAP-1**	Human endogenous retrovirus L	AL590235	114430-115010	*Homo*
*HSAP-2**	Human endogenous C type retrovirus	AC078899	151820-152410	*Homo*

*Name used only in this study.

variation from that of *Mmr10_IAP* elements. Lengths of ORR1 *MalR*s peak sharply at 2,300 bp and those of MT *MalR*s at 1,980 bp, with very few elements in either case differing from these peak frequencies by more than 100 bp (<1%). Moreover, most copies of *Mmr10_IAP*, from the shortest to the longest, are preponderantly represented by copies with a high level of LTR-LTR identity (>99%), a finding consistent with recent transposition. The ability of internally truncated *Mmr10_IAP*s to complete their replication cycle is consistent with the fact that a number of *Mmr10_IAP* copies bearing the same 1,800-bp deletion (affecting the polyprotein ORF) were found in our survey on a variety of different mouse chromosomes. A similar dispersed distribution of lengths was observed in two other families *Mmr19_MusD* and *Mmr1_MmERV*. Comparison of a VL30 element (AF486451) with our data revealed a high degree of LTR-LTR similarity (>90%) to elements in family *Mmr1_MmERV* and therefore are members of that family (VL30s are non-autonomous and cannot be compared with other elements on the basis of RT similarity).

Interspecific considerations

Certain families of mouse LTR retrotransposons are more closely related to elements present in other species than to other classes of mouse elements. For example, murine Class I elements are more similar to viruses in gibbon, pig, cat, and

Table 2

Exemplars of mouse LTR retrotransposon families characterized in this study

Family	Accession number	Location	Chromosome number	LTR length	Element length (bp)	TSD	LTR-LTR identity (%)
Mmr1_MmERV	AC116580	60869-69866	18	562	8,998	GATG	99.1
Mmr2_MuLV	AC122266	144706-153433	8	523	8,728	AGCT	99.8
Mmr3_MuRRS	AC131730	135746-141194	5	482	5,468	TGTG	97.6
Mmr4	AC129291	52257-60643	6	431	8,391	GCTG	ND
Mmr5	AC125146	55312-65867	2	458	10,556	CCTTGT	96.0
Mmr6	AL645686	82031-82609*	13	ND	ND	ND	ND
Mmr7	AL669907	109127-109663*	11	ND	ND	ND	ND
Mmr8	AL63044	52153-57800	11	415	5,648	GCTCAA	ND
Mmr9	AC093445	57410-58100*	1	ND	ND	ND	ND
Mmr10_IAP	AC066688	63525-70600	6	336	7,076	ATAACT	99.7
Mmr11_MMTV	AC122322	95423-105323	6	1328	9,901	TTGTAC	100.0
Mmr12	AL669825	36552-43387	11	398	6,836	CTTCAT	90.0
Mmr13	AC122304	117988-118560*	18	ND	ND	ND	ND
Mmr14	AC127274	11141-11509	17	380	8,969	AGAAAG	ND
Mmr15	AL669827	49044-57291	11	306	8,248	CAGAGA	96.0
Mmr16	BX294008	113859-114576*	X	ND	ND	ND	ND
Mmr17	AC090008	169300-176476	2	351	7,177	GCCTCT	93.0
Mmr18	AC093341	96667-101604	5	359	4,938	GGGATC	94.4
Mmr19_MusD	AC24426	12212-13012*	13	ND	ND	ND	ND
Mmr20_MuERV_L	AF481949	811-7241	12	494	6,331	GTCGG	100.0
Mmr21_MaLR	AL672246	35744-37735	X	492	1,992	GTCAC	ND

*Endpoints given are for RT not the whole element. ND, not determined.

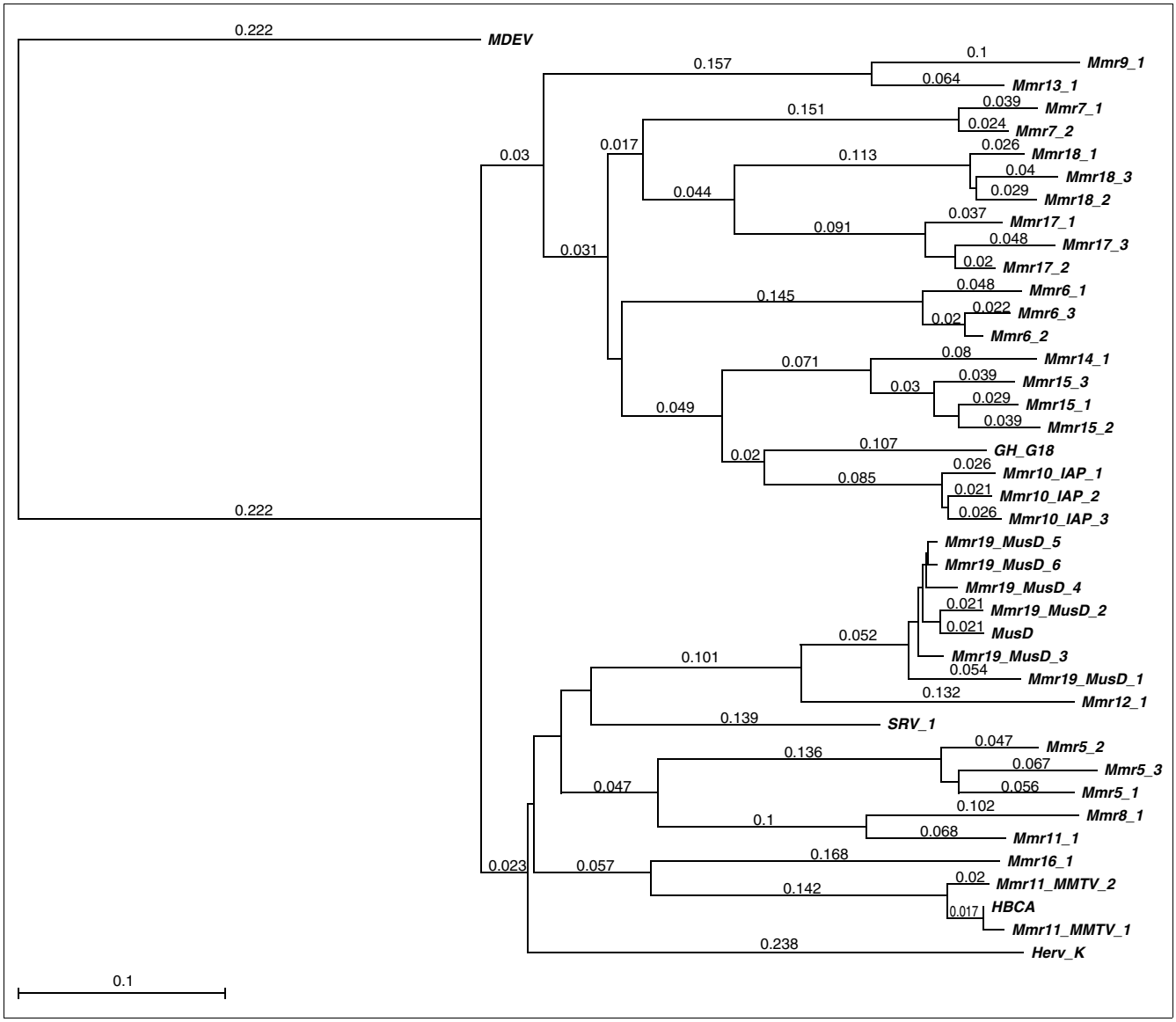

Figure 3
RT-based neighbor-joining tree for Class II murine retrotransposons. The distances (uncorrected 'p') appear next to each of the branches. RT sequences from host species other than mouse are included for comparison. The outgroup is the Class I element *MDEV* (from house/rice field mouse, *M. dunni*; see Table 3 and Figure 2).

koala, than to murine retrotransposons of Classes II or III (Figure 2). Among Class II murine endogenous retroviruses (Figure 3), family *Mmr10_IAP* is more closely related to the golden hamster element *GH-G18* than it is to any other family of murine retroviral elements. Similarly, the amino acid sequences (RT ORFs) of members of *Mmr20_MuERV_L* (mouse Class III elements, Figure 4) differ from a human element (for example, *HSAP-1*, Table 3) by only 15%, but differ from those of any non-Class III element by more than 60%. Such findings suggest that horizontal transfer may have been a source of new mouse LTR retrotransposon families over evolutionary time.

Conclusions

All autonomous retrotransposons identified in our study were retroviral-like elements (of Classes I, II, and III). At least 21 distinct families of murine LTR retrotransposons exist. Families *Mmr4*, *Mmr5*, *Mmr6*, *Mmr7*, *Mmr8*, *Mmr9*, *Mmr12*, *Mmr13*, *Mmr14*, *Mmr15*, *Mmr16*, *Mmr17*, and *Mmr18* have not been previously recognized, 13 families in all. These new families are all Class II elements (with the exception of *Mmr4*, which belongs to Class I) and are thus akin to immune deficiency viruses such as simian retrovirus SRV-1, to mouse mammary tumor virus (MMTV), and to IAP elements.

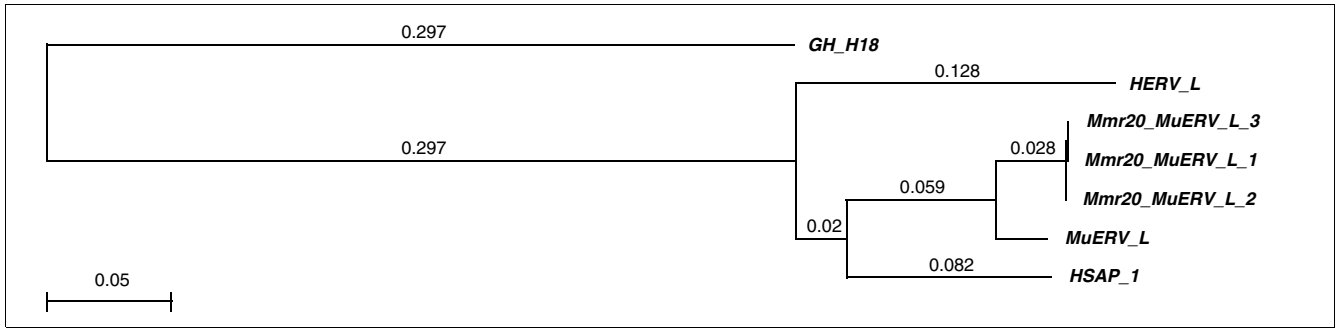

Figure 4
RT-based neighbor-joining tree for Class III murine retrotransposons. Distances (uncorrected 'p') appear next to each of the branches. RT sequences from host species other than mouse are included for comparison. The outgroup is the Class II element *GH-G18* (from golden hamster, *Mesocricetus auratus*; see Table 3 and Figure 3).

Our purpose in using LTR_STRUC to begin our survey of the mouse genome was to obtain a broadly representative sample of murine retrotransposons. Since the algorithm it employs is not dependent upon sequence homology, as in standard search methods such as BLAST, the initial results of our survey presumably were not biased toward a particular set of queries. Also, since the current version of LTR_STRUC now categorizes the elements it locates and assigns a new name to any element that differs sufficiently from any found earlier in the search, the chances of overlooking low-copy families has been reduced. The thoroughness of our BLAST search can only have been augmented by using LTR_STRUC because, in the BLAST phase of our survey, the queries used were a combination of those element types already recognized, prior to our investigation, with those found by LTR_STRUC. We believe this approach is the reason we were able to identify the 13 previously unreported families listed above.

Table 3

Known RTs used for comparison in phylogenies

Name	Name of retrovirus	Accession number/citation	Host genus
GALV	Gibbon ape leukemia virus	AAA46810	*Hylobates*
PERV	Porcine endogenous retrovirus ERV-PK15	AF038601	*Sus*
BLV	Bovine Leukemia Virus	P03361	*Bos*
HERV-K	Human endogenous retrovirus K	P10266	*Homo*
*HBCA**	Human breast cancer associated	AAG18012	*Homo*
HERV-L	Human endogenous retrovirus L	Z72519	*Homo*
*GH_H18**	Golden hamster intracisternal A-particle H18	GNHYIH	*Mesocricetus*
FeLV	Feline leukemia virus	L06140	*Felis*
RERV	Rabbit endogenous retrovirus	AAM81191	*Oryctolagus*
*GH-G18**	Golden hamster intracisternal type-A	P04026	*Cricetus*
SRV-1	Simian SRV-1 type D retrovirus	M11841	*Macaca*
MPMV	Mason-Pfizer Monkey Virus	GNLJMP	*Macaca*
MuLV	Moloney murine leukemia virus	AF033811	*Mus*
MuERV-L	Murine endogenous retrovirus ERV-L	T29097	*Mus*
MusD	Murine type D-like endogenous retrovirus MusD1	AF246632	*Mus*
HERV-C	Human endogenous retrovirus type C oncovirus	AAA73090	*Homo*
*Phasco**	Koala type C endogenous virus	AAF15098	*Phascolarctos*
*MDEV**	*M. dunni* endogenous virus	AAC31805	*Mus*
MMTV	Mouse mammary tumor virus	NC_001503	*Mus*
MmERV	*M. musculus* endogenous retrovirus	[13]	*Mus*

*Name used only in this study.

Materials and methods

Using a new data-mining program, LTR_STRUC [6], we have mined the Ensembl mouse (*M. musculus*) dataset [5] for LTR retrotransposons. We have used elements found in this initial search, as well as murine LTR retrotransposons identified by previous workers, to conduct BLAST searches of the GenBank mouse database.

Automated characterization of LTR retrotransposons

The methods used in our survey of the mouse genome are essentially the same as those used in our earlier study of the rice genome and are described elsewhere [8]. Briefly, we began our survey by using a new computer program, LTR_STRUC, which identifies new LTR retrotransposons based on the presence of characteristic retroelement features [6]. Additional elements were identified by BLAST searches using the RTs, both of elements located by LTR_STRUC and of ones previously recognized in earlier studies by previous researchers.

Datasets scanned

Initial scans with LTR_STRUC were conducted on a dataset consisting of the 2.9 Gbp of *M. musculus* sequence data available in the Ensembl database at the time of the initial scan (December 2002). The dataset (EMD) was obtained from the Ensembl website [5]. In an effort to identify additional elements not picked up in the initial survey with LTR_STRUC, we have used representative sequences from each retrotransposon family identified in this study as queries to conduct BLAST searches against the GenBank mouse database (GBMD). Thus, the results reported here constitute a reasonably unbiased survey of LTR-retrotransposon diversity in mouse. RT sequences were identified according to previously described criteria [16,17].

Multiple sequence alignments and phylogenetic analyses

The RT domains of the various *Mmr* elements were aligned, as described elsewhere [8], with previously reported RT sequences (Table 3). In the case of elements lacking an RT sequence because of fragmentation or internal truncation, the LTR sequences were used to assign them the proper family.

References

1. SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL: **Nested retrotransposons in the intergenic regions of the maize genome.** *Science* 1996, **274:**765-768.
2. Flavell RB: **Repetitive DNA and chromosome evolution in plants.** *Philos Trans R Soc Lond B Biol Sci* 1986, **312:**227-242.
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409:**860-921.
4. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, *et al.*: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420:**520-562.
5. **Mouse Genome Server** [http://www.ensembl.org/Mus_musculus/]
6. McCarthy EM, McDonald JF: **LTR_STRUC: a novel search and annotation program for LTR retrotransposons.** *Bioinformatics* 2003, **19:**362-367.
7. Bowen N, McDonald JF: ***Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside.** *Genome Res* 2001, **11:**1527-1540.
8. McCarthy EM, Liu J, Gao L, McDonald JF: **Long terminal repeat retrotransposons of *Oryza sativa*.** *Genome Biol* 2002, **3:**research0053.1-0053.11.
9. Mager DL, Freeman JD: **Novel mouse Type D endogenous proviruses and ETn elements share long terminal repeat and internal sequences.** *J Virol* 2000, **74:**7221-7229.
10. Jiang N, Jordan IK, Wessler SR: **Dasheng and RIRE2: a non-autonomous long terminal repeat element and its putative autonomous partner in the rice genome.** *Plant Physiol* 2002, **130:**1697-1705.
11. Witte CP, Hien L, Bureau T, Kumar A: **Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring the host-plant genome.** *Proc Natl Acad Sci USA* 2001, **98:**13778-13783.
12. Bowen N, McDonald JF: **Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral-like elements.** *Genome Res* 1999, **9:**924-935.
13. Bromham L, Clark F, McKee JJ: **Discovery of a novel murine type C retrovirus by data mining.** *J Virol* 2001, **75:**3053-3057.
14. Brulet P, Kaghad M, Xu YS, Croissant O, Jacob F: **Early differential tissue expression of transposon-like repetitive DNA sequences of the mouse.** *Proc Natl Acad Sci USA* 1983, **80:**5641-5645.
15. Kuff EL, Lueders KK: **The intracisternal A-particle gene family: structure and functional aspects.** *Adv Cancer Res* 1988, **51:**183-276.
16. Xiong Y, Eickbush TH: **Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns.** *Mol Biol Evol* 1988, **5:**675-690.
17. Xiong Y, Eickbush TH: **Origin and evolution of retroelements based upon their reverse-transcriptase sequences.** *EMBO J* 1990, **9:**3353-3362.