Original Contributions

Incorporating Mouse Genome
Mammalian Genome
Genes and Phenotypes

B1 insertions as easy markers for mouse population studies

Pavel Munclinger,[1] Pierre Boursot,[2] Barbara Dod[2]

[1]Biodiversity Research Group, Department of Zoology, Faculty of Science, Charles University, Viničná 7, CZ 128 44 Praha 2, Czech Republic
[2]Laboratoire Génome, Populations, Interactions, Université Montpellier II, F-34095 Montpellier, France

Received: 22 November 2002 / Accepted: 5 February 2003

Abstract

Few simple, easy-to-score PCR markers are available for studying genetic variation in wild mice populations belonging to *Mus musculus* at the population and subspecific levels. In this study, we show the abundant B1 family of short interspersed DNA elements (SINEs) is a very promising source of such markers. Thirteen B1 sequences from different regions of the genome were retrieved on the basis of their high degree of homology to a mouse consensus sequence, and the presence of these elements was screened for in wild derived mice representing *M. spretus*, *macedonicus* and *spicilegus* and the different subspecies of *M. musculus*. At five of these loci, varying degrees of insertion polymorphism were found in *M. m. domesticus* mice. These insertions were almost totally absent in the mice representing the other subspecies and species. Six other B1 elements were fixed in all the *Mus* species tested. At these loci, polymorphism associated with three restriction sites in the B1 consensus sequence was found in *M. musculus*. Most of these polymorphisms appear to be ancestral as they are shared by at least one of the other *Mus* species tested. Both insertion and restriction polymorphism revealed differences between five inbred laboratory strains considered to be of mainly *domesticus* origin, and at the six restriction loci a surprising number of these strains carried restriction variants that were either not found or very infrequent in *domesticus*. This suggests that in this particular group of loci, alleles of far Eastern origin are more frequent than expected.

Although the house mouse (*Mus musculus*) is one of the most important mammalian laboratory models, there are still surprisingly few easy-to-score markers that can be used in population genetics studies or to differentiate among its different subspecies. Because of their high degree of polymorphism and the homoplasy (alleles of the same size resulting from independent mutation events) associated with their mode of evolution, microsatellite loci are difficult markers to use in such studies and there is still relatively little comparative sequence data for the different subspecies of *M. musculus* that can be used to design SNP markers. Another potential source of markers are the orthologous SINEs, which have a very low probability of being inserted more than once in the same site. At the species level, young unfixed SINE loci have proved to be very useful for studying population structure and intraspecific demography (Shedlock and Okada 2000).

The youngest members of the Alu subfamilies that are currently active in the human genome have been shown to be a good source of markers for human population genetics (Batzer et al. 1996), and systematic genomic database mining (Roy-Engel et al. 2001) has been used successfully to identify the candidate Alu elements in the genomic sequences provided by the Human Genome Project. The B1 repeat family is the mouse equivalent of the Alu family, and here we assess the utility of using a similar strategy to find insertion polymorphisms of B1 elements within *M. musculus* that can be used as intra or intersubspecific markers.

The B1 SINE family in the mouse consists of more than 100,000 copies that have accumulated by retroposition throughout the course of rodent evolution (Rogers 1985). Like the human Alu repeats, they are derived from 7SL RNA (Ullu and Tschudi 1984) and are copies of a limited number of active master or source copies that have inserted into the genome by retroposition. Quentin (1989) showed that throughout the evolution of the rodent lineages, B1 elements have been retroposed from a succession

Correspondence to: P. Munclinger; E-mail: muncling@natur.cuni.cz

of variant progenitor sequences that have left subfamilies of related elements in the mouse genome. The subfamily generated by a given progenitor copy can be identified by the presence of shared variant nucleotide position(s) relative to the B1 consensus sequence. Kass et al. (2000) provide evidence that the progenitors of the four most recent subfamilies found in *M. musculus*, known as A, B, C, and D, have been active concomitantly in the lineage of the genus *Mus* that gave rise to *M. musculus*, *M. spretus*, *M. spicilegus*, and *M. macedonicus*.

Our knowledge of the dynamics of B1 insertion in *M. musculus* is not as well documented as the Alu family in humans. Kass et al. (2000) showed that polymorphic B1 insertions associated with recent integrations exist in *M. musculus*, but there is no information about how widespread they are in the mouse genome. In this study, we undertook a population survey involving 250 wild mice belonging to the different *M. musculus* subspecies, to investigate the degree of insertion polymorphism associated with 13 B1 insertions of presumably recent origin that show a high degree of homology to the consensus sequence.

Materials and methods

Mice. The reference panel of mouse DNA, isolated from wild derived strains or wild mice representing a substantial part of the *M. musculus* species range, came from the mouse DNA collections of the authors' laboratories in Montpellier and Prague. Most of the individuals represent the *M. m. domesticus* and *musculus* populations, but a few wild derived strains representing *M. m. castaneus* (Taiwan, Thailand), *M. m. molossinus*, *M. spretus* (France and Spain), *M. macedonicus* (Bulgaria, Turkey, and Syria), and *M. spicilegus* (Moldavia, Austria, and Ukraine) were also included. The detailed list of individuals is available on request.

B1 insertions. Three hundred and ninety seven sequences showing significant alignments with a consensus B1 sequence (Quentin 1989) were found in a Blast search (Altschul et al. 1997) of 7606 gene-containing genomic sequences (a total of 2.8 Mb) extracted from GenBank. Thirteen highly conserved B1 elements present in or close to mapped genes or pseudogenes were chosen from among these sequences. Their accession numbers and position in the sequence are summarized in Table 1. For the sake of simplicity, these PCR loci are referred to by the name of the nearest gene or pseudogene.

Table 1. The presence (p) or absence (a) of the 13 B1 elements in *M. musculus* (*M. mus.*) and *M. spretus, macedonicus,* and *spicilegus* (*M. spr., mac.,* and *spi.*). The nearest gene or pseudogene to the B1 insertions and its chromosome position are indicated

Nearest gene to B1 sequence	Chr	cM	Presence (p) or absence (a) of the B1 element in			GenBank accession No.	Nucleotide position of the B1 sequence		Primer sequences and the size of the amplified fragment containing the B1 element		
			M. spr. mac. spi.	M. mus.			Begins	Ends	Forward	Reverse	bp
Psmb5	14	20	p	p		AB003306	2888	3022	TGAAGCGGTTAAGTGTGGTG	GTGGTGCACAGTCATACATGC	348
Mtm1	X	27.8	p	p		AF125314	148121	147987	AGCCACAACCCCAACTTTTT	AAATGCTTGTAAGATGGCAAGG	331
Unp	9	60	p	p		AF026469	7873	7739	ACCTCCTTGTTCCCAGTCCT	CCTGTCGAGTTCCAGACCA	349
Igf2r	17	7.3	p	p		AJ249895	63037	62903	TCATGGGTTCCACTCAGCTA	TGGAGAAGGAAGAGGGAACTT	345
Tpardl	5	42	p	p		AF146793	145023	145157	CCTGAGGCACACATAAACACA	GGAAATGTCCCACTGTTTCAGG	375
Renbp	X	29.5	p	p		AF133093	197928	197794	GGACATGTGTCTTGGCATCA	GCAAGCGGATCTCTGTGAAT	389
Pirb	7	1	a	a		AF038149	3726	3593	GCGTACGAAGGAGGCAGA	CTGTCCCGAAAAACCAAAAA	348
Ercc2	7	4	a	p		L47235	10928	11062	ACCCTCGTCTGGTGCATCT	CCGGAGTGAGTGCTGAGATT	340
Brd2	17	18.5	a	a/p		AF100956	211061	211195	CAAACAAAACACAGGAGCTGAG	GTGGCACCAGCCATATTTTC	362
Nktr	9	71	a	a/p		U63544	7114	6980	TGCATTAGTGTTTTGCCTGAA	AGAAAAGGAATCATGGAACTGG	345
Pafaha-ps1	3	48.2	a	a/p		AF030883	6626	6759	TTGAACTCAACTCGGGACCT	TGGAAGAGGCTGGAAGAGAG	316
Tsx	X	42.1	a	a/p		X99946	47825	47959	GCTGGTGATTGTGACCTGAA	GCTGAAGCGTGTCACCTAGA	381
Btk	X	51	a	a/p		U58105	77464	77329	AATGGGCTAGCGTAGTGCAG	AGGGGACGTACACTCAGCTTT	342

Primers and PCR conditions. Table 1 shows the PCR primers flanking these B1 elements sequences, which were defined by using the program PRIMER3 (Rozen and Skaletsky 2000). PCR reactions (10 μL) were performed in the presence of 2 mM $MgCl_2$ and 0.25 units of *Taq* polymerase. After an initial incubation at 95°C for 3 min, the reactions were amplified for 35 cycles at 95°C for 45 s, 60°C for 45 s, and 72°C for 45 s. The last step was a final extension at 72°C for 10 min. The PCR products were run on 2% agarose gels, and the presence or absence of the given B1 element was deduced from the size of the amplified DNA product.

RFLP analysis. Certain amplified products were digested with the enzymes *Msp*I and *Taq*I. The reaction was done in a final volume of 20 μL containing a 10-μL aliquot of the amplification product to be digested, 2 U of restriction enzyme, and the corresponding digestion buffer. The products were analyzed in 1.5 or 2% agarose gels.

Results

Search for insertion polymorphism of B1 elements within M. musculus. The presence or absence of the 13 conserved B1 sequences in a panel of DNA samples representing the different *M. musculus* subspecies and *M. spretus*, *spicilegus*, and *macedonicus* is given in Table 1. Six of these insertions (loci *Psmb5*, *Mtml*, *Unp*, *Igf2r*, *Tpardl*, and *Renbp*) were present in all the samples tested, so their insertion must have occurred before the divergence of this group of species. The insertion at locus *Pirb* was not found in any of the wild mice tested. The B1 element in the gene *Ercc2* was not found in any of individuals representing *M. spretus*, *macedonicus*, and *spiciligus*, although it is apparently fixed in all the subspecies of *M. musculus*. However, there is considerable variation at this locus, almost certainly because of the $(GAAA)_n$ microsatellite just after its poly(A) tail.

The five remaining loci show varying degrees of presence/absence of polymorphism within *M. musculus* (Table 2). The two elements in the loci *Btk* and *Tsx* were found in the majority of the *domesticus* mice we tested. The distributions of the insertions found in the loci *Pahah1b1-ps1* and *Nkrt* suggest the presence of geographical frequency gradients, as the former appears to be close to fixation in Northern Europe and is much less frequent in the more Southern parts of the range, while the latter is very common in North and Western Europe but much less frequent in SE Europe, the Middle East, and N Africa. The element close to *Brd2* is found at low frequency throughout the *domesticus* range.

Search for RFLPs in the fixed B1 elements. In general, the CpG dinucleotides present in the B1 elements have a higher tendency to show changes relative to the consensus sequence than the other positions (Quentin 1989), and this holds for the conserved sequences we analyzed in this study (Fig. 1). Restriction enzymes with sites that include these dinucleotides are more likely to reveal polymorphisms than enzymes with sites in the more conserved regions of the B1 sequence. We therefore used *Msp*I, which has one site, and *Taq*I, which has two sites [*Taq*I(1) and *Taq*I(2)] in the consensus sequence (see Fig. 1) to look for restriction polymorphisms associated with the six B1 elements that were present in all the *Mus* species tested. The panel of mice used to investigate the restriction polymorphism in the amplified fragments obtained at these loci was essentially the same as the one used to search for insertion polymorphisms. Haplotypes for all three sites of interest could be reconstructed in all cases because none of the individuals tested possessed two polymorphic sites. The distribution of these haplotypes is given in Table 3, and they are indicated as absent (a) or present (p) in the order of the potential sites in the B1 sequence. Other polymorphic sites in the regions flanking the B1 repeats are indicated in italics (one *Taq*I site 5′ of the *Renbp* insertion and one *Msp*I site 3′ of the *Tpardl* insertion).

All the haplotypes found in *M. musculus*, except for the haplotype pap at the locus *Unp*, also occur in at least one of the other three species analyzed. This means that most of the polymorphisms we found in *M. musculus*, *spretus*, *macedonicus*, and *spiciligus* must have already been present in the common lineage leading to these species. Although it is possible that certain of these shared haplotypes are the result of different mutational events, it is very unlikely that most of them are homoplasies.

Even though all the haplotypes found in *M. musculus* are shared by two or more of the subspecies, strikingly different haplotype frequencies between *domesticus* and *musculus* are found at four of the loci. At the loci *Psmb* and *Tpardl* a different haplotype is almost completely fixed in *musculus* and *domesticus*, and at the loci *Mtm* and *Igf2r* the haplotypes fixed in *domesticus* are found only in 15% and 35% of the *musculus* samples respectively. The sampling of *castaneus* and *molossinus* is too small for a reliable estimation of their haplotype frequencies to be made.

Although the same *Renbp* haplotype (ppp*a*) is present in most of the mice belonging to these two subspecies, an insertion of 10–20 bp, which is absent in all the other mice tested, was found in both the frequent and infrequent *domesticus* haplotypes.

Table 2. The insertion frequency in various geographical location of the 5 B1 elements that were only found in wild mice belonging to *M. musculus*. The figures in brackets give the number of mice tested when it differs from the indicated number. NT is not tested

Geographical region		*Localities*	*Mice*	*Brd2*	*Nktr*	*Pafaha-ps1*	*Tsx*	*Btk*
M. m. domesticus								
N Europe	Jutland	8	60	0	0.49	0.92	0.85	1
	Westphalia, Germany	4	8	NT	0.82	0.88	NT	NT
	N Bavaria, Germany	9	13	0.35	0.8	1	1	1
	England	1	6	0	1	0	0.67	1
S & SW Europe	Switzerland	1	3	0	0.83	0	1	1
	N Italy	3	7	0.07	0.86	0.29	1	1
	S France	8	19	0.3	0.76	0.32	0.68	1
	Spain	3	3	0.03	0.83	0	1	1
SE Europe &	S Bulgaria	2	2	0	0.25	0.25	1	1
Near East	N Greece	1	5	0	0.30	0	1	1
	E. Turkey	5	10	0.20	0.35	0	0.80	1
	Israel	3	4	0.13	0	0.38	1	0.60
	Egypt	1	1	0	0	0	1	1
N & W Africa	Libya	5	16	0.13	0.34	0.09	0.24	1
	Tunisia	1	1	0	0.50	0	1	1
	Algeria	1	1	1	0.50	0	1	1
	Maroco	2	2	0	0	0	1	1
	Senegal	1	1	0	0	0.50	1	1
Pacific	Oceania	1	1	0	1	0.50	1	1
	Australia	1	3	0	0.67	0.67	1	1
	New Zealand	1	4	0	0.75	1	1	1
M. m. musculus								
NW Europe	Jutland	9	40	1(1)	0.06	0	0	0
C Europe	Czech Republic	9	17	0.03	0	0	0.03	0
	Slovakia	3	4	0	0	0	0	0
	Austria	1	1	0	0	0	0	0
	Slovenia	1	1	0	0	0	0	0
SE Europe	Serbia	1	1	0	0	0	0	0
	Bulgaria	4	4	0	0.25	0	0	0
	Rumania	1	1	0	0	0	0	0
M. m. castaneus								
	Taiwan	1	1	0	0	0	0	0
	Thailand	2	2	0	0	0	0	0
M. m. molossinus								
	Japan	1	1	0	0	0	0	0
Inbred strains								
	AKR			1	0	1	1	1
	BALB/c			0	1	0	1	1
	C57BL/6			1	1	1	1	1
	DDK			1	1	1	1	1
	SJL			0	1	0	1	1

This insertion, which occurs in the region that flanks 3′ end of the B1 element, could be related to the $(CAAA)_3$ repeat that follows the poly(A) tail of the repeat in the GenBank sequence. The size difference is hardly perceptible when the intact amplified fragments are compared, but is clearly detectable in 2% agarose gels after digestion with *Taq*I.

Polymorphism between classical inbred strains. Five inbred strains were screened for the same B1 insertion and restriction polymorphisms as the wild mice populations. It can be seen in Table 2 that the three loci with B1 insertion polymorphisms in *domesticus* also reveal differences between the inbred strains we tested. Neither BALB/c or SJL carried the insertion found in the loci *Brd2* and *Pafaha-ps1*, and the element in the locus *Nktr* is absent in AKR. However, all five strains carry the B1 elements that are nearly fixed in the *domesticus* populations *(Btk* and *Tsx)*. The B1 insertion in *Pirb*, which is present in the inbred strain 129 but absent in all the wild derived mice we tested, is also absent in these strains.

The restriction haplotypes found in these strains are given in Table 3, and it can be seen that

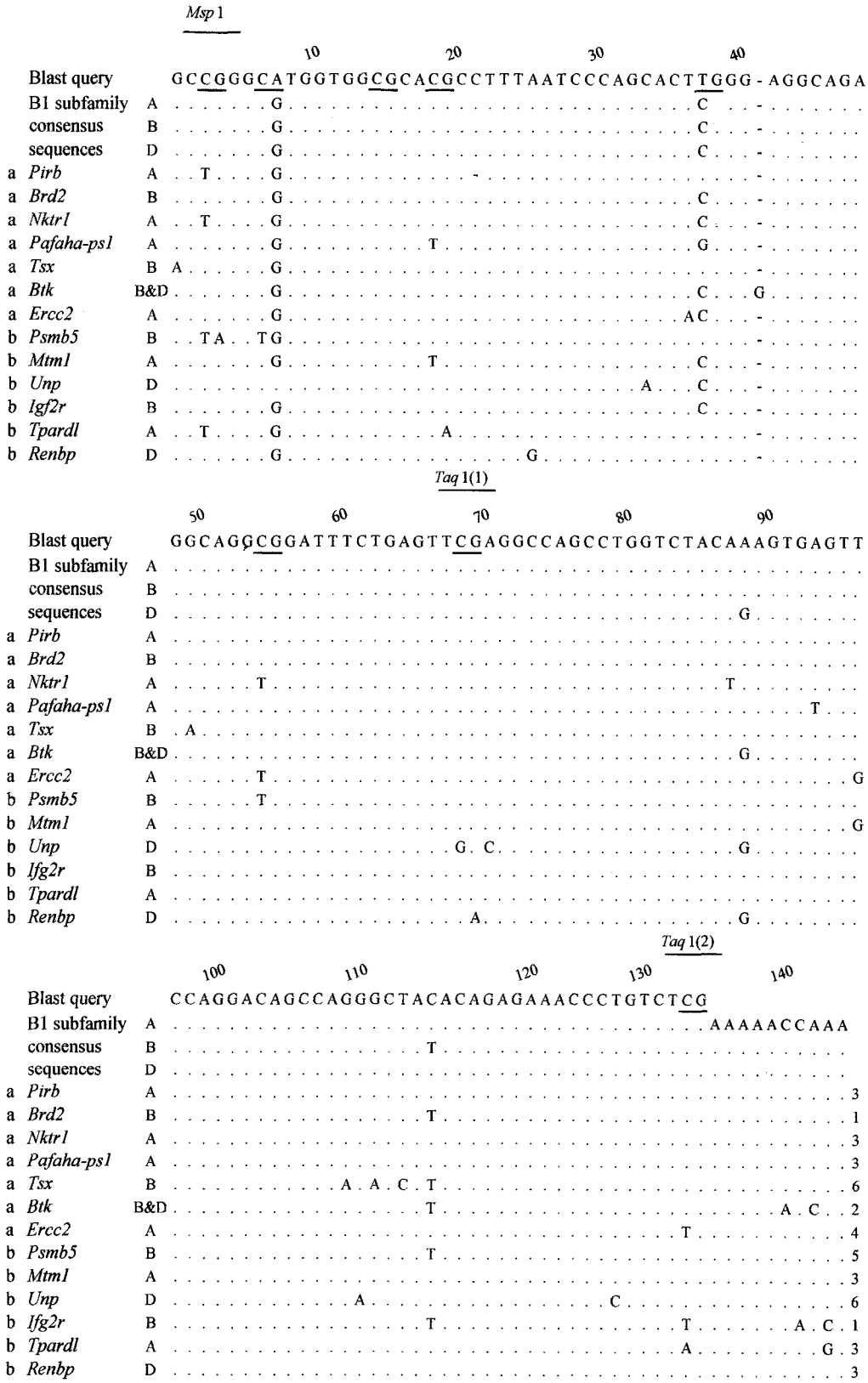

Fig. 1. The sequences of the 13 B1 elements studied aligned under the Blast query sequence. The consensus sequences corresponding to the A, B, and D subfamilies are indicated. The positions of the restriction sites for *Msp*I and *Taq*I are indicated above the sequence, and the sites containing the CpG dinucleotides in the subfamily consensus sequences are underlined. The number at the end of each sequence corresponds to the number of changes that occur outside the sites defining the subfamilies, and a and b in the first column indicate respectively the B1 elements that are found only in *M. musculus* and those fixed in *M. musculus, spretus, macedonicus,* and *spicilegus.*

Table 3. Frequency of the restriction haplotypes associated with the six B1 elements that are fixed in all the *Mus* species tested. They are defined as the absence (a) or presence (p) of the *MspI*, *TaqI*(1) and *TaqI*(2) sites that occur in the B1 consensus sequence. The additional sites in flanking sequences are given in italics.The number of different localities screened in each species or subspecies is indicated in brackets. The presence of the insertion in the 5′ flanking region of *Renbp* indicated by ¤

	Locus:	*Psmb*		*Mtm*		*Unp*		*Igf2r*		*Tpardl*			*Renbp*		
	haplotype:	app	apa	ppp	pap	pap	aap	ppa	apa	aap*p*	aaa*p*	paa*p*	*a*ppp	*a*app	*p*ppp
M. m. domesticus	(34)	0.94	0.06	1	0	0.94	0.06	1	0	0.95	0.05	0	0.81¤	0.19¤	0
M. m. musculus	(20)	0.04	0.96	0.15	0.85	1	0	0.35	0.65	0	1	0	0.91	0.09	0
M. m. castaneus	(2)	0.5	0.5	1	0	0	1	1	0	0	1	0	0	1	0
M. m. molossinus	(1)	0	1	0	1	0	1	0	1	1	0	0	0	1	0
M. macedonicus	(1)	0	1	1	0	1	0	1	0	0	0	1	1	0	0
M. spicilegus	(3)	0	1	1	0	1	0	0	1	0	1	0	0	0	1
M. spretus	(1)	1	0	0	1	1	0	1	0	0	0	1	0	1	0
AKR		1	0	0	1	1	0	1	0	0	1	0	0	1¤	0
BALB		1	0	0	1	1	0	1	0	0	1	0	0	1¤	0
C57/B6		1	0	0	1	1	0	1	0	0	1	0	0	1¤	
DDK		0	1	0	1	0	1	1	0	0	1	0	0	1¤	0
SJL		1	0	1	0	0	1	1	0	0	1	0	0	1¤	0

differences also occur at three of these loci. At the locus *Psmb*, DDK has the most frequent *musculus* haplotype, which is also found in *molossinus* and one of the *castaneus* strains tested. At the locus *Mtm*, only SJL carries the haplotype fixed in *domesticus*, and all the others share the haplotype that is very frequent in *musculus* and present *molossinus*. Unlike AKR, BALB/c, and C57BL/6, neither DDK nor SJL shares the *Unp* haplotype that occurs most frequently in the European mice belonging to both the *domesticus* and *musculus* subspecies. No differences were found among the five strains at the other three loci, but it is interesting to note that neither the *Tardl* nor *Renbp* haplotypes correspond to the most frequent *domesticus* ones.

Sequence comparisons. Figure 1 shows the 13 B1 elements aligned under the B1 consensus sequence used as the Blast query. This consensus is derived from 51 mouse B1 elements that were inserted in the genome over a wide period of evolutionary time (Quentin 1989). The sequences corresponding to the presumed progenitor sequences of the recent A, B, and D subfamilies of B1 elements (Zietkiewicz and Labuda 1996) are also indicated. The number of changes, including insertions and deletions, that occur outside the sites defining the three subfamilies (positions 89 and 116 respectively for the D and B subfamilies) are given at the end of each sequence. The B1 sequences that are fixed in all the species tested are presumably older than those found only in *M. musculus*, so one might expect to find a correlation between the time a given element has been present in the genome and its degree of homology relative to the consensus sequence of the subfamily to which it belongs. However, we found no significant difference in the ranges of nucleotide variation between the fixed and unfixed groups of repeats (one to six variant positions). This overlap suggests that it will be difficult to predict the presence of an insertion polymorphism that is close to fixation within *domesticus* simply on the basis of homology to a subfamily consensus sequences.

Another feature of all the sequences we studied is the presence of an $(A)_5CC$ motif at their 3′ ends. This motif, which was found by Zietkiewicz and Labuda (1996) to be frequent in the recent B1 sequences in different rodent species, is not present in the majority of the B1 sequences in the mouse genome. It therefore appears to be labile, and its presence could be a useful criterion for choosing candidate loci to screen for polymorphic B1 insertions.

Discussion

This study shows that B1 elements that differ from their presumed subfamily consensus sequence at between one and six positions (see Fig. 1) provide a good potential source of markers that can be used to study genetic variation in mouse populations belonging to *M. musculus*. The 13 B1 sequences were retrieved from a very small portion of the genome (a total 2.8 Mb), and since we started this study, the amount of genome sequence available through the public mouse genome sequencing effort has increased enormously. As most of the B1 elements we tested show some degree of either insertion or restriction polymorphism within *M. musculus*, systematic genomic data mining for conserved B1

sequences, combined with the rapid screening strategy detailed above, should now make it possible to cover a large part of the genome. It should also be possible to increase the number of informative restriction polymorphisms associated with the B1 elements present in a given region of the genome by extending the search to slightly more diverged B1 sequences and including more restriction enzymes in the screening protocol. For example, *Ava*I and *Nla*III both have either a restriction site or a site minus one nucleotide that includes other labile CpG dinucleotides in the B1 consensus sequence.

Most of the present-day laboratory inbred strains have mosaic genomes that are mainly of *domesticus* origin, but they also have contributions from the Japanese and Chinese mice of *M. m. musculus*, *castaneus*, and *molossinus* origin (Bonhomme et al. 1987), the most notable example being the Asian variant of the Y Chr that is carried by many of these strains (Bishop et al. 1985; Nagamine et al. 1992; Tucker et al. 1992). All but one of the B1 elements showing insertion polymorphism in *M. musculus* were essentially specific to *domesticus*, which is to be expected for B1 elements that were all originally found in an inbred strain (most of the time, derived from the strain 129). It is possible, however, that the B1 element present in the *Pirb* locus of 129, which was not found in any of the mice we tested, originated in mouse populations from eastern Asia that were not well represented in the sample used in this study.

There is an increasing demand for markers that can contribute to the understanding of the genealogy of the inbred mouse strains (Beck et al. 2000). Our preliminary results on just a few inbred strains suggest that loci containing fixed but recent B1 elements are also likely to be a very useful source of such markers. Not only do half of the six loci we tested for restriction polymorphism reveal differences between strains, but they also provide clues about the subspecies origin of the gene carrying the insertion. None of the strains in the panel we analyzed (AKR, BALB/c, and C57BL/6 derived from Castle's Mice; SJL from the Swiss mice; and DDK from old Japanese stock) carry the most frequent *domesticus* haplotype at more than three of the loci we tested (Table 3). If the haplotypes present in the majority of the strains are considered to be of *domesticus* origin, half of them would have to have been derived from *domesticus* mice carrying infrequent variants. It seems rather unlikely that all these 'atypical' haplotypes are of *domesticus* origin, and certain of them were probably drawn from far Eastern populations of *musculus*, *castaneus*, or *molossinus* mice. It is interesting to note that in DDK the most frequent *domesticus* haplotype was found only at one locus (*Igf2r*). Females of this strain are almost completely infertile when crossed with other laboratory strains (Wakasugi et al. 1967), but are fully fertile in crosses with two strains of Asian origin, CAS and MOM (Zhao et al. 2002). Our results suggest that, although this strain is considered to be of mainly *domesticus* origin, contributions from the Far Eastern subspecies might be quite important.

Unlike the study of Kass et al. (2000), our search for recent B1 elements did not target the recent B1 subfamilies, but used a sequence consensus derived from a much wider range of B1 elements that does not represent any of these subfamilies. Even so, it can be seen from Fig. 1 that all the sequences we retrieved possessed a nucleotide variant that defines one of recent A, B, or D subfamilies (Quentin 1989). Like Kass et al. (2000), we found sequences that can be classed in the A and B subfamilies in both the group of B1 elements present only in *M. musculus* and in the group of fixed B1 elements. Our results therefore provide further evidence that different progenitors were active concomitantly in the lineages leading to *M. musculus*, *M. spretus*, *M. spicilegus*, and *M. macedonicus*. However, certain features of the sequences we analyzed suggest that the B1 subfamily structure is more complex. For instance, the B1 element in the locus *Btk* that is found only in *domesticus* carries the nucleotide positions that define both the supposedly recent B and the older D subfamilies. As we found this nucleotide combination in two other fixed B1 sequences (unpublished observations), it probably defines another related subfamily. Similarly, the G at position 97 that is shared by the B1 sequences in *Ercc2* and *Mtm1* could be the signature of yet another variant source copy that was active during the same period. The presence among the recently inserted Alu elements in humans of a variety of small subfamilies related to the major Alu subfamilies that have been active over a longer period of evolutionary time has led to the suggestion that certain insertions acquire the ability to retropose for a limited period of time (Roy-Engel et al. 2001). These act as source copies that generate small subfamilies of daughter copies carrying the same sequence differences. Our data suggest that a similar situation may exist in the mouse, and the presence of unidentified subfamilies could be one of the reasons why the estimated degree of homology to the presumed consensus sequence does not discriminate between the sequences that are polymorphic in *M. musculus* and the older B1 elements that were inserted before the divergence of *M. musculus*, *spretus*, *macedonicus*, and *spicilegus*.

As B1 elements tend accumulate preferentially in the GC-rich areas of the genome, suitable B1 elements are not likely to be retrieved frequently from the AT-rich regions (Smit 1999). The L1 long interspersed repeats, which are another very abundant retroposon family in the mouse (Loeb et al. 1986), are preferentially associated with the AT-rich isochores. Although complete L1 elements are too long to be used as practical markers, most of them are variably truncated sequences that are incomplete reverse transcripts of the full RNA transcript. Several L1 lineages have been active recently in the mouse genome (Hardies et al. 2000), so it should be possible to apply a similar strategy to the one we used in this study to identify insertion or restriction polymorphisms associated with short L1 repeats.

Acknowledgments

The participation of P. Munclinger was supported by the Grant Agency of the Czech Republic (Grant No. 206/01/0989), the Grant Agency of the Czech Academy of Sciences (Grant No. A6045902), and by the Barrande (Czech-French Bilateral Co-operation in Research and Development) Project No. 99043. This study was partly supported by the CNRS GDR 1928 (Evolution des génomes dans les populations).

References

1. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402
2. Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH et al. (1996) Genetic variation of recent Alu insertions in human populations. J Mol Evol 42, 22–29
3. Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT et al. (2000) Genealogies of mouse inbred strains. Nat Genet 24, 23–25
4. Bishop CE, Boursot P, Bonhomme F, Hatat D (1985) Most classical *Mus musculus domesticus* laboratory mouse strains carry a *Mus musculus musculus* Y chromosome. Nature 315, 70–72
5. Bonhomme F, Guénet J-L, Dod B, Moriwaki K, Bulfield G (1987) The polyphyletic origin of laboratory inbred mice and their rate of evolution. Biol J Linn Soc 30, 51–58
6. Hardies SC, Wang L, Zhou L, Zhao Y, Casavant NC et al. (2000) LINE-1 (L1) lineages in the mouse. Mol Biol Evol 17, 616–628
7. Kass DH, Raynor ME, Williams TM (2000) Evolutionary history of B1 retroposons in the genus *Mus*. J Mol Evol 51, 256–264
8. Loeb DD, Padgett RW, Hardies SC, Shehee WR, Comer MB et al. (1986) The sequence of a large L1Md element reveals a tandemly repeated 5′ end and several features found in retrotransposons. Mol Cell Biol 6, 168–182
9. Nagamine CM, Nishioka Y, Moriwaki K, Boursot P, Bonhomme F et al. (1992) The musculus-type Y chromosome of the laboratory mouse is of Asian origin. Mamm Genome 3, 84–91
10. Quentin Y (1989) Successive waves of fixation of B1 variants in rodent lineage history. J Mol Evol 28, 299–305
11. Rogers JH (1985) The origin and evolution of retroposons. Int Rev Cytol 93, 187–279
12. Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV et al. (2001) Alu insertion polymorphisms for the study of human genomic diversity. Genetics 159, 279–290
13. Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, (eds.) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. (Totowa, NJ: Humana Press), pp 365–386
14. Shedlock AM, Okada N (2000) SINE insertions: powerful tools for molecular systematics. BioEssays 22, 148–160
15. Smit AF (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev 9, 657–663
16. Tucker PK, Lee BK, Lundrigan BL, Eicher EM (1992) Geographic origin of the Y chromosomes in "old" inbred strains of mice. Mamm Genome 3, 254–261
17. Ullu E, Tschudi C (1984) Alu sequences are processed 7SL RNA genes. Nature 312, 171–172
18. Wakasugi N, Tomita T, Kondo K (1967) Differences of fertility in reciprocal crosses between inbred strains of mice. J Reprod Fertil 13, 41–50
19. Zhao WD, Ishikawa A, Yamagata T, Bolor H, Wakasugi N (2002) Female mice of DDK strain are fully fertile in the intersubspecific crosses with *Mus musculus molossinus* and *M. m. castaneus*. Mamm Genome 13, 345–351
20. Zietkiewicz E, Labuda D (1996) Mosaic evolution of rodent B1 elements. J Mol Evol 42, 66–72