

Identification and characterization of *t* haplotypes in wild mice populations using molecular markers

BARBARA DOD, CATHY LITEL, PATRICK MAKOUNDOU,
ANNIE ORTH AND PIERRE BOURSOT*

Laboratoire Génome Populations Interactions, UMR 5000 CNRS-Université Montpellier II, F-34095 Montpellier, France

(Received 27 September 2002 and in revised form 2 December 2002)

Summary

As part of a population genetics survey of the hybrid zone between mouse subspecies *Mus musculus domesticus* and *M. m. musculus*, we identified and characterized the *t* haplotypes in 1068 mice from 186 different populations in a 2500 km² area in central Jutland. On the basis of two *t*-specific PCR markers, 130 mice possessed this haplotype. The allele frequencies at six microsatellites on the third and fourth chromosomal inversions of the *t* region were sufficiently different between *t*-bearing and non-*t*-bearing mice, and linkage disequilibria sufficiently marked on the *t* haplotype, to be able to reconstitute the genotype of most *t* haplotypes. A total of three frequent and 15 rarer haplotypes were identified. These haplotypes resemble each other more than they resemble a panel of known haplotypes from a wide range of geographical regions, except for *tw73*, which was also extracted from Jutland. The patterns of variation at the microsatellite loci suggest that the Jutland haplotypes were derived from a small number of haplotypes, followed by recombination between complementing haplotypes. Further evidence of recombination came from complementation tests that we performed, showing the lack of concordance between the degrees of complementation and of molecular resemblance between haplotypes. This study shows that it is possible to characterize the presence and variation of *t* haplotypes by a population genetics approach using simple molecular markers. However recombination between *t* haplotypes has occurred frequently enough to obscure the links between this variation and the biological properties of distortion and lethality of the haplotypes that originally colonized Jutland.

1. Introduction

The *t* haplotype is a variant form of the proximal region of chromosome 17 that is found at low frequencies in wild mice belonging to all the subspecies of the house mouse, *Mus musculus*. The most striking property of *t* haplotypes is the ability of heterozygous males to transmit their *t* chromosome to up to 99% of their offspring. A series of distorter loci (*trd*) impair the flagella function of wild-type and *t* spermatozoa, and the *t* sperm but not the wild-type sperm is rescued by interaction with a responder locus (*ter*) that is only present on the *t* chromosome (reviewed by Lyon, 2000). This complex system of genes behaves as a single unit because it is prevented from recombining with wild-type chromosomes by the presence of four

neighbouring inversions (Fig. 1) (Hammer *et al.*, 1989; Herrmann *et al.*, 1987). Most *t* haplotypes carry recessive lethal mutations affecting a variety of genes (Klein *et al.*, 1984) but not all carry the same set of such mutations, so that combinations of *t* haplotypes complement each other to different extents (for a review, see Bennett, 1975). However, the male *t/t* offspring produced in such crosses are sterile. These negative properties of *t* haplotypes counterbalance the advantage derived from their ability to distort their transmission ratio and has prevented them from becoming fixed.

An ongoing project has been designed to investigate the possible contribution of genes that map to the *t*-complex region to the partial genetic isolation of the two subspecies *Mus musculus musculus* and *M. m. domesticus* in their hybrid zone. As part of this

* Corresponding author. e-mail: boursot@univ-montp2.fr

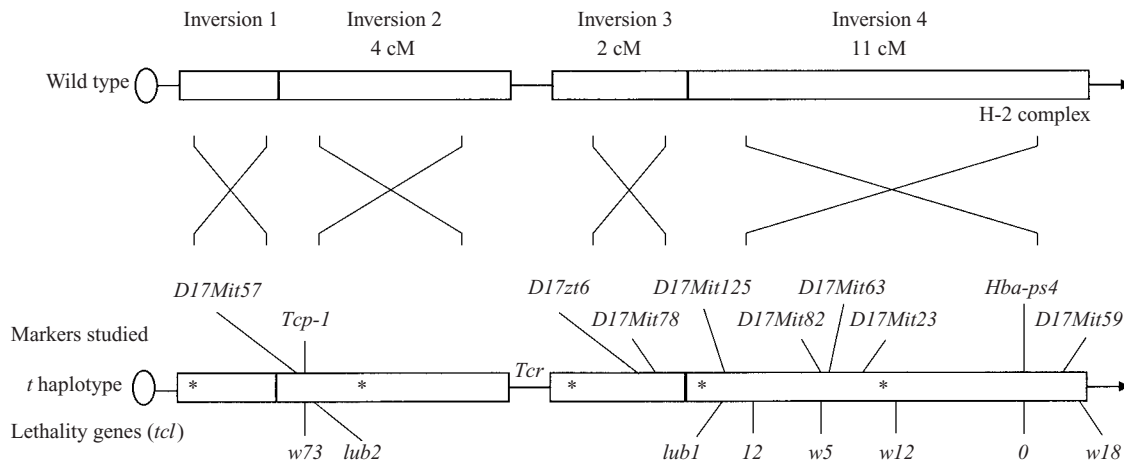


Fig. 1. Schematic map of the four inversions between wild-type and *t* haplotype forms of mouse proximal chromosome 17. We indicate on the *t*-haplotype map the names and approximate locations of the markers studied (above) and of the known lethal genes (below). The asterisks show the approximate locations of the known transmission distortion loci. *Tcr* is the responder locus.

project, we need to identify and characterize the *t* haplotypes present in a sample of wild populations from an area of $\sim 2500 \text{ km}^2$ in eastern Jutland (Denmark). The limited introgression of the sex chromosomes across this hybrid zone suggests that they play a major role in the genetic incompatibilities between the two subspecies (Dod *et al.*, 1993; Tucker *et al.*, 1992). Certain autosomal genes such as the hybrid sterility gene *Hst1* (Forejt & Ivanyi, 1975) and perhaps other genes that map to the region of the *t* complex (Fig. 1) are also likely to contribute to the hybrid dysgenesis. However, before the roles of these genes can be understood, the *t* haplotypes present in the population have to be identified, because they have properties that differ from either of the two wild-type versions of chromosome 17.

Although the ancestral *t* haplotypes appear to be considerably older than the divergence of the *M. musculus* subspecies, all the present-day *t* haplotypes appear to be derived from a recent common ancestor that spread across the species range after the formation of the present-day subspecies (Delarbre *et al.*, 1988; Hammer & Silver, 1993; Morita *et al.*, 1992; Silver *et al.*, 1987; Willison *et al.*, 1986). These studies have revealed a number of loci that distinguish between the wild-type and *t* chromosomes. Here, we identified individuals carrying putative *t* haplotypes using PCR markers for two of these loci: *Hba-ps4* (Schimenti, 1990) on the distal and *Tcp1* (Morita *et al.*, 1993) on the proximal end of the *t* chromosome.

We also need to characterize the different *t* haplotype variants segregating in the natural populations on either side of the hybrid zone. Polymorphisms that distinguish between different *t* haplotypes are relatively infrequent (Figuroa *et al.*, 1988; Fox *et al.*, 1985; Horiuchi *et al.*, 1992; Neufeld *et al.*, 1991; Schimenti *et al.*, 1987; Silver *et al.*, 1987). A number

of related *H2* serotypes are associated with different *t* haplotypes (Nizetic *et al.*, 1984) and, although studies of the molecular organization of the major histocompatibility complex (MHC) have revealed substantial differences between *t* haplotypes (Artzt *et al.*, 1985; Figuroa *et al.*, 1985; Golubic *et al.*, 1987; Uehara *et al.*, 1991; Uehara *et al.*, 1987), much of the variation in this region appears to have been generated by recombination between complementing *t* haplotypes. Polymorphisms that distinguish between individual *t* haplotypes have been found at two highly mutable repeated regions of the *t* chromosome: a dinucleotide repeat in the *Tcp10* gene cluster (Ardlie & Silver, 1996), and the *TSE* clusters that are unique to the *Tla* region of the *H2* complex of *t* haplotypes (Uehara *et al.*, 1990). Microsatellite loci that differentiate between certain *t* haplotypes appear to be more frequent (Ebersole *et al.*, 1992; Lai & Artzt, 1992). Because they are abundant in the genome and easy to score, they are well suited to large-scale population surveys, unless convergent mutations blur our capacity to discriminate between wild-type and *t*-specific alleles. Here, we show that this is not the case and that, by using a series of microsatellite markers, it is possible to characterize the diversity of most *t* haplotypes segregating in natural populations, at least in a restricted geographic region. Although the number of different variants found in the Jutland populations sampled was surprisingly high, much of the polymorphism appears to have been generated by recombination between a limited number of complementing haplotypes. Except for *tw73*, which also comes from Jutland, the *t* haplotypes in our sample were much more closely related to each other than they were to any of the known *t* chromosomes from other geographical regions that we also analysed.

2. Material and methods

(i) Mice

We studied a sample of 1068 house mice caught in 135 different farms or buildings in eastern Jutland from 1984 to 1998, in an ~2500 km² area of the hybrid zone between *M. m. musculus* and *M. m. domesticus*. More details about the geographical origins of these mice are not relevant here. The panel of DNAs from a number of known wild-derived *t* haplotypes was kindly provided by K. Artzt and K. Ardlie.

(ii) Genetic analysis

DNA from spleen or tail tissue was digested in 50 mM Tris pH 8, 100 mM EDTA, 100 mM NaCl, 1% SDS and proteinase K. The DNA was extracted using a standard phenol-chloroform procedure. The two diagnostic PCR markers, *Hba-ps4* and *Tcp1*, were typed using the protocols of Schimenti & Hammer (1990) and Morita *et al.* (1993), respectively. The microsatellite markers used were *D17Mit57*, *D17Zt6* (Gregorova *et al.*, 1996), *D17Mit78*, *D17Mit125*, *D17Mit82*, *D17Mit63* and *D17Mit23* (Dietrich *et al.*, 1994). ~25 ng of DNA was amplified in a 10 μ l PCR reaction with 0.25 units of Taq polymerase in the presence of 1.5 mM of MgCl₂, except for *D17Zt6* and *D17Mit78*, which required 2 mM MgCl₂. One of the primers was labelled with the fluorophore Cy5 (Pharmacia). The reactions were amplified for 35 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 30 seconds, and analysed on 6% denaturing polyacrylamide gels in an automated DNA sequencer (Pharmacia). For the sake of convenience, the sizes of the alleles found in the wild mice were measured relative to the *BALB/c* allele that was attributed the arbitrary size of 50 bp at each locus. The real size of allele 50 at each locus is 320, 140, 98, 154, 146, 158 and 140 bp, respectively.

(iii) Phylogenetic analysis

In order to quantify the differences between the *t* haplotypes revealed by the microsatellite data, we constructed a molecular distance that takes into account both the number of different alleles and the size differences between the alleles. The distance between two haplotypes *i* and *j* is defined as follows

$$D_{ij} = \alpha \frac{1}{L} \sum_{k=1}^L (1 - \delta_{a_i^k a_j^k}) + (1 - \alpha) \times \left\{ \left[\frac{\sum_{k=1}^L |a_i^k - a_j^k|}{\max_{i,j} \left(\sum_{k=1}^L |a_i^k - a_j^k| \right)} \right] \right\},$$

where a_i^k is the allele size at locus *k* in haplotype *i*, *L* is the number of loci, and δ_{xy} is Kronecker's δ ($\delta_{xy} = 1$ when $x = y$; $\delta_{xy} = 0$ otherwise). The first term of the sum is thus the proportion of loci at which haplotypes

i and *j* differ. The second term is the sum of allele size differences over loci for this pair of haplotypes, normalized by the maximum value of this difference among all possible pairs of haplotypes. The weighting parameter α is introduced to vary the effect of the number of loci at which the haplotypes differ (irrespective of the amplitude of the allele size differences) relative to the effect of the cumulated amplitude in allele size differences between haplotypes (irrespective of the number of loci concerned). The matrix of distances between all pairs of haplotypes was used to build a tree using the neighbour-joining method, as implemented in the computer program MEGA (Kumar *et al.*, 2001).

(iv) Distortion and complementation properties

The distortion and complementation properties of a few *t* haplotypes that had been maintained alive after capture were investigated by crossing them to *BALB/c* mice in our animal facilities and by crossing the resulting *t/w* F1 mice between each other. The *t* chromosomes present in the offspring were identified by typing DNA from tail snips for the locus *Hba-ps4*. The likelihood of the observed numbers of the different genotypes born in each cross was evaluated under simple models. In distortion tests (male *t/w* \times female *w/w*), a single parameter *d* (the transmission rate of *t* by *t/w* males) was considered. In viability and complementation tests (*t/w* \times *t/w*), a second parameter *v* (viability of *t/t* offspring) was introduced. We determined the parameter values that maximized the likelihood of the data, as well as the lowest and highest parameter values in the two-log likelihood interval below this maximum.

3. Results

(i) Identification of mice carrying *t* haplotypes

Among 1068 mice tested, 143 possessed the *t*-specific alleles of *Hba-ps4*. Of these, 130 could be typed with *Tcp1*, and the two markers gave concordant results: 129 mice were heterozygote for the *t* allele at both loci, and one appeared to be homozygous at both loci. Because the *t*-specific alleles at these two loci are associated with complete *t* chromosomes and are only infrequently found in the non-*t* mice (Erhart *et al.*, 1989; Hammer *et al.*, 1991), it is highly likely that these mice carry *t* haplotypes. The female mouse that was *t/t* for both these loci presumably carried two complementing *t* haplotypes or a single semi-lethal haplotype.

(ii) Characterization of the different *t* variants with microsatellite markers in the third and fourth inversions

It was possible to characterize a ~4 cM region on the third and fourth inversions of the *t* haplotype that

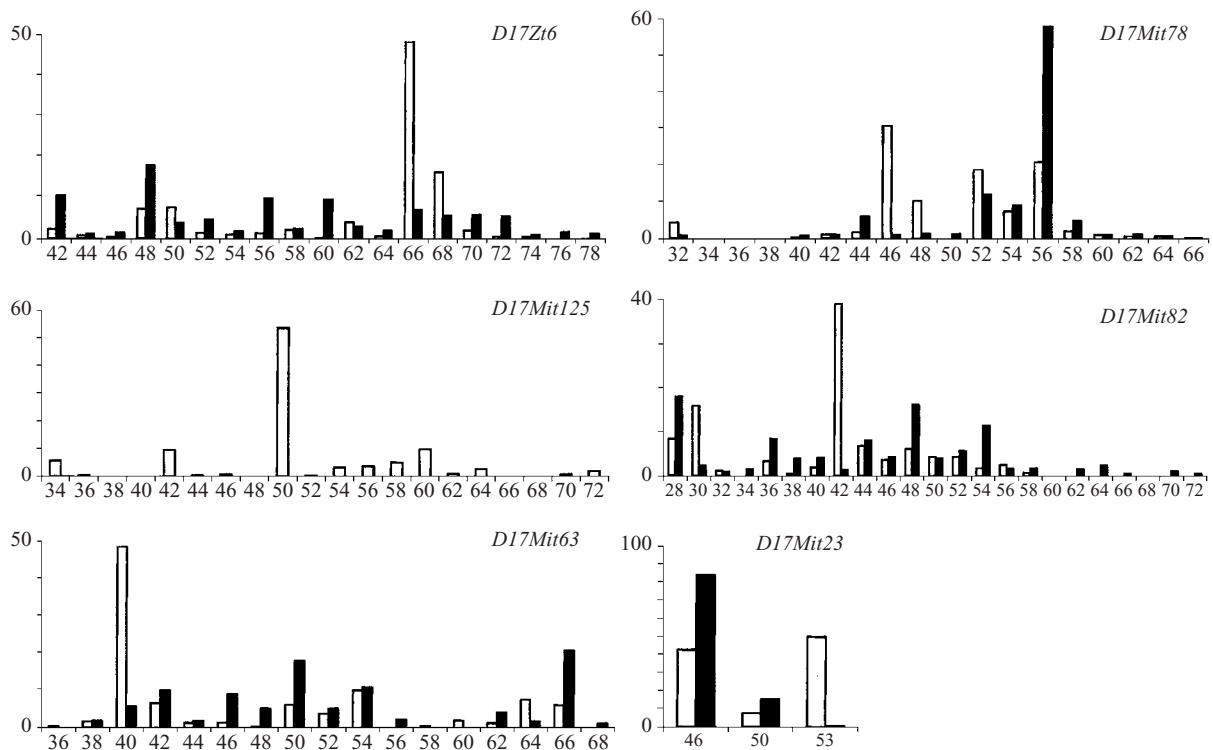


Fig. 2. Frequency distribution of alleles at six microsatellite loci in wild populations from Jutland, in *t*-bearing mice (open bars) and, when available, in a sample of non-*t*-bearing mice from the same populations (closed bars).

flanks the MHC using six microsatellite loci: *D17Zt6*, *D17Mit78*, *D17Mit125*, *D17Mit82*, *D17Mit63* and *D17Mit23* (Fig. 1). The relative allele frequencies at each locus found in these *t*-carrying mice are given in Fig. 2 and, for certain loci, the alleles found in subsets of wild-type mice are also included. As is to be expected with loci that have an exceptionally high mutation rate, the microsatellite alleles associated with *t* chromosomes were not necessarily different from those found in wild-type chromosomes and the attribution of an allele to a given *t* chromosome was not always simple. The different factors we took into account in order to attribute the most likely allele to the *t* chromosome of a given *t*-carrying mouse are described below. The haplotypes deduced from this analysis are summarized in Table 1a, which describes the genotypes of the 18 different *t* haplotypes (named J1 to J18) that we were able to infer.

For two loci the difference between the frequencies of a particular allele found in the wild-type and the *t*-carrying mice is sufficiently important to be able to identify it as the allele associated with the *t* chromosome. It can be seen in Fig. 2 that alleles 40 and 53 of the loci *D17Mit63* and *D17Mit23* respectively are very frequent in the mice identified as carrying *t* haplotypes but are rare in the wild-type mice tested (only 12 out of 208 and 6 out of 530, respectively). Both these alleles were present in all the putative *t/w* mice except for one that did not have allele 40 at the locus *D17Mit63*. The alleles associated with *t* chromosomes are more

difficult to identify at the other loci, as more than one *t* allele is needed to characterize all the *t*-carrying individuals (Fig. 2). Several different types of evidence were used to identify the allele most likely to be associated with a given *t* chromosome. The alleles that occur as homozygotes (Table 2) can be used to identify alleles that are associated with certain *t* chromosomes. Because these alleles are more often than not preferentially associated with certain alleles at the other loci, it is possible to define the most probable group of alleles defining a particular *t* chromosome. More specifically, alleles 30 and 42 of *D17Mit82* occur as homozygotes in five and ten individuals, respectively. As only one individual carries both alleles and one or the other is found in practically all the *t*-bearing individuals, there is little doubt that they define most of the *t* chromosomes in the sample. At each of the other three loci, which are all on the proximal side of the MHC, even more alleles are needed to define all the *t* variants found. The number of different homozygotes shown in Table 2 indicates that there is considerably more variation between the different *t* chromosomes in this region. However, because alleles 66 and 68 of *D17Zt6* and alleles 50 and 60 of *D17Mit125* all occur frequently as homozygotes, it is easy to identify two combinations of *D17Zt6*, *D17Mit125* and *D17Mit82* alleles (66-50-42 or 68-60-30) that, between them, can define most of the *t* chromosomes. In turn, when the alleles of *D17Mit78* are considered, alleles 46 and 50 are clearly much more frequently associated with the

Table 1. The allele size determined at six microsatellite loci (a) in the 18 haplotypes inferred from our Jutland wild population sample, with their frequency of occurrence, and (b) from a panel of laboratory-maintained t haplotypes from different complementation groups (tcl)

Haplotype	Origin	tcl group	D17tZ6	D17Mit78	D17Mit125	D17Mit82	D17Mit63	D17Mit23	Number of mice	Number of localities
a) Jutland haplotypes										
J1	Jutland		66	46	50	42	40	53	57	26
J2	Jutland		66	48	50	42	40	53	1	1
J3	Jutland		66	52	50	42	40	53	2	1
J4	Jutland		66	50	50	42	40	53	22	10
J5	Jutland		68	48	50	42	40	53	1	1
J6	Jutland		62	46	50	42	40	53	4	3
J7	Jutland		62	50	50	42	40	53	3	1
J8	Jutland		66	46	60	42	40	53	1	1
J9	Jutland		66	46	54	42	40	53	1	1
J10	Jutland		66	46	50	30	40	53	1	1
J11	Jutland		68	48	50	30	40	53	2	2
J12	Jutland		68	48	60	30	40	53	15	10
J13	Jutland		68	50	60	30	40	53	1	1
J14	Jutland		48	48	60	30	40	53	1	1
J15	Jutland		66	46	42	30	40	53	5	4
J16	Jutland		68	46	42	30	40	53	2	1
J17	Jutland		66	48	null	30	40	53	2	1
J18	Jutland		68	48	null	30	40	53	3	1
b) Reference panel										
tw73	Jutland	w73	66	50	50	42	40	53		
tlub4	Italy	Lub1 Lub4	68	50	50	30	42	53		
tTuw27	S. Germany	S L	58	46	60	30	42	53		
tw5	USA	w5	58	42	50	30	46	53		
tw12	USA	w 1	68	46	50	42	42	53		
tw8	USA	S L	68	46	54	48	42	53		
twPA1	France	Pal	68	46	50	48	42	53		
t0	France	0	70	46	50	48	44	50		
t6	lab derived	partial	72	46	50	42	42	50		
t12 & tw32	USA	12	58	46	50	48	36	50		
tw2	USA	S L	68	52	36	48	42	50		
tw1	USA	w 1	68	46	50	44	42	50		
tw130	Chile	w5	58	46	50	44	42	50		
tlub1	Italy	Lub1	66	46	46	44	42	50		
tTuw11	Chile	S L	58	46	46	44	42	50		
tTuw28	S. Germany	S L	72	52	50	56	40	50		
tw121	Tajikistan	w73	68	50	50	62	72	50		

Table 2. The alleles found homozygous at four microsatellite loci in *t*-bearing mice, with the numbers of occurrences indicated in brackets

Locus	Alleles (number of homozygous mice)				
<i>D17Mit82</i>	30 (5)	42 (10)			
<i>D17Mit125</i>	42 (3)	50 (12)	60 (6)	64 (2)	
<i>D17Mit78</i>	46 (2)	48 (1)	50 (3)		
<i>D17Zt6</i>	48 (1)	50 (1)	62 (2)	66 (29)	68 (16)

first combination, whereas allele 48 of *D17Mit78* is nearly always associated with the second. There is, therefore, little doubt that the three most frequent *t* variants found in our sample are defined by the allele combinations 66-46-50-42, 66-50-50-42 and 68-48-60-30 of *D17Zt6*, *D17Mit78*, *D17Mit125* and *D17Mit82*, respectively (see variants J1, J4 and J12 in Table 1). In the few cases where alleles 46 and 50 of *D17Mit78* occur in the same individual, the allele associated with the *t* variant carried by the other mice coming from the same building was considered to be the *t* allele, because they are very likely to be members of the same deme and to carry the same *t* chromosome.

These three major variants account for >75% of the *t* chromosomes found in the sample. Variants J5 and J10 could each have been derived in a direct recombination between mice carrying the variants J1 and J12. Four others (J2, J8, J11 and J13) only differ from one of the three most frequent haplotypes at one locus but could not have been generated in a single recombination event between any of them.

Two variants defined by allele 42 of *D17Mit125* in association with alleles 46 and 30 of *D17Mit78* and *D17Mit82*, respectively, were found in a limited number of localities. These variants (J15 and J16) only differ at locus *D17Zt6*. Similarly, allele 62 of *D17Zt6* appears to define another small group of haplotypes (J6 and J7; Table 1). It has been formally identified (after breeding, see below) as the *t* allele of one J6 individual, and three other mice from two other localities probably carry this haplotype. Although allele 62 is clearly associated with the *t* chromosomes of the three mice from one locality, because it is homozygous in two of the three mice from this locality, these mice all share allele 50 at *D17Mit78* rather than allele 46, which defines the variant J6. This defines haplotype J7.

The most parsimonious explanation for the presence of a homozygote for allele 42 of *D17Mit125* and four homozygotes for allele 64 found in two samples from the same locality is the presence of a null allele at this locus in the *t* chromosomes of these mice. If these homozygotes are considered to represent null alleles, one obtains two new variants, J17 and J18. The haplotype J17 could have been obtained in a single recombination event between a variant carrying the null

allele (J18) and a haplotype related to the common J1 variant that was also found in the locality. If both alleles 42 and 64 are considered to be associated with *t* chromosomes, three instead of two new variants that only occur in this locality have to be postulated.

The homozygous alleles that define three other variants pose a similar problem but, as each one occurs in a single locality, there is no way of differentiating between the presence of new *t* alleles or null alleles. These variants are J3, J9 and J14, which we have considered to be defined by alleles 52 at *D17Mit78*, 54 at *D17Mit125* and 48 at *D17Zt6*, respectively.

The *t* haplotypes carried by the *t/t* mouse are very likely to be complementary because they are clearly very different. One could be the same as the J1 variant at all the loci making up the haplotype except *D17Zt6*. In this case, the other haplotype would be defined by 58 or 64 at *D17Zt6* and 48, 54, 48, 50 and 47, respectively, at the following five loci. Most of these alleles are associated with the *t* haplotype of at least one other mouse in our Danish sample or the reference panel of *t* haplotypes (see below). Only alleles 64 of *D17Zt6* and 47 of *D17Mit23* are not found in any of the other *t* haplotypes we analysed. It was not possible to assign alleles to two other *t* chromosomes, because they were not sufficiently related to any of the other *t* variants found in Jutland. They could either represent isolated *t* haplotypes or mosaic haplotypes derived from rare non-homologous recombination events between *t* haplotypes and wild-type chromosomes that are found in certain populations (Artzt *et al.*, 1985; Erhart *et al.*, 1989; Hammer *et al.*, 1991).

(iii) Comparison of the *t* variants from Jutland with a panel of known *t* haplotypes

A reference panel of wild-derived *t* haplotypes from various parts of the world were typed for the same loci we used to characterize the Jutland haplotypes (Table 1b). It was possible to identify the *t* haplotypes formally because either they were present on known genetic backgrounds or their genetic backgrounds were obviously similar to those of the other *t* haplotypes (*tTuw11*, *tTuw27*, *tTuw28* and *tlub4*).

At the three distal loci (*D17Mit82*, *D17Mit63* and *D17Mit23*), more alleles were found in the reference panel of haplotypes than in the samples from a small area of Jutland, which is to be expected for a sample representing such a wide geographical range. All the Jutland alleles at these loci are also found in one or the other reference haplotype. At the three proximal loci (*D17Zt6*, *D17Mit78* and *D17Mit125*), more variation was found and the Jutland samples display almost as many alleles as the reference panel, including a few that were not found in this panel. Even so, the number of alleles associated with *t* haplotypes of such

diverse origins is limited compared with the numerous alleles found in the wild-type chromosomes in the restricted area of Jutland we surveyed (Fig. 2).

This relatively small number of *t*-associated alleles contrasts with the large number of haplotypes that are defined by combinations of these alleles. In the reference panel, only two haplotypes (*tw32* and *t12*) have the same microsatellite haplotype. This provides further confirmation for the earlier suggestion that they were both derived from the same laboratory-derived chromosome (Ardlie & Silver, 1996; Artzt *et al.*, 1985; Uehara *et al.*, 1990). The fact that J4 in the Jutland sample and *tw73* have the same haplotype therefore suggests that they could have a recent common ancestor. This is quite plausible because *tw73* was extracted from a mouse caught in north-west Jutland (Dunn & Bennett, 1971) on the *M. m. musculus* side of the hybrid zone. This haplotype carries the lethal gene (*tclw73*) that is associated with all the *t* haplotypes that have been isolated so far in the regions occupied by *M. m. musculus* (Forejt *et al.*, 1988; Klein *et al.*, 1984; Ruvinsky *et al.*, 1991).

(iv) *The degree of polymorphism associated with the second t inversion*

We were not able to find a microsatellite marker on the second *t* inversion for which all the *t* alleles could be identified with a reasonable degree of confidence but the locus *D17Mit57*, which maps to the beginning of this inversion on the *t* haplotype (Fig. 1) allowed us to detect differences between the Jutland variants. Although we found a total of 14 alleles at this locus, most of the *t*-carrying mice appeared to be homozygotes. Because more than half the alleles contributed to these apparent homozygotes (Table 2), the locus is probably either absent or modified in most *t* chromosomes. This is also the case for most of the reference *t* haplotypes. However, a minority of the *t* haplotypes carry an amplifiable allele; in Jutland, this is either 38 or 44. These do not appear to be partial haplotypes because the complete *t* haplotypes *tw12* and *tw73* also have amplifiable alleles (44 and 50, respectively).

It was not possible to estimate the overall frequency of the haplotypes with amplifiable alleles homozygosity for alleles 38 and 44 occurred in 5 and 24 mice, respectively. Even so, sufficient unambiguous attributions could be made to establish that allele 44 was associated with a minority of J1, J6 and J12 variants. Allele 38 was only found in certain J4 variants.

(v) *Phylogenetic analysis*

It is evident from Table 1 that the phylogenetic content of the microsatellite data is rather poor and that

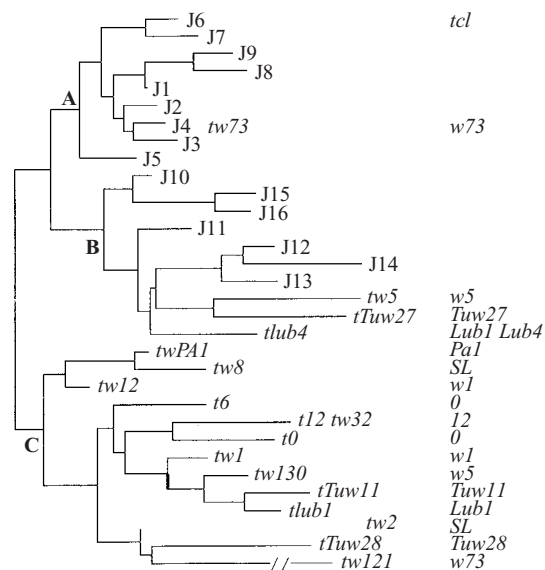


Fig. 3. Neighbour-joining tree built from the distance between haplotypes on the basis of their genotype at six microsatellite loci (Table 1).

the frequent incompatibilities between loci would generate homoplasies in a phylogenetic inference. However, the tree that we built using our measure of distances between haplotypes defines three major groups of haplotypes, designated A, B and C in Fig. 3, and these groups are robust to variations of the parameter α that defines the relative weight that is given to the proportion of differing loci and the size differences between alleles at each locus (for instance $\alpha = 0, 0.5$ or 1 , not shown). In fact, this trichotomy is entirely due to the alleles found at the last three loci that define the haplotype (Table 1). Group A is defined by allele 42 of *D17Mit82* and contains Jutland haplotypes J1–J9, and *tw73*, also from Jutland. The haplotypes in group B include all the remaining Jutland haplotypes (J10–J16), *tw5* from the USA, *Tuw27* from western Germany and *tlub4* from northern Italy, which all have allele 30 at *D17Mit82*. Variants J17 and J18, which appear to be related to group B, were not included in this tree because there is no obvious way of treating the null allele that is thought to occur at locus *D17Mit125*. The haplotypes in the third group are all of *domesticus* origin (from Europe and America) except for *tw121*, which was sampled in Tajikistan and is only distantly related to all other haplotypes. Thus, the tree obtained shows some geographic coherence because the 16 Jutland haplotypes, which come from a restricted geographical region, hardly mix with the panel of haplotypes of diverse geographical origin. This geographic coherence could result from a common origin of the Jutland haplotypes or from a common history of recombination between a small number of founders that colonized Jutland (or both).

Table 3. Estimating the distortion (d) and viability (v), and the bounds of the two-log likelihood intervals below the maximum, in crosses to test distortion (a), lethality (b) and complementation (c)

Parents: t variant (numbers of mice)			Offspring			d (two-log-likelihood bounds)	v (two-log-likelihood bounds)
Male	Female	N	t/t	t/w	w/w		
(a) Distortion tests (male $t/w \times$ female w/w)							
J4 (16516)	BALB/c	9	–	6	3	0.67 (0.34–0.90)	
J4 (16518)	BALB/c	41	–	31	10	0.76 (0.61–0.87)	
J12 (16618)	BALB/c	41	–	35	6	0.85 (0.73–0.94)	
J12 (16619)	BALB/c	46	–	32	14	0.70 (0.56–0.81)	
J6 (16584)	BALB/c	58	–	54	4	0.93 (0.85–0.97)	
(b) Lethality tests (male $t1/w \times$ female $t1/w$)							
J4 (16518)	J4 (16516)	28	0	25	3	0.88 (0.66–0.97)	0 (0.00–0.09)
J12 (16618)	J12 (16619)	11	0	11	0	1.00 (0.81–1.00)	0 (0.00–0.19)
J12 (16619)	J12 (16619)	35	0	26	9	0.65 (0.28–0.83)	0 (0.00–0.12)
(c) Complementation tests (male $t1/w \times$ female $t2/w$)							
J4 (16518)	J12 (16619)	27	0	17	10	0.41 (0.00–0.74)	0 (0.00–0.29)
J12 (16618)	J4 (16518)	45	0	40	5	0.87 (0.71–0.95)	0 (0.00–0.05)
J12 (16619)	J4 (16516)	10	0	9	1	0.89 (0.40–0.99)	0 (0.00–0.27)
J12 (16619)	J10 (16600)	20	0	18	2	0.89 (0.61–0.98)	0 (0.00–0.13)
J6 (16584)	J4 (16518)	57	15	33	9	0.73 (0.55–0.86)	0.63 (0.34–1.00)
J6 (16584)	J10 (16600)	15	6	8	1	0.87 (0.54–0.99)	0.86 (0.34–1.00)
J6 (16584)	J12 (16619)	26	10	15	1	0.93 (0.73–0.99)	0.71 (0.34–1.00)

(vi) *Transmission distortion and complementation properties of the t variants available for breeding experiments*

It should first be mentioned that the t chromosomes tested in the crossing experiments described in this section became available on a BALB/c genetic background, so it was possible to identify their microsatellite haplotypes formally. We were therefore able to confirm that, in these particular cases, the allele designations we deduced using the procedure described above were correct. The transmission ratios of the t haplotypes by five males carrying one of three different Jutland variants (J4, J6 and J12) were measured in a series of crosses with BALB/c females (Table 3a). To obtain statistically significant results, the offspring obtained in the crosses with BALB/c between both the original male and its male t/w F1 sons were pooled. The t haplotypes tested show transmission ratios ranging from 67% to 93% and the distortion is significant in all crosses but one, which led too few offspring. Table 3b shows the results of the viability tests that could be performed on haplotypes J4 (from two different wild mice) and J12 (either from the same or two different mice; Table 3). No t/t offspring were observed, and the two-log likelihood intervals on the estimate of the viability of t/t offspring (v) indicate that it is likely to be very low if not null. Therefore, both haplotypes J4 and J12 presumably carry lethal genes. Unfortunately, for technical reasons, we were unable to perform viability tests on J6.

We were also able to perform complementation tests between a certain number of pairwise combinations of haplotypes (each extracted from either a single or several original wild mice; Table 3c). J4 and J12 were confronted in three different crosses, and no t/t offspring were observed in a total of 82, which reinforces the presumption that both these haplotypes are lethal. It also provides strong evidence that they share a lethal allele for at least one gene. No t/t individuals were observed among the progeny of the J10 \times J12 cross, for which fewer offspring were obtained (20), so it is possible that these two haplotypes also share a lethal allele. By contrast, J6 was found to complement with J4, J10 and J12. Although the estimates of the viabilities of the t/t homozygotes are not very accurate, they are all relatively high. In addition, the lethality and complementation tests confirmed that all haplotypes are transmission distorters (a single exception is one complementation test involving a J4 male, but J4 is clearly a distorter in several other crosses).

Overall, we can conclude from these crosses that (i) all haplotypes tested (J4, J6 and J12) are distorters; (ii) J4, J10 and J12 are lethal; (iii) J12 shares one lethal allele with J4 and one with J10; (iv) J6 does not share a lethal allele with either J4, J10 or J12.

4. Discussion

These results show that it is not only possible to identify the t haplotypes in a population of wild

mice using two *t*-specific PCR markers but that it is also possible to characterize different *t*-haplotype variants in this population using nonspecific microsatellite loci, provided that a sufficient number of individuals are analysed. Our approach, which is based on the presence of obvious linkage disequilibria between groups of alleles at a series of different loci, was possible because the number of alleles associated with *t* haplotypes is relatively restricted compared with the total number found in the population as a whole. In these conditions, we were able to characterize with a reasonable degree of probability all but four of the *t* haplotypes identified by the two diagnostic markers. However, although we have seen that the range of alleles that are associated not only with the Jutland haplotypes but also with the panel of known haplotypes we analysed is relatively limited, the huge variation prevailing among the wild-type chromosomes make it unlikely that microsatellite loci can be used to characterize isolated *t* haplotypes drawn from a large geographical area unless they are first introduced into a known genetic background.

The fact that the Jutland haplotypes show little variation at the three distal loci (Table 1) is unlikely to be due to a general lack of variability of these loci because *D17Mit82* and *D17Mit63* show extensive variation in both the panel of *t* haplotypes (Table 1) and the wild Jutland mice (Fig. 2). The simplest explanation is that the Jutland haplotypes are derived from a small number of founders, but this has to be reconciled with the substantial variation observed at the three proximal loci. It is possible that some of this variation was generated *de novo* by mutation, especially at the loci *D17Mit78* and *D17Zt6*, at which most shared alleles differ by only two or four base pairs. However, repeated mutation is unlikely to have generated the alleles of very different sizes observed at the third locus, *D17Mit125*. It is therefore more parsimonious to assume that recombination between complementing haplotypes is responsible for at least part of the observed haplotypic variation. The increase in haplotypic variation that is obtained when the locus *D17Mit57* on the second inversion is added the central haplotype reinforces this idea. We also found variation among the different Jutland haplotypes at the locus *D17Mit59* on the distal end of the fourth inversion (results not shown). At this locus, more than three alleles are needed to define all the *t* chromosomes, which means that the region including the loci *D17Mit63* and *D17Mit23* that is shared by all the Jutland haplotypes does not extend to the end of the fourth inversion.

Another indication that recombination played a role in the history of these Jutland haplotypes comes from the complementation experiments. Although haplotype J12 shares a lethal gene with J4 and J10, the

former falls into group A of haplotypes, whereas the other two are in group B (Fig. 3). By contrast, J6 and J4, which both belong to the A group and have quite similar haplotypes, clearly belong to a different complementation groups. Thus, both the homoplasmy at the microsatellite loci and the lack of correspondence between the molecular and complementation classifications of haplotypes are strong indications of a history of recombination between the original haplotypes introduced in this region. It follows that the classification on the basis of the molecular markers, such as summarized in Fig. 3, unfortunately does not reflect the biological properties of the haplotypes. Under these conditions, it could also be that more variability originally existed at the three distal loci but was selectively lost after it was shuffled by such recombination events.

Many of the *t* chromosomes in the reference panel were typed in earlier studies for a series of different markers that are able to detect polymorphisms between certain *t* haplotypes. These include the regional class1 *H2* serological haplotypes defined by Artzt *et al.* (1985) and the RFLP found at the loci *Cryal* and *Pim1* (Neufeld *et al.*, 1991), which all map to the same region as the six microsatellite loci we used. When these three different types of markers, which do not have the same modes of mutation, are combined to give the composite haplotypes shown in Table 4, they provide further strong evidence that frequent recombination between complementing haplotypes regularly shuffles the mutations that occur in different *t*-chromosomes lineages. Their most striking feature is the overlapping patchwork of homologies that occur between the different *t* haplotypes. These are not necessarily correlated with the lethal genes they carry and, depending on the part of the haplotype considered, a given haplotype can share homologies with a series of different *t* chromosomes that carry the same or different lethal genes. For example, the haplotypes *tw1* and *tw12*, which both belong to the complementation group *w1*, have an identical *H2* regional haplotype but differ at three out the next four loci. The haplotype *tw1* has the same haplotype as *Lub1* from the middle of the *H2* region and *D17Mit23*, whereas *tw12* and *t6* are the same in the region covering *D17Mit78* to *D17Mit63*.

The many related *t* variants found in the Jutland population suggest that recombination occurs relatively frequently in wild populations and that the recombined haplotypes often persist in populations in which different lethal genes come into contact. Another factor that might have contributed to the presence of a large number of variants in the transect across the hybrid zone we sampled is the probable presence of the *musculus* lethal gene *tclw73* in some of the variants. Unlike the lethal genes found in *domesticus* populations that nearly all map to the fourth

Table 4. The genotypes of 12 *t* haplotypes characterized at 11 loci in the third and fourth chromosomal inversions

Haplotype	Origin	<i>tcl</i>	Regional H2 haplotype							<i>D17Mit63</i>	<i>Cryal</i>	<i>D17Mit23</i>	<i>Pim1</i>
			<i>D17Z16</i>	<i>D17Mit78</i>	<i>D17Mit125</i>	D region	Middle	K region	<i>D17Mit82</i>				
<i>tw2</i>	USA	S L	68	52	36	A	E	H	48	42	t	50	t
<i>tw8</i>	USA	S L	68	46	54	A	E	H	48	42	t	53	b
<i>twPA1</i>	France	<i>Pa1</i>	68	46	50	A	E	H	48	42	t	53	-
<i>t0</i>	France	0	70	46	50	A	E	H	48	44	a	50	t
<i>t6</i>	Laboratory	0	72	46	50	B	E	I	42	42	a	50	b
<i>tw1</i>	USA	<i>w1</i>	68	46	50	B	E	I	44	42	a	50	b
<i>tw12</i>	USA	<i>w1</i>	68	46	50	B	E	I	42	42	t	53	b
<i>tlub1</i>	Italy	<i>Lub1</i>	66	46	46	A	E	I	44	42	a	50	t
<i>t12, tw32</i>	USA	<i>12</i>	58	46	50	A	E	G	48	36	t	50	t
<i>tw73</i>	Jutland	<i>w73</i>	66	50	50	C	E	I	42	40	t	53	b
<i>tw121</i>	Tajikistan	<i>w73</i>	68	50	50	C	E	H	62	72	b	50	-
<i>tw5</i>	USA	<i>w5</i>	58	42	50	D	F	J	30	46	a	53	b

inversion, it is situated on the second inversion of the *t* haplotype (Fig. 1). The distance between these two groups of lethal genes would increase the probability of a recombination occurring in between them. Complex *t* haplotypes carrying the *musculus* lethal gene and one of a number of *domesticus*-type lethal genes (*tcl0*, *tclw1*, *tclPa1* or *tclw5*) have been reported in different regions of the *M. m. musculus* territory (Dunn et al., 1973; Ruvinsky et al., 1991) and it is interesting to note that one of them, *tw74* (*tclw73* and *tclw5*) was found to the north of the Jutland hybrid zone transect analysed in this study (Dunn & Bennett, 1971). However, not all the wild-derived complex *t* haplotypes are associated with *tclw73*. In the haplotype *tw75*, from central Germany, the lethal gene *tclw5* is associated with *tclw1* (Artzt et al., 1979) and in *tw120*, isolated in the state of New York, it is associated with *tclw0* (Artzt et al., 1985). Another complex haplotype, *tLub4* from northern Italy carries the lethal genes *tclLub1* and *tclLub4* (Golubic et al., 1985) and the lethal genes *tclTuw25* and *tclTuw2* are found in *tTuw2* extracted from *domesticus* populations in southern Germany (Klein et al., 1984). There is still relatively little information available about the frequency of such exchanges in natural populations, because their detection requires laborious progeny testing. Because no molecular markers are available for the different lethal genes, molecular characterization of the *t* haplotypes coming from a wide variety of wild mice populations belonging to the different subspecies followed by complementation studies between representative variants drawn from each of these populations are needed in order to answer this question.

The transmission distortion of a given *t* haplotype results from the interaction between a receptor gene and a series of distorter genes spread over the entire *t* chromosome, and so crossovers between two complementing *t* haplotypes within the *t* complex will often lead to an exchange of genes involved in the complex distortion mechanism. The tendency for certain of the less-frequent *t* variants we found in Jutland to be localized to particular areas of the transect suggests that recombination is able to generate new haplotypes that are able to compete with the haplotypes originally present in the area and to establish themselves in the population. The geographical distribution of the different *t* variants, the degree of infection of the different localities sampled and their position on the genetic cline between the two subspecies will be presented and discussed in a subsequent paper.

K. Artzt and K. Ardlie kindly provided the DNAs in the panel of known *t* haplotypes. K. Belkhir provided invaluable help by setting up the genotype database. The continued help of T. Nielsen made it possible to obtain the mice in Denmark. This study was partly funded by the CNRS GDR1928 ('Evolution des g nomes dans les populations').

References

- Ardlie, K. G. & Silver, L. M. (1996). Recent evolution of mouse *t* haplotypes at polymorphic microsatellites associated with the *t* complex responder (*Tcr*) locus. *Genetical Research* **67**, 1–10.
- Artzt, K., Barbiarz, B. & Bennett, D. (1979). A *t*-haplotype (*t^{w75}*) overlapping two complementation groups. *Genetical Research* **33**, 279–285.
- Artzt, K., Shin, H. S., Bennett, D. & Dimeotalento, A. (1985). Analysis of major histocompatibility complex haplotypes of *t*-chromosomes reveals that the majority of diversity is generated by recombination. *Journal of Experimental Medicine* **162**, 93–104.
- Bennett, D. (1975). The *t* locus of the mouse: a review. *Cell* **6**, 441–454.
- Delarbre, C., Kashi, Y., Boursot, P., Beckmann, J. S., Kourilsky, P., Bonhomme, F. & Gachelin, G. (1988). Phylogenetic distribution in the genus *Mus* of *t*-complex-specific DNA and protein markers: inferences on the origin of *t*-haplotypes. *Molecular Biology and Evolution* **5**, 120–133.
- Dietrich, W. F., Miller, J. C., Steen, R. G., Merchant, M., Damron, D., Nahf, R., Gross, A., Joyce, D. C., Wessel, M., Dredge, R. D., Marquis, A., Stein, L. D., Goodman, N., Page, D. C. & Lander, E. S. (1994). A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genetics* **7**, 220–245.
- Dod, B., Jermiin, L. S., Boursot, P., Chapman, V. M., Nielsen, J. T. & Bonhomme, F. (1993). Counterselection on sex chromosomes in the *Mus musculus* European hybrid zone. *Journal of Evolutionary Biology* **6**, 529–546.
- Dunn, L. C. & Bennett, B. D. (1971). Polymorphisms for lethal alleles in European populations of *Mus Musculus*. *Evolution* **25**, 451–453.
- Dunn, L. C., Bennett, B. D. & Cookingham, J. (1973). Polymorphisms for lethal alleles in European populations of *Mus Musculus*. *Journal of Mammalogy* **54**, 822–830.
- Ebersole, T., Lai, F. & Artzt, K. (1992). New molecular markers for the distal end of the *t*-complex and their relationships to mutations affecting mouse development. *Genetics* **131**, 175–182.
- Erhart, M. A., Phillips, S. J., Bonhomme, F., Boursot, P., Wakeland, E. K. & Nadeau, J. H. (1989). Haplotypes that are mosaic for wild-type and *t* complex-specific alleles in wild mice. *Genetics* **123**, 405–415.
- Figueroa, F., Golubic, M., Nizetic, D. & Klein, J. (1985). Evolution of mouse major histocompatibility complex genes borne by *t* chromosomes. *Proceedings of the National Academy of Sciences of the USA* **82**, 2819–2823.
- Figueroa, F., Neufeld, E., Ritte, U. & Klein, J. (1988). *t*-specific DNA polymorphism among wild mice from Israel and Spain. *Genetics* **119**, 157–160.
- Forejt, I., Gregorova, S. & Jansa, P. (1988). Three new *t*-haplotypes of *Mus musculus* reveal structural similarities to *t*-haplotypes of *Mus domesticus*. *Genetical Research* **51**, 111–119.
- Forejt, J. & Ivanyi, P. (1975). Genetic studies on male sterility of hybrids between laboratory and wild mice (*Mus musculus* L.). *Genetical Research* **24**, 189–206.
- Fox, H. S., Martin, G. R., Lyon, M. F., Herrmann, B., Frischauf, A. M., Lehrach, H. & Silver, L. M. (1985). Molecular probes define different regions of the mouse *t* complex. *Cell* **40**, 63–69.
- Golubic, M., Budimir, O., Schoepfer, R., Kasahara, M., Mayer, W. E., Figueroa, F. & Klein, J. (1987). Nucleotide sequence analysis of class II genes borne by mouse *t* chromosomes. *Genetical Research* **50**, 137–146.
- Golubic, M., Figueroa, F., Tosi, M. & Klein, J. (1985). Restriction fragment length polymorphism of C4 genes in mice with *t* chromosomes. *Immunogenetics* **21**, 247–256.
- Gregorova, S., Mnukova-Fajdelova, M., Trachtulec, Z., Capkova, J., Loudova, M., Hoglund, M., Hamvas, R., Lehrach, H., Vincek, V., Klein, J. & Forejt, J. (1996). Sub-milliMorgan map of the proximal part of mouse chromosome 17 including the *hybrid sterility 1* gene. *Mammalian Genome* **7**, 107–113.
- Hammer, M. F., Bliss, S. & Silver, L. M. (1991). Genetic exchange across a paracentric inversion of the mouse *t* complex. *Genetics* **128**, 799–812.
- Hammer, M. F., Schimenti, J. & Silver, L. M. (1989). Evolution of mouse chromosome 17 and the origin of inversions associated with *t* haplotypes. *Proceedings of the National Academy of Sciences of the USA* **86**, 3261–3265.
- Hammer, M. F. & Silver, L. M. (1993). Phylogenetic analysis of the alpha-globin pseudogene-4 (*Hba-ps4*) locus in the house mouse species complex reveals a stepwise evolution of *t* haplotypes. *Molecular Biology and Evolution* **10**, 971–1001.
- Herrmann, B. G., Barlow, D. P. & Lehrach, H. (1987). A large inverted duplication allows homologous recombination between chromosomes heterozygous for the proximal *t* complex inversion. *Cell* **48**, 813–825.
- Horiuchi, Y., Agulnik, A., Figueroa, F., Tichy, H. & Klein, J. (1992). Polymorphisms distinguishing different mouse species and *t* haplotypes. *Genetical Research* **60**, 43–52.
- Klein, J., Sipos, P. & Figueroa, F. (1984). Polymorphism of *t*-complex genes in European wild mice. *Genetical Research* **44**, 39–46.
- Kumar, S., Tamura, K., Jakobsen, I. & Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.
- Lai, F. & Artzt, K. (1992). Map positions of four dinucleotide repeats in the mouse *t* complex. *Mammalian Genome* **3**, 476–477.
- Lyon, M. F. (2000). An answer to a complex problem: cloning the mouse *t*-complex responder. *Mammalian Genome* **11**, 817–819.
- Morita, T., Kubota, H., Murata, K., Nozaki, M., Delarbre, C., Willison, K., Satta, Y., Sakaizumi, M., Takahata, N., Gachelin, G. & Matsushiro, A. (1992). Evolution of the mouse *t* haplotype: recent and worldwide introgression to *Mus musculus*. *Proceedings of the National Academy of Sciences of the USA* **89**, 6851–6855.
- Morita, T., Murata, K., Sakaizumi, M., Kubota, H., Delarbre, C., Gachelin, G., Willison, K. & Matsushiro, A. (1993). Mouse *t* haplotype-specific double insertion of B2 repetitive sequences in the *Tcp-1* intron 7. *Mammalian Genome* **4**, 58–59.
- Neufeld, E., Vincek, V., Figueroa, F. & Klein, J. (1991). Limits of the distal inversion in the *t* complex of the house mouse: evidence from linkage disequilibria. *Mammalian Genome* **1**, 242–248.
- Nizetic, D., Figueroa, F. & Klein, J. (1984). Evolutionary relationships between the *t* and *H-2* haplotypes in the house mouse. *Immunogenetics* **19**, 311–320.
- Ruvinsky, A., Polyakov, A., Agulnik, A., Tichy, H., Figueroa, F. & Klein, J. (1991). Low diversity of *t* haplotypes in the eastern form of the house mouse, *Mus musculus* L. *Genetics* **127**, 161–168.
- Schimenti, J., Vold, L., Socolow, D. & Silver, L. M. (1987). An unstable family of large DNA elements in the center of the mouse *t* complex. *Journal of Molecular Biology* **194**, 583–594.

- Schimenti, J. a. H. M. (1990). Rapid identification of mouse *t* haplotypes by PCR polymorphism. *Mouse Genome* **87**, 108.
- Silver, L. M., Hammer, M., Fox, H., Garrels, J., Bucan, M., Herrmann, B., Frischauf, A.-M., Lehrach, H., Winking, H., Figueroa, F. & Klein, J. (1987). Molecular evidence for the rapid propagation of mouse *t* haplotypes from a single, recent, ancestral chromosome. *Molecular Biology and Evolution* **4**, 473–482.
- Tucker, P. K., Sage, R. D., Warner, J., Wilson, A. C. & Eicher, E. M. (1992). Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution* **46**, 1146–1163.
- Uehara, H., Abe, K., Flaherty, L., Bennett, D. & Artzt, K. (1991). Molecular organization of the *D-Qa* region of *t*-haplotypes suggests that recombination is an important mechanism for generating genetic diversity of the major histocompatibility complex. *Mammalian Genome* **1**, 92–99.
- Uehara, H., Abe, K., Park, C. H., Shin, H. S., Bennett, D. & Artzt, K. (1987). The molecular organization of the *H-2K* region of two *t*-haplotypes: implications for the evolution of genetic diversity. *EMBO Journal* **6**, 83–90.
- Uehara, H., Ebersole, T., Bennett, D. & Artzt, K. (1990). Submegabase clusters of unstable tandem repeats unique to the *Tla* region of mouse *t* haplotypes. *Genetics* **126**, 1093–1102.
- Willison, K. R., Dudley, K. & Potter, J. (1986). Molecular cloning and sequence analysis of a haploid expressed gene encoding *t* complex polypeptide 1. *Cell* **44**, 727–738.