Mitochondrial DNA Variation and the Evolution of Robertsonian Chromosomal Races of House Mice, *Mus domesticus*

Michael W. Nachman,* Sarah N. Boyer,* Jeremy B. Searle[†] and Charles F. Aquadro*

*Section of Genetics and Development, Biotechnology Building, Cornell University, Ithaca, New York 14853, and
[†]Biology Department, University of York, Heslington, York YO1 5DD, England

ABSTRACT

The house mouse, *Mus domesticus*, includes many distinct Robertsonian (Rb) chromosomal races with diploid numbers from $2n = 22$ to $2n = 38$. Although these races are highly differentiated karyotypically, they are otherwise indistinguishable from standard karyotype (*i.e.*, $2n = 40$) mice, and consequently their evolutionary histories are not well understood. We have examined mitochondrial DNA (mtDNA) sequence variation from the control region and the ND3 gene region among 56 *M. domesticus* from Western Europe, including 15 Rb populations and 13 standard karyotype populations, and two individuals of the sister species, *Mus musculus*. mtDNA exhibited an average sequence divergence of 0.84% within *M. domesticus* and 3.4% between *M. domesticus* and *M. musculus*. The transition/transversion bias for the regions sequenced is 5.7:1, and the overall rate of sequence evolution is approximately 10% divergence per million years. The amount of mtDNA variation was as great among different Rb races as among different populations of standard karyotype mice, suggesting that different Rb races do not derive from a single recent maternal lineage. Phylogenetic analysis of the mtDNA sequences resulted in a parsimony tree which contained six major clades. Each of these clades contained both Rb and standard karyotype mice, consistent with the hypothesis that Rb races have arisen independently multiple times. Discordance between phylogeny and geography was attributable to ancestral polymorphism as a consequence of the recent colonization of Western Europe by mice. Two major mtDNA lineages were geographically localized and contained both Rb and standard karyotype mice. The age of these lineages suggests that mice have moved into Europe only within the last 10,000 years and that Rb populations in different geographic regions arose during this time.

A N important goal in evolutionary genetics is to understand the nature of genetic differences that distinguish recently separated populations or species. Chromosomal rearrangements may be of particular interest because of their potential for reducing gene flow and thereby promoting speciation (for reviews, see WHITE 1978; SITES and MORITZ 1987). One extreme example of intraspecific chromosomal differentiation involves the Western European house mouse, *Mus domesticus* (also known as *Mus musculus domesticus*; see AUFFRAY *et al.* 1990). The standard karyotype of this species has 40 acrocentric chromosomes. However, there are numerous populations with lower diploid numbers as a result of Robertsonian (Rb) whole-arm translocations (also referred to as fusions). The first Rb population was discovered in Switzerland and described as the Tobacco mouse, *Mus poschiavinus* (GROPP *et al.* 1969). It is now evident that the Tobacco mouse is just 1 of over 20 Rb populations from Western Europe and the Mediterranean region.

Most Rb populations occupy fairly restricted geographic areas and most individuals within these populations are homozygous with respect to their Rb rearrangements. Rb populations are typically surrounded by populations of $2n = 40$ (*i.e.*, standard karyotype) mice,

and hybrid zones are formed where the two come into contact (SPIRITO *et al.* 1980; CORTI *et al.* 1990; SEARLE 1991; SAID and BRITTON-DAVIDIAN 1991). Extensive laboratory studies have documented substantially decreased fertility in the progeny of crosses between Rb and standard karyotype mice (*e.g.*, GROPP and WINKING 1981), although it is unclear whether this level of underdominance extends to natural hybrids (WINKING *et al.* 1988; SEARLE 1993). Some fusions are widely distributed among geographically isolated Rb populations, although most Rb populations also contain one to several unique fusions. There are also several systems in which nearby Rb populations share common sets of fusions, although individual populations within such systems usually have some unique fusions (CAPANNA 1982). Altogether, 79 fusions have been documented in natural populations (BAUCHAU 1990), involving all autosomes except number 19.

The fossil record for *M. domesticus* is unusually detailed and suggests that house mice reinvaded Western Europe from the Middle East following the spread of agriculture within the last 8,000 years (AUFFRAY *et al.* 1990). Because Rb populations are not found in the Middle East, they are hypothesized to have originated since the time mice reentered Europe. The appearance
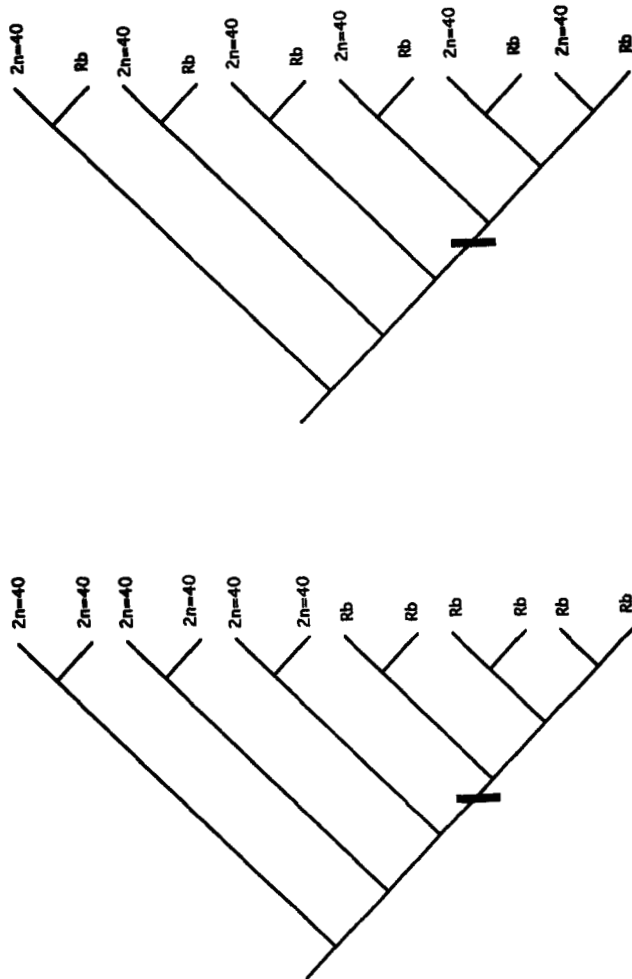
FIGURE 1.—Competing hypotheses for the origin of Rb populations. (Top) Multiple independent origins would result in a clade with both Rb and standard karyotype animals. (Bottom) Single origin would result in a clade consisting solely of Rb animals.

of so many fusions in such a short time represents an unusual pattern of chromosomal change and suggests that there may be some mechanism (*e.g.*, a mutator locus or transposable element) responsible for generating these changes.

Hypotheses to explain the evolutionary origin of Rb races have taken two forms (Figure 1), although the true history of Rb mice may lie somewhere between these extremes. The traditional view is that many or most Rb populations arose independently of each other, each presumably from a nearby standard karyotype population (SAGE 1981; BRITTON-DAVIDIAN *et al.* 1989). The fact that most populations have a unique set of fusions is consistent with this view. This hypothesis also accommodates the widespread and disjunct geographic distribution of different Rb populations. However, if each Rb population had a unique origin, then identical fusions have presumably evolved independently many times. This hypothesis suggests that the mechanism responsible for producing the Rb translocations is present throughout the species.

An alternative view is that many or most Rb populations arose in just one or a few places from which they spread to their present locations (*e.g.*, TICHY and VUCAK 1987; WINKING *et al.* 1988; BAUCHAU 1990). This hypothesis requires dispersal over long distances. However, such movements may be quite common in house mice because they are commensal with humans (SAGE 1981). Because many Rb populations carry unique fusions, this view also implies that the mice which gave rise to present day Rb populations carried either few fusions (those that are widely shared) or no fusions (but still carried the fusion-producing mechanism). This hypothesis suggests that all Rb populations may be more closely related to each other than any are to a standard karyotype population. Consequently, this hypothesis implies that the Rb-producing mechanism resides in a single lineage within *M. domesticus*.

Distinguishing between these hypotheses has been difficult, in part because of the lack of substantial genetic differentiation between Rb and standard karyotype populations. Support for the multiple origin hypothesis is provided by surveys of allozyme variation (BRITTON-DAVIDIAN *et al.* 1989) which show that Rb populations as a group do not exhibit substantially less variation than standard karyotype populations. Evidence suggesting that some distinct Rb populations may be more closely related to each other than to standard karyotype populations comes from a phylogenetic study based on *Y* chromosome RFLPs (TUCKER *et al.* 1989) which showed that several, but not all, geographically disjunct Rb populations fall into a single clade.

Here we investigate how mitochondrial DNA (mtDNA), a rapidly evolving genetic marker, may help resolve this issue. Earlier mtDNA studies of mice have focused primarily on the standard karyotype populations of *M. domesticus* (FERRIS *et al.* 1983; SAGE *et al.* 1990) or on the relationships between *M. domesticus* and *M. musculus* (PRAGER *et al.* 1993), although these studies also included a few presumptive Rb mice and found that they did not cluster on one mtDNA clade. In this paper we report on sequence variation from the control region and from the ND3 gene from 56 *M. domesticus*, representing 15 Rb populations and 13 standard karyotype populations, and from 2 *Mus musculus*. These data are used to address the pattern of mtDNA evolution in mice, the colonization and evolution of mice in Europe, and the origin of Rb races.

MATERIALS AND METHODS

Specimens: Mice used in this study came from field trapping or from laboratory colonies derived from wild mice. *M. domesticus* were sampled from 15 Rb populations and 13 standard karyotype populations. In all cases, two mice per locality were sampled. In most cases, populations were sampled in pairs representing one Rb population and a nearby standard
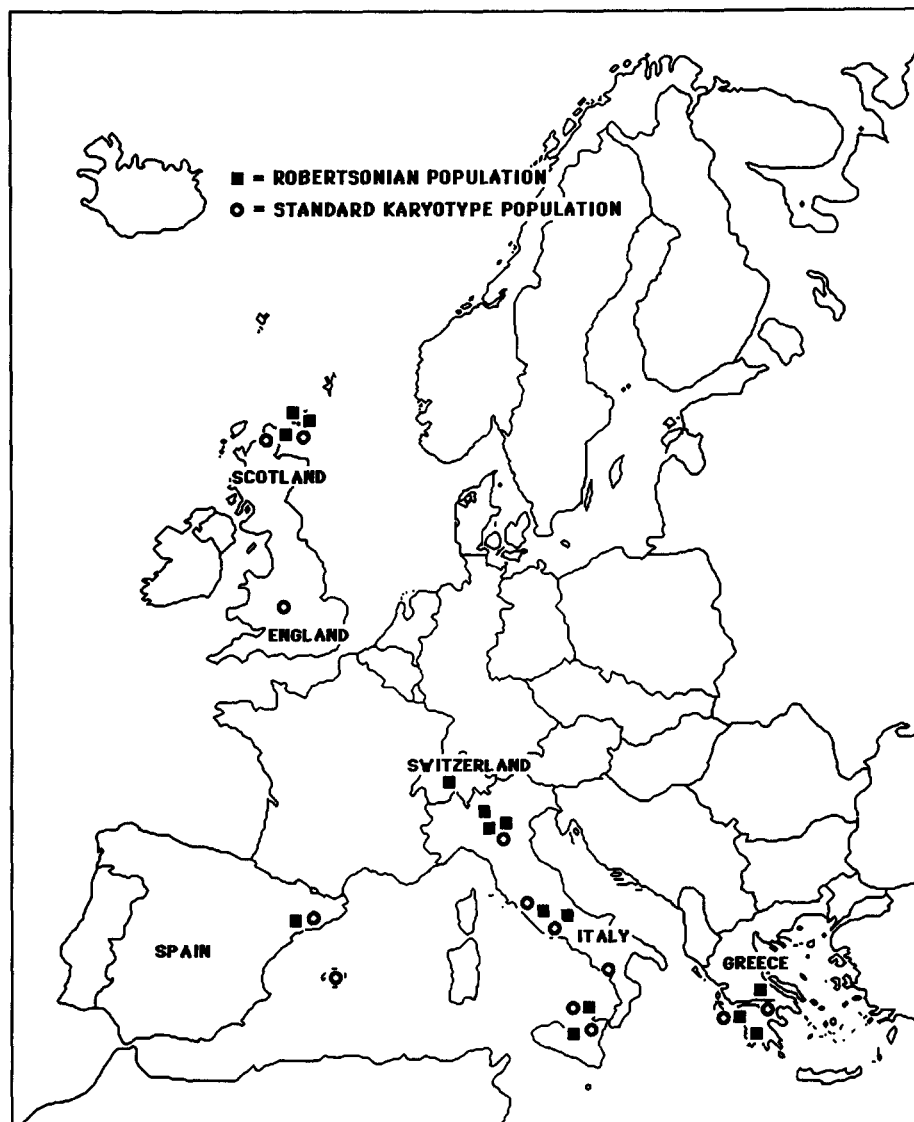
FIGURE 2.—Collecting localities for animals included in this study.

karyotype population (Figure 2). Two *M. musculus*, collected in Prague, were used as an outgroup. The original collecting localities, numbers of individuals, and specimen identification numbers are given in Table 1. All specimens were deposited in the collections of the Museum of Zoology at The University of Michigan.

Chromosome analysis: Mice were captured alive and karyotypes were prepared in the field from bone marrow suspensions following PATTON (1967). To identify the individual chromosome arms involved in the fusions, G-banded chromosomes were prepared following LEVERSHA *et al.* (1980).

DNA preparation: Genomic DNA was prepared from frozen spleen or liver tissue following SAMBROOK *et al.* (1989) with modifications. Tissues were ground in liquid nitrogen and the resulting powder was suspended in 10 ml of extraction buffer (10 mM Tris, pH 8.0, 100 mM EDTA, 0.5% sodium dodecyl sulfate, 20 µg/ml RNase) and incubated at 55° for 1 hr. Proteinase K (100 µg/ml) was added, and the solution was incubated for up to 12 hr at 55° and extracted three times with phenol/chloroform. DNA was ethanol-precipitated, dried and resuspended in Tris-EDTA (pH 8.0) to a final concentration of 500 µg/ml.

Polymerase chain reaction (PCR) amplification: Amplification of double-stranded DNA was performed using PCR (SAIKI

et al. 1986, 1988). Two different mitochondrial DNA regions were amplified: a 1100-bp fragment encompassing the control region, and a 534-bp fragment including the ND3 gene and portions of the adjacent tRNApro and tRNAphe genes. Amplification primers were based on the "universal" primers of KOCHER *et al.* (1989) and slightly modified to correspond to the mouse mitochondrial sequence (BIBB *et al.*, 1981). The letters in the primer designations that follow correspond to the light (L) or heavy (H) strand of mitochondrial DNA, and the numbers correspond to the position of the 3' base of each primer in the sequence of BIBB *et al.* (1981). The control region was amplified with L15320 (5'-ATAAACATTACTCTGG-TCTTGTAAACC-3') and H00072 (5'-ATTAATTATAAGGC-CAGGACCAAACCT-3'). The ND3 gene was amplified with L9385 (5'-ACGTCTCCATTTATTGATGAGG-3') and H9876 (5'-GAGGTTGAAGAAGGTAGATGGC-3'). DNA was amplified in 35 cycles of 94° for 1 min, 55° for 1 min, and 72° for 3 min. Taq polymerase (Perkin-Elmer Cetus) was used in a 100-µl reaction volume with conditions as specified by the supplier, and the reaction mixture was overlayed with mineral oil. Following the reaction, oil was removed using a chloroform extraction, and the double-stranded PCR product was precipitated with 33 µl ammonium acetate (10 M) and 133 µl 100% cold ethanol, washed once in 80% ethanol, and resuspended

TABLE 1

Collecting localities and specimens of *M. domesticus* analyzed

Scotland
   John O'Groats, Northern Scotland ($2n = 32$); 1356 m, 1357 f
   Eday, Orkney Islands ($2n = 34$); 1334 m, 1335 f
   Westray, Orkney Islands ($2n = 36$); 1345 m, 1346 f
   Papa Westray, Orkney Islands ($2n = 40$); 1328 m, 1329 f
   Ribigill, Northern Scotland ($2n = 40$); 1322 m, 1323 f
England
   Birmingham ($2n = 40$); 1316 m, 1317 f
Switzerland
   Preonzo ($2n = 28$); 1366 m, 1367 f
Northern Italy
   Sernio, Upper Valtellina Valley ($2n = 24$); 1008 m, 1009 f
   Cremona ($2n = 22$); 1372 f, 1373 m
   Binasco ($2n = 24$); 1368 m, 1369 f
   Canova, Staffora Valley ($2n = 40$); 1036 m, 1037 m
Central Italy
   Bonefro, 6 km W. of town, Region Molise ($2n = 22$); 1128 f, 1129 m
   Torrita, Tronto Valley, Region Lazio ($2n = 24$); 1067 m, 1068 f
   Sette Vene, 3.5 km SW of town, Region Lazio ($2n = 40$); 1081 f,
      1083 m
   Cassino, 11.4 km WNW of train station, Region Lazio ($2n = 40$);
      1103 m, 1104 f
   Policastro, Region Campania ($2n = 40$); 1204 m, 1205 f
Southern Italy
   Lipari, Eolean Islands ($2n = 26$); 1225 m, 1226 f
   Castelbuono, 4 km N of town ($2n = 24$); 1262 m, 1263 f
   Salina, Eolean Islands ($2n = 40$); 1248 m, 1249 f
   Milazzo, 6 km S of town ($2n = 40$); 1213 m, 1214 f
Greece
   Niforeka, Vicinity of Patras, Peloponissos ($2n = 24$); 1150 f, 1151 m
   Patras, University of Patras, Peloponissos ($2n = 32$); 1173 f, 1174 m
   Thiva, 6 km SW of town, Sterea Elada ($2n = 28$); 1176 m, 1177 f
   Korinthos, 19 km W of town, Peloponissos ($2n = 40$); 1192 m,
      1193 f
   Zakinthos Island, 1 km SSW of Laganas ($2n = 40$); 1199 m, 1200 f
Spain
   Avinyonet, Catalunya ($2n = 30$); 1310 m, 1311 f
   La Roca (del Valles), Catalunya ($2n = 40$); 1286 m, 1287 f
   Mallorca Island, 6.5 km N of Colonia Sant Jordi ($2n = 40$); 1272 m,
      1273 f

Specimens and collector's notes have been deposited in the Museum of Zoology, University of Michigan. Numbers in parentheses refer to diploid number of chromosomes. Specimen numbers refer to collector's numbers; m, male; f, female.

in 15 µl double distilled H_2O for direct double-stranded sequencing.

DNA sequencing: Direct sequencing of double-stranded PCR products was done using the dideoxy chain termination method (SANGER *et al.* 1977) with Sequenase 2 enzyme and kit (BRL) according to the protocol supplied by the manufacturer with slight modifications. The sequencing primer was annealed to template DNA (annealing step of Sequenase protocol) by heating at $100°$ for 3 min, freezing directly in dry ice/ethanol bath, and allowing to thaw in presence of enzyme and elongation reagents. Sequencing primers were 17 mers and the following numbers correspond to the position of the 3′ base of each primer in the sequence of BIBB *et al.* (1981): control region, L15420, L15520, L15654, L15761, L15881, L16007, L16143; ND3 region, L9520, L9631, L9763. One strand only was sequenced in each individual (L strand) and compared to the published mouse sequence (BIBB *et al.* 1981). All polymorphisms were scored by running DNAs from each individual twice on a 5% acrylamide gel, once with the standard arrangement of four lanes (A, G, C, T) adjacent to each other, and once with each of the 4 bases from different individuals grouped together (Figure 3). The latter method provides an unambiguous way of efficiently detecting all polymorphisms in a sample.
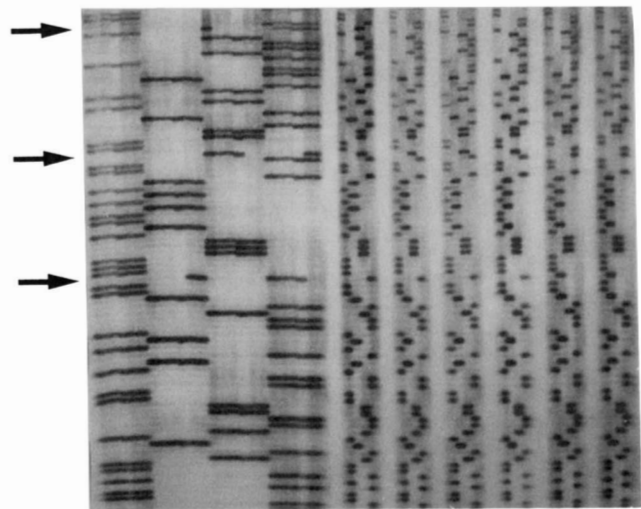

FIGURE 3.—DNA sequencing gel with individuals 1–6 scored for polymorphisms two different ways: grouped by base (left) and grouped by individual (right). Arrows mark polymorphic sites.

Data analysis: Sequences were aligned by hand, and the numbers and frequencies of all polymorphic sites were counted. Two different estimates of nucleotide variability (π and θ) were calculated. Under neutral, equilibrium conditions for mtDNA, π and θ are equal to the parameter for $2N\mu$, where N is the effective population size of females and µ is the neutral mutation rate. The average number of differences between any two sequences drawn from a sample, on a per site basis, called nucleotide diversity (NEI and LI 1979), is given by:

$$\pi = \sum_{ij} x_i x_j \pi_{ij}$$

where x_i and x_j are the frequencies of the ith and jth type of DNA sequences, and π_{ij} is the number of nucleotide differences between the ith and jth type of sequences divided by the length of the sequence (LI and GRAUR 1991). The parameter θ is given by:

$$\theta = p/A, \quad \text{where} \quad A = \sum_{i=1}^{n-1} 1/i$$

and p is the number of polymorphic sites in the sample divided by the length of the sequence (NEI 1987). Estimates of variability were based on the assumption of small levels of divergence, and no corrections for multiple hits were used.

Aligned sequences were analyzed cladistically using the PAUP (Phylogenetic Analysis Using Parsimony) program (version 3.0s; SWOFFORD 1991), using two *M. musculus* sequences as an outgroup. This program provides a hypothesis of relationships by searching for the network that requires the fewest number of character state changes to link all taxa. Heuristic searches were employed to find minimal length trees, and strict consensus trees were constructed from all minimal length trees. The program was run with equal character weightings and with different weights applied to insertion/deletions *vs.* base substitutions and with different weights for transitions *vs.* transversions. Consistency indices (KLUGE and FARRIS 1969) were calculated for insertion/deletions, transitions and transversions. The consistency index (CI) is defined as the number of character states minus the number of characters in a data set, divided by the total number of changes in

TABLE 2

Karyotypes of mice in this study with Robertsonian rearrangements

Scotland												
John O'Groats (2n = 32)				4.10		6.13			9.12		11.14	
Eday (2n = 34)			3.14	4.10					9.12			
Westray (2n = 36)						6.14			9.12			
Switzerland												
Preonzo (2n = 28)					5.13	6.7		8.15	9.14	10.12		16.17
Northern Italy												
Sernio (2n = 24)	1.3	2.8		4.6	5.15				9.14	10.12	11.13	16.17
Cremona (2n = 22)	1.6	2.8	3.4		5.15		7.18		9.14	10.12	11.13	16.17
Binasco (2n = 24)		2.8	3.4		5.15	6.7			9.14	10.12	11.13	16.17
Central Italy												
Bonefro (2n = 22)	1.18	2.17	3.13	4.11	5.15	6.7		8.14	9.16	10.12		
Torrita (2n = 24)	1.2		3.9	4.17	5.13	6.16		8.14		10.12	11.15	
Southern Italy												
Lipari (2n = 26)	1.2		3.9	4.13	5.14	6.16		8.12		10.15		
Castelbuono (2n = 24)		2.15	3.4		5.13	6.12	7.11	8.17	9.16	10.14		
Greece												
Niforeka (2n = 24)	1.3	2.5		4.6				8.12	9.16	10.14	11.17	13.15
Patras (2n = 32)								8.12	9.16	10.14		13.15
Thiva (2n = 28)		2.15		4.14	5.12	6.9		8.17		10.13		
Spain												
Avinyonet (2n = 30)				4.14	5.15	6.10			9.11			12.13

Two mice were sampled from each population. In all cases, mice within populations had the same karyotype. Numerical designations refer to autosomes involved in Rb rearrangements. For example, 1.3 indicates that chromosomes *1* and *3* are fused. All rearrangements were present as homozygotes. Autosomes not listed in the table were not involved in Rb rearrangements.

character states on a given tree. This index is a measure of the amount of homoplasy (convergent, reversal or parallel evolutionary changes) for a particular character (or set of characters) on a given tree, and as such, provides information about the underlying mutational process for different types of characters. A value of 1 is obtained when there is no homoplasy on the tree, and the CI approaches 0 as the amount of homoplasy increases. Trees were also constructed using the Fitch, UP-GMA, and Neighbor-Joining algorithms in Phylip (version 3.0; FELSENSTEIN 1989). These algorithms estimate relationships based on the mutational distances between sequences rather than on character state changes.

RESULTS

Karyotypes: The translocations found in the 15 Rb populations that were sampled are shown in Table 2. The individuals included in this study were homozygous with respect to chromosomal rearrangements. The diploid numbers of the Rb populations ranged from $2n = 22$ (9 fusions) to $2n = 36$ (2 fusions). Eleven of the 15 populations were fixed for six or more fusions. No previously undescribed fusions were detected.

Each population has a unique set of chromosomal rearrangements, although there are some groups of translocations that are shared among several populations from the same geographic region. One example of such a chromosomal system can be seen in Northern Italy where the three populations share six fusions (2.8, 5.15, 9.14, 10.12, 11.13 and 16.17). We also note that there are some fusions that are shared among two or more geographically isolated populations. For example, the 5.15 fusion is found in Northern Italy, Central Italy and Spain.

Mitochondrial DNA sequence variability: A total of 1449 bases were sequenced in each of 56 *M. domesticus* and 2 *M. musculus*, including both the control region and the ND3 gene region. The distribution of variation within *M. domesticus* and between *M. domesticus* and *M. musculus* is shown in Table 3.

Within *M. domesticus*, 42 haplotypes were present among the 56 individuals sampled, and no widespread haplotype was found. Eighty-two nucleotide substitution polymorphisms and six insertion/deletion polymorphisms were observed. At most sites, only two nucleotides were observed segregating within *M. domesticus*; however, at two positions in the control region (15363 and 15530), three different nucleotides were segregating, providing direct evidence of multiple recent mutations at the same site. The distribution of variation within *M. domesticus* exhibits a high degree of population substructure (Table 4). Nucleotide diversity (π) measured between populations (0.85%) is approximately four times as great as nucleotide diversity measured within populations (0.19%). The high degree of sequence similarity observed within populations may reflect sampling of relatives since mice were sometimes collected from the same barn or room. Average nucleotide diversity among individuals within major geographic regions (United Kingdom, Italy, Spain and Greece) is 0.69%. Fixation indices calculated from these data (population $F_{ST} = 0.776$; geographic region $F_{ST} = 0.216$) are in reasonable agreement with average F_{ST} values calculated from 28 allozyme loci (population $F_{ST} = 0.535$, geographic region $F_{ST} = 0.247$; BRITTON-DAVIDIAN 1990).

The two *M. musculus* mtDNA sequences were identical (Table 3). Between *M. domesticus* and *M. musculus*, the average level of sequence divergence is 3.4%, or roughly four times greater than the variation found among populations of *M. domesticus*.

TABLE 3

DNA sequence variability in the ND3 and D-loop regions among 2 *M. musculus* and 56 *M. domesticus*

```
           0                                              1                                                                                                    1
           9                                              5                                                                                                    6
           4444444455555555555  5566666666667777888 88    333333444444445555555555555555555555555555555555556789999000000000000001111112222222222
           4467788900000112334  790234445690239011 55    6678990014479900111122334444445556677788889999312123901556678999024478044455667 8
           1318948713469138094  895454573528185088 56    3580140271343424134586905123458012780238028014778325859267392723987899861480268 17
                                                                                                                                          a  b
con꜀       TACTCTTACAAGATCCTAAAATTATTATACTTTTGAGA-AGA      GTCGTTATACGTAAAAGTCATTTCAATCAATTATCTAACTACTATGTTA-TCAAAATGCCA-C-CACAAAAT-TGGACCTC
Bibbᵈ      ...............C.---C.................-.C      ...............T..............................C..TC................A.....T..
1378ᵉ      G.T.....T.GAG.........C..CG..TC....A.A.A.      A....C..TT....T.AAT.CC.T..CTG.C--AAC.T-C.TCT.AC..............A...TG.C.C.....T
1379       G.T.....T.GAG.........C..CG..TC....A.A.A.      A....C..TT....T.AAT.CC.T..CTG.C--AAC.T-C.TCT.AC..............A...TG.C.C.....T
1334       ..............T.T.............C.CA......      .................CT.G.........T.............C............A......
1335       ..............T.T.............C.CA......      .................CT.G.........T.............C............A......
1345       ..............T.T.............C.CA...A..      .................CT.G.........T.............C............A......
1346       ..............T.T.............C.CA...-..      .................CT.G.........T.............C............A......
1356       ...............T.............C.CA......      .................CT.G.........T.............C............A......
1357       ...............T.............C.CA......      .................CT.G.........T.............C............A......
1328       ...............T.............C.CA......      .................CT...........T.............C............A......
1329       ...............T.............C.CA......      .................CT...........T.............c............A......
1036       .........................C.............      ....................................C.....T.....................
1037       .........................C.............      ....................................C.....T.....................
1372       .........................C.......A...      ....................................C.....T.....C...............
1373       .........................C.............      ....................................C.....T.....C...............
1204       .........................C.C.......A.      ....................................C.....T.....C...............
1205       .........................C.C...........      ....................................C.....T.....C...............
1128       .........................C.........-..      ....................................C.....T.....................
1129       .........................C.........-..      ....................................C.....T.....................
1213       .........................C.........-..      ...............C....................C.....T.....C...............
1214       .G.......................C.........-..      ....................................C.....T.....................
1249       .G.......................C.......A...      ....................................C.....T.....................
1262       .........................C.............      A...................................C.....T.....................
1263       .G.......................C.............      ..T.................................C.....T..................A..
1272       ...........G..........T...G.............      ...........A........C...........T.......T....C...........A.....
1273       ...........G..........T...G.............      ...........A........C...........T.......T....C...........A.....
1311       ...........G..........T...G.............      ....................C...........T.......T....C...-.....T.......A.....
1286       ......C...G..........T.............-..      ....................T...........T.......T....C....C.............
1322       ......C...G..........T.............-..      ....................C...........T.......TG...C.................
1323       ......C...G..........T.............-..      ....................C...........T.......TG...C.................
1083       ...........G..........T.............-..      ..........................G.............T....C.................
1192       ..T.....................C..........-..      A................C.........C..T........T.CT......G............
1193       ..T.....................C..........-..      A................C.........C..T........T.CT......G............
1173       ......................................      .................G....C--..............T.....T..............AAG....
1174       ......................................      .................G....C--..............T.....T..............AAG....
1310       ......................................      ......................C--..............T....T..............AG...
1366       ...CT.....................TA...        C.................C....T...--C.........T..T.GT.........T.........AG....
1367       ...CT.....................TA...        C.................C....T...--C.........T..T.GT.........T.........AG....
1103       ......G.............................      .C..G..................................T.............T.....G.-........GT...
1104       ......G.............................      .C...G.................................T.............T.....G.-........GT...
1369       .G..................................      ...A........G.........--...............T.....A.............GT.C.
1316       .............A.C....C...............      ............T.........A.......--...........C..C................T...
1317       .............A.C....C...............      ............T.........A.......--...........C...C................T..
1176       .............C....C...............        .........................................C...C................T..
1177       ..........C....C...........-..        ....C.........................C.C...C................T..
1008       ......................................      ..................T...C.....---...T...T....C.......C...........
1009       ......................................      ..................T...C.....--...T...T....C.......C...........
1287       ...........C...............G..A...        .................C.........---...T.......C......C........G.......
1067       ......................................      ..................A....C...T...........C...C...........
1068       ......................................      ..................A....C...T...........C...C...........
1081       .....C...G.............-..        A.............G......TG....C...T.........T...C................
1150       ......................................      ..................................T...........C..............A.....
1151       ......................................      ..................................T...........C..............A.....
1225       ...........................-..        ............C.........T...T...C...T..C...................
1248       ...........................A...        ............C.........T...T...C...T..C...................
1226       .........................C........      ........................T...T...C...T..C................
1199       ...........................-..        ................G...T...T...C........C...T...............
1368       ...........................-..        ................G......T....C.C.....G.C..T...............
1200       ......................................      ................G.......T...T...C........C..T...............
```

ᵃ Animal 1311 has the following 11-bp insertion after nucleotide 16072, represented by a single dash: TTTTAACTCTC.
ᵇ Animals 1103 and 1104 have the following 5-bp deletion after nucleotide 16148, represented by single dash: AATAT.
ᶜ Consensus sequence of all individuals; dots indicate identity to consensus; dashes indicate deletions.
ᵈ Published sequence of BIBB *et al.* (1981); numbers at top correspond to nucleotide positions from this sequence. Deletions in the BIBB *et al.* (1981) sequence come after the indicated nucleotide position.
ᵉ Numbers refer to individual mice (see Table 1); *M. musculus*: 1378, 1379; *M. domesticus*: all others.

Overall, the ND3 gene and the entire control region exhibited similar levels of variability (Table 4). Much of the variation at ND3 is attributable to a high rate of non-synonymous substitution (M. W. NACHMAN, S. N. BOYER and C. F. AQUADRO, unpublished data). Of the 24 polymorphic sites within the coding region, 11 result in amino acid changes.

The data presented here (Table 3) also demonstrate that the control region is divided into three distinct regions with different levels of variability, as previously re-

Levels of nucleotide diversity (π) and population substructure (F_{ST})

	Control region	ND3 region	Combined
Nucleotide diversity (π)			
Within populations (H_{WP})	0.0019	0.0018	0.0019
Between populations (H_{BP})	0.0092	0.0073	0.0085
Within major geographic regions[a] (H_{WR})	0.0077	0.0056	0.0069
Between major geographic regions (H_{BR})	0.0094	0.0075	0.0088
Within total *M. domesticus* sample	0.0091	0.0072	0.0084
Between *M. domesticus* and *M. musculus*	3.0050	3.0030	3.0040
Population substructure (F_{ST})[b]			
Population $F_{ST} = (H_{BP} - H_{WP})/H_{BP}$	0.793	0.753	0.776
Geographic region $F_{ST} = (H_{BR} - H_{WR})/H_{BR}$	0.181	0.253	0.216

[a] Major geographic regions are Italy, Greece, Spain, and the United Kingdom.
[b] F_{ST} calculated as in HUDSON *et al.* (1992; equation 3).

ported for house mice (PRAGER *et al.* 1993) and other species (*e.g.*, HOELZEL *et al.* 1991). The first region is hypervariable and spans roughly from base position 15490 to 15600 (in the sequence of BIBB *et al.* 1981), the second region is extremely conserved and spans from 15600 to 15990, and the third region is moderately variable and spans 15990–16295. These differences are apparent within *M. domesticus* as well as between *M. domesticus* and *M. musculus*.

Transition/transversion bias: The bias in the number of transition *vs.* transversion base substitutions was calculated by (1) a "site" method, where the number of sites at which transitions occur and the number of sites at which transversions occur are simply counted, and (2) a "tree" method, where the number of transitions and transversions which occur along the branches of a phylogenetic tree are counted. The site method is likely to underestimate frequently occurring mutations since it does not account for convergent, parallel, or reversal mutations (*i.e.*, homoplasy); however, it is independent of any hypothesis of the relationship of haplotypes. The tree method accounts for homoplasy, however its accuracy rests on finding the correct evolutionary relationship of haplotypes. We inferred evolutionary relationships of haplotypes using PAUP (SWOFFORD 1991) with the heuristic search option and with transversions weighted four times transitions (these weights were based on the bias estimated from the site method). The transition/transversion ratios obtained with these two approaches are 4:1 (site method) and 5.7:1 (tree method) for comparisons within *M. domesticus*.

Since the tree method estimates the number of mutational steps occurring along the branches of a minimal length tree, it depends on the topology of the tree. However, the tree may be based on some idea of the relative weights of transitions and transversions. A tree based on an equal weight for these two types of characters may result in a very different estimate of bias than one based on different weights. We investigated this problem by measuring the transition/transversion ratio on trees constructed with a spectrum of different weights applied

to transversions *vs.* transitions. If equal weights are used, the resulting transition/transversion ratio is 4.9:1. If weights from 2:1 to 20:1 are used, the resulting ratio is 5.7:1.

In comparisons between *M. domesticus* and *M. musculus*, the transition/transversion ratios are 3.8:1 (site method) and 4.4:1 (tree method). The observation of a greater relative number of transversions at greater evolutionary distances may reflect multiple transition mutations at the same site, as previously reported in other taxa (*e.g.*, BROWN *et al.* 1982). The hypothesis that transition homoplasy is more prevalent than transversion homoplasy is also supported by the consistency indices for these two types of mutations derived from a minimum length tree based on equal weightings. The CI for transitions is 0.702 and the CI for transversions is 0.885.

For both the site method and the tree method, a smaller transition/transversion ratio was observed in the control region (3.4:1 and 5.2:1) than in the ND3 gene region (6:1 and 7:1) in comparisons within *M. domesticus*.

Insertion/deletions: In Table 3, there are eight polymorphic insertion/deletion sites compared to 108 polymorphic sites for base substitutions among all individuals (including both *M. musculus* and *M. domesticus*). Nonetheless, it appears that there are more convergent changes ("multiple hits") for insertion/deletion mutations than for point mutations. A parsimony tree based on equal weightings for all characters was constructed to estimate the amount of convergence in insertion/deletion mutations. The CI on this tree for insertion/deletions is 0.267 and the CI for all base substitutions is 0.732. Thus, although insertion/deletions occur at relatively few places, where they do occur, they appear to occur repeatedly and frequently.

Phylogenetic relationships of mtDNA haplotypes: Cladistic analysis of the 56 mitochondrial sequences from *M. domesticus* using the two *M. musculus* sequences as an outgroup yielded 520 equally parsimonious trees of length 300 using the program PAUP (SWOFFORD 1991) in 10 replicates of a heuristic search with stepwise addition of taxa in a random order. The consistency index of each of these trees, over all characters,

TABLE 5

Distribution of characters used in phylogenetic analysis

	Control region	ND3 region	Total
Nucleotides	974	475	1449
Total variable sites	80	40	120
Variable in *M. domesticus*	59	31	90
Informative sites	48	25	73

TABLE 6

Distribution of numbers of mutational steps for transversions, transitions, and insertion/deletions

	No. of mutational steps on tree in Figure 4									
	1	2	3	4	5	6	7	8	9	10
Transversions	21	1								
Transitions	64	16	5	1	3					
Insertions/ deletions	4			1	1	1				1

For example, there are 21 transversion sites at which a single mutation is inferred in a parsimony analysis (Figure 4), and there is one transversion site at which two mutations are inferred.

was 0.649. The numbers and distribution of characters for this analysis are shown in Table 5, and the distribution of character state changes for transversions, transitions, and insertion/deletions is shown in Table 6.

The strict consensus tree of all 520 trees is shown in Figure 4. A strict consensus tree is a conservative summary of hypotheses of relationships since it includes only the clades (groups) which are present in all equally parsimonious trees. While many equally short trees were found, the differences among these trees were minor and resulted mainly from changes in the placement of a few taxa. This can be seen by the large number of clades (38) which remain resolved in the strict consensus tree. Additionally, the consensus tree contains six major clades which include most of the variation (54 of the 56 *M. domesticus*). The trees from which the consensus tree was derived were obtained by weighting transversions six times transitions. However, the same consensus is obtained using transversion/transition weightings from 2:1 to 20:1. When an equal weight is given to transitions and transversions, the resulting consensus tree has one less major clade but otherwise remains unchanged (clade VI in Figure 4 becomes a basal polytomy). Insertion/deletion changes were given the same weight as transitions; however, the same consensus tree is obtained if insertion/deletion changes are eliminated from the analysis. Likewise, an analysis of the control region by itself yields a similar though slightly less resolved consensus tree. These results suggest that the strict consensus tree presented here is robust and represents an accurate, if conservative, estimate of the evolutionary relationships of the mitochondrial haplotypes.

Further evidence supporting the relationships obtained in the cladistic analysis comes from three distance-based tree building algorithms. The six major clades shown in Figure 4 were also present in trees obtained with UPGMA (Unweighted Pair Group Method with Arithmetic mean), Neighbor-Joining, and Fitch analyses using the Phylip program of FELSENSTEIN (1989). While the membership and internal topology of these six clades were identical in all analyses, the positions of these clades relative to each other showed minor differences among the analyses.

Geographic variation: There is limited concordance between the mitochondrial phylogeny and the geographic origin of the haplotypes represented in the phylogeny. For example, mtDNAs from Spain are distrib-

uted on three of the six major clades. This discordance between phylogeny and geography could, in theory, result from either current gene flow or ancestral polymorphism or both. To determine if ancestral polymorphism contributes to the observed pattern, the time of the reentry of *M. domesticus* into Western Europe (AUFFRAY *et al.* 1990) has been plotted on the mitochondrial phylogeny (Figure 5). In this tree, branch lengths are approximately proportional to the number of substitutions occurring along the lineage, and divergence dates were obtained by calibrating the mitochondrial molecular clock according to SHE *et al.* (1990). This reveals that much of the diversity present in the mitochondrial phylogeny predates the movement of mice into Europe. Four of the six major clades contain lineages that are older than the time of colonization (clades III, IV, V and VI in Figure 5). Each of these four clades also contain individuals from distant geographic regions. For example, mice from England and some of the mice from Greece are grouped in one clade. Conversely, two of the six clades only contain lineages that are younger than the time of colonization (clades I and II in Figure 5). These same two clades are the only ones that are localized to single geographic regions (one in Scotland, one in Italy). Thus the age of a clade relative to the time of colonization appears sufficient to explain the observed discrepancy between geography and phylogeny.

If the recent colonization of Europe from the Middle East was accompanied by a rapid population expansion, then the distribution of pairwise differences among all mtDNA sequences in the sample is expected to be nearly Poisson (DI RIENZO and WILSON 1991; SLATKIN and HUDSON 1991; ROGERS and HARPENDING 1992). The distribution of differences among the 56 *M. domesticus* sampled is shown in Figure 6 and is significantly different from a Poisson distribution (chi-square, $\chi^2 = 243$; $P < 0.001$). The distribution is bimodal with a major peak at 12 differences and a minor peak at 1 difference. To test whether this pattern is a result of sampling over a large geographic region (*i.e.*, much of Western Europe), the distribution of pairwise differences was also plotted for individuals from (1) Great Britain, (2) Italy, (3) Greece and (4) Spain (data not shown). The distribution of pair-
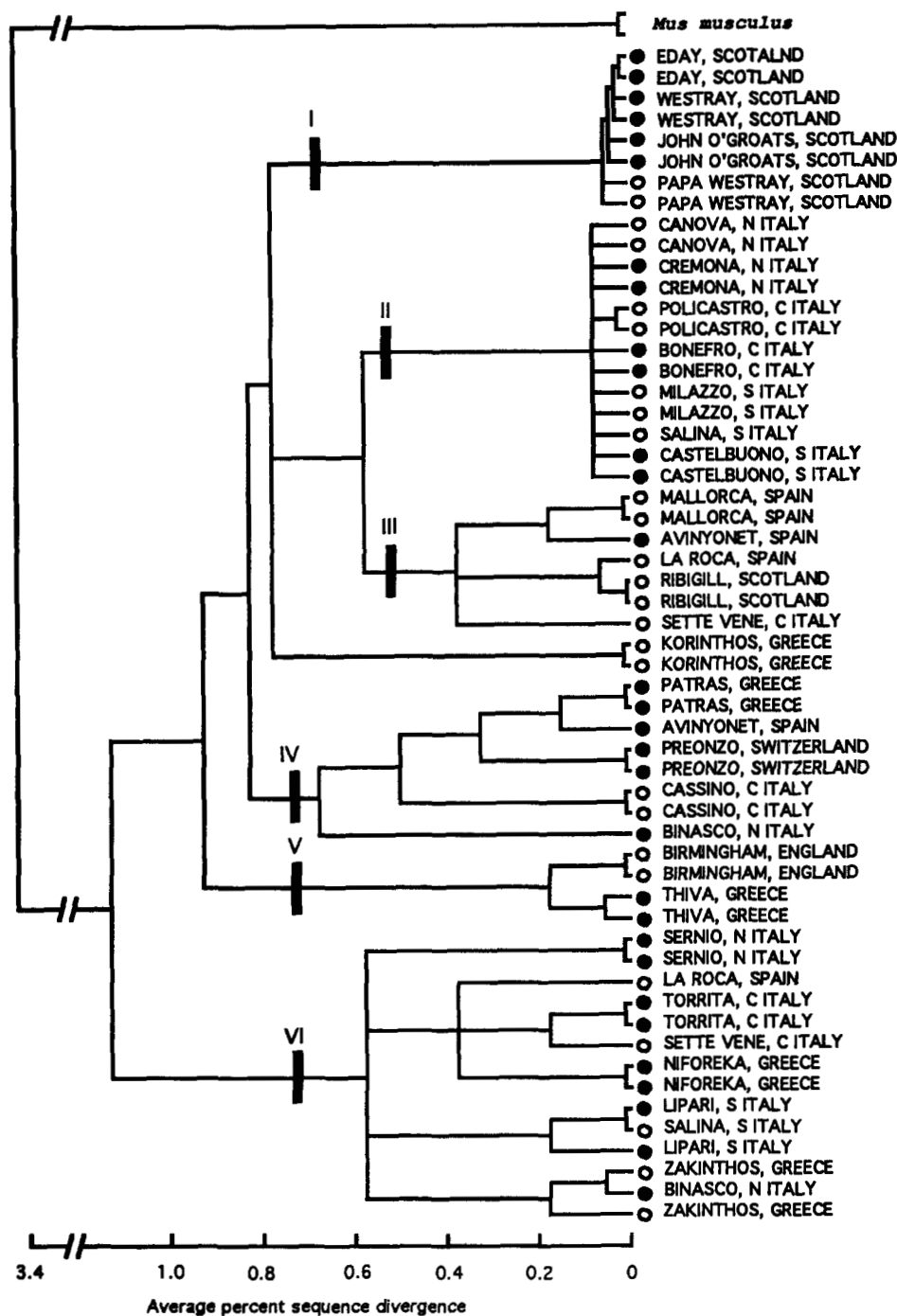
FIGURE 4.—Strict consensus tree of 520 equally parsimonious solutions obtained using the heuristic search option in PAUP (see text for details). The six major clades are numbered I–VI. Each terminal lineage represents one individual (same order from top to bottom as in Table 3). The top two lineages are *M. musculus*, all others are *M. domesticus*. Closed circles represent Rb mice, open circles represent standard karyotype (*i.e.*, $2n = 40$) mice. Collecting localities are listed at right. Rb and standard karyotype mice are found on each of the 6 major clades.

wise differences was also plotted for a sample including all populations, but only a single individual per population. The pattern in each of these plots was similar to the overall pattern, and in all cases was significantly different from a Poisson distribution (chi-square tests for all cases, $P < 0.001$). These results do not provide evidence of recent exponential population growth (but see also discussion), and are consistent with the deep branches and ancestral polymorphism observed in the phylogenetic tree.

Evolution of Robertsonian populations: The distribution of Robertsonian and standard karyotype individuals

on the strict consensus tree is shown in Figure 4. Robertsonian and standard karyotype individuals, often from the same geographic region, are found on each of the six major clades. Thus, there is no evidence for a single mitochondrial clade which contains all Robertsonian individuals. To evaluate the robustness of this result, we have also constructed phylogenetic trees in which all Robertsonian individuals are constrained to a single clade. The shortest such tree requires 364 mutational steps, or 64 mutations more than the tree in Figure 4, and is, therefore, a far less parsimonious explanation of the data. These results
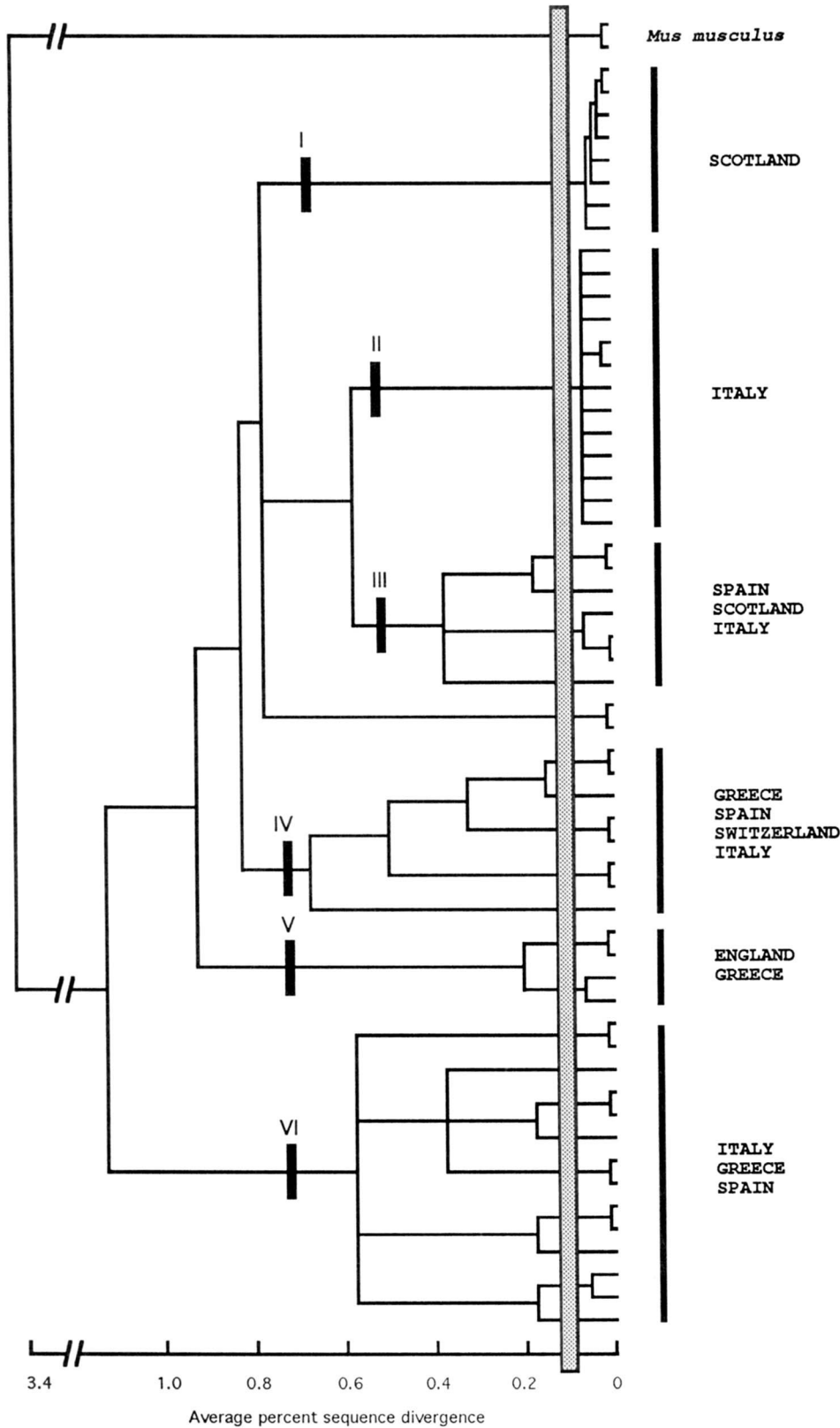
FIGURE 5.—Strict consensus tree as in Figure 4, with hypothesized time of colonization of Western Europe by mice indicated by a vertical bar. See text for calibration of molecular clock and timing of movement of mice into Europe.

are consistent with a hypothesis of independent origin of Rb populations, although mitochondrial gene flow between Rb and standard karyotype populations could also contribute to the observed pattern (see DISCUSSION).

Single and independent origin hypotheses also differ in their predictions concerning the distribution of genetic variation among Rb and standard karyotype animals. A single origin hypothesis predicts that Rb animals (as a group) will exhibit less variation than standard
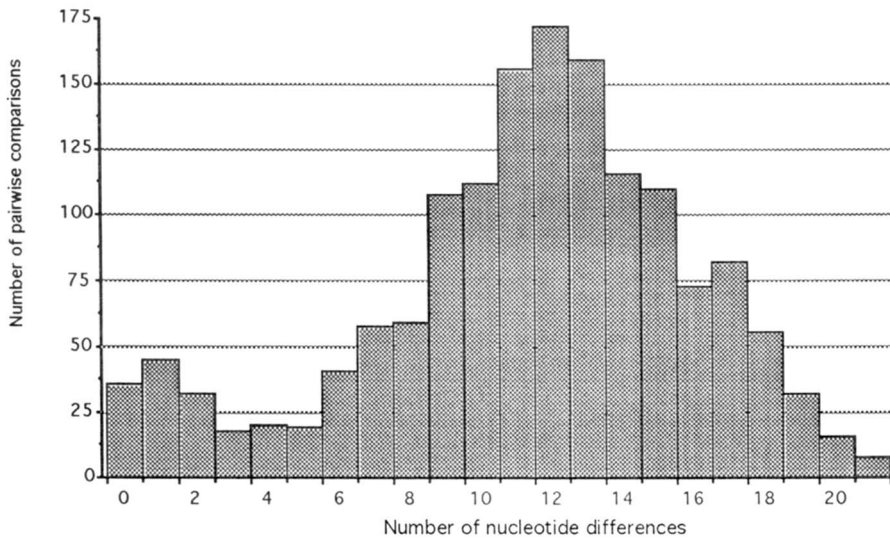
FIGURE 6.—Frequency distribution of all pairwise differences among mitochondrial sequences within *M. domesticus*.

karyotype animals, while a multiple origin hypothesis predicts that both Rb and standard karyotype animals will exhibit similar overall levels of genetic variation. Levels of nucleotide variability were measured for Rb animals ($\pi = 0.8\%$, $\theta = 1.2\%$) and for standard karyotype animals ($\pi = 0.8\%$, $\theta = 1.3\%$). For both of these measures of genetic variability, the two groups appear the same or nearly identical.

Several fusions appear in two or more geographically isolated regions, and these have been mapped onto the strict consensus tree. The number of clades on which such fusions appear is shown in Table 7. The majority of these fusions are not localized to a single clade, consistent with an independent origin hypothesis (but see caveats below).

DISCUSSION

Mitochondrial DNA evolution in mice: Levels of sequence variability were found to be quite similar in the ND3 gene and in the control region. The consistency index for base substitutions at ND3 (CI = 0.90) was higher than in the control region (CI = 0.70), suggesting that multiple hits are more prevalent in the control region. These observations generally corroborate earlier results from restriction fragment length polymorphism (RFLP) data (FERRIS *et al.* 1983) and substantiate the general accuracy of RFLP data for revealing underlying levels and patterns of DNA sequence variation.

The distribution of polymorphic sites in the control region reported in this study agrees well with previous findings from cetaceans (HOELZEL *et al.* 1991), primates (BROWN *et al.* 1982; AQUADRO and GREENBERG 1983; FORAN *et al.* 1988), rats (BROWN *et al.* 1986), and mice (PRAGER *et al.* 1993). Two hypervariable regions at the ends of the control region surround the central, more conserved region. As in other taxa, the first hypervariable region is less conserved than the second.

TABLE 7

Number of major clades on which geographically widespread Rb fusions occur

		No. of major clades[a]		
	Fusion	1	2	3
Present in two geographic regions[b]	1.2	×		
	1.3	×		
	2.15		×	
	3.4			×
	3.9	×		
	4.6	×		
	4.14			×
	6.7			×
	6.16	×		
	8.12		×	
	8.17		×	
	9.14			×
	10.14			×
	16.17			×
Total		5	3	6
Present in three geographic regions	5.13			×
	5.15			×
	9.16			×
	10.12			×
Total		0	0	4

[a] Major clades refer to the six clades shown in Figure 4.
[b] Geographic regions as in Table 2.

Our measures of sequence divergence between *M. domesticus* and *M. musculus* allow us to estimate the rate of mtDNA sequence evolution. SHE *et al.* (1990) have used a 10-million year divergence time for the *Mus-Rattus* split, estimated from paleontological evidence, to calibrate a molecular clock based on single-copy nuclear DNA hybridization data. From pairwise comparisons of species within the genus *Mus*, they estimate the time of the *musculus-domesticus* split at 350,000 years ago ± 1 SD of 290,000 years. From our divergence estimate of 3.4%, we estimate a rate of sequence evolution of approximately 9.7% divergence per million years

(minimum 5.3%, maximum 56.7%). This estimate is very similar for the control region alone (10% per million years) and for the ND3 gene region alone (9.4% per million years). Both values are considerably higher than the 2% per million years rate of evolution reported for a variety of taxa for the mtDNA molecule as a whole (Wilson *et al.* 1985), but are in rough agreement with other mtDNA studies on mice (Prager *et al.* 1993). We note that the 2% value reported by Wilson *et al.* (1985) has never been confirmed for rodents using a calibration of the molecular clock based on the rodent fossil record. Thus the difference between our rate estimate and the rate reported by Wilson *et al.* (1985) may reflect an actual difference in the rate of mouse mtDNA evolution compared to other groups, or it may reflect uncertainties in the dates estimated from the fossil record. However, it is probably not due to differences in rates of evolution between the regions we sequenced and the rates obtained from the entire molecule. The overall depth of our tree within *M. domesticus* (approximately 1.1% sequence divergence) is nearly identical to that reported by Sage *et al.* (1990) and Ferris *et al.* (1983) based on RFLP data from the entire molecule. This emphasizes that while portions of the control region evolve very rapidly, the overall rate of evolution for the control region may not be substantially different from the rate of evolution of the entire mtDNA molecule.

The transition/transversion ratios for the control region reported in this study (from 3.4:1 to 5.2:1) are considerably smaller than previously reported for the control region in humans at a comparable level of sequence divergence (Aquadro and Greenberg 1983; Kocher and Wilson 1991; Di Rienzo and Wilson 1991; Vigilant *et al.* 1991), but are in general agreement with the ratios reported for kangaroo rats (Thomas *et al.* 1990) and mice (Prager *et al.* 1993).

The number of positions at which insertion/deletion events occur (8) is far fewer than the number of sites at which base substitutions occur (108), and thus sites of insertions/deletions are relatively rare compared to those of base substitutions. However, where they do occur, insertion/deletions occur often, as evidenced by the high number of independent origins of the same insertion or deletion (Table 6) and by the lower CI for insertion/deletions (0.250) relative to base substitutions (CI = 0.732) on the tree in Figure 4. This suggests that very different mutational processes underlie these two kinds of changes. As reported elsewhere for mouse mtDNA (Prager *et al.* 1993), insertion/deletion mutations appear particularly common in short mono- or dinucleotide runs.

Geographic distribution of mtDNA haplotypes: The distribution of mtDNA haplotypes is consistent with the recent colonization of Western Europe from populations in the Middle East as proposed by Auffray *et al.* (1990). Four of the six major clades in Figure 4 include

individuals from distant geographic areas. This general lack of concordance between phylogeny and geography is expected if ancestral polymorphism is still segregating in present-day populations. Tajima (1983) has investigated this process with a coalescent model and the simple situation where one population becomes subdivided into two populations, each of N diploid individuals. If two sequences are sampled from each of these two populations, then the probability that the allele phylogeny will accurately reflect the population phylogeny reaches 0.95 only after $8N$ generations since the subdivision (or $4N$ generations for haploids). Given a conservatively low estimate of mouse female effective population size of 10^4 and one generation per year, then 40,000 years would need to elapse before phylogeny would coincide with geography under even this simple two-population model. This is considerably longer than the 8,000 years during which mice are believed to have been in Europe (Auffray and Britton-Davidian 1992). It is therefore reasonable to interpret the lack of correlation between phylogeny and geography in terms of ancestral polymorphism. An alternative explanation is that extensive gene flow is responsible for this pattern; however, under such a scenario there is no expectation for younger clades to exhibit a substantially different geographic distribution than older clades. Because such a relationship is evident (Figure 5), we feel that ancestral polymorphism may be largely responsible for the observed pattern.

If the colonization of Western Europe by mice resulted from a rapid expansion from a small founder population, then the pairwise differences among sequences drawn from the population are expected to be approximately Poisson distributed (Di Rienzo and Wilson 1991; Slatkin and Hudson 1991; Rogers and Harpending 1992). The distribution of pairwise differences that we observed is not Poisson, but does have a clear, major peak at 12 differences. One interpretation of such a peak is that it reflects population growth (or recovery from a bottleneck) at some time in the past (Rogers and Harpending 1992). The timing of this putative expansion can be estimated from a molecular clock (12 differences = 0.83% sequence divergence or approximately 85,000 years). However, because simulations of populations under neutral equilibrium conditions produce a wide variety of distributions of pairwise differences, including distributions with a single major peak (Slatkin and Hudson 1991), our data do not provide convincing evidence for population growth. This pattern may be contrasted with the situation in humans, in which non-African populations exhibit Poisson distributions of pairwise differences among mtDNA haplotypes (Di Rienzo and Wilson 1991). It is noteworthy that among the mtDNA sequences reported here, there is no single widespread haplotype; no haplotype was found in more than two individuals. It is possible that the

colonization of Europe by mice did not follow from a small founder population which then underwent a rapid expansion; indeed, mouse population sizes may have remained relatively stable during the range expansion into Europe. This idea could be tested further by looking at the distributions of pairwise differences from several different nuclear genes (in regions of low recombination). If several genes all showed similar distributions, each with a single major peak, this would provide stronger evidence for changes in population size.

There has been some debate recently about when mice colonized Europe (*e.g.*, AUFFRAY and BRITTON-DAVIDIAN 1992; SAGE *et al.* 1990). The traditional and more widely accepted view is that house mice followed the spread of agriculture from the Fertile Crescent in the Middle East westward and northward into Europe during the last 8,000 years, and perhaps only as recently as the last 2,000–4,000 years (AUFFRAY *et al.* 1990; BRITTON-DAVIDIAN 1990). This view is supported by an unusually detailed archaeozoological record of *M. domesticus* from the Epipaleolithic to the late Iron Age in sites throughout Europe and the Mediterranean region. These sites reveal a steady westward progression of *M. domesticus* from Israel to Great Britain as one moves forward in time from the Epipaleolithic. This view is further supported by allozyme surveys which suggest a rapid and recent pattern of colonization (BRITTON-DAVIDIAN 1990). An alternative hypothesis, proposed by SAGE *et al.* (1990) based on RFLP data from the entire mtDNA molecule, is that colonization of Europe north of the Alps took place much earlier (roughly 30,000–70,000 years ago), perhaps coincident with the spread of anatomically modern *Homo sapiens*.

The mtDNA results presented here allow us to further address this issue. By looking at the amount of sequence divergence that has occurred in clades that are restricted to a single geographic region, it is possible to estimate how long the mice have been in that region (SAGE *et al.*, 1990). This estimate relies on the assumptions that (1) the observed divergence in a geographically restricted clade arose *in situ* and not in some other region, and (2) it is possible to estimate divergence times from a molecular clock. Our data reveal two such clades, one in Italy and one in Scotland. The average depth of the Italian clade is 0.1% and the average depth of the Scottish clade is 0.07%. Given our estimated rate of sequence divergence of approximately 9.7% per million years (above), the Italian clade can be dated at roughly 10,000 years and the Scottish clade at roughly 7,000 years. These dates must certainly be taken as very approximate as they depend on accurately calibrating the molecular clock. Nonetheless they are in general agreement with the paleontological evidence in suggesting a very recent arrival for mice in Europe. It is also noteworthy that the date of the Scottish clade is younger than the date of the more southern, Italian clade, as might be expected of colo-

nization spreading northward. The discrepancy between our results and the estimate derived from the mtDNA RFLP data of SAGE *et al.* (1990) derives from two sources. First, the levels of divergence among mice within the geographically restricted clades in this study (0.07% for Scotland and 0.1% for Italy) are two to three times less than the level of divergence within the one geographically restricted European clade in their study (0.24% for Germany), even though the overall depths of both trees are very similar. This difference may reflect sampling biases in one or both studies (sample sizes were nearly identical in both studies). Second, a different calibration of the mtDNA molecular clock was used in this study. SAGE *et al.* (1990) used an estimate of 2–4% change per million years based on a variety of taxa for the entire mtDNA molecule (reviewed by WILSON *et al.* 1985). Our estimate of 9.7% per million years is based specifically on the ND3 and control regions of the mtDNA molecule which we surveyed in *M. domesticus* as well as our own data on divergence between *M. domesticus* and *M. musculus*.

Evolution of Rb races: The distribution of mtDNA variation within and among Rb populations allows us to make several inferences concerning their origin. The evidence for a recent arrival of mice in western Europe corroborates the view that Rb populations have evolved very quickly. Evidence that Rb populations arose only since mice moved into their present range comes from the geographic distribution of the Rb fusions. There is a strong correspondence between geography and a phylogeny of Rb populations based on the chromosomes themselves as characters (LARSON *et al.* 1984) as would be expected if the fusions arose in their present locations. The geographic localization of related karyotypes can also be seen in Table 2. A total of 79 different fusions are known from natural populations in western Europe and the Mediterranean region, and it appears reasonable to infer that these have all arisen within the last 10,000 years.

Further evidence that Rb populations may be young comes from the small amount of genetic differentiation between Rb populations and standard karyotype populations. For example, in Scotland, the Papa Westray population ($2n = 40$) and the John O'Groats population ($2n = 32$) are separated by a single mtDNA mutation, corresponding to a divergence of 0.06%, or roughly 6,000 years.

This rapid rate of chromosomal evolution raises the issue of whether particular population processes exist which favor the establishment of new variants. Theoretical studies (LANDE 1979; HEDRICK 1981; WALSH 1982) have shown that small founder populations will greatly facilitate the fixation of new chromosomal mutations, especially if they are underdominant. This expectation is in accord with the geographic location of many Rb populations, which are often found in relatively isolated

mountain valleys or archipelagos. We looked for evidence of smaller population size within Rb populations compared to standard karyotype populations by calculating nucleotide diversity for each. The average value of π where both individuals in a pairwise comparison come from the same Rb population (0.181%) is slightly smaller than the corresponding value for individuals from standard karyotype populations (0.213%), although these values are not significantly different from each other (Mann-Whitney $U = 94.5$, $P > 0.10$). This suggests that at most a modest reduction in population size may have accompanied the fixation of new rearrangements.

The apparent rapid rate of chromosomal evolution in *M. domesticus* also raises the question of whether there is some intrinsic genetic mechanism, such as a mutator allele, responsible for producing these rearrangements, a possibility discussed by a number of previous authors (*e.g.*, SAGE 1981; LARSON *et al.* 1984; WINKING *et al.* 1988). One hypothesis is that a mutator is present in a single lineage of *M. domesticus* which has given rise to all or most Rb populations (WINKING *et al.* 1988). In support of this idea is the observation that laboratory stocks which contain a single Rb fusion often give rise to subsequent fusions (reviewed by WINKING *et al.* 1988).

The phylogeny in Figure 4 allows us to address the issue of whether all Rb populations derive from a single maternal lineage. The topology of this tree in general and the membership of the six major clades in particular appear well established. Essentially the same topology is obtained with four different tree building methods. This topology is also obtained with many different weighting schemes for transitions, transversions, or insertions/deletions. Finally, the topology is insensitive to omissions of taxa or characters.

While the topology of this tree is probably a good estimate of mtDNA relationships, making inferences about the relationships of populations from this tree is more problematic. Rb and standard karyotype individuals are present on each of the six major clades. This pattern is consistent with earlier mtDNA studies of *M. domesticus* which included a few presumptive Rb mice (FERRIS *et al.* 1983; SAGE *et al.* 1990; PRAGER *et al.* 1993), and appears to contradict the hypothesis of an origin of all Rb individuals from a single maternal lineage. However, this interpretation depends on concordance between the mtDNA phylogeny and the phylogeny of populations. This is a general problem with intraspecific phylogenetics: we cannot be sure whether similarities among individuals within a population are due to shared ancestry or current (or recent) gene flow. Significant gene flow and ancestral polymorphism may result in different phylogenies for different gene regions; under such circumstances, there is no single phylogeny for populations (AVISE *et al.* 1987). In particular for this study, gene flow between Rb and standard karyotype populations has the potential to obscure the origin

of Rb races. Numerous laboratory studies have shown that crosses between Rb and standard karyotype animals often result in F_1s with greatly reduced fertility (*e.g.*, GROPP and WINKING 1981), and therefore gene flow between these populations in nature has been traditionally viewed as limited (CAPANNA 1982). More recently, however, several results have challenged this view, including the finding of a wide hybrid zone between Rb and standard karyotype mice (SEARLE 1991), the demonstration that certain naturally occuring Rb heterozygotes have nearly normal fertility (WINKING *et al.* 1988; BRITTON-DAVIDIAN *et al.* 1990; VIROUX and BAUCHAU 1992; WALLACE *et al.* 1992), and the differential spread of mtDNA and Rb chromosomes when introduced into a standard karyotype population (BERRY *et al.* 1991; SCRIVEN 1992).

One approach to the problem is to ask how much gene flow would be required to obtain the tree in Figure 4 if all Rb individuals in fact arose from a single founding female. This is similar to the approach used by SLATKIN and MADDISON (1989) to obtain an estimate of gene flow from a phylogeny of non-recombining alleles sampled from different populations. Their model simulates two populations which derive from a single ancestral population at a time $50N$ generations ago, where N is the size of each population. Without migration between the two populations, each population will become monophyletic with respect to the gene genealogy during this time. By introducing migration events in a coalescent model, they derive values of the parameter s which correspond to different values of the parameter Nm, where m is the migration rate, and s is the minimum number of migration events required to make each population in a given tree monophyletic. In phylogenetic terms, the geographic location from which each allele is sampled is treated as one state of a multistate character. A parsimony criterion applied to the evolution of this character on a given tree gives the minimum number of migration events (s) consistent with the phylogeny. Because their model simulates monophyletic populations, we can use their results to provide a minimum estimate of s that is consistent with the tree in Figure 4 if in fact all Rb individuals derive from a single maternal lineage. Treating presence or absence of Rb fusions as a two-state character provides a minimum number of migration events between Rb and standard karyotype animals that is consistent with our data. Using this approach we obtain from Figure 4, $s = 11$, and from SLATKIN and MADDISON (1989; interpolating from Table 1, p. 609), Nm is approximately 10. While approximate, this value of Nm is roughly an order of magnitude higher than measured for other mammals (SLATKIN 1985; HARTL and CLARK 1989), including house mice (BRITTON-DAVIDIAN 1990).

Because the level of gene flow required under this scenario is unusually high, not all relationships depicted in Figure 4 are easily explained by a single origin and

subsequent migration between Rb and standard karyotype populations. A more likely explanation is that at least some of the Rb populations have arisen independently. While our data allow us to reject the hypothesis that all Rb individuals arose from a single female, estimating the number of independent origins is currently not possible.

Another way to address the problem of gene flow in intraspecific phylogenies is to use several markers. The *Y* chromosome has the potential to be useful in a comparison with mtDNA because the *Y* is inherited from father to son. Because male and female mice disperse differently (LIDICKER and PATTON 1987), agreement between a mtDNA tree and a *Y* chromosome tree may be attributed to shared ancestry, whereas differences are likely to reflect different patterns of gene flow. Unfortunately with *M. domesticus*, significant ancestral polymorphism (and a different pattern of ancestral polymorphism for the two markers) is likely to make it difficult to interpret conflicting patterns, and may account for some of the differences between our results and the monophyly of some Rb mice deduced from *Y* chromosome RFLPs (TUCKER *et al.* 1989).

Further evidence concerning the origin of Rb races is obtained from the distribution of genetic variation. If all Rb populations arose from a single common lineage then they should contain a subset of the total variation found in *M. domesticus*. However, our results show that levels of mtDNA variability are the same among different Rb populations as among different standard karyotype populations, as expected if Rb races arose multiple times. This distribution of genetic variation is consistent with the distribution reported for allozymes (BRITTON-DAVIDIAN *et al.* 1989).

Our results do not allow us to make inferences about the origin of particular fusions. A number of fusions, such as 5.13, are present in three different populations and also map to three different mtDNA clades (Table 7). It may therefore appear that such fusions have arisen independently three times. However, because the fusions are not linked to the mtDNA marker, there is no reason to believe they will always evolve together. Differential gene flow of either a Rb chromosome or a mtDNA haplotype may be invoked as an alternative explanation to any specific case of independent origin. It is noteworthy that seven different fusions (2.17, 4.15, 6.13, 8.17, 10.14, 11.14 and 12.13) have arisen independently in laboratory strains of mice and in natural populations (WINKING *et al.* 1988). This clearly shows that translocations sometimes arise independently involving the same chromosome arms. Whether or not this also occurs in nature remains an open question.

LITERATURE CITED

AQUADRO, C. F., and B. D. GREENBERG, 1983 Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics **103:** 287–312.

AUFFRAY, J.-C., and J. BRITTON-DAVIDIAN, 1992 When did the house mouse colonize Europe? Biol. J. Linn. Soc. **45:** 187–190.

AUFFRAY, J.-C., F. VANLERBERGHE and J. BRITTON-DAVIDIAN, 1990 The house mouse progression in Eurasia: a palaeontological and archaeozoological approach. Biol. J. Linn. Soc. **41:** 13–25.

AVISE, J. C., J. ARNOLD, R. M. BALL, E. BERMINGHAM, T. LAMB *et al.*, 1987 Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. **18:** 489–522.

BAUCHAU, V., 1990 Phylogenetic analysis of the distribution of chromosomal races of *Mus musculus domesticus* Rutty in Europe. Biol. J. Linn. Soc. **41:** 171–192.

BERRY, R. J., G. S. TRIGGS, P. KING, H. R. NASH and L. R. NOBLE, 1991 Hybridization and gene flow in house mice introduced into an existing population on an island. J. Zool. Lond. **225:** 615–632.

BIBB, M. J., R. A. VAN ETTEN, C. T. WRIGHT, M. W. WALBERG and D. A. CLAYTON, 1981 Sequence and gene organization of mouse mitochondrial DNA. Cell **26:** 167–180.

BRITTON-DAVIDIAN, J., 1990 Genic differentiation in *M. m. domesticus* populations from Europe, the Middle East and North Africa: geographic patterns and colonization events. Biol. J. Linn. Soc. **41:** 27–45.

BRITTON-DAVIDIAN, J., J. H. NADEAU, H. CROSET and L. THALER, 1989 Genic differentiation and origin of Robertsonian populations of the house mouse (*Mus musculus domesticus* Rutty). Genet. Res. **53:** 29–44.

BRITTON-DAVIDIAN, J., H. SONJAYA, J. CATALAN and G. CATTANEO-BERREBI, 1990 Robertsonian heterozygosity in wild mice: fertility and transmission rates in Rb 16.17 translocation heterozygotes. Genetica **80:** 171–174.

BROWN, G. G., G. GADALETA, G. PEPE, C. SACCONE and E. SBISA, 1986 Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. J. Mol. Biol. **192:** 503–511.

BROWN, W. M., E. M. PRAGER, A. WANG and A. C. WILSON, 1982 Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18:** 225–239.

CAPANNA, E., 1982 Robertsonian numerical variation in animal speciations: *Mus musculus*, an emblematic model, pp. 155–177 in *Mechanisms of Speciation*, edited by C. BARIGOZZI. Alan R. Liss, New York.

CORTI, M., C. M. CIABATTI and E. CAPANNA, 1990 Parapatric hybridization in the chromosomal speciation of the house mouse. Biol. J. Linn. Soc. **41:** 203–214.

DI RIENZO, A., and A. C. WILSON, 1991 Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc. Nat. Acad. Sci. USA **88:** 1597–1601.

FELSENSTEIN, J., 1989 Phylip (Phylogeny Inference Package) Version 3.4. Department of Genetics, University of Washington, Seattle.

FERRIS, S. D., R. D. SAGE, E. M. PRAGER, U. RITTE and A. C. WILSON, 1983 Mitochondrial DNA evolution in mice. Genetics **105:** 681–721.

FORAN, D. R., J. E. HIXSON and W. M. BROWN, 1988 Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res. **16:** 5841–5862.

GROPP, A., and H. WINKING, 1981 Robertsonian translocations: cytology, meiosis, segregation patterns and biological consequences of heterozygosity. Symp. Zool. Soc. Lond. **47:** 141–181.

GROPP, A., U. TETTENBORN and E. VON LEHMANN, 1969 Chromosomenuntersuchen bei der Tabakmaus (*M. poschiavinus*) und bei Hybriden mit der Laboratoriums Maus. Experientia **25**: 875–876.

HARTL, D. L., and A. G. CLARK, 1989 *Principles of Population Genetics*. Sinauer, Sunderland, Mass.

HEDRICK, P. W., 1981 The establishment of chromosomal variants. Evolution **35**: 322–332.

HOELZEL, A. R., J. M. HANCOCK and G. A. DOVER, 1991 Evolution of the cetacean mitochondrial D-loop region. Mol. Biol. Evol. **8**: 475–493.

HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics **132**: 583–589.

KLUGE, A. G., and J. S. FARRIS, 1969 Quantitative phyletics and the evolution of anurans. Syst. Zool. **18**: 1–32.

KOCHER, T. D., and A. C. WILSON, 1991 Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and a protein-coding region, pp. 391–413 in *Evolution of Life. Fossils, Molecules, and Culture*, edited by S. OSAWA and T. HONJO. Springer-Verlag, Tokyo.

KOCHER, T. D., W. K. THOMAS, A. MEYER, S. V. EDWARDS, S. PAABO *et al.*, 1989 Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA **86**: 6196–6200.

LANDE, R., 1979 Effective deme sizes during long-term evolution estimated from rates of chromosomal rearrangement. Evolution **33**: 234–251.

LARSON, A., E. M. PRAGER and A. C. WILSON, 1984 Chromosomal evolution, speciation, and morphological change in vertebrates: the role of social behaviour. Chromosomes Today **8**: 215–228.

LEVERSHA, M., C. SINFIELD and G. WEBB, 1980 Rapid and reliable methods for the G- and C-banding of human and other mammalian chromosomes. Aust. J. Med. Lab. Sci. **1**: 139–143.

LI, W.-H., and D. GRAUR, 1991 *Fundamentals of Molecular Evolution*. Sinauer, Sunderland, Mass.

LIDICKER, W. Z., JR., and J. L. PATTON, 1987 Patterns of dispersal and genetic structure in populations of small rodents, pp. 144–161 in *Mammalian Dispersal Patterns; the Effects of Social Structure on Population Genetics*, edited by B. D. CHEPKO-SADE and Z. T. HALPIN. University of Chicago Press, Chicago.

NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

NEI, M., and W.-H. LI, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76**: 5269–5273.

PATTON, J. L., 1967 Chromosome studies of certain pocket mice, genus *Perognathus* (Rodentia: Heteromyidae). J. Mamm. **48**: 27–37.

PRAGER, E. M., R. D. SAGE, U. GYLLENSTEN, W. K. THOMAS, R. HUBNER *et al.*, 1993 Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. Biol. J. Linn. Soc. **50**: 85–122.

ROGERS, A. R., and H. HARPENDING, 1992 Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. **9**: 552–569.

SAGE, R. D., 1981 Wild mice, pp. 39–90 in *The Mouse in Biomedical Research*, edited by H. L. FOSTER, J. D. SMALL and J. G. FOX. Academic Press, New York.

SAGE, R. D., E. M. PRAGER, H. TICHY and A. C. WILSON, 1990 Mitochondrial DNA variation in house mice, *Mus domesticus* (Rutty). Biol. J. Linn. Soc. **41**: 105–123.

SAID, K., and J. BRITTON-DAVIDIAN, 1991 Genetic differentiation and habitat partition of Robertsonian house mouse populations *Mus musculus domesticus* of Tunisia. J. Evol. Biol. **4**: 409–428.

SAIKI, R. K., T. L. BUGAWAN, G. T. HORN, K. B. MULLIS and H. A. ERLICH, 1986 Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. Nature **324**: 163–166.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. HIGUCHI *et al.*, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**: 487–491.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SANGER, F., S. NICKLON and A. R. COULSON, 1977 DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**: 5463–5467.

SCRIVEN, P. N., 1992 Robertsonian translocations introduced into an island population of house mice. J. Zool. Lond. **227**: 493–502.

SEARLE, J. B., 1991 A hybrid zone comprising staggered chromosomal clines in the house mouse (*Mus musculus domesticus*). Proc. R. Soc. Lond. B **246**: 47–52.

SEARLE, J. B., 1993 Chromosomal hybrid zones in eutherian mammals, in *Hybrid Zones and the Evolutionary Process*, edited by R. G. HARRISON. Oxford University Press, New York.

SHE, J. X., F. BONHOMME, P. BOURSOT, L. THALER and F. CATZEFLIS, 1990 Molecular phylogenies in the genus *Mus*: comparative analysis of electrophoretic, scnDNA hybridization, and mtDNA RFLP data. Biol. J. Linn. Soc. **41**: 83–103.

SITES, J. W., and C. MORITZ, 1987 Chromosome evolution and speciation revisited. Syst. Zool. **36**: 153–174.

SLATKIN, M., 1985 Gene flow in natural populations. Annu. Rev. Ecol. Syst. **16**: 393–430.

SLATKIN, M., and R. R. HUDSON, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129**: 555–562.

SLATKIN, M., and W. P. MADDISON, 1989 A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics **123**: 603–613.

SPIRITO, F, A. MODESTI, P. PERTICONE, M. CRISTALDI, R. FEDERICI *et al.*, 1980 Mechanisms of fixation and accumulation of centric fusions in natural populations of *Mus musculus*. I. Karyological analysis of a hybrid zone between two populations in the Central Apennines. Evolution **34**: 453–466.

SWOFFORD, D. L., 1991 PAUP (Phylogenetic Analysis Using Parsimony) Version 3.0s. Illinois Natural History Survey, Champaign, Ill.

TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics **105**: 437–460.

THOMAS, W. K., S. PAABO, F. X. VILLABLANCA and A. C. WILSON, 1990 Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. J. Mol. Evol. **31**: 101–112.

TICHY, H., and I. VUCAK, 1987 Chromosomal polymorphism in the house mouse (*Mus domesticus*) of Greece and Yugoslavia. Chromosoma **95**: 31–36.

TUCKER, P. K., B. K. LEE and E. M. EICHER, 1989 Y chromosome evolution in the subgenus *Mus* (genus *Mus*). Genetics **122**: 169–179.

VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES and A. C. WILSON, 1991 African populations and the evolution of human mitochondrial DNA. Science **253**: 1503–1507.

VIROUX, M-C., and V. BAUCHAU, 1992 Segregation and fertility in *Mus musculus domesticus* (wild mice) heterozygous for the Rb (4.12) translocation. Heredity **68**: 131–134.

WALLACE, B. M. N., J. B. SEARLE and C. A. EVERETT, 1992 Male meiosis and gametogenesis in wild house mice (*Mus musculus domesticus*) from a chromosomal hybrid zone; a comparison between "simple" Robertsonian heterozygotes and homozygotes. Cytogenet. Cell Genet. **61**: 211–220.

WALSH, J. B., 1982 Rate of accumulation of reproductive isolation by chromosome rearrangements. Am Nat. **120**: 510–532.

WHITE, M. J. D., 1978 *Modes of Speciation*. W. H. Freeman, San Francisco.

WILSON, A. C., R. L. CANN, S. M. CARR, M. GEORGE, U. B. GYLLENSTEN *et al.*, 1985 Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. **26**: 375–400.

WINKING, H., B. DULIC and G. BULFIELD, 1988 Robertsonian karyotype variation in the European house mouse, *Mus musculus*; survey of present knowledge and new observations. Z. Saugetierkd. **53**: 148–161.