DNA Sequence Variation in the Mitochondrial Control Region of Red-Backed Voles (Clethrionomys)

Cole W. Matson[1] *and Robert J. Baker**

*Department of Biological Sciences, Texas Tech University

The complete mitochondrial DNA (mtDNA) control region was sequenced for 71 individuals from five species of the rodent genus Clethrionomys both to understand patterns of variation and to explore the existence of previously described domains and other elements. Among species, the control region ranged from 942 to 971 bp in length. Our data were compatible with the proposal of three domains (extended terminal associated sequences [ETAS], central, conserved sequence blocks [CSB]) within the control region. The most conserved region in the control region was the central domain (12% of nucleotide positions variable), whereas in the ETAS and CSB domains, 22% and 40% of nucleotide positions were variable, respectively. Tandem repeats were encountered only in the ETAS domain of *Clethrionomys rufocanus*. This tandem repeat found in *C. rufocanus* was 24 bp in length and was located at the 5′ end of the control region. Only two of the proposed CSB and ETAS elements appeared to be supported by our data; however, a "CSB1-like" element was also documented in the ETAS domain.

Introduction

Several studies (Brown et al. 1986; Saccone, Attimonelli, and Sbisà 1987; Mignotte et al. 1990; Biju-Duval et al. 1991; Arnason and Johnsson 1992; Arnason et al. 1993; Ghivizzani et al. 1993; Hoelzel, Hancock, and Dover 1993; Hoelzel et al. 1994; Stewart and Baker 1994a, 1994b; Fumagalli et al. 1996; Sbisà et al. 1997; Wilkinson et al. 1997; Pesole et al. 1999) have provided a detailed characterization of the overall structure of the control region in the mitochondrial genome of mammals and other vertebrates. The mammalian control region is organized into three major regions, or domains, including the extended terminal associated sequence (ETAS), central, and conserved sequence block (CSB) domains (Sbisà et al. 1997; Pesole et al. 1999; fig. 1). Although this general structure is maintained across divergent groups of mammals, the functional significance of each region is less clear. Several authors (Doda, Wright, and Clayton 1981; Sbisà et al. 1997; Shadel and Clayton 1997) have suggested that the ETAS elements provide a termination signal for replication, whereas the CSB elements presumably function in priming H-strand replication (Walberg and Clayton 1981; Chang, Hauswirth, and Clayton 1985; Ghivizzani et al. 1994; Sbisà et al. 1997; Shadel and Clayton 1997) and the central domain is proposed as a site for the origin of replication (Brown et al. 1986; Saccone, Attimonelli, and Sbisà 1987; Sbisà et al. 1997). More recent studies of mammalian control regions have identified tandem repeats within both the ETAS and CSB domains (Mignotte et al. 1990; Biju-Duval et al. 1991; Wilkinson and Chapman 1991; Arnason et al. 1993; Ghivizzani et al. 1993; Hoelzel et al. 1993, 1994; Stewart and Baker 1994a; Fumagalli et al. 1996; Wilkinson et al. 1997). These tandem repeats pro-

vide a source of length heteroplasmy within individuals and species of particular mammalian taxa (Wilkinson and Chapman 1991; Hoelzel et al. 1993; Fumagalli et al. 1996; Wilkinson et al. 1997), and both their specific origin and function are unclear.

Comparative studies of control region variation in mammals have revealed a mosaic pattern of variation, with some regions evolving rapidly (e.g., the ETAS and CSB domains; fig. 2; Brown et al. 1986; Sbisà et al. 1997; Pesole et al. 1999) and others retaining high sequence similarity across divergent taxa (e.g., the central domain; fig. 2; Brown et al. 1986; Sbisà et al. 1997; Pesole et al. 1999). This apparent among-sites rate heterogeneity within the control region has been used to study patterns of genetic variation at several different levels of divergence within mammals. For instance, more variable domains within the control region (the ETAS and CSB domains; fig. 2) have been used for detailed studies of both population structure (Hoelzel et al. 1993; Goldberg and Ruvolo 1997; Ishibashi et al. 1997; Stacy et al. 1997; Bickham et al. 1998; Slade et al. 1998; Rosel et al. 1999; Ehrich et al. 2000; Kerth, Mayer, and Konig 2000; Matson et al. 2000) and phylogeographic structure (Arctander, Johansen, and Coutellec-Vreto 1999; Castro-Campillo et al. 1999; Ehrich et al. 2000; Koh, Lee, and Kocher 2000), whereas more conserved regions (e.g., the central domain; fig. 2) may prove to be useful in phylogenetic studies of more divergent taxa. All of the so-called hypervariable regions (HVRs) within the mtDNA control region are included within the ETAS and CSB domains (Vigilant et al. 1991; Tamura and Nei 1993; Excoffier and Yang 1999; Lutz et al. 2000; fig. 2). With the exception of earlier comparisons of control region structure in Mus and Rattus and other orders of mammals (Brown et al. 1986; Saccone, Attimonelli, and Sbisà 1987; Sbisà et al. 1997), few detailed studies of control region structure and patterns of variation have been provided for other muroid rodents.

This paper provides a detailed characterization of the control region structure of the rodent genus Clethrionomys. Clethrionomys was chosen as a model system because these species are widely distributed and easily
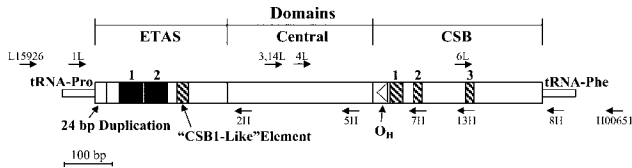
FIG. 1.—Schematic diagram of the mtDNA control region of *Clethrionomys*. Arrows identify the location and direction of amplification and sequencing primers. Locations of extended termination associated sequence (ETAS), central conserved, and conserved sequence block (CSB) domains are also presented. All CSB (diagonal striped box) and ETAS (black box) elements present in *Clethrionomys* are mapped and numbered accordingly. The location of the 24-bp tandem repeat of *Clethrionomys rufocanus* is represented by the gray box. The three-stranded D-loop structure extends from the O_H in the CSB domain to near the ETAS elements; the entire central domain is within the D-loop (Doda, Wright, and Clayton 1981).

collected and have been used in several biological investigations, including for an investigation of the effects of logging on small mammal populations (Nordyke and Buskirk 1991), as a biomarker of pollution (Fallon et al. 1997; Matson et al. 2000), and for virus research (Plyusnin et al. 1994; Horling et al. 1996; Labuda et al. 1997; Lundkvist et al. 1997; Tryland et al. 1998). Patterns of variation across the major control region domains were examined both within and between species. Finally, the overall organization of the control region was compared with similar regions in other rodent and mammal species.

Materials and Methods
Specimens Examined

The control region was sequenced for 71 *Clethrionomys* (53 from the Museum of Texas Tech University and 18 from other institutions; table 1). The control region sequence of a single individual of *Clethrionomys rufocanus* from GenBank was also included (accession number D42091). To characterize microgeographic variation over a limited region, 48 *Clethrionomys glareolus* from northern Ukraine were examined. Samples from three localities were used: Red Forest (51°22′N, 30°05′E), Glyboke Lake (51°26′N, 30°04′E), and Oranoe field station (51°02′N, 30°07′E). Localities were defined as an area of continuous habitat with a radius of less than 150 m. Vouchers for all specimens included in the population analysis have been deposited in the Museum of Texas Tech University, Natural Science Research Laboratory. Identification numbers of specimens included in the population analysis are as follows: TK 81178, TK 81026, TK 50088, TK 50056, TK 50110, TK 81106, TK 81019, TK 74289, TK 81148, TK 74138, TK 50031, TK 74093, TK 74172, TK 74082, TK 74120, TK 74099, TK 81152, TK 74306, TK 74210, TK 81151, TK 74342, TK 81058, TK 50076, TK 74292, TK 81119, TK 81162, TK 74008, TK 81087, TK 74227, TK 81150, TK 81110, TK 74129, TK 74010, TK 50078, TK 74251, TK 74207, TK 50059, TK 50107, TK 50086, TK 81101, TK 81118, TK 81115, TK 81017, TK 81166, TK 50074, TK 50079, TK 50108, and TK 50103. To characterize intraspecific variation, three species were used, including 54 *C. glareolus* (all 48 specimens from the Ukraine
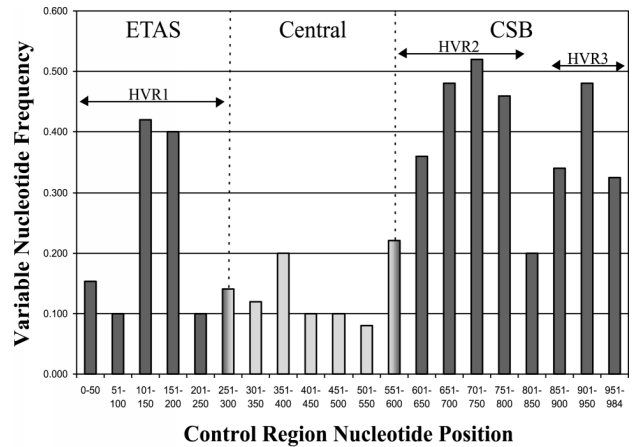

FIG. 2.—Distribution and frequency of variable nucleotides across the mtDNA control region of *Clethrionomys*. The three domains are demarcated by dotted vertical lines. The ETAS and CSB domain bars are shaded black, whereas the central domain bars are light gray. Approximate locations of the three human control region hypervariable regions are also noted.

population), 7 *Clethrionomys gapperi*, and 6 *Clethrionomys rutilus*. To characterize interspecific variation within the genus *Clethrionomys*, 26 specimens were used. These included the specimens used in the intraspecific characterization portion (only two representatives from the Ukraine population included to avoid bias) of this study, in addition to 3 *C. rufocanus* and 2 *Clethrionomys californicus* (table 1).

DNA Extraction and mtDNA Sequence Analysis

Genomic DNA was isolated using a proteinase K/phenol/chloroform extraction method (Longmire, Maltbie, and Baker 1997). Using the polymerase chain reaction (PCR), the entire mitochondrial DNA (mtDNA) control region plus the tRNAPro and tRNAPhe genes were directly amplified from genomic DNA with the primers L15926 and H00651 (Kocher et al. 1989). PCR conditions included 35 cycles of denaturation at 95°C for 15 s, annealing at 55°C for 10 s, and extension at 72°C for 1 min. Following the last cycle, an additional step at 72°C for 7 min was performed. PCR products were then purified using the QIAquick PCR Purification Kit (Qiagen, Chatsworth, Calif.) following the manufacturer's instructions. The complete mtDNA control region (942–971 bp) was sequenced using 10 sequencing primers (Stacy et al. 1997; this study, table 2). Sequencing primers were designed using Oligo 4.0 (National Biosciences, Inc., Plymouth, Minn.) primer analysis software. Sequencing reactions were performed using the ABI dRhodamine Dye Terminator Ready Reaction Mix (PE Applied Biosystems, Foster City, Calif.) and purified following the manufacturer's instructions. DNA sequences were analyzed on an ABI PRISM 310 Genetic Analyzer (PE Applied Biosystems). Sequences were verified and aligned using Sequencher 3.1 software (Gene Codes Corp., Ann Arbor, Mich.) and were then imported and analyzed with the PAUP software package, version 4.0b2a (Swofford 1998). GenBank acces-

Table 1
List of Specimens Included in Intra- and Interspecific Analyses

Species and Museum No.	Locality	Origin of Tissues[a]	GenBank Accession No.
Clethrionomys californicus			
TK 49845.................	Oregon	OSU	AF367180
TK 49846.................	Oregon	OSU	AF367179
Clethrionomys gapperi			
NK 1134	New Mexico	UNM-MSB	AF367182
NK 29177	New Mexico	UNM-MSB	AF367183
RMZ 14002...............	Washington	WSU	AF367184
TK 26677.................	Maine	TTU-NSRL	AF367185
TK 32369.................	Maryland	TTU-NSRL	AF367187
TK 32370.................	Maryland	TTU-NSRL	AF367186
TK 52286.................	Alberta, Canada	TTU-NSRL	AF367181
Clethrionomys glareolus			
MVZFC 4752	Berkshire, England	MVZ	AF367195
MVZFC 4858	Niederosterreich, Austria	MVZ	AF367199
MVZFC 5225	Grisons Canton, Switzerland	MVZ	AF367198
MVZFC 5498	Bayern, Germany	MVZ	AF367196
TK 11834.................	Avon, England	TTU-NSRL	AF367197
TK 74138.................	Chernobyl, Ukraine	TTU-NSRL	AF364667
TK 81101.................	Chernobyl, Ukraine	TTU-NSRL	AF364681
AF 3306..................	Sotkamo, Finland	UAM	AF367194
Clethrionomys rufocanus			
D42091	Hokkaido, Japan	GenBank	D42091
NK 6289	Hokkaido, Japan	UNM-MSB	AF367200
NK 6290	Hokkaido, Japan	UNM-MSB	AF367201
Clethrionomys rutilus			
NK 6361	Hokkaido, Japan	UNM-MSB	AF367188
NK 6372	Hokkaido, Japan	UNM-MSB	AF367189
AF 3172..................	Kilpisjarvi, Finland	UAM	AF367190
AF 21811.................	Alaska	UAM	AF367192
AF 15079.................	Alaska	UAM	AF367191
AF 21914.................	Alaska	UAM	AF367193

[a] Institutional abbreviations are as follows: TTU-NSRL, Museum of Texas Tech University, Natural Science Research Laboratory; UNM-MSB, University of New Mexico, Museum of Southwestern Biology; UAM, University of Alaska Museum; MVZ, University of California, Berkeley, Museum of Vertebrate Zoology; OSU, Oregon State University, Department of Forest Science; WSU, Washington State University, Department of Zoology.

sion numbers are AF364667, AF364681, and AF367179–AF367201 (table 1). Control region domains and elements were identified and demarcated based on sequence data presented for several mammals by Sbisà et al. (1997).

Table 2
Amplification and Sequencing Primers

Primer Name	Primer Sequence (5′–3′)
L15926[a]........	TCAAAGCTTACACCAGTCTTGTAAACC
HN00651[a]......	TAACTGCAGAAGGCTAGGACCAAACCT
stacy1L[b]........	TCCCCACCATCAGCACCCAAAGC
stacy2H[b]	TGGGCGGGTTGTTGGTTTCACGG
cwm3L	CTGGTTCTTACTTCAGGGTCATCC
cwm4L	CCCCTTAAATAAGACATCTCG
cwm5H	CGGCTATGCTGAGTGATACCATCC
cwm6L	CCGAATCTGCCTATCCTTACCAAAC
cwm7H	GGGGGTTGGTAGGTGGAAAAAGC
cwm8H	GCATTTTCAGTGCTTTGCTTTACATAAG
cwm13H	TTGGGGTTTGGYAAGGATAGG
cwm14L	CATCTGGTTCTTACYTCAGGG

[a] Primers described by Kocher et al. (1989) were used only for amplification.

[b] Primers were not given names in the original description by Stacy et al. (1997). Names have been assigned for convenience.

Results
Control Region Length Variation

For the 72 specimens examined, control region length ranged from 942–971 bp. No two individuals shared the same haplotype, with the exception of the specimens included in the populational analysis. A total of 18 haplotypes were present within the 48 specimens from Ukraine (GenBank accession numbers AF364667–AF364684). All individuals examined in interspecific analyses exhibited unique control region sequences. The range in control region length in *C. glareolus, C. gapperi, C. californicus,* and *C. rutilus* was 942–947 bp. *Clethrionomys rufocanus* had a longer control region (970–971 bp), which was the result of a 24-bp duplication (tandem repeat) at the 5′ end of the control region in the ETAS domain (fig. 1). This duplication was present in all *C. rufocanus* examined, as well as a GenBank sequence (accession number D42091). The length of the control region varied in all species examined (*C. rufocanus,* 970–971 bp; *C. glareolus,* 945–946 bp; *C. gapperi,* 945–947 bp; *C. californicus,* 943–944 bp; *C. rutilus,* 942–944 bp). When representatives from all species were included in a single alignment, the overall

Table 3
Base Composition and Nucleotide Polymorphism Heterogeneity

	DOMAIN		
	ETAS	Central	CSB
Base frequency by domain[a]			
Adenine	0.362	0.237	0.357
Cytosine	0.201	0.275	0.269
Guanine	0.075	0.195	0.100
Thymine	0.362	0.293	0.273
No. of variable nucleotides within domain			
Clethrionomys glareolus[b] (n = 54) . . .	10	2	18
Clethrionomys gapperi[b] (n = 7)	34	11	72
Clethrionomys rufocanus[b] (n = 2) . . .	7	0	10
Clethrionomys rutilus[b] (n = 6)	17	6	19
Clethrionomys californicus (n = 2) . .	3	3	3
Average intraspecific frequency[c]	0.046	0.010	0.059
Interspecific (*Clethrionomys*)[b]	61	37	159
Interspecific frequency	0.221	0.118	0.403

[a] Base compositions of the three domains were significantly different from one another (χ^2, $P < 0.0001$).

[b] Numbers of variable nucleotides were significantly different between domains (χ^2, $P < 0.05$).

[c] Calculated using a weighted average, thus taking into account different sample sizes.

length was 984 nucleotide positions, resulting from small gaps used to maximize sequence homology (alignment available in GenBank).

Domain length variation was examined by comparing the unaligned sequence lengths for each individual. These values were compared with all individuals of a given species to the ETAS domain, ranging in length from 251 to 276 bp. This domain contained the 24-bp duplication of *C. rufocanus*. Excluding this duplication, the size range for this domain was 251–252 bp. *Clethrionomys gapperi* was the only species with length variation in this domain (251–252 bp). The central domain ranged in length from 311 to 312 bp. Only two species examined (*C. gapperi* and *C. californicus*) had length variation in this domain. The third domain, CSB, ranged in length from 379 to 384 bp. The CSB domain was variable in length in all but one species (*C. californicus*). However, this lack of length variation in *C. californicus* was probably due to the small number of individuals (n = 2) included in the analysis. The length variation in the CSB domain was caused by several single-nucleotide indels which were not present in all individuals.

Base Composition Heterogeneity

Base composition varied significantly among domains of the *Clethrionomys* control region (table 3). All taxa showed similar composition biases of A = T > C > G in the ETAS domain, T > C > A > G in the central domain, and A > T > C > G in the CSB domain. *Clethrionomys* follows the general mammalian mtDNA control region pattern of (A+T) > (C+G) in all domains (Sbisà et al. 1997).

Variable Nucleotide Frequency Heterogeneity

The number of variable sites differed significantly across the control region, with fewer substitutions in the

central domain and a higher number in the CSB domain relative to the ETAS domain (table 3 and fig. 2). This pattern was present in four of the species examined; however, *C. californicus* had the same number of variable nucleotide positions in each domain. This difference noted within *C. californicus* was most likely the consequence of a small sample size (n = 2) and not a significant pattern change. *Clethrionomys gapperi* had a significantly larger number of intraspecific variable nucleotide positions in all domains than did the other four species.

To determine the type and pattern of mutations, comparisons were made within species and populations. The restriction of this method to intraspecific variation was necessary to reduce the number of assumptions resulting from the extensive variation in this molecule. At many nucleotide positions, there were three or more character states present. Therefore, we were unable to specifically characterize each position for interspecific comparisons. Variation in the control region was classified as pyrimidine or purine transitions (ts-pyr, ts-pur), transversions (tv), or insertions/deletions (indels). Pyrimidine : purine transition (ts-pyr : ts-pur) and transition : transversion (ts : tv) ratios were calculated for all species examined (table 4). Domains differed significantly in frequency of variable nucleotide positions; however, substitution ratios (i.e., ts : tv : indel or ts-pyr : ts-pur) did not differ significantly between domains.

Distributions of variable nucleotide positions were also determined for a population of *C. glareolus* from Ukraine (table 4). There were 7 variable nucleotides (6 ts-pyr, 1 ts-pur) in the ETAS domain and 11 variable nucleotide positions (7 ts-pyr, 1 ts-pur, 2 tv, 1 indel) in the CSB domain. There were no variable nucleotide positions present in the central domain.

Intraspecific Genetic Distance

Genetic distance (Tamura and Nei 1993) ranges were calculated for the three species of *Clethrionomys* that were represented by more than two individuals. Values presented represent the genetic distances between the two most closely related individuals and the two most distantly related individuals within a species. *Clethrionomys glareolus* had a range of distance values of 0.002–0.014. Within our sample, the genetic distance of *C. rutilus* ranged from 0.004 to 0.032, and that of *C. rufocanus* ranged from 0.009 to 0.018. *Clethrionomys gapperi* had the largest genetic distance values, ranging from 0.013 to 0.078. It was not possible to determine a range of distance values for *C. californicus* because of the small sample size (n = 2). The genetic distance for *C. californicus* was 0.008.

Conservation of Structural Elements

There were two ETAS elements in the control region. ETAS1 was located approximately 25–50 bp from the tRNAPro gene at the 5′ end of the control region. This element was highly conserved within *Clethrionomys*. Only 7 of the 57 (12%) nucleotide positions were variable within the ETAS1 element, which was signifi-

Table 4
Intraspecific Polymorphism Frequencies and Ratios

	SPECIES-LEVEL MUTATIONS					POPULATION-LEVEL MUTATIONS
	C. glar (n = 54)	C. gapp (n = 7)	C. rufo (n = 3)	C. ruti (n = 6)	C. cali (n = 2)	C. glar (Ukraine) (n = 48)
ts-pyr	21	67	15	22	5	13
ts-pur	5	23	3	16	0	2
Total ts	26	90	18	38	5	15
ts-pyr/pur	4.20[b]	2.91[b]	5.00[b]	1.38	5.00[b]	6.50[b]
tv	4	26	2	4	3	2
ts/tv	6.50[b]	3.46[b]	9.00[b]	9.50[b]	1.67	7.50[b]
Indels	1	14	1	2	1	1
Variable bp[a]	30	117	21	42	9	18
Length (bp)	945–946	945–947	970–971	942–944	943–944	945–946

NOTE.—*C. glar* = *Clethrionomys glareolus*; *C. gapp* = *Clethrionomys gapperi*; *C. rufo* = *Clethrionomys rufocanus*; *C. ruti* = *Clethrionomys rutilus*; *C. cali* = *Clethrionomys californicus*; ts = transitions; tv = transversions; ts-pyr = pyrmidine transitions; ts-pur = purine transitions.

[a] The number of variable nucleotides is smaller than the sum of mutations in some taxa because of the presence of multiple mutation types at single nucleotide positions.

[b] Ratio significantly different from 1:1 (χ^2, $P < 0.05$).

cantly lower than the 24.7% average variable nucleotide frequency for the remaining ETAS domain (χ^2, $P = 0.030$). In addition, the consensus sequence for ETAS1 in *Clethrionomys* differed from Mus (Sbisà et al. 1997) at only four sites. Unlike ETAS1, ETAS2 was highly variable in *Clethrionomys*, with approximately 25 of the 54 (47%) nucleotide positions being variable. This frequency was significantly higher than that of the remaining ETAS domain, which had only 16% variable nucleotide positions (χ^2, $P < 0.0001$).

There were three CSB elements found in the control region of *Clethrionomys*. The first was CSB1, which was moderately conserved. Within this element, 10 of the 25 (40%) nucleotide positions were variable. This level of conservation (60%) was the same as that found in the remaining CSB domain (60%). CSB2 and CSB3 were the other two CSB elements found in *Clethrionomys*; however, CSB2 was not well conserved. CSB2 had only seven conserved nucleotide positions out of its 19-bp length (60% variable), which made it significantly more variable than the remaining CSB domain (χ^2, $P = 0.032$). CSB3 was the most conserved of all of the structural elements found in the control region. The CSB3 element was significantly more conserved than the remaining CSB domain (χ^2, $P = 0.009$). Only 2 of the 18 (11%) nucleotide positions found within this element were variable. In fact, we found no intraspecific variation in the CSB3 element. In addition to the previously documented CSB elements, we also identified a "CSB1-like" element. However, this element was not in the CSB domain. The CSB1-like element was found 14 bp downstream of the ETAS2 element in the ETAS domain. The consensus sequence of this element for *Clethrionomys* was 5′-AAMYATTAATGYTYDAHAGACATA. When all specimens were included in a single alignment, only 6 of the 24 (25%) nucleotide positions were variable. When compared with the CSB1 element of *Mus musculus* (Sbisà et al. 1997), the consensus sequence of this CSB1-like element differed at three nucleotide positions and was polymorphic at five additional positions. Although the level of conservation in this

CSB1-like element was higher than that found in the CSB1 element in the CSB domain, this level of conservation was not higher than the 22% average for the remaining ETAS domain (55 of 252 positions variable; χ^2, $P = 0.704$).

Discussion

The control region is unique because of a faster rate of evolution as compared with the rRNA and protein-coding genes of the mitochondrial genome. This region in *Clethrionomys* follows the general structure previously described (Brown et al. 1986; Saccone, Attimonelli, and Sbisà 1987; Fumagalli et al. 1996; Sbisà et al. 1997; Pesole et al. 1999). The presence of the three domains that were documented by these studies is supported by our data. Nucleotide composition data and the frequency of variable nucleotide positions for *Clethrionomys* lend additional support for the separation of the control region into three domains.

There are differences in base composition between all three domains. Base composition heterogeneity has previously been documented in several other mammalian taxa by Sbisà et al. (1997). In addition, there are differences in the frequencies of variable nucleotide positions within the three domains (table 3 and fig. 2). When all species of *Clethrionomys* are included, 159 out of the 395 (40%) nucleotide positions within the CSB domain are variable. In the ETAS domain, 61 of the 276 (22%) nucleotide positions are variable. However, in the central domain, only 37 of the 313 (11.3%) nucleotide positions are variable. This conserved central domain has levels of interspecific variation similar to those of mitochondrial rRNA and tRNA genes, as well as nonsynonymous sites of protein-coding genes in several mammalian groups (Pesole et al. 1999), whereas the ETAS and CSB domains have much faster evolutionary rates. This interspecific pattern of polymorphic nucleotide position heterogeneity varies slightly across taxonomic levels. When the frequency of variable sites within a species is calculated, the percentages of variable

nucleotide positions in the CSB, ETAS, and central domains were 5.9%, 4.6%, and 1.0%, respectively (table 3). Finally, the distribution of tandem repeats and short indels is heterogeneous among domains. The only tandem repeat is found in the ETAS domain of *C. rufocanus*. Only one species (*C. gapperi*) has an indel in the ETAS domain, two species have indels in the central domain (*C. gapperi* and *C. californicus*), and four of the five species included have indels in the CSB domain. The CSB domain contains significantly more indels than the other two domains (χ^2, $P = 0.009$). In fact, 14 of the 19 intraspecific indels occurred in the CSB domain.

The amounts of variation found in some of the structural elements were higher than expected based on the level of variation found within the same domain. The fact that some of these elements are conserved across different orders (Brown et al. 1986; Saccone, Attimonelli, and Sbisà 1987; Sbisà et al. 1997) suggests that they should be highly conserved within a single genus. The CSB3 and ETAS1 elements were highly conserved within *Clethrionomys* (>87%), providing additional support for their existence and functionality. The remaining elements were at least as variable as the domains in which they were found, thus raising doubts about the functionality of those elements. Functions have been proposed for both CSB and ETAS elements. Saccone, Attimonelli, and Sbisà (1987) suggest that CSB and ETAS elements are associated with start and stop sites, respectively, for D-loop strand synthesis. Doda, Wright, and Clayton (1981) suggest that displacement-loop strands terminated near specific sequences within the ETAS domain of humans and mice. The fact that many of these elements are not well conserved in the generic analysis suggests that these elements have one or more of the following characteristics. First, the elements may not have a crucial role in replication. Second, only one of the CSB elements and one of the ETAS elements may be needed for mitochondrial DNA replication to function properly. If only one of each type of element is necessary, any additional elements could be variable without detrimental effects. Third, the elements may be able to function properly even with multiple variable nucleotide combinations. If these elements function primarily via secondary structure, high levels of variation may not significantly affect function.

Sbisà et al. (1997) suggest that even though CSB1 was the least conserved sequence block in their study, it is functionally the most important element. They base their conclusion on the observation that CSB1 has been identified in all mammals examined, while CSB2 and CSB3 are sometimes absent. In the current study, CSB3 was the most conserved sequence block, followed by CSB1 and, finally, CSB2, which was the least conserved. A CSB1-like element identified in the ETAS domain of *Clethrionomys* was highly conserved, with only three fixed differences when compared with the CSB1 element of Mus.

Functions have also been attributed to both ETAS elements. ETAS1 may contain a recognition signal for termination of the nascent DNA chain. ETAS2 may contain binding sites for termination factors (Sbisà et al.

1997). If ETAS2 has a critical role in mtDNA replication, its function must not be compromised by multiple nucleotide combinations. However, the high level of conservation found within the ETAS1 element supports the hypothesis that this element may be conserved to function in mtDNA replication.

The single tandem repeat found in *C. rufocanus* is the only documented tandem repeat present in the ETAS domain of rodents. This likely dates the origin of this unique feature as subsequent to the divergence of *C. rufocanus* from the remaining species of *Clethrionomys*. We used the proposed time of the Microtus/*Clethrionomys* divergence (7–8 MYA; Martin et al. 2000) to estimate an overall control region rate of evolution for arvicoline rodents (3.4%–3.9%/Myr), although this rate varies among domains, as expected (ETAS, 3.6%–4.2%; central, 1.5%–1.7%; CSB, 5.8%–6.6%). We then used this overall rate to estimate the date of the divergence of *C. rufocanus* from the remaining *Clethrionomys* species. It appears that *C. rufocanus* diverged approximately 4.2–4.8 MYA. If these dates are accurate, then *C. rufocanus* probably evolved this repeat in the last 4.8 Myr. However, it is also possible that other species of *Clethrionomys* had this tandem repeat or several repeats at one time, but the repeats were subsequently lost in all of the species except *C. rufocanus*.

Species-level analyses revealed a nucleotide bias within all species examined. A bias towards transitions over both transversions and indels was apparent. Ts : tv ratios were as high as 9.5 to 1 (*C. rutilus*). These results provide additional support for the conclusions of Brown et al. (1986), who noted a similar bias when comparing different species of Rattus. In addition, we also documented a significant bias toward pyrimidine (C↔T) transitions over purine (A↔G) transitions in four of the five species (table 4). For example, *C. californicus* exhibited a pyrimidine : purine transition ratio of 5 to 1. This pyrimidine transition bias differs from the results of Brown et al. (1986), who were unable to show any pyrimidine transitional bias when comparing two different species of Rattus. Although we were able to document a bias in nucleotide substitution type within the control region, we were unable to document any substantial differences in transition substitution bias between domains.

Concerning the selection of an appropriate section of the control region for maximum variability, the ETAS and CSB domains provide the greatest variability for population analyses. The ETAS has been shown to be effective for population analysis (Hoelzel et al. 1993; Goldberg and Ruvolo 1997; Ishibashi et al. 1997; Stacy et al. 1997; Bickham et al. 1998; Slade et al. 1998; Rosel et al. 1999; Ehrich et al. 2000; Kerth, Mayer, and Konig 2000; Matson et al. 2000). If small indels are a concern, the CSB region should be avoided. However, if indels are of interest for the analysis, the CSB domain is an ideal region for examination. The central domain might prove useful in interspecific, intergeneric, or even family-level studies. For most interspecific analyses, the ETAS and central domains combined provide adequate resolution. The mtDNA control region can provide an

appropriate region for examination for a number of different types of studies. However, the use of a single domain may provide all of the resolution necessary for a specific analysis, thus reducing the cost of a research project by reducing the required amount of DNA sequencing.

Acknowledgments

LITERATURE CITED

ARCTANDER, P., C. JOHANSEN, and M.-A. COUTELLEC-VRETO. 1999. Phylogeography of three closely related African bovids (Tribe Alcelaphini). Mol. Biol. Evol. **16**:1724–1739.

ARNASON, U., A. GULLBERG, E. JOHNSSON, and C. LEDJE. 1993. The nucleotide sequence of the mitochondrial DNA molecule of the grey seal, *Halichoerus grypus,* and a comparison with mitochondrial sequences of other true seals. J. Mol. Evol. **37**:323–330.

ARNASON, U., and E. JOHNSSON. 1992. The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina.* J. Mol. Evol. **34**:493–505.

BICKHAM, J. W., T. R. LOUGHLIN, D. G. CALKINS, J. K. WICKLIFFE, and J. C. PATTON. 1998. Genetic variability and population decline in Steller sea lions from the Gulf of Alaska. J. Mammal. **79**:1390–1395.

BIJU-DUVAL, C., H. ENNAFAA, N. DENNEBOUY, M. MONNEROT, F. MIGNOTTE, R. C. SORIGUER, A. EL GAÄIED, A. EL HILI, and J.-C. MOUNOLOU. 1991. Mitochondrial DNA evolution in lagomorphs: origin of systematic heteroplasmy and organization of diversity in European rabbits. J. Mol. Evol. **33**:92–102.

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

CASTRO-CAMPILLO, A., H. R. ROBERTS, D. J. SCHMIDLY, and R. D. BRADLEY. 1999. Systematic status of *Peromyscus boylii ambiguus* based on morphologic and molecular data. J. Mammal. **80**:1214–1231.

CHANG, D. D., W. W. HAUSWIRTH, and D. A. CLAYTON. 1985. Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. EMBO J. **4**:1559–1567.

DODA, J. N., C. T. WRIGHT, and D. A. CLAYTON. 1981. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA **78**:6116–6120.

EHRICH, D., V. B. FEDEROV, N. C. STENSETH, C. J. KREBS, and A. KENNEY. 2000. Phylogeography and mitochondrial DNA (mtDNA) diversity in North American collared lemmings (*Dicrostonyx groenlandicus*). Mol. Ecol. **9**:329–337.

EXCOFFIER, L., and Z. YANG. 1999. Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol. Biol. Evol. **16**:1357–1368.

FALLON, B., P. BARRETT, O. CARROLL, P. HENDERSON, J. DONLON, and J. FAIRLEY. 1997. Hepatic phenylalanine hydroxylase of bank voles *Clethrionomys glareolus* Schreber as a biomarker of pollution. Proc. R. Irish Acad. B **97**:115–119.

FUMAGALLI, L., P. TABERLET, L. FAVRE, and J. HAUSSER. 1996. Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. Mol. Biol. Evol. **13**:31–46.

GHIVIZZANI, S. C., S. L. D. MACKAY, C. S. MADSEN, P. J. LAIPIS, and W. W. HAUSWIRTH. 1993. Transcribed heteroplasmic repeated sequences in the porcine mitochondrial DNA D-loop region. J. Mol. Evol. **37**:36–47.

GHIVIZZANI, S. C., C. S. MADSEN, M. R. NELEN, C. V. AMMINI, and W. W. HAUSWIRTH. 1994. In organello footprint analysis of human mitochondrial DNA: human mitochondrial transcription at the origin of replication. Mol. Cell. Biol. **14**:7717–7730.

GOLDBERG, T. L., and M. RUVOLO. 1997. The geographic apportionment of mitochondrial genetic diversity in east African chimpanzees, *Pan troglodytes schweinfurthii.* Mol. Biol. Evol. **14**:976–984.

HOELZEL, A. R., J. HALLEY, S. J. O'BRIAN, C. CAMPAGNA, T. ARNBOM, B. LE BOEUF, K. RALLS, and G. A. DOVER. 1993. Elephant seal genetic variation and the use of simulation models to investigate historical population bottlenecks. J. Hered. **84**:443–449.

HOELZEL, A. R., J. M. HANCOCK, and G. A. DOVER. 1993. Generation of VNTRs and heteroplasmy by sequence turnover in the mitochondrial control region of two elephant seal species. J. Mol. Evol. **37**:190–197.

HOELZEL, A. R., J. V. LOPEZ, G. A. DOVER, and S. J. O'BRIAN. 1994. Rapid evolution of a heteroplasmic repetitive sequence in the mitochondrial DNA control region of carnivores. J. Mol. Evol. **39**:191–199.

HORLING, J., A. LUNDKVIST, M. JAAROLA, A. PLYUSNIN, H. TEGELSTROM, K. PERSSON, H. LEHVASLAIHO, B. HORNFELDT, A. VAHERI, and B. NIKLASSON. 1996. Distribution and genetic heterogeneity of Puumala virus in Sweden. J. Gen. Virol. **77**:2555–2562.

ISHIBASHI, Y., T. SAITOH, S. ABE, and M. C. YOSHIDA. 1997. Sex-related spatial kin structure in a spring population of grey-sided voles *Clethrionomys rufocanus* as revealed by mitochondrial and microsatellite DNA analyses. Mol. Ecol. **6**:63–71.

KERTH, G., F. MAYER, and B. KONIG. 2000. Mitochondrial DNA (mtDNA) reveals that female Bechstein's bats live in closed societies. Mol. Ecol. **9**:793–800.

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

KOH, H. S., W. J. LEE, and T. D. KOCHER. 2000. The genetic relationships of two subspecies of striped field mice, *Apodemus agrarius coreae* and *Apodemus agrarius chejuensis.* Heredity **85**:30–36.

LABUDA, M., O. KOZUCH, E. ZUFFOVA, E. ELECKOVA, R. S. HAILS, and P. A. NUTTALL. 1997. Tick-borne virus trans-

mission between ticks cofeeding on specific immune natural rodent hosts. Virology **235**:138–143.

LONGMIRE, J. L., M. MALTBIE, and R. J. BAKER. 1997. Use of ''lysis buffer'' in DNA isolation and its implication for museum collections. Occas. Pap. Mus. Tex. Tech Univ. **163**:1–3.

LUNDKVIST, A., Y. CHENG, K. B. SJOLANDER, B. NIKLASSON, A. VAHERI, and A. PLYUSNIN. 1997. Cell culture adaptation of Puumala hantavirus changes the infectivity for its natural reservoir, *Clethrionomys glareolus,* and leads to accumulation of mutants with altered genomic RNA S segment. J. Virol. **71**:9515–9523.

LUTZ, S., H. WITTING, H.-J. WEISSER et al. (11 co-authors). 2000. Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? Forensic Sci. Int. **113**:97–101.

MARTIN, Y., G. GERLACH, C. SCHLOTTERER, and A. MEYER. 2000. Molecular phylogeny of European muroid rodents based on complete cytochrome *b* sequences. Mol. Phylogenet. Evol. **16**:37–47.

MATSON, C. W., B. E. RODGERS, R. K. CHESSER, and R. J. BAKER. 2000. Genetic diversity of *Clethrionomys glareolus* populations from highly contaminated sites in the Chornobyl region. Environ. Toxicol. Chem. **19**:2130–2135.

MIGNOTTE, F., M. GUERIDE, A.-M. CHAMPAGNE, and J.-C. MOUNOLOU. 1990. Direct repeats in the non-coding region of rabbit mitochondrial DNA. Eur. J. Biochem. **194**:561–571.

NORDYKE, K. A., and S. W. BUSKIRK. 1991. Southern red-backed vole, *Clethrionomys gapperi,* populations in relation to stand succession and old-growth character in the central Rocky Mountains. Can. Field Nat. **105**:330–334.

PESOLE, G., C. GISSI, A. DE CHIRICO, and C. SACCONE. 1999. Nucleotide substitution rate of mammalian mitochondrial genomes. J. Mol. Evol. **48**:427–434.

PLYUSNIN, A., O. VAPALAHTI, K. ULFVES, H. LEHVASLAIHO, N. APEKINA, I. GAVRILOVSKAYA, V. BLINOV, and A. VAHERI. 1994. Sequences of wild Puumala virus genes show a correlation of genetic variation with geographic origin of the strains. J. Gen. Virol. **75**:405–409.

ROSEL, P. E., S. C. FRANCE, J. Y. WANG, and T. D. KOCHER. 1999. Genetic structure of harbour porpoise *Phocoena phocoena* populations in the northwest Atlantic based on mitochondrial and nuclear markers. Mol. Ecol. **8**:S41–S54.

SACCONE, C., M. ATTIMONELLI, and E. SBISÀ. 1987. Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. J. Mol. Evol. **26**:205–211.

SBISÀ, E., F. TANZARIELLO, A. REYES, G. PESOLE, and C. SACCONE. 1997. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene **205**:125–140.

SHADEL, G. S., and D. A. CLAYTON. 1997. Mitochondrial DNA maintenance in vertebrates. Annu. Rev. Biochem. **66**:409–435.

SLADE, R. W., C. MORITZ, A. R. HOELZEL, and H. R. BURTON. 1998. Molecular population genetics of the southern elephant seal Mirounga leonina. Genetics **149**:1945–1957.

STACY, J. E., P. E. JORDE, H. STEEN, R. A. IMS, A. PURVIS, and K. S. JAKOBSEN. 1997. Lack of concordance between mtDNA gene flow and population density fluctuations in the bank vole. Mol. Ecol. **6**:751–759.

STEWART, D. T., and A. J. BAKER. 1994*a*. Patterns of sequence variation in the mitochondrial D-loop region of shrews. Mol. Biol. Evol. **11**:9–21.

———. 1994*b*. Evolution of mtDNA D-loop sequences and their use in phylogenetic studies of shrews in the subgenus *Otisorex* (*Sorex*: Soricidae: Insectivora). Mol. Phylogenet. Evol. **3**:38–46.

SWOFFORD, D. L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.

TAMURA, K., and M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10**:512–526.

TRYLAND, M., T. SANDVIK, R. MEHL, M. BENNETT, T. TRAAVIK, and O. OLSVIK. 1998. Serosurvey for orthopoxviruses in rodents and shrews from Norway. J. Wildl. Dis. **34**:240–250.

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.

WALBERG, M. W., and D. A. CLAYTON. 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res. **9**:5411–5421.

WILKINSON, G. S., and A. M. CHAPMAN. 1991. Length and sequence variation in evening bat d-loop mtDNA. Genetics **128**:607–617.

WILKINSON, G. S., F. MAYER, G. KERTH, and B. PETRI. 1997. Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. Genetics **146**:1035–1048.