Unequal Transmission of Mitochondrial Haplotypes in Natural Populations of Field Mice with XY Females (Genus *Akodon*)

Hopi E. Hoekstra*

Department of Zoology and Burke Museum, University of Washington, Seattle, Washington 98195

Submitted November 12, 2001; Accepted June 29, 2002; Electronically published December 30, 2002

ABSTRACT: In species with fertile XY females, such as South American field mice (genus *Akodon*), there are two types of mitochondrial DNA (mtDNA), one passing from XX females and one from XY females. The XX mothers pass their mtDNA to their XX daughters. The XY mothers, however, produce both XX and XY daughters. Because of this breeding scheme, the XY mtDNA remains isolated whereas the XX lineage is continuously invaded by XY mtDNA haplotypes. Using a set of recursion equations, I predicted that XY mtDNA haplotypes should rapidly spread through entire populations composed of both XX and XY females. I examined patterns of nucleotide polymorphism and divergence from the mtDNA control region as well as phylogenetic patterns for evidence of an mtDNA sweep. I compared patterns in two sister species, *Akodon boliviensis* and *Akodon azarae,* that are composed of 35% and 10% XY females, respectively. *Akodon boliviensis* XY females are found in all clades of a phylogenetic mtDNA tree consistent with the spread of mtDNA haplotypes. In addition, *A. azarae* mtDNA haplotypes showed no deviations from neutrality. These results, in combination with high levels of mtDNA nucleotide diversity in XY females, suggest an ancient origin (>10^4 generations) of XY females in both *A. boliviensis* and *A. azarae.*

Keywords: mitochondrial DNA, natural selection, meiotic drive, XY females, sex chromosomes, *Akodon.*

Many recent studies have identified regions of the genome that are influenced by balancing, purifying, or directional selection (e.g., Hughes and Nei 1989; Tanaka and Nei 1989; McDonald and Kreitman 1991). Such studies use neutral evolution as a null hypothesis to statistically detect selection or the unequal transmission of alleles at the molecular level (Kimura 1968, 1983). Mitochondrial DNA (mtDNA) in particular has been the focus of many studies of nonneutral evolution (e.g., Nachman et al. 1996). In many cases deviations from neutrality have been detected in the mtDNA of a variety of organisms (Nachman 1998; Rand and Kann 1998). However, in general the cause and strength of selection or unequal transmission is obscure and the effects of genotypic differences on individual fitness are often unknown. Generally such explanations of selection are ad hoc. Ideally, one would predict a priori that patterns at the level of the organism should result in unequal transmission of alleles at a particular locus and then test that prediction using empirical data. Here I examined a system in which such predictions are possible as well as quantifiable, and deviations from neutral evolution are expected because of the unique breeding systems of XX and XY females. Ultimately, these theoretical expectations can be used to estimate the age of XY females.

In several species of South American field mice (genus *Akodon*), fertile XY females exist along with wild-type XX females in natural populations. In mammals, the male sex is determined by the presence of a functional Y chromosome whereas the default phenotype is female. In this case, XY females result from a mutation to the Y chromosome, which in females is referred to as Y* to indicate its inability to function properly in male sex determination (Lizarralde et al. 1982; Vitullo et al. 1986; Bianchi et al. 1993; Espinosa and Vitullo 1996). XY* females have evolved independently in each of six species known to contain XY* females and occur at varying frequencies along with normal XX females (Hoekstra and Edwards 2000). These two genotypic races of females produce two types of mtDNA haplotypes: haplotypes derived from XX ancestors and those derived from the XY* ancestor. XX mothers pass their mtDNA haplotypes to their XX daughters, but XY* mothers pass their mtDNA to both genotypes because they produce both XY* and XX daughters (fig. 1). While litter size is the same for both XX and XY* females, XY* females produce female-biased litters because of the loss of YY* zygotes (Lizarralde et al. 1982; Espinosa 1991;

* Present address: Department of Ecology and Evolutionary Biology, BioSciences West Building, University of Arizona, Tucson, Arizona 85721; e-mail: hopi@u.arizona.edu.

Am. Nat. 2003. Vol. 161, pp. 29–39.

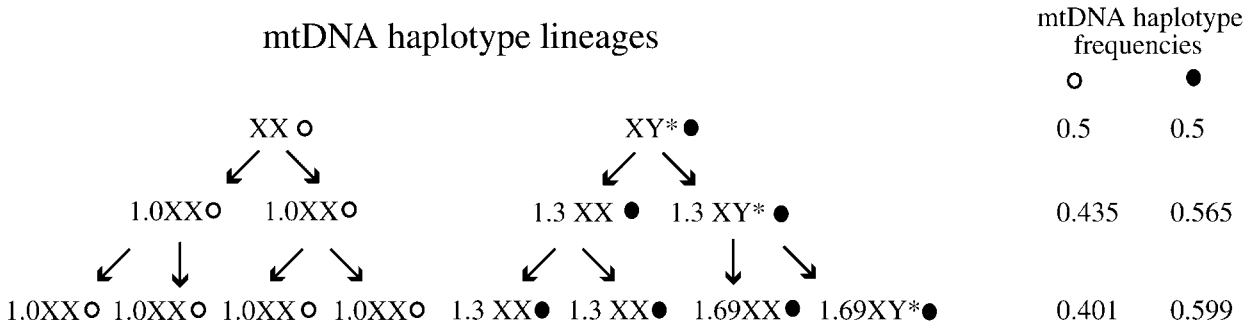

Figure 1: Diagram of breeding scheme and frequencies of females and mtDNA haplotypes. Open circles represent XX mtDNA haplotype; filled circles represent XY* mtDNA haplotype. Numbers represent relative frequency of females. The table on the right shows the change in frequency of mtDNA haplotypes each generation given the frequencies of the XX and XY* females and equal starting frequencies.

Espinosa and Vitullo 1996; Hoekstra and Hoekstra 2001); this provides a transmission advantage to mtDNA haplotypes because XY* mothers produce relatively more daughters (fig. 1). Finally, XY* females have a higher reproductive output relative to their XX counterparts (Espinosa and Vitullo 1996; Hoekstra and Hoekstra 2001). Because of this breeding scheme, the XX lineage is continuously invaded by mtDNA haplotypes originating from XY* females.

In this study I examined two sister species, *Akodon boliviensis* and *Akodon azarae*, in which XY* females occur at 35% and 10%, respectively (Hoekstra and Edwards 2000). On the basis of the transmission dynamics of two competing types of mtDNA, I used a recursion model to predict the dynamics of mtDNA evolution in the XX and XY* mtDNA lineages. I then tested these predictions with empirical data from natural populations to determine how breeding dynamics translate to variation at the molecular level. Specifically, I examined sequences of the mtDNA control region and used both patterns of nucleotide polymorphism and phylogenetic analysis to detect deviations from neutral expectations. These molecular patterns may ultimately be used to estimate the age of XY* females in each species.

Model Predictions

I explored the change in frequencies of XX and XY* mtDNA haplotypes following the origin of the Y* chromosome using a set of recursion equations. It is important to note that this model considers both compensation for loss of YY* zygotes and the relative reproductive fitness difference between XX and XY* females as observed in *Akodon azarae*. XY* females "compensate" for YY* zygotes that are inviable by producing more ovules than implant, equalizing litter sizes between XX and XY* mothers (Espinosa and Vitullo 1996). As a consequence, XY* mothers produce on average four-thirds times as many daughters as an XX mother. When we consider maternally inherited mtDNA haplotypes, males represent an evolutionary "dead end" and are therefore not considered here. Importantly, XX daughters produced by an XY* mother will produce half daughters and half sons (just as an XX daughter derived from an XX mother), so the XY* haplotype will not be passed down multiplicatively in this lineage (fig. 1). Because XY* mothers produce more daughters than XX females, the XY* mtDNA haplotype has a transmission advantage. In addition to producing proportionately more daughters because of compensation, XY* mothers have an added reproductive advantage: XY* mothers have a longer reproductive life and the interval between litters is reduced (Espinosa 1991; Espinosa and Vitullo 1996).

Previous work suggests that meiotic drive may also be operating in this system (Hoekstra and Hoekstra 2001). Transmission bias favoring the Y* chromosome over the X results in unequal transmission of the XY* mtDNA in XY* litters because those loci are jointly inherited. A value of meiotic drive found in XY* females favoring the Y* chromosome over the X ($\delta = 0.264$; i.e., the Y* chromosome is transmitted 76.4% of the time and the X chromosome 23.7% of the time) and a value for male meiotic drive favoring the Y chromosome over the X ($\gamma = 0.1$; i.e., the Y chromosome is transmitted 60% of the time and the X chromosome 40% of the time), which equalizes the sex ratio, are incorporated into the model. Similarly, a relative reproductive fitness estimate ($\omega = 1.15$) favoring XY* females was used. Similar values were estimated using reproductive data from a laboratory colony of *A. azarae* and are reported in Hoekstra and Hoekstra (2001). It is important to note that the prediction of the XY* mtDNA spread holds regardless of the exact parameter values; rather, the parameters influence the rate of the sweep.

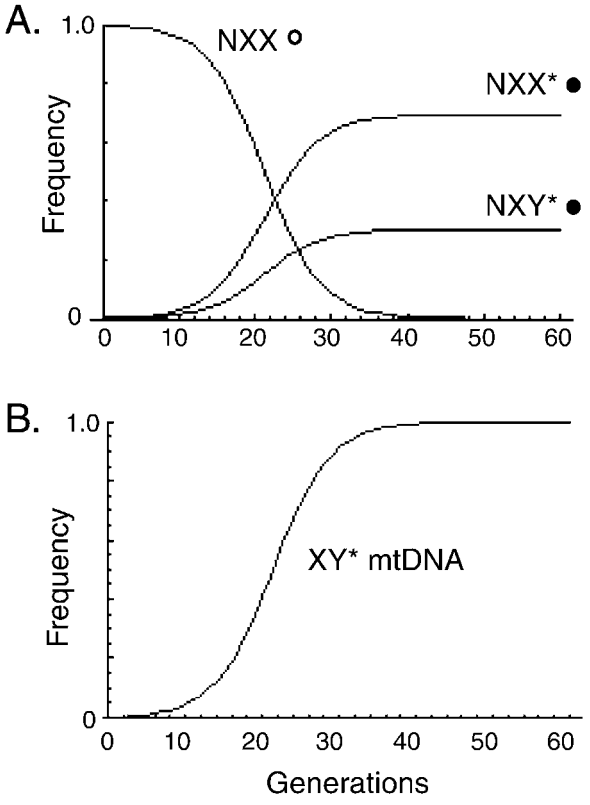

Figure 2: Dynamics and rate of XY* mtDNA haplotype spread. *A* shows the change in the frequency of females over time: N_{XX} = XX females with XX mtDNA; N_{XX^*} = XX females with XY* mtDNA; N_{XY^*} = XY* females with XY* mtDNA. Open circle represents XX mtDNA haplotype; filled circles represent XY* mtDNA haplotype. After 50 generations, XX and XY* females are maintained at equilibrium frequencies. *B* shows the change in frequency of XY* mtDNA haplotype over time. The XY* haplotype reaches fixation (frequency = 1) within 50 generations.

Using these values, however, the model maintains a constant proportion of XY* females to XX females (1 : 9 XY* to XX) and maintains a constant sex ratio (1 : 0.8 M to F); both patterns are consistent with observations from the wild.

The model consists of a set of linear recursion equations that can be solved analytically or numerically. In this model, populations can grow or decline exponentially, but because frequencies of XX and XY* mtDNA haplotypes are calculated, population size does not matter. The equations are N_{XX} = number of XX females with the XX mtDNA haplotype, N_{XX^*} = number of XX females with the XY* mtDNA haplotype, N_{XY^*} = number of XY* females with the XY* mtDNA haplotype, and N_{XY} = number of males (independent of mtDNA haplotype). The parameters are defined as R_{XX} = reproductive rate of XX females, R_{XY^*} = reproductive rate of XY* females (relative to XX females), δ = meiotic drive in XY* females favoring the Y* chromosome (transmission frequency > 0.5), and γ = meiotic drive in XY males favoring the Y chromosome (transmission frequency > 0.5).

$$N_{XX}(t+1) = \frac{1}{2}(1-\gamma)R_{XX} \times N_{XX}(t), \quad (1)$$

$$N_{XX^*}(t+1) = \frac{(1-\delta-\gamma+\delta\gamma)}{(3-\delta-\gamma-\delta\gamma)}R_{XY} \times N_{XX}(t), \quad (2)$$

$$N_{XY^*}(t+1) = \frac{(1+\delta-\gamma-\delta\gamma)}{(3-\delta-\gamma-\delta\gamma)}R_{XY} \times N_{XY^*}(t). \quad (3)$$

These numbers (eqq. [1]–[3]) can then be normalized by dividing by the number of all females to obtain relative frequencies of the three genotypes.

$$N_{XY}(t+1) = \frac{(1-\delta+\gamma-\delta\gamma)}{(3-\delta-\gamma-\delta\gamma)}R_{XY} + \frac{1}{2}(1+\gamma)R_{XX} \times N_{XY}(t) \quad (4)$$

By dividing the number of males (eq. [4]) by the total number of individuals in the population, sex ratio can be determined over time.

This model has several assumptions. First, it is assumed that there is no differential mortality between XX and XY* females. While there are no observed differences in mortality in the lab between these genotypes (A. Vitullo, personal communication), data from field populations are lacking. Second, I assumed that the first XY* individual has a unique mtDNA haplotype. If the initial population of XX females was polymorphic for mtDNA, it could be that some frequency of XX individuals had this same mtDNA haplotype as in the founding XY* individual, thus reducing the time to fixation of this haplotype. Therefore, these equations report the maximum time to fixation. Finally, I assumed no new mutations to the mtDNA haplotypes within the XY* lineage, or identity by descent. However, mutations within the lineage do not affect one's ability to detect a sweep; instead, the sweep is by a clade of related lineages with similar haplotypes rather than a single genotype.

Because XY* mothers produce proportionally more daughters and have more offspring over their lifetimes than XX mothers (Lizarralde et al. 1982), it is expected that XY* mtDNA haplotypes will be unequally transmitted to subsequent generations. This model predicts that the mtDNA haplotype found in a single, ancestral XY* female should, in fact, rapidly increase in frequency and reach fixation in the population (fig. 2). Because of the difference in breeding schemes of XX and XY* females, the frequen-

Table 1: Time-specific expectation for the frequency spectrum of mtDNA nucleotide polymorphisms and shape of the mtDNA phylogeny for species with XY* females

Time	Frequency distribution	Shape of phylogeny
Before origin of XY* females	No skew in allele frequency	Neutral
During mtDNA sweep	No skew to excess of rare alleles	Two clades: XX females (XX mtDNA) and XY*-XX females (XY* mtDNA)
Immediately following sweep	Excess of rare alleles	Star shape
After *N* generations	No skew in allele frequency for XY* females	Neutral for XY* females; shorter XX branches

cies of three female genotypes (XX females with the XX mtDNA, XX females with the XY* mtDNA, and XY* females with the XY* mtDNA) change with time (fig. 2*A*). Ultimately the XY* mtDNA haplotype should reach fixation in a population within 50 generations of the first appearance of XY* females ($P > .99$; fig. 2*B*).

Empirical Predictions

The theoretical predictions described above can be tested using empirical data from mtDNA of species with XY* females. These predicted effects on patterns of nucleotide polymorphism and phylogeny vary with the time since the origin of the first XY* female (table 1). If the origin of XY* females is very recent within a species, patterns of nucleotide variation should reflect the rapid spread of mtDNA haplotypes, which specifically results in a reduced variation and a skew toward an excess of rare sites in the frequency distribution of polymorphisms. The spread of the XY* mtDNA mimics a selective sweep of a beneficial mutation, but it is important to note that the sweep described here is due to the breeding dynamics of the XX-XY* system and not due to any selective fitness advantage associated with XY* mtDNA haplotype. After the spread of *N* generations (where N = population size), the patterns of polymorphism in XY* females should return to neutral expectations. It is also important to note that this model predicts that the spread of the XY* female mtDNA will be rapid and that current variation in mtDNA sequences (in a population with XY* females) has arisen after the origin of XY* females in that population.

The spread of mtDNA haplotypes should also influence the shape of the mtDNA phylogeny and the distribution of XX and XY* females at the branch tips (fig. 3). After the origin of the first XY* female and initial stage of paraphyly, two distinct clades should appear representing the two mtDNA types (XX mtDNA and the invading XY* mtDNA). At this time there will be a nonneutral distribution of females on the branches: one clade should contain the remnant XX clade, and a second clade will consist of XY* females and XX females derived from XY* females (fig. 3). Because of the predicted efficiency of the sweep, it is likely that this pattern will be short-lived. Immediately following the sweep, the XX mtDNA haplotype (and the pure XX female clade) will be wiped out, leaving XY* and XX females randomly distributed throughout the tree. In addition the phylogeny should be star shaped (most mutations are recent and fall on the tips of the branches). After *N* generations the distribution of XY* and XX females at the tips should approximate random because all the XX females carry the XY* mtDNA haplotype. Branches leading to XY* female mtDNA haplotypes should be longer relative to XX females because XX female mtDNA will always be derived from XY* mtDNA haplotypes.

Methods

DNA Samples

Frozen tissue samples of *Akodon azarae* were collected in Argentina (Departamento Buenos Aires, 5 km S of La Plata) by me in the spring of 1998. *Akodon boliviensis* samples from Peru (Departamento Puno, 12 km S of Santa Rosa; 1995–1996) were provided by the Museum of Vertebrate Zoology, Berkeley, California (see appendix). Samples were taken from a single locale for each species in order to avoid complications of potential population subdivision. The DNA was extracted using Quiagen (Valencia, Calif.) tissue kits. Seventeen female individuals (five XY* females) of *A. azarae* and 23 female individuals (six XY*) of *A. boliviensis* are included in this study. The XY* females were detected by polymerase chain reaction (PCR) assays (Hoekstra and Edwards 2000).

Amplification and Sequencing

The entire mtDNA control region, flanking tRNAs, and approximately 100 base pairs (bp) of 12SRNA (corresponding to positions 15330 to 160 in *Mus musculus* mtDNA for *A. azarae* and 15336 to 205 for *A. boliviensis*) were amplified at a 52°C annealing temperature (following Hoekstra and Edwards 2000). Both strands were sequenced on a 377 ABI automated sequencer. A total of 1,187 bp in *A. azarae* and 1,244 bp in *A. boliviensis* were analyzed. There was no evidence to suggest a nuclear copy. Sequences were aligned using Sequencher 4.0 (GeneCodes) and confirmed by eye.

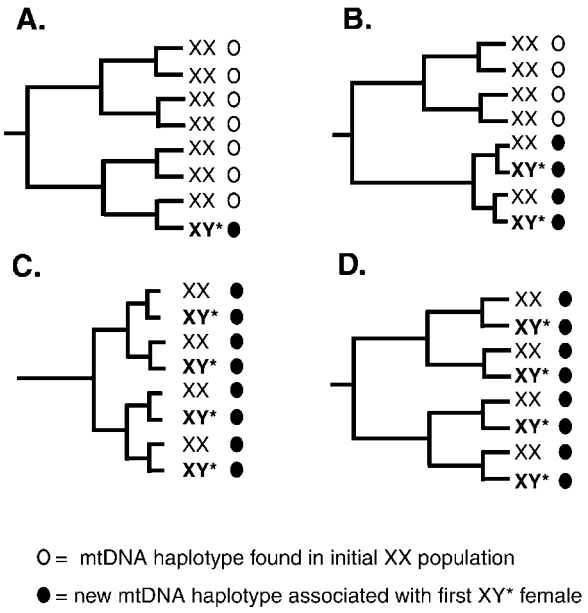

Figure 3: Predicted patterns of the phylogenetic distribution of XX and XY* females over time mapped onto mtDNA gene trees. *A* shows the appearance of the first XY* female with unique mtDNA haplotype; *B* shows initial spread of XY* mtDNA haplotype, and XY* females are restricted to a portion of the topology; *C* shows the complete sweep of XY* mtDNA haplotype through all individuals and shortened branch lengths; and *D* shows the return to neutral expectations. Filled circles represent XY* mtDNA haplotype; open circles represent XX haplotype.

RepeatMasker 2.0 (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) was used to detect any tandem repeats in the sequences.

Phylogenetic Analysis

Phylogenetic analysis was completed using the program PAUP* 4.01b (Swofford 1999). Both neighbor-joining (NJ; Saitou and Nei 1987) and maximum likelihood (ML; Felsenstein 1981) algorithms were used to reconstruct historical relationships. To account for mutation rate heterogeneity as often observed for control regions (Meyer et al. 1999), gamma-distributed rates were applied (estimates are transition/transversion (ti/tv) ratio = 2.69, $\kappa = 5.66$, proportion of invariable sites = 0.405, and gamma distribution $\alpha = 0.72$ for *A. boliviensis*; ti/tv ratio = 3.23, $\kappa = 6.73$, prop. of invariable sites = 0.692, and gamma distribution $\alpha = 0.73$ for *A. azarae*). Randomly chosen out-group sequences from the reciprocal sister species were used to root the trees: *A. azarae* sequence UWBM 72355 was the out-group for the *A. boliviensis* tree and *A. boliviensis* sequence MVZ 172906 was the out-group for the *A. azarae* tree. The program PHYLIP (Felsenstein 1993) was used to conduct a Kashino-Hasegawa test to compare different tree topologies. Differences in coalescent times were calculated using the 95% confidence intervals on branch lengths. The distribution of XY* females across the phylogeny tips was compared with random distributions using Monte Carlo simulations (with a single tree topology) to test for structure in the distribution of XY* females. In addition I estimated the probability that XY* females will cluster in a single clade, given a .5 probability that an XY* female could fall in either clade.

Deviations from Neutrality

The number of segregating sites, number of mutations, and nucleotide diversity were calculated using the program DnaSP (Rozas and Rozas 1999). Fu and Li's (1993) test was used to detect deviations from neutrality. This test uses phylogenetic information to compare the number of mutations found on external branches versus internal branches. Directional selection should produce an excess of mutations on external branches and a negative test statistic, whereas balancing selection should result in an excess of mutations along internal tree branches and a positive test statistic (Fu and Li 1993).

Results

Patterns of Polymorphism

Sequences from the mtDNA control region from both *Akodon boliviensis* and *Akodon azarae* showed similar levels of nucleotide diversity (fig. 4; table 2). Average pairwise nucleotide difference was 5.677 in *A. boliviensis* and 8.866 in *A. azarae*. Unlike in several mammalian taxa, no repeated sequences were found that might complicate evolutionary analysis (Stewart and Baker 1994; Prager et al. 1996; Douzery and Randi 1997). As expected for vertebrate control region, both species showed conserved central domains of the control region (Saccone et al. 1987, 1991).

The number of singletons or nucleotide sites that differed in only one sequence was high in *A. boliviensis* ($n = 15$ excluding gaps and $n = 21$ including gaps; fig. 4*A*). In *A. boliviensis* there were 13 different haplotypes in 17 individuals. Nine different haplotypes were responsible for the 21 singletons, which suggests that a single sequence (e.g., a rare migrant individual) is not responsible for the excess of singletons. *Akodon azarae* sequences had many fewer singletons ($n = 8$ excluding gaps and $n = 10$ including gaps; fig. 4*B*). In *A. azarae*, 13 haplotypes were observed for 23 individuals, and six sequences were responsible for the 10 singletons.

When XY* female mtDNA alone was considered, both species showed high levels of nucleotide diversity. Within

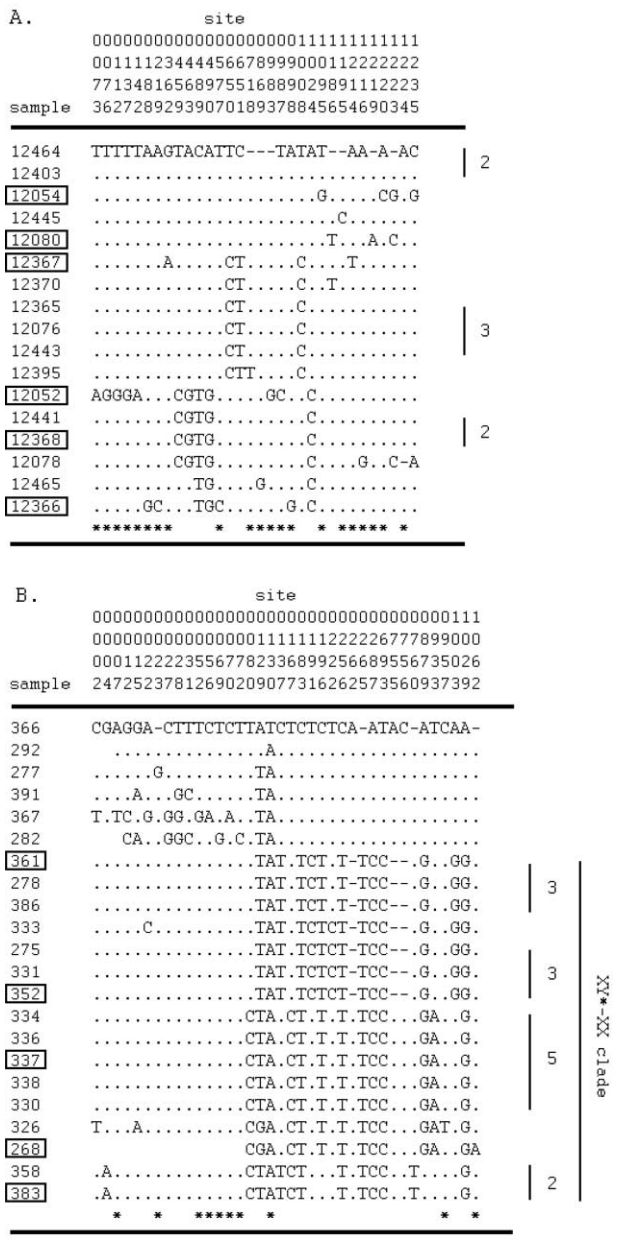

Figure 4: Polymorphic sites for (*A*) *Akodon boliviensis* ($n = 17$) and (*B*) *Akodon azarae* ($n = 23$) mtDNA control region sequences (1,244 bp and 1,187 bp, respectively). Sample names in boxes represent XY* females. Bars on the right indicate the number of shared haplotypes. Dots represent identity with the first sequence. Dashes indicate gaps in the alignment. Asterisks indicate singletons.

the six XY* females of *A. boliviensis*, there are 23 segregating sites (S) and nucleotide diversity (π) is 0.740. Nucleotide diversity of *A. azarae* XY* females was not as high as *A. boliviensis* but still showed relatively high levels compared with the entire *A. azarae* sample ($S = 11$; $\pi = 0.471$).

Phylogenetic Relationships of XX and XY Female mtDNA*

Phylogenetic methods were used to determine the relationships of mtDNA haplotypes from XX and XY* females. If a single XY* mtDNA haplotype had spread throughout the population, XY* females should be scattered throughout the tree. This was the pattern seen in the *A. boliviensis* mtDNA haplotype topology using NJ algorithms (fig. 5*A*). All major clades were retained under ML algorithms (following methods of Voelker and Edwards [1998]). The XY* females were found in all clades of the tree.

If XY* females are of relatively recent origin, the mtDNA phylogeny is expected to have relatively short branches with little interior (or basal) structure. Branch lengths in the observed phylogeny are not short, and the phylogeny does not have a star shape; there are many mutations on internal branches ($\eta_i = 10$). Branch lengths of XY* females appeared longer than those leading to XX females ($P < .10$), which is expected if XX females are derived from XY* female ancestors.

A slightly different pattern emerged from *A. azarae* mtDNA relationships. In NJ topologies there are two ancient clades (characterized by different proportions of XY* females): one with exclusively XX females and a second with a combination of XY* females and XY*-derived XX females (fig. 5*B*). These clades are supported by ML bootstrap values of 68 and 76, respectively. Moving any XY* female into the XX clade results in a significantly worse tree ($P < .05$). However, there is not strong statistical support suggesting that these two clades are different from random expectations. Monte Carlo simulations suggest that the exclusion of the five XY* females from the XX clade is not different from a random distribution ($P > .10$). In addition, Fisher's Exact test ($P = .27$) also does not support a nonrandom distribution of XY* females in the two clades.

When comparing the mtDNA phylogenies from both species, branch lengths in the *A. azarae* topology were on average larger than branches in the *A. boliviensis* topology (fig. 5). Within the *A. azarae* topology, coalescent times were larger in the XX clade than in the XY*-XX clade ($P < .05$).

Tests of Neutrality

Results from the Fu and Li (1993) test suggest that there is a deviation from a model of neutral evolution within *A. boliviensis* (table 2; $D = -2.331$, $F = -2.273$). Of the 25 mutations found in the dataset, 15 were found only in

Table 2: Estimates of nucleotide variability for *Akodon boliviensis* and *Akodon azarae* mtDNA control region sequences

Species	n	bp	S	η	η(s)	k	π (%)	D statistic	P-value	F statistic	P-value
Akodon boliviensis (all)	17	1,244	24	25	15	5.677	.456	−2.331	<.03	−2.273	<.05
A. boliviensis (XY* only)	6	1,244	23	23	18	8.868	.740	−.831	ns	−.890	ns
Akodon azarae (all)	22	1,187	31	34	8	8.866	.747	.386	ns	.457	ns
A. azarae (XY* only)	5	1,187	11	11	5	5.600	.471	.596	ns	.629	ns
A. azarae XY*-XX clade	17	1,187	15	16	3	5.000	.421	.277	ns	.497	ns

Source: D and F statistics are from Fu and Li (1993).

Note: Summary statistics exclude sites with gaps. n = number of sequences/individuals sampled; bp = length of sequence; S = number of segregating sites; η = total number of mutations; η(s) = number of singletons; k = average pairwise nucleotide differences; π (%) = nucleotide diversity; ns = nonsignificant.

a single individual (singletons). Thus, most mutations were at a lower frequency than predicted under neutrality. However, many of the singletons were found along XY* female branches and were not scattered over the tips of the tree as expected following a (selective) sweep. When XY* females branches were removed from the analysis, there was no significant deviation from neutrality and XY* female mtDNA alone did not show a significant skew in the frequency spectrum (table 2).

Sequences from *A. azarae* showed no statistically significant deviations from neutrality (table 2; $D = 0.386$, $F = 0.457$). In this species, mutations were more evenly distributed throughout the genealogy. This pattern is not due to lack of variation, as nucleotide diversity is higher in *A. azarae* than in *A. boliviensis*. Within the XY*-XX clade of *A. azarae* alone ($n = 17$), there were no statistical deviations from neutrality (table 2; $D = 0.277$, $F = 0.497$). Similar to XY* mtDNA patterns from *A. boliviensis*, XY* female mtDNA from *A. azarae* showed no significant deviations from neutral expectations (table 2).

Discussion

Results of the recursion equations, which account for litter sex ratio (including Y* meiotic drive) and relative reproductive fitness of XX and XY* females, predicted the rapid spread of mtDNA haplotypes associated with XY* females in *Akodon* species (figs. 1, 2). This pattern suggests that the presence of a single breeding XY* female should rapidly wipe out existing mtDNA genetic variation and that new variation will accumulate over time. This dynamic allows for the estimation of the age of XY* females in different populations and species. If XY* females originated very recently within the population, the frequency distribution of polymorphisms and the shape of the mtDNA genealogy should reflect a signature of the recent mtDNA sweep (table 1). If there is no evidence of a sweep, which suggests XY* females are not of recent origin, estimation of the present genetic variability of mtDNA haplotypes and the application of molecular clock can provide an estimate of more ancient origins of XY* females. Using this approach, the age of XY* females in *Akodon boliviensis* and *Akodon azarae* can be estimated.

Patterns of mtDNA Variation in Akodon boliviensis

The mtDNA haplotypes from *A. boliviensis* females showed a complex pattern of nucleotide evolution. First, *A. boliviensis* mtDNA sequences had an excess of singletons (external mutations), a hallmark of a directional sweep. However, the singleton mutations were clustered along XY* female branches as opposed to clustering on the external branches of all individuals. Second, the mtDNA sequences appeared to be evolving neutrally when XY* females were removed from the analysis. These patterns of nucleotide variation do not support the very recent spread of XY* haplotypes. Similarly, mtDNA haplotypes from XY* females were phylogenetically interspersed with haplotypes found in XX females (fig. 5*A*). While this pattern suggests a sweep has occurred, the relatively long branches between mtDNA haplotypes, particularly between XY* females, suggest an ancient origin of XY* females in this population. It is of interest to note that in *A. boliviensis* several of the longest branches lead to XY* females (fig. 5*A*). This pattern, which includes several shorter XX female external branches, is consistent with neutral expectations (>N generation; table 1) because XX female mtDNA is expected to originate from the XY* mtDNA lineage that persists since the origin of the first XY* female. In *A. boliviensis* the patterns of nucleotide variation and phylogenetic signal together suggest that XY* female mtDNA has spread through this population but greater than N generations ago.

Patterns of mtDNA Variation in Akodon azarae

The second species, *A. azarae,* also showed little evidence of a recent mtDNA sweep. First, patterns of nucleotide polymorphism showed no detectable deviation from neutral expectations (table 2), which suggests that if a sweep has occurred, sufficient time has passed to return to neutral expectations. However, the mtDNA phylogeny showed two

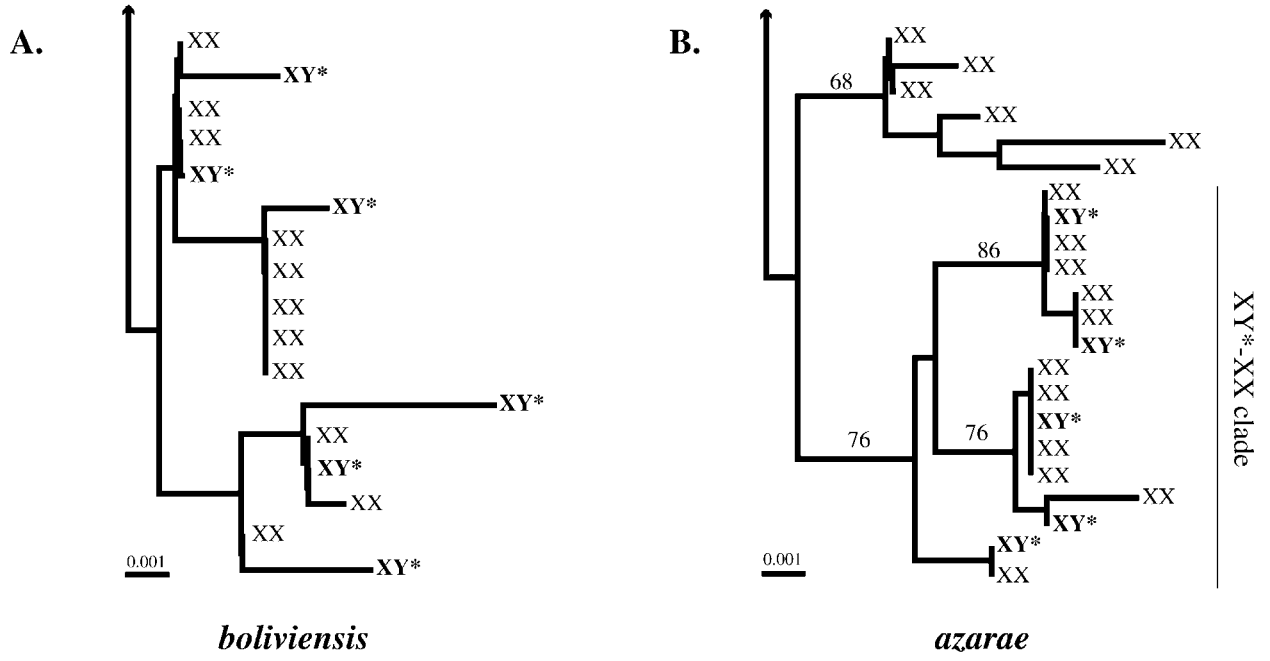

Figure 5: Neighbor-joining phylogenies of mtDNA control region sequences. XX indicates females with XX sex chromosomes, and XY* indicates those with XY* sex chromosomes. *A* shows *Akodon boliviensis* phylogeny of 10 XX and six XY* females rooted with *Akodon azarae*. *B* shows *A. azarae* phylogeny of 18 XX and five XY* females rooted with *A. boliviensis*. The XY*-XX clade is indicated. Bootstrap support above 50% is shown at given nodes (1,000 replicates). Figures were drawn using TreeView (Page 1996).

clades, one of exclusively XX individuals and a second of XY* females and XX females presumably derived from an XY* female ancestor (fig. 5*B*). While this dichotomy is not statistically significant, it is important to discuss how two distinct clades may arise and what impact this may have on understanding the evolution of XY* females in this species.

There are several hypotheses as to how two distinct clades could emerge. First, the spread of the XY* mtDNA haplotype may not yet have reached completion. However, if XY* mtDNA has not finished spreading through the *A. azarae* population, the XY* females must be extremely recent in origin (<50 generations; fig. 2), and this scenario is unlikely. A second hypothesis is that selection is acting differentially on XX and XY* mtDNA to slow the spread of the XY* mtDNA haplotype. In other words, XX females derived from an XY* mother would be less fit than an XX derived from an XX mother. I tested this scenario using the model, and a stable equilibrium of XX and XY*-derived haplotypes can be reached if XY* mtDNA had a relative fitness value of 76.4% of its original fitness (or XY* mtDNA is 1.032 times more fit than XX haplotypes). Additional exploration of the effects of natural selection on the spread of mtDNA haplotypes is warranted. Third, females from the pure XX clade may be either reproductively isolated (e.g., a sibling species) or have recently migrated into the population. This hypothesis can be tested with additional molecular markers such as autosomal microsatellites. Finally, a fourth alternative is simply that incomplete sampling has resulted in a pure XX clade and that additional sampling will reveal an XY* female in this clade.

While it remains unclear if these two clades will be supported after additional rigorous sampling and analysis, the neutral frequency spectrum and the phylogenetic signal from the XY*-XX clade support an ancient origin of XY* females in *A. azarae*. To be conservative, additional estimates of the age of XY* females were conducted using only the XY*-XX clade.

Estimating the Age of XY Females*

Predictions from the model suggest that both skew in the frequency spectrum and phylogenetic patterns of XY* female mtDNA should return to neutral expectations after N generations following an mtDNA sweep, where N is the (effective) population size (table 1). The XY* mtDNA from both *A. boliviensis* and *A. azarae* does not show deviations from neutral expectations (table 2). If a sweep has occurred, then it is likely that XY* females originated over

N generations ago. Although there is no estimate for the effective population size (N_e) for these *Akodon* species, the average estimate of N_e for rodents is approximately 10^4 (e.g., Nachman 1997). Therefore, empirical observations from XY* mtDNA sequences of both species suggest that approximately 10^4 generations or more have passed since the origin of the first XY* female. This rough approximation suggests that XY* females are relatively old (>10^4 generations) in both *A. boliviensis* and *A. azarae*.

A second approach, independent and more precise, to estimating the age of XY* females is based on the nucleotide variation of XY* mtDNA sequences. Because of the unique breeding dynamics of XY* females, the origin of XY* females in a population is expected to erase all existing mtDNA variation as the new XY* mtDNA haplotype spreads throughout the population. Therefore, in populations with XY* females, the present variation has arisen since the origin of XY* females and can be used to estimate the age of XY* female origin on the basis of molecular clock data.

In order to estimate the age of XY* females in these two species, I used two measures of variation: the average percent pairwise difference between XY* female mtDNA haplotypes (derived from table 2) and the percent difference between the two most divergent XY* female mtDNA haplotypes (in order to estimate a maximum age of the clade). Thus, in *A. azarae* only individuals from the XY*-XX clade were considered. The average pairwise difference between XY* haplotypes was 0.740% in *A. boliviensis* and 0.471% in *A. azarae*. Prager et al. (1993) report an approximate rate of evolution of 20% pairwise difference per million years for entire *Mus musculus* control region sequences. This estimate is similar to rate estimates from shrews but is a faster estimate compared with other vertebrates (Stewart and Baker 1994). Using this 20% per million year rate and assuming a molecular clock, the estimated age of the first XY* female is 37,000 yr in *A. boliviensis* and 23,550 yr in *A. azarae*. *Akodon* mice have several litters per year, and breeding age is reached at approximately 3 mo (Espinosa 1995; Espinosa and Vitullo 1996). If on average there are three generations per year, then approximately 1.11×10^5 and 7.07×10^4 generations have passed since the origin of XY* females in *A. boliviensis* and *A. azarae,* respectively. I used the maximum percent difference between XY* mtDNA sequences to estimate the age of the mtDNA clade in each species. The most divergent haplotypes were 1.29% different in *A. boliviensis* and 0.93% different in *A. azarae*. This translates to 64,500 yr and 46,500 yr (or 1.93×10^5 and 1.40×10^5 generations), respectively. If more conservative estimates of divergence rate were used, the age of XY* females would be even older. Using both approaches, the results suggest that XY* females originated at similar times in each species (on the basis of these samples) and that the origin of XY* females is relatively ancient.

It is important to recognize, however, that both the above analyses assume that all mtDNA haplotypes sampled are descended from a single XY* mtDNA haplotype. The introduction of additional XY* mtDNA haplotypes may affect these estimates. If it is assumed that a single XY* female arose within the species, the inclusion of additional XY* females from neighboring populations may result in a larger yet more accurate estimate for the age of XY* females for the species. However, if XY* females originated multiple times within the species via new Y chromosome mutations (in individuals with divergent mtDNA haplotypes), the age of XY* females may be greatly overestimated. Thus, it will be important to identify the molecular changes underlying the XY* female genotype (presumably on the Y* chromosome) within species to determine if all XY* females within the species evolved from a single XY* female. Ultimately, identification of the underlying Y* chromosome mutations can also be determined across species, where it is likely that XY* females have evolved independently (Hoekstra and Edwards 2000).

Conclusions

This study presents a system in which the dynamics at the level of the organism—namely, the different breeding schemes of the two genotypic races, XX and XY* females—affect patterns of variation at the nucleotide level. Theoretical predictions in combination with estimates of nucleotide diversity can be used to estimate the age of XY* females across species of *Akodon*. In two species, *Akodon boliviensis* and *Akodon azarae*, there were no strong deviations from neutrality or phylogenetic signal, which suggests that XY* females were not of recent origin in these populations. This result was confirmed by the high degree of genetic variation in XY* female mtDNA sequences. Additional work on the molecular basis of Y* chromosomes in these two species, as well as other *Akodon* species, will lead to a more complete understanding of the evolution of XY* females across this unique group of South American field mice.

Acknowledgments

I wish to thank C. T. Bergstrom, K. Dyer, S. V. Edwards, D. Garrigan, J. M. Hoekstra, J. Jaenike, J. G. Kingsolver, M. Kuhner, M. W. Nachman, H. Walsh, and one anonymous reviewer for helpful discussion and comments on the manuscript. G. M. Odell was extremely helpful with coding the model. Molecular work was conducted in the lab of S. V. Edwards. Travel to Argentina was supported by G. J. Kenagy and the Burke Museum, University of

Washington. Fieldwork was greatly aided by C. Galliari and A. Minn. Tissue samples of *Akodon boliviensis* were kindly provided by the Museum of Vertebrate Zoology, University of California, Berkeley. I was supported by a Howard Hughes Predoctoral Fellowship. Research was supported by a National Science Foundation Dissertation Improvement Grant (DEB 9902337).

APPENDIX

Akodon azarae: University of Washington, Burke Museum. Collected in Argentina, Dept. Buenos Aires, 5 km S of La Plata (UWBM 72201, 72242, 72249, 72259, 72265, 72296, 72300, 72301, 72303, 72304, 72306, 72308, 72320, 72326, 72331, 72332, 72347, 72350, 72355, and HEH 282, 361).

Akodon boliviensis: University of California Berkeley, Museum of Vertebrate Zoology. Collected in Peru, Dept. Puno, 12 km S of Santa Rosa (MVZ 171603, 171605, 171618, 171620, 171622, 172901–3, 172906, 172922, 172927, 172931, 172929, 173299, 173348–49).

Literature Cited

Bianchi, N. O., M. S. Bianchi, G. Baillet, and A. de la Chapelle. 1993. Characterization and sequencing of the sex determining region Y gene (*Sry*) in *Akodon* (Cricetidae) species with sex reversed females. Chromosoma 102:389–395.

Douzery, E., and E. Randi. 1997. The mitochondrial control region of Cervidae: evolutionary patterns and phylogenetic content. Molecular Biology and Evolution 14: 1154–1166.

Espinosa, M. B. 1991. Hembras hetero y homogameticas en el roeder cricetido *Akodon azarae*: su desempeno reproductivo y mantemimiento en condiciones de laboratorio. Ph.D. diss. Universidad de Buenos Aires.

———. 1995. *Akodon azarae* (Rodentia-Cricetidae): breeding, management and reproductive performance in laboratory conditions. Revista Brasileira de Biologia 55:201–206.

Espinosa, M. B., and A. D. Vitullo. 1996. Offspring sex-ratio and reproductive performance in heterogametic females of the South American field mouse *Akodon azarae*. Hereditas (Lund) 124:57–62.

Felsenstein, J. 1981. Evolutionary tree from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17:368–376.

———. 1993. PHYLIP: phylogenetic inference package. University of Washington, Seattle.

Fu, Y.-X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693–709.

Hoekstra, H. E., and S. V. Edwards. 2000. Multiple origins of XY females (genus *Akodon*): phylogenetic and chromosomal evidence. Proceedings of the Royal Society of London B, Biological Sciences 267:1825–1831.

Hoekstra, H. E., and J. M. Hoekstra. 2001. An unusual system of sex determination in South American field mice (genus *Akodon*): the role of mutation, selection and meiotic drive in maintaining XY females. Evolution 55:190–197.

Hughes, A. L., and M. Nei. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proceedings of the National Academy of Sciences of the USA 86:958–962.

Kimura, M. 1968. Evolutionary rate at the molecular level. Nature 217:624–626.

———. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Lizarralde, M. S., N. O. Bianchi, and M. S. Merani. 1982. Cytogenetics of South American akodont rodents (Cricetidae). VII. Origin of sex chromosome polymorphism in *Akodon azarae*. Cytologia 47:183–193.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351: 652–654.

Meyer, S., G. Weiss, and A. von Haeseler. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152:1103–1110.

Nachman, M. W. 1997. Patterns of DNA variability at X-linked loci in *Mus domesticus*. Genetics 147:1303–1316.

———. 1998. Deleterious mutations in animal mitochondrial DNA. Genetica 103:61–69.

Nachman, M. W., W. M. Brown, M. Stoneking, and C. F. Aquadro. 1996. Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics 142: 953–963.

Page, R. D. M. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences 12:357–358.

Prager, E. M., R. D. Sage, U. Gyllensten, W. K. Thomas, R. Hubner, C. S. Jones, L. Noble, et al. 1993. Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. Biological Journal of the Linnean Society 50:85–122.

Prager, E. M., H. Tichy, and R. D. Sage. 1996. Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*: comparison with other house mice and report of a 75-bp tandem repeat. Genetics 143:427–446.

Rand, D. M., and L. M. Kann. 1998. Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. Genetica 103:393–407.

Rozas, J., and R. Rozas. 1999. DnaSP version 3: an integrated program for molecular population genetics and

molecular evolution analysis. BioInformatics 15: 174–175.

Saccone, C., M. Attimonelli, and E. Sbisa. 1987. Structural elements highly preserved during the evolution of the D-loop containing region in vertebrate mitochondrial DNA. Journal of Molecular Evolution 26:205–211.

Saccone, C., G. Pesole, and E. Sbisa. 1991. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. Journal of Molecular Evolution 33:83–91.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406–425.

Stewart, D. T., and A. J. Baker. 1994. Patterns of sequence variation in the mitochondrial d-loop region of shrews. Molecular Biology and Evolution 11:9–21.

Swofford, D. L. 1999. PAUP*: phylogenetic analysis using parsimony. Sinauer, Sunderland, Mass.

Tanaka, T., and M. Nei. 1989. Positive Darwinian selection observed at the variable-region genes of immunoglobulins. Molecular Biology and Evolution 6:447–459.

Vitullo, A. D., M. S. Merani, O. A. Reig, A. E. Kajon, O. Scaglia, M. B. Espinosa, and A. Perezzapata. 1986. Cytogenetics of South American akodont rodents (Cricetidae): new karyotypes and chromosomal banding patterns of Argentinean and Uruguayan forms. Journal of Mammalogy 67:69–80.

Voelker, G. A., and S. V. Edwards. 1998. Can weighting improve bushy trees? models of cytochrome *b* evolution and the molecular systematics of pipits and wagtails (Aves: Motacillidae). Systematic Biology 47:589–603.

Associate Editor: Austin Burt