Independent Occurrences of Multiple Repeats in the Control Region of Mitochondrial DNA of White-Tailed Deer

JAMES R. PURDUE, TARAS K. OLEKSYK, AND MICHAEL H. SMITH

From the Illinois State Museum, 1101 East Ash Street, Springfield, IL 62704 (Purdue); Laboratory of Genomic Diversity, National Cancer Institute-Frederick, Frederick, MD (Oleksyk); Basic Research Program, SAIC-Frederick, Inc., National Cancer Institute-Frederick, Frederick, MD 21702 (Oleksyk); and Savannah River Ecology Laboratory, Drawer E, Aiken, SC 29802 (Smith).

Address correspondence to T. K. Oleksyk at the address above, or e-mail: oleksyk@ncifcrf.gov.

Abstract

Deer in the genera Mazama and Odocoileus generally have two copies of a 75–base-pair (bp) repeat in the left domain of the control region of the mitochondrial DNA (mtDNA). Phylogenetic analyses further suggest an ancient origin for the duplication supporting a previously stated contention that this event occurred before the separation of Mazama and Odocoileus. However, white-tailed deer (Odocoileus virginianus) had three or four copies of a 75-bp repeat in the control region of their mtDNA in 7.8% of the individuals analyzed, and all of these animals were from the coastal plain of the southeastern United States. When copy 3 is present, it is very similar in sequence to copy 2, but variation suggests that copy 3 probably evolved multiple times from copy 2. The pattern of phylogenetic clustering of the haplotypes from across the coastal plain also suggests that phenotypes with three or four copies of the repeat have originated multiple times. The 44 observed haplotypes showed strong spatial subdivision across the area with subpopulations frequently showing complete shifts in haplotype frequencies from others taken from nearby areas. Many of the subpopulations right along the coast or on adjacent barrier islands have a limited number of haplotypes as would occur in populations undergoing drift because of small numbers of breeding females and limited female dispersal.

Survival and evolution of genes in space and time is the central question of phylogeography (Avise 2000), and population dynamics has important effects on fluctuations of genetic landscapes (Rhodes et al. 1996). While evolution of genes at the macrogeographic scale is well understood, gene dynamics at a finer scale are not so well studied, especially for large, highly mobile wildlife. Use of DNA markers has simplified the study of such spatial genetic patterns, but the choice of an appropriate genetic system depends on the scale of the study. In general, the finer the geographic scale the more variable the marker must be to detect the influence of local population dynamics on spatial genetic patterns (Avise 1994).

The control region, or D-loop, of mitochondrial DNA (mtDNA) is highly variable in mammals. It is noncoding but important for the initiation of replication of the molecule (Clayton 1992; Douzery and Randi 1997; Wilkinson et al. 1997). The control region is divided into three domains, a central conserved region flanked by left and right domains, which tend to be variable (Douzery and Randi 1997). Both the left and right domains can have variable numbers of tandem repeats (VNTRs). Some mammals, such as bats (Wilkinson and Chapman 1991), shrews (Stewart and Baker 1994), cats (Lopez et al. 1996), and sheep (Wood and Phua 1996), have long VNTRs as a common feature in the left domain. In the Cervidae, only Cervus nippon and species of Mazama and Odocoileus show a repeat in the left domain, specifically two copies of a 75–base-pair (bp) sequence in an area known as the RS2 region (Douzery and Randi 1997) (Figure 1A). Because C. nippon clearly falls in a subfamily different from that for Mazama and Odocoileus, Douzery and Randi (1997) concluded that the 75-bp repeat evolved independently in each subfamilial lineage. They also suggest that the repeat in RS2 had an ancient origin. More recently, some white-tailed deer (Odocoileus virginianus) have been shown to have three or four copies of the repeat, suggesting that further evolution in the RS2 region is occurring at the present time (Purdue et al. 2000).
Even though the D-loop is noncoding, increased sequence length, such as that introduced by multiple copies of the repeat in the RS2 region, may slow the rate of replication of mtDNA (Wilkinson et al. 1997). Within the variable regions, variable number of tandem repeat sequences that have been described for numerous species of mammals, including deer for more than a decade (Hoelzel 1993; Hoelzel et al. 1994; Mahmut et al. 2002) may impact survival (Wilkinson et al. 1997). Selection may act against those individuals with multiple copies of the repeat in the RS2 region and constrain the evolutionary pathways taken by the D-loop. The mtDNA genes are important in cell metabolism and the rate of mtDNA transcription. The southeastern coastal plain of the United States may be an ideal region to look for such phenomena in local whitetail populations because the region has not been subject to large-scale migrations of species caused by the movement of glaciers (Hewitt 2001) and has been relatively free from the reintroductions of deer that have occurred during the last century (Blackard 1971).

In addition, there are no absolute barriers to gene flow in the region, although females are very philopatric (Purdue et al. 2000). We have observed deer swimming across major rivers in the area and in the intercoastal waterway. Male dispersal to the barrier islands off the coast must be relatively common because the populations on some of these islands have normal levels of genetic variability for allozymes (Hillestad 1984; Rowland 1989). Whitetails are among the most genetically variable mammals (Breshears et al. 1988) and also show strong spatial and temporal heterogeneity in gene frequencies (Chesser et al. 1982; Smith et al. 1990).

Our primary objective was to document the spatial pattern for different numbers of copies of the repeat in a large sample of white-tailed deer from the southeastern coastal plain of the United States. To put this pattern into a larger geographical context, we have included whitetails from other regions of North America. We also sequenced the control region for a subsample of deer, especially from localities where individuals display varying numbers of copies of the repeat. This allowed us to estimate the order of events that led to the various copies. Finally, we tested the idea that whitetails show fine-scale spatial structuring over the southeastern coastal plain as suggested by Purdue et al. (2000) and compared this degree of structuring to that occurring on a larger geographical scale.

**Materials and Methods**

Muscle samples were collected from hunter-killed white-tailed deer from 30 localities in the United States and Guatemala (Table 1). Total DNA from each sample was extracted with a high salt-alcohol precipitation protocol (Medrano et al. 1990). The left domain of the control region was amplified by the polymerase chain reaction (PCR) using forward primer (LGL283: 5'-TACACTGGTCTTGTAA-ACC-3') set in the flanking tRNA-Pro gene. The reverse primer (ISM015: 5'-ATGGCCCTGTAGAAAGAAC-3')...
was located in the central conserved region of the control region. PCR was accomplished using Taq polymerase. Generally, 50 or 100 ml of PCR product was produced by the following constituents and parameters: 0.5 l l DNA template added to 1/C2 PCR buffer (supplied by manufacturer); 0.2 l M of each primer; 1 mM each of the four deoxynucleotide triphosphates (dNTPs): dATP, dCTP, dGTP, and dTTP; and 1.25 units of Taq. Usually, 32–34 thermal cycles (denature, 95°C for 60 s; anneal, 54°C for 40 s; and extend, 70°C for 60 s, with 3 s added per cycle) were necessary to complete PCR. For success with some samples, slight alterations were necessary in the amount of template, annealing temperature, or number of cycles. The RS2 region was incorporated in the resulting PCR product, and the number of copies of the 75-bp repeat was easily determined by comparing the position of the product in a 1% agarose electrophoretic gel to that of a standard.

Genetic variants (haplotypes) among the PCR samples were identified by the restriction fragment length polymorphism (RFLP) technique. Haplotypes were determined from the positions of bands in 2.5% Metaphor (FMC, Corporation, Rockland, ME) electrophoretic gels after independent digestions with the restriction enzymes AluI, AseI, BfaI, DdeI, and RsaI (Purdue et al. 2000). A representative of each RFLP haplotype with more than two copies of the 75-bp repeat was sequenced for the entire length of the control region. Also sequenced were representatives of all haplotypes found with the more

| Localitya | Clade | Number of RFLP haplotypesb | Copies | Total |
|-----------|-------|----------------------------|--------|-------|
| Illinois  |       |                            |        |       |
| Brown Co. (BRW) | — | 5 | 32 | 0 | 0 | 32 |
| DeKalb Co. (DEK) | — | 4 | 30 | 0 | 0 | 30 |
| Effingham Co. (EFF) | — | 9 | 26 | 0 | 0 | 26 |
| Johnson Co. (JOH) | — | 3 | 36 | 0 | 0 | 36 |
| Piatt Co. (PIA) | — | 7 | 30 | 0 | 0 | 30 |
| Randolph Co. (RAN) | — | 6 | 35 | 0 | 0 | 35 |
| Williamson Co. (WIL) | — | 3 | 30 | 0 | 0 | 30 |
| South Carolina |       |                            |        |       |
| Bull Island, Charleston Co. (BU) | — | 2 | 15 | 0 | 0 | 15 |
| Cedar Knoll Hunt Club, Allendale Co. (CKL) | B, D | 7 | 26 | 8 | 1 | 35 |
| Francis Marion National Forest, Berkeley and Charleston counties (FM) | — | 1 | 106 | 0 | 0 | 106 |
| Halls Island, Beaufort Co. (HI) | A, D | 4 | 2 | 7 | 17 | 26 |
| Hilton Head and Pinckney islands, Beaufort Co. (HH) | — | 1 | 40 | 0 | 0 | 40 |
| Kiawah Island, Charleston Co. (KI) | — | 1 | 13 | 0 | 0 | 13 |
| Palmetto Bluff Plantation, Beaufort Co. (POB) | D | 4 | 10 | 13 | 0 | 23 |
| Savannah River Site, Aiken, Barnwell, and Allendale counties (SRS) | B, C | 8 | 202 | 13 | 0 | 215 |
| Spring Island, Beaufort Co. (SPR) | D | 3 | 42 | 2 | 0 | 44 |
| Weeb Wildlife Center, Hampton Co. (WC) | B, D | 3 | 13 | 18 | 0 | 31 |
| South Carolina |       |                            |        |       |
| Blackbeard and Sapelo islands, McIntosh Co. (BB) | — | 1 | 65 | 0 | 0 | 65 |
| Cumberland Island, Camden Co. (CUM) | C | 2 | 52 | 1 | 0 | 53 |
| Fort Stewart Military Reservation, Bryan and Liberty counties (FS) | — | 3 | 35 | 0 | 0 | 35 |
| Harris Neck NWR, Liberty Co. (HN) | — | 1 | 31 | 0 | 0 | 31 |
| Jekyll Island, Glynn Co. (JKL) | C | 2 | 3 | 3 | 0 | 6 |
| Kings Bay Submarine Support Unit, Camden Co. (KB) | — | 1 | 27 | 0 | 0 | 27 |
| Ossabaw Island, Bryan Co. (OSS) | — | 1 | 30 | 0 | 0 | 30 |
| St. Catherines Island, Liberty Co. (STC) | — | 2 | 23 | 0 | 0 | 23 |
| Paulk’s Pasture WMA, Glynn Co. (PPA) | C | 4 | 31 | 5 | 0 | 36 |
| Wassaw Island, Chatham Co. (WAS) | — | 4 | 37 | 0 | 0 | 37 |
| Other Localities |       |                            |        |       |
| Guatemala (GTM) | — | 3 | 3 | 0 | 0 | 3 |
| South Dakota, western portion (SOD) | — | 4 | 11 | 0 | 0 | 11 |
| Texas, southwest portion (TEX) | — | 3 | 11 | 0 | 0 | 11 |
| Totals (%) |       |                            |        |       |

|        | 2     | 3     | 4     | Total |
|--------|-------|-------|-------|-------|
|        | 1047  | 70    | 18    | 1135  |

a See Figure 2 for location of populations across the range of Odocoileus virginianus and Figure 3 for the phylogenetic tree.
b RFLP haplotypes were determined using protocols described in Purdue et al. (2000).
c Abbreviations used elsewhere in the paper.
than two copies-variants at the locality as well as selected haplotypes from a variety of other populations. A total of 33 individuals were sequenced. For sequencing, forward primer LGL283 was paired with reverse primer CST-39 (5′-GGGTCGGAAAGGCCTGACAAAC-3′) set in the flanking RNAPlus gene (Strobeck 1992). Usually, several internal primers were needed to produce a complete sequence (forward primers: ISMO01, 5′-GCATATTAGTTCTTAAATACC-3′; ISMO05, 5′-GATCCGGTCCC-YTAGATCACCAC-3′; and ISMO16, 5′-CATCTCGATGGACTAATGAC-3′ and reverse primers: ISMO02UGA, 5′-GATTTTACTATTGCTATG-3′ and IMS017, 5′-CCAAAACCCTATGTGTTTATGG-3′). Sequencing was accomplished on automated sequencers in the Molecular Genetics Instrumentation Facility of The University of Georgia in Athens. Sequence alignments were accomplished by eye using the ESEE (The Eyeball Sequence Editor ver 1.09d) Sequence Editor (Cabot 1989). Alignments were accomplished easily, except for a long string of guanines (Gs) in the right domain that tended to disrupt the sequencing reaction; the segment of Gs was not included in subsequent analyses. The sequence numbering system used by Douzery and Randi (1997) was adopted in our analysis. Phylogenetic analyses were accomplished with the beta version of PAUP* (Swofford 1998) using maximum likelihood (ML). For each analysis, default parameters were assumed: equal rates for transitions versus transversions. Particulars of each analysis are outlined in text and figure captions.

Results

We examined the length of the left domain of the mtDNA for 1,135 deer (Table 1). RFLP analysis revealed much sequence variation with 44 haplotypes identified. All deer examined had at least two copies of the 75-bp repeat in the RS2 region. However, some individuals (7.8%) displayed heteroplasmy indicated by two copies of the repeat. PCR product for a few individuals was double banded, which suggested that heteroplasmy might be present. However, slight alterations of one or more parameters of the PCR nearly always produced a single-banded product. The individuals with additional copies of the 75-bp repeat occurred in populations either along or near the Savannah River drainage or along the coast of southern Georgia (Figure 2). In five populations, haplotypes with three or four copies were the predominant variant present. No instances of added repeats were detected in haplotypes in deer from localities outside the southeastern United States (N = 244).

Gene trees based on the sequences of the control region were constructed using maximum parsimony, ML, and Neighbor-Joining algorithms (Swofford 1998). All analyses produced a common set of well-supported major clades (results not shown), but only the ML tree is illustrated here (Figure 3). One major branch, depicted in Figure 3 by dashed lines, contains haplotypes that occur in the southeastern United States, Illinois, and Guatemala. All haplotypes in this lineage have only two copies of the repeat in their sequences.

Another major clade, shown by solid lines in Figure 3, consists of variants that have two, three, or four copies of the repeat in the RS2 region. Within this lineage, haplotypes with three or four copies appear in at least four separate, well-supported clades (Figure 3). Within these clades, most haplotypes with extra copies of the repeat are paired with sister haplotypes that contain only two copies. Two haplotypes displayed four copies of the repeat, and both types occurred in the same clade along with variants that display two or three copies of the repeat. A geographic trend is apparent in the gene tree. Clade D (Figure 3) is composed of haplotypes found in populations CKL, HI, POB, SPR, and WC (Table 1), which are located along the southern portion of the Savannah River or on the lower coastal plain of a nearby river drainage to the north (Figure 2). The geographic distribution of Clade B, a star phylogeny of unresolved relationships, overlaps that of Clade D but is offset north along the Savannah River. Haplotypes included in Clade C are found in southern Georgia. Nodes within Clade C are well supported with three lineages, two of which are found only on barrier islands [Cumberland (CUM) and Jekyll (JKL) islands in Figure 2]. Haplotypes from a mainland locality, PPA, represent the third lineage in Clade C. Surprisingly, haplotype SRS-d, which is found on the Savannah River Site about 240 km north of PPA, is included in the lineage. Finally, Clade A, which contains just two haplotypes, appears restricted to HI.

Copies of the repeat in the RS2 region are not identical. For convenience, repeats are numbered from the 3′ to the 5′ end of the control region, which presumably represents the order of evolution (Douzery and Randi 1997). For three clades in Figure 3 that included haplotypes with more than two copies, we considered each copy of a repeat as an independent operational taxonomic unit and examined their phylogenetic relationships (Figure 3, panels B, C, and D). For Clade B, the first copies of all haplotypes form an unresolved group with SRS-c, copy 2, which is separated with strong support from a loose group of second and third copies. A similar pattern exists for clades C and D, where first copies group together separate from another supported cluster of all the second, third, and fourth (Clade D only) copies.

Discussion

Our data suggest that white-tailed deer with three or more copies of the 75-bp repeat in the RS2 region occur only in one major lineage of the species, restricted among our populations to the southeastern United States. Deer were sampled from a large geographic area, ranging from South Dakota south to Guatemala and southeast to South Carolina and Georgia. Only in South Carolina and Georgia did we find haplotypes with more than two copies of the repeat. Other population-based analyses have examined mtDNA variability in white-tailed deer (Carr et al. 1986; Cronin et al. 1991; Ellsworth et al. 1994), but the techniques used in these studies were not sensitive enough to detect the length variation observed in the control region. Even at a smaller geographic scale, a spatial pattern in the occurrence of more than two
copies is also evident. In the southeastern United States, longer-than-normal haplotypes appear along and near the Savannah River and again in southern Georgia, but the two areas are interrupted in northern coastal Georgia by a suite of haplotypes with the customary two copies of the repeat. Because female deer are not strong dispersers (Purdue et al. 2000) and because mtDNA is transmitted only through the female line, the geographic distribution of haplotypes with more than two copies of the repeat suggests that independent events have been involved in the evolution of the molecule.

The phylogenetic analysis of the control region supports strongly the notion that extra copies of the RS2 repeat have evolved multiple times in the major lineage of haplotypes that occur in the southeastern United States. Extirpation and subsequent transplanting efforts complicate the picture, but the counties along the Savannah River in the upper and lower coastal plain and the lower coastal plains of South Carolina and Georgia only rarely received relocated deer (Blackard 1971). In contrast, virtually all of deer in Illinois derive from restocking efforts (Pietsch 1954). Northern stocks provided most of the introduced animals in Illinois (Pietsch 1954) and parts of Georgia (Blackard 1971), but apparently these lineages only carry two copies of the repeat. The sequence of the deer from Illinois (RAN-a in Figure 3) is divergent from those representing southeastern deer (SRS-e, which is similar to RAN-a, is thought to stem from introduced deer that probably spread from Georgia to South Carolina). In Illinois, 11 RFLP haplotypes have been identified (Purdue unpublished), and based on the similarities in RFLP banding patterns, all would appear to be similar to RAN-a. Unlike the lineage of deer in the southeast, there is no instance known of haplotypes with more than two copies of the repeat.

**Figure 2.** Map showing localities sampled and geographic distribution of haplotypes with more than two copies of the 75-bp RS2 repeat. Abbreviations for locations are given in Table 1. Solid circles represent populations that contain haplotypes that contain only two copies of RS2. Open circles are used to represent populations with at least one haplotype containing more than two repeats. Letters within circles indicate clades defined in Figure 3. Shaded area indicates the recognized extent of the North American range of white-tailed deer *Odocoileus virginianus* (Wilson and Ruff 1999).
occurring in animals from Illinois or elsewhere, even though many mtDNA variants have been identified. Both Neigel and Avise (1993) and Templeton et al. (1995) note the tendency for closely related mtDNA variants to be found in geographic proximity.

Four clades containing haplotypes with more than two copies of the repeat are identified in Figure 3. All have high levels of support (bootstrap values are 87 for Clade A, 85 for Clade B, 61 for Clade C, and 80 for Clade D), although internal nodes within Clade C are better supported than the basal node (Figure 3). This may indicate that within Clade C, third copies of the repeat have evolved independently three times. Elaborating on the situation for Cumberland Island, haplotypes CUM-a and -b are unique to the island and form a well-differentiated branch on which one haplotype has two and the other three copies of the repeat. A similar situation is present on Jekyll Island, where the JKL-series haplotypes form a distinct clade with variants with two and

Figure 3. Unrooted bootstrap consensus tree using ML based on 1,207 bp of the control region of the mtDNA in white-tailed deer. The tree was generated in beta version 4.0 PAUP* (Swofford 1998). Heuristic search and bootstrap options were exercised, the latter for 100 replicates. Each sequenced haplotype is identified by a two- or three-digit locality code given in Table 1, followed by a variant notation. All of the haplotypes on the dashed branches have no more than two repeats. In addition, three independent gene trees (panels B, C, and D) depict relationships between individual copies of the 75-bp repeat for three out of four clades defined (B, C, and D). Each individual copy of the repeat within each haplotype was considered an independent operational taxonomic unit and analyzed in a separate ML tree. The number of the copy shows their order as in Figure 1 and is represented by different symbols. Copy 1 (dark circles) always forms a separate group with high level of bootstrap support (68%–99%). The remaining clade (A) represents two sequences from a single location: Halls Island, Beauford county, GA (HI).
three copies in the RS2 region. Also on the mainland, another well-supported clade includes haplotypes with two and three copies. The presence of three haplotypes with three copies of the repeat in Clade C is probably the result of convergent evolution.

Unlike that for Clade C, the structure within Clade B is totally unresolved. Within the clade, there are six haplotypes, half of which has two copies of the repeat and the other half has three. Although no internal structure was decipherable, the overall distinctness of Clade B is well supported and thus likely represents at least one other instance of the independent evolution of a third copy of the repeat. Although fewer haplotypes are involved, a similar situation and conclusion exist for Clade A.

Clade D is more complex, with the presence of five haplotypes (Figure 3). One of the haplotypes carries two copies of the repeat, whereas two haplotypes display three copies and the remaining two haplotypes contain four copies. The branching pattern within the clade is only supported moderately. Also, the haplotypes within the clade are distributed over a fairly wide geographic area along drainages of the middle and lower Savannah River and the northerly adjacent Broad River. However, no geographic patterning is evident for Clade D, and resolution is lacking to evaluate the particulars of evolution within the clade. Probably the events that led to formation of haplotypes with extra copies of the repeat within the clade occurred convergent with the evolution of variants with added copies elsewhere in the southeast.

Among the enumerated clades, there is a continuum of internal support: Clade C has well-defined internal nodes, whereas Clade B has none and Clade D is somewhere in between. This feature may reflect the timing of the formation of the third and fourth copies of the repeat. The unresolved Clade B may stem from relatively recent events, especially when compared to those that produced the well-structured Clade C. Clade D, with its intermediate level of internal support, may have evolved after Clade C but before Clade B. The presumed timing of the mutational events supports the contention that some copies of the repeat have evolved independently.

The phylogenetic comparison of copies in separate panels represented next to the corresponding plates of Figure 3 addresses the order of events for the evolution of the RS2 region. Within each of the three clades analyzed, the first copies form a cluster distinctly different from a second cluster that comprised second, third, and fourth (Clade D only) copies (represented by the darker circles in panels B, C, and D of Figure 3). Only in Clade B does a second copy (SRS-6, copy 2) appear in a first copy cluster, but even then it is quite different from any of the copy 1 sequences in the cluster (bootstrap support is at 68, Figure 3B). Douzery and Randi (1997) concluded that the second copy of the repeat in the RS2 region for Mazama and Odocoileus evolved early in the history of the Odocoileini. Our data support this view, given that all the deer we examined carried at least two copies of the repeat and the differentiation between sequences of the first and second copies was always distinct. However, copies 3 and 4, when they occur, seem to represent more recent events. Our data suggest that the third copy derives from the second copy, given the similarities of the sequences. The origin of a fourth copy is unresolved from our data, but the fourth probably stems from the third.

Repeats are probably formed by intra- or intermolecular recombination (Rand and Harrison 1989), strand slippage, or competitive displacement during replication (Levinson and Gutman 1987; Wilkinson et al. 1997). Strong secondary structure may stabilize the repeat region (Wilkinson et al. 1997), which may account for the apparent antiquity of the RS2 repeats in cervids (Douzery and Randi 1997). All the repeats show high levels of sequence conservation as compared to Mazama sp., its likely closest relative of Odocoileus that also has multiple RS2 copies (Figure 1B). On the other hand, these RS2 repeats are characterized by the presence of a 22-bp-long sequence called RS-XXII, which has the ability to form a single-strand hairpin (Douzery and Randi 1997). This hairpin may indicate the region of DNA slippage during replication, which may be the origin of the instability RS2 repeats in white-tailed deer. Because more than two copies of the repeat can occur within white-tailed deer (Purdue et al. 2000) and domestic sheep (Wood and Phua 1996), the addition of new copies of the repeat may not be rare. New combinations with additional repeats are found in a number of different areas on the southeastern coastal plain. This situation may indicate one way in which new DNA variants can become established during the evolution of a species.

Wilkinson et al. (1997) suggest that the number of copies present in the left domain of the control region in mammals is the result of a balance between selection and mutation. They note further that there may be a fixed probability that a repeat in an array will fold and either be duplicated or deleted during replication. If that were the case, then length mutation rates would increase additively with the number of copies of the repeat. Deer, like many other mammals, have three conserved sequence elements (TAS, mt5, and mt6) in each copy of the RS2 repeat (Wilkinson et al. 1997). Apparently, the order of these elements is critical for the stable secondary structure needed for the binding of a regulatory protein crucial for replication. Douzery and Randi (1997) also note that a robust secondary structure could stabilize the repeat region in Mazama and Odocoileus. While it is unclear how the secondary structure of the RS2 region and regulatory proteins work together during replication of the mtDNA, the added copies of the repeat could provide redundancy in the event when binding is hampered by a point mutation, which apparently can be a common event (Wilkinson et al. 1997). However, selection may favor short sequences in the RS2 region because overall replication of shorter mtDNA would be faster (Wilkinson et al. 1997). Consequently, selection for rapid replication of the mtDNA genome would tend to keep the number of copies of the RS2 repeat low.

The major lineage of deer native to the southeastern United States apparently is more susceptible to the mutational events that produce additional copies of the 75-bp repeat. Alternatively, it may be under selection that favors additional redundancy and therefore greater retention of the added copies. If, however, the balance between redundancy
and control region length is shifted toward more copies of the repeat, then selection must not be uniform across our study area or there is also a stochastic component involved in the appearance of the additional repeats. Variants with more than two copies of the repeat are distributed in at least two discreet geographic pockets, which may imply that the direction of selection has a spatial component. If this is the case, it is unlikely that such conditions only occur on the coastal plains of South Carolina and Georgia. The situation may be more apparent than real due to our intensive sampling effort in the southeastern United States.

The spatial pattern of occurrence for the additional repeats may be due to stochastic population processes. The characteristics of the base-pair sequences in the central region of the D-loop in the major lineage of southeastern coastal plain deer may allow the establishment of an occasional addition of another copy of the repeat. Once present, the copy remains until random drift results eventually in either the elimination or the fixation of the variant in a population. Among our study localities, we found diversity in the frequency of individuals with more than two copies of the repeat. In three populations (HI, POB, and WC), haplotypes with more than two copies were dominant and conceivably could be drifting toward fixation. Selection and drift interact, and their interaction could be the most important cause for the observed spatial pattern (Wright 1978). A shifting balance between selection and drift could be caused by dynamics within the metapopulation of white-tailed deer on the coastal plain (Hanski 1996). Allozyme studies have also documented strong spatial heterogeneity in allele frequencies for a variety of loci (Chesser et al. 1982; Scribner et al. 1997; Smith et al. 1990). These results along with the high degree of spatial divergence in gene frequencies observed in this study and that of Purdue et al. (2000) for mtDNA suggest that a metapopulation structure is characteristic for this species on the southeastern coastal plain. However, this structure may be modified where the metapopulation has been heavily impacted by reintroductions.

This paper gives an example of how the repeat in the RS2 region could have evolved. Many attempts at establishing the second copy of this repeat may have been attempted before a more stable combination became established. Because of female philopatry and the maternally inherited nature of mtDNA, drift likely occurred in local deer populations within the overall metapopulation. Many of the combinations were probably lost due to chance before fixation occurred. Whether this will occur for the third and fourth copies is questionable. There are undoubtedly limits to the amount of extra material that can be added to the control region because of its effects on the speed of replication for the mtDNA. This scenario is very similar to the way in which Wright (1978) envisioned evolution occurring in his shifting balance model.

Acknowledgments

We thank the hunters in Georgia, South Carolina, and elsewhere who allowed their deer to be sampled. Tissues were collected with the assistance of

C. Bales, D. Bauman, J. P. Davis, P. A. Dratch, D. Etter, D. Forster, G. Garris, J. Hamilton, R. Hayes, C. Hericks, P. E. Johns, P. Moore, C. M. Nixon, J. Robinette, F. Rogers, P. Schornfeld, H. Simmons, T. Swanyham, D. H. Thomas, and R. J. Warren. Laboratory assistance was provided by J. Kepp, K. J. Little, A. Painter, F. Rogers, M. Stevens-Sommer, and D. White. A draft of the manuscript was critiqued by E. Randi and A. L. Rosa. Support was provided by the St. Catherine’s Island Foundation and the U.S. Department of Energy contract (DE-FCO9-96SR18546) with The University of Georgia and by federal funds from the National Institutes of Health/NCI (NO1-CO-12400) Intramural Research Program. The content of this publication necessarily reflects the views or policies of the Department of Health and Human Services or the U.S. Department of Energy and the mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

References

Avise JC. 1994. Molecular markers, natural history and evolution. New York: Chapman and Hall.

Avise JC. 2000. Phylogeography: the history and formation of species. Cambridge, MA: Harvard University Press.

Blackard JJ. 1971. Restoration of the white-tailed deer in the southeastern United States (MS thesis). Baton Rouge, LA: Louisiana State University.

Breshears DD, Smith MH, Cothran EG, and Johns PE. 1988. Genetic variability in white-tailed deer. Heredity 60:139–146.

Cabot EL and Beckenbach AT. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Computer Application in the Biosciences 5(3):233–234.

Carr SM, Ballinger SW, Derr JN, Blankenship LH, and Bickham JW. 1986. Mitochondrial DNA analysis of hybridization between sympatric white-tailed deer and mule deer in west Texas. Proc Natl Acad Sci 83: 9576–9580.

Chesser RK, Smith MH, Johns PE, Manlove MN, Straney DO, and Baccus R. 1982. Spatial, temporal and age-dependent heterozygosity of beta-hemoglobin in white-tailed deer. J Wildl Manage 46:983–990.

Clayton DA. 1992. Transcription and replication of animal mitochondrial DNAs. In: Mitochondrial genomes (Wolstenholme DR and Jeon KW, eds). International Review of Cytology, vol 141. San Diego, CA: Academic Press; 217–232.

Cronin MA, Nelson ME, and Pac DF. 1991. Spatial heterogeneity of mitochondrial DNA and allozymes among populations of white-tailed deer and mule deer. J Hered 82:118–127.

Douzery E and Randi E. 1997. The mitochondrial control region of Cervidae: evolutionary patterns and phylogenetic content. Mol Biol Evol 14:1154–1166.

Ellsworth DL, Honeycutt RL, Silvy NJ, Smith MH, Bickham JW, and Klimstra WD. 1994. White-tailed deer restoration to the southeastern United States: evaluating genetic variation. J Wildl Manage 58:686–697.

Hanski I. 1996. Metapopulation ecology. In: Population dynamics in ecological space and time (Rhodes OE Jr, Chesser RK, and Smith MH, eds). International Review of Cytology, vol 141. San Diego, CA: Academic Press; 13–43.

Hewitt GM. 2001. Speciation, hybrid zones and phylogeography—or seeing genes in space and time. Mol Ecol 10:537–549.

Hosezel AR. 1993. Evolution by DNA turnover in the control region of vertebrate mitochondrial DNA. Curr Opin Genet Dev 6:891–895.

Hosezel AR, Lopez JV, Dover GA, O’Brien SJ. 1994. Rapid evolution of a heteroplasmic repetitive sequence in the mitochondrial DNA control region of carnivores. J Mol Evol 39:191–199.

Hillestad HO. 1984. Stocking and genetic variability of white-tailed deer in the southeastern United States. (PhD dissertation). Athens, GA: University of Georgia; 112.
Levinson G and Gutman GA, 1987. Slipped-strand mis-pairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221.

Lopez JV, Cevario S, and O’Brien SJ, 1996. Complete nucleotide sequences of the domestic cat (Felis catus) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. Genomics 33(2):229–246.

Mahmut H, Masuda R, Onuma M, Takahashi M, Nagata J, Suzuki M, and Ohtaishi N, 2002. Molecular phylogeography of the red deer (Cervus elaphus) populations in Xinjiang of China: comparison with other Asian, European, and North American populations. Zoolog Sci 19(4):485–495.

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, and Dubchak I, 2000. VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16(11):1046–1047.

Medrano JF, Aasen E, and Sharow L, 1990. DNA extraction from nucleated red blood cells. Biotechniques 8:43.

Neigel JE and Avise JC, 1993 Application of a random walk model to geographic distributions of animal mitochondrial DNA variation. Genetics 135:1209–1220.

Pietsch L, 1954. White-tailed deer populations in Illinois. Springfield, IL: Illinois Natural History Survey Biological Note 34.

Purdue JR, Smith MH, and Patton JC, 2000. Female philopatry and extreme spatial genetic heterogeneity in a large mammal. J Mammal 81:179–185.

Rand DM and Harrison RG, 1989. Molecular population genetics of mtDNA size variation in crickets. Genetics 121:551–569.

Rhodes OE Jr, Chesser RK, and Smith MH, 1996. Population dynamics in ecological space and time. Chicago, IL: University of Chicago Press.

Rowland RD, 1989. Population genetics of white-tailed deer on Cumberland Island, Georgia. (MS thesis). Athens, GA: University of Georgia; 42.

Scrbner KT, Smith MH, and Chesser RK, 1997. Spatial and temporal variability of microgeographic genetic structure in white-tailed deer. J Mammal 78:744–755.

Smith MH, Willis KB, and Johns PE, 1990. Spatial-genetic variation in a white-tailed deer herd. In: Transactions XX Congress Game Biologists, vol. 1. Trondheim, Norway; 80–84.

Stewart DT and Baker AJ, 1994. Patterns of sequence variation in the mitochondrial D-loop of shrews. Mol Biol Evol 11:9–21.

Strobeck C. 1992. Molecular genetic research and DNA repository: Final Report. Department of Zoology, University of Alberta, Edmonton, Alberta, Canada.

Swofford DL., 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, MA: Sinauer Associates.

Templeton AR, Routman R, and Phillips CA, 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, Ambystoma tigrinum. Genetics 140:767–782.

Wilkinson GS and Chapman AM, 1991. Length and sequence variation in evening bat D-loop mtDNA. Genetics 128:607–617.

Wilkinson GS, Mayer F, Kerth G, and Petri B, 1997. Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. Genetics 146:1035–1048.

Wilson DE and Ruff S, 1999. The Smithsonian book of North American mammals. Washington, DC: Smithsonian Institution Press.

Wood NJ and Phua SH, 1996. Variation in the control region sequences of the sheep mitochondrial genome. Anim Genet 27:25–33.

Wright S, 1978. Evolution and the genetics of populations, volume 4: variability within and among natural populations. Chicago, IL: University of Chicago Press.

Received July 2, 2005
Accepted February 27, 2006

Corresponding Editor: Stephen J. O’Brien