Phylogeny and evolution of Cervidae based on complete mitochondrial genomes

W.-Q. Zhang and M.-H. Zhang

College of Wildlife Resources, Northeast Forestry University, Harbin, China

Corresponding author: M.-H. Zhang
E-mail: zhangminghai2004@126.com

ABSTRACT. Mitochondrial DNA sequences can be used to estimate phylogenetic relationships among animal taxa and for molecular phylogenetic evolution analysis. With the development of sequencing technology, more and more mitochondrial sequences have been made available in public databases, including whole mitochondrial DNA sequences. These data have been used for phylogenetic analysis of animal species, and for studies of evolutionary processes. We made phylogenetic analyses of 19 species of Cervidae, with *Bos taurus* as the outgroup. We used neighbor joining, maximum likelihood, maximum parsimony, and Bayesian inference methods on whole mitochondrial genome sequences. The consensus phylogenetic trees supported monophyly of the family Cervidae; it was divided into two subfamilies, Plesiometacarpalia and Telemetacarpalia, and four tribes, Cervinae, Muntiacinae, Hydropotinae, and Odocoileinae. The divergence times in these families were estimated by phylogenetic analysis using the Bayesian method with a relaxed molecular clock method; the results were consistent with those of previous studies. We concluded that the evolutionary structure of the family Cervidae can be reconstructed by phylogenetic analysis based on whole mitochondrial genomes; this method could be used broadly in phylogenetic evolutionary analysis of animal taxa.

Key words: Cervidae; Mitochondria genome; Bayesian method; Evolution analysis

INTRODUCTION

The family Cervidae, which includes 40 species of deer, constitutes the second most populous family of artiodactyls (Bouvrain et al., 1989). The species of Cervidae are distributed widely throughout the world, including the northern hemisphere as well as South America and Southeast Asia, and they have adapted to various climates and environments. As a result, these species vary widely in size, habitat, and behavior, making their classification difficult (Kuwayama and Ozawa, 2000). Researchers have adopted many synapomorphies to identify genera of deer, for example, the presence of antlers in males. Grubb (1993) has developed an identification system for Cervidae that classifies deer genera into four subfamilies: Cervinae, Muntiacinae, Hydropotinae, and Odocoileinae. Other morphological classifications and phylogenies have been developed that agree with the monophyly of Cervidae retained by Grubb (1993), but many inconsistencies remain in the intergeneric relationships, for example, the intergeneric relationships of subfamily Odocoileinae.

Owing to the use of different character matrices for different morphological phylogenies of Cervidae, disagreements have arisen among the various monophyly of Cervidae. Therefore, it is difficult to evaluate the reliability of classification results because a clear and authoritative methodological approach is lacking. Moreover, owing to a high number of homologues in morphological characters, the phylogeny of Cervidae deciphered using morphological characters is questionable. Therefore, a molecular approach should be used to classify Cervidae.

Several molecular approaches have been used for identification in family Cervidae, and most of the genes used for species determination are from the mitochondrial genome (Douzery and Randi, 1997; Ludt et al., 2004). Randi et al. (1998) used cytochrome b (Cyb) for effective construction of the monophyly of Plesiometacarpalia. Hassanin and Douzery, (2003) have used multigene analysis for the major Plesiometacarpalia/Telemetacarpalia dichotomy, and Amato et al. (2000) have used several mitochondrial genes or regions (D-loop, Cyb, 12S and 16S ribosomal RNA [rRNA], ND4, and ND4L) to determine the monophyly of subfamily Muntiacinae, which provided strong support for the classification of Muntiacus and determined interspecific relationships. Within Cervinae, the monophyly has been well developed through molecular analysis of Cyb and the D-loop region (Kuwayama and Ozawa, 2000). These techniques have also been used for subfamily Odocoileinae (Hundertmark et al., 2002; Hundertmark and Bowyer, 2004). Molecular investigations have greatly helped in delimiting the phylogenetic tree within Cervidae, but owing to insufficient genetic information for classification, the interpretation of the evolution of the family has been impeded.

The mitochondrial genome is a small and circular molecule of 15-20 kb and is conserved in vertebrates (Grande et al., 2008; Flegontov et al., 2011). Therefore, it is regarded as the marker of choice for population genetics, reconstruction of phylogenetic relationships among vertebrates, and molecular evolution analysis (Grande et al., 2008; Flegontov et al., 2011). The mitochondrial genome contains 37 genes, including 13 protein-coding genes, 22 transfer RNA genes, two rRNA genes, and the control region (D-loop). The excellent characteristics of the mitochondrial genome, such as the small size of the molecule, its abundance in animal tissues, the strict orthology of encoded genes, and its uniparental inheritance make mitochondrial DNA (mtDNA) a reliable and easy-to-use phylogenetic marker (Chan et al., 2010; Kan et al., 2010; Krajewski et al., 2010; Li et al., 2010; Md-Zain et al., 2010; Morin et al., 2010). In animals, many phylogenetic trees have been re-

constructed from the sequence of single genes or elements-for example, Cyb or the D-loop region. The information provided by single genes or genetic elements is poor, however, and such phylogenetic analyses cannot avoid bias.

Currently, mtDNA sequences are released quickly in public databases, and many mitochondrial genomes of animals are available in GenBank, EMBL-Bank, and the DNA Databank of Japan. In family Cervidae, mtDNA of approximately 19 species from four subfamilies have been sequenced, and they are available for researchers, making phylogenetic analysis based on the whole mitochondrial genome feasible. Therefore, in this study, we present phylogenetic analyses based on the complete mitochondrial genomes of family Cervidae retrieved from GenBank. We used the molecular data to examine hypotheses regarding the evolution and classification of the family, focusing on the evolutionary relationships of subfamilies.

MATERIAL AND METHODS

Phylogenetic and evolution analysis of Cervidae

All mitochondrial genome sequences of family Cervidae were searched and downloaded from the National Center for Biotechnology Information website. The access numbers of the mitochondrial genomes of 19 species of Cervidae are listed in Table 1. All of the protein coding regions, transfer RNA folding regions, and rRNA regions were identified through annotation information. The mitochondrial genome of *Bos taurus* was also downloaded for use as an outgroup. All of the sequences were aligned using ClustalW2 (Larkin et al., 2007) with manual adjustment. The multiple sequence alignments of 20 mitochondrial genomes were input into Molecular Evolutionary Genetics Analysis (MEGA) version 4 for evolution analysis. Neighbor-joining (NJ) analysis with the bootstrap test (1000 replicates) was performed in MEGA using pairwise deletion for gaps/missing data (Tamura et al., 2007). Owing to the unavailability of data partitioning and the general time reversible

Table 1. Species and accession numbers of mitochondria genome sequences used in this study.

Species	Accession number
Cervus elaphus	NC_007704
Cervus elaphus songaricus	NC_014703
Cervus elaphus xanthopygus	NC_013836
Cervus elaphus yarkandensis	NC_013840
Cervus eldi	NC_014701
Cervus nippon centralis	NC_006993
Cervus nippon hortulorum	NC_013834
Cervus nippon taiouanus	NC_008462
Cervus nippon yakushimae	NC_007179
Cervus nippon yesoensis	NC_006973
Cervus unicolor swinhoei	NC_008414
Elaphodus cephalophus	NC_008749
Hydropotes inermis	NC_011821
Muntiacus crinifrons	NC_004577
Muntiacus muntjak	NC_004563
Muntiacus reevesi	NC_004069
Muntiacus reevesi micrurus	NC_008491
Odocoileus virginianus	NC_015247
Rangifer tarandus	NC_007703
Bos taurus	AF492351

model in MEGA, however, the non-partitioned dataset and the best available Tamura-Nei model were used for the NJ reconstruction.

A Bayesian approach implemented within the program MrBayes 3.1.2 was used to analyze the nucleotide and amino acid datasets using the general time reversible model with gamma-distributed rate variation (Ronquist and Huelsenbeck, 2003). Four Metropolis-coupled Markov chain Monte Carlo analyses were run twice for 5,000,000 generations and sampled every 100 generations. After the first 2500 trees (25%) were discarded as burn-in, the remaining 7500 sampled trees were used to estimate the 50% majority rule consensus tree and the Bayesian posterior probabilities. The results were visualized within Mesquite.

Estimation of divergence times in Cervidae

Following previous research, three calibration points were used for the analysis: the first was the oldest fossil of Cervidae (20 ± 2 mya), the second was the oldest fossil attributed to the subfamily Muntiacinae (8 ± 1 mya), and the third was to the oldest fossil of tribe Odocoileini (5 ± 1 mya; Dong et al., 2004). Using such constraints involves the assumption that the age of the oldest fossil attributed to a node is a good approximation of the minimum age of the node. The divergence times within family Cervidae were estimated using the Bayesian method implemented in the Bayesian evolutionary analysis sampling trees (BEAST) program with a relaxed molecular clock method (Drummond and Rambaut, 2007). The BEAST program was set as follows: the substitution models were implemented with the model $GTR + I + \Gamma$ and 1,000,000 generations of Metropolis-coupled Markov chain Monte Carlo steps sampled every 5000 generations with two independent runs. The log output files were combined using LogCombiner 1.5.3, and the effective sample size values (>350) were adequate for all parameters. Convergence was assessed in Tracer 1.5, and the first 1000 samples (10,000,000 generations) were excluded as burn-in. The chronological tree files were analyzed and visualized using TreeAnnotator 1.5.3 and FigTree 1.3.1.

RESULTS

Monophyly of family Cervidae

The NJ and Bayesian phylogenetic trees, which are based on mitochondrial genome sequences, have similar topology, as shown in Figure 1. From bottom to top, there were two subfamilies, Plesiometacarpalia and Telemetacarpalia. Several tribes such as Cervinae, Muntiacinae, Hydropotinae, and Odocoileinae have been classified in these subfamilies by Gilbert et al. (2006). The monophyly of Cervidae was well supported, with a bootstrap value of 100% in NJ and posterior probabilities of 1.00 in Bayesian analysis. The phylogenetic tree showed that the tribes Hydropotinae and Odocoileinae diverged earlier, and Muntiacinae and Cervinae later. The phylogenetic tree showed not only the phylogeny of the whole family but also the monophyly of tribes, such as tribe Cervinae. The phylogenetic tree suggested evolution information about *Cervus nippon*, *C. elaphus*, *C. eldi*, and *C. unicolor*. In addition, it could determine classification information within species-for example, the division of *C. elaphus* into three ecotypes: *C. elaphus yarkandensis*, which was adapted in

Europe, and *C. elaphus xanthopygus* and *C. elaphus songaricus*, which were Asian ecotypes (Pitra et al., 2004; Gilbert et al., 2006).
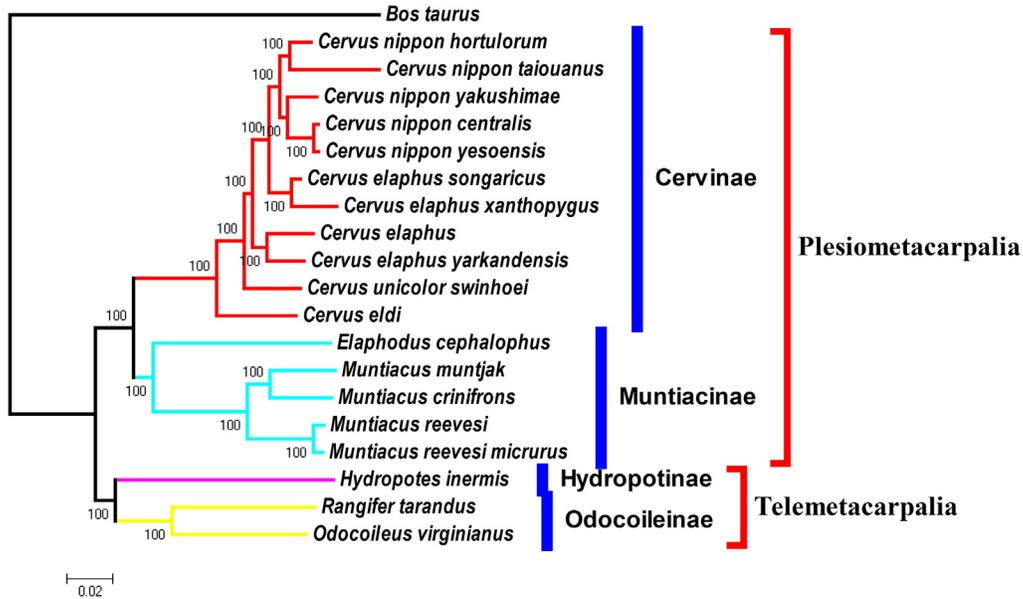
Figure 1. Phylogenetic tree of the relationships among family Cervidae based on whole mitochondria genomes. *Bos taurus* served as outgroup.

Estimation of divergence times in family Cervidae

Three pairs of calibration points were used for estimating divergence times. For example, Cervidae (20 ± 2 mya), subfamily Muntiacinae (8 ± 1 mya), and tribe Odocoileini (5 ± 1 mya) were combined to estimate divergence times (Dong et al., 2004). The estimated divergence times of family Cervidae are shown in Figure 2. The results suggested that Hydropotinae and Odocoileiae diverged from each over a long time period, whereas other subfamilies, such as subfamily Cervinae, diverged during a very short time frame. According to the phylogenetic tree, the origin of family Cervidae was estimated to be a mean value of 7.17 mya, indicating that the true divergence times of family Cervidae was later than the fossil evidence suggested. The results agreed with dates estimated from molecular phylogenetic analysis, such as those by Gilbert et al. (2006), who suggested that family Cervidae was present at 7.7-9.6 mya. The origin of American deer has been estimated at 6.15 mya, which was also close to the 8.9-8.7 mya estimated by researchers. The subfamilies Plesiometacarpalia and Telemetacarpalia became divided at same time, 6.13 and 6.15 mya, respectively, whereas the four tribes, Cervinae, Muntiacinae, Hydropotinae, and Odocoileinae, were divided at different times. The divergence times of some species of Cervidae were very small; for example, the divergence time between *C. nippon centralis* and *C. nippon yesoensis* was just 0.17 mya, which suggested that they may not be two species but two varieties of *C. nippon*.
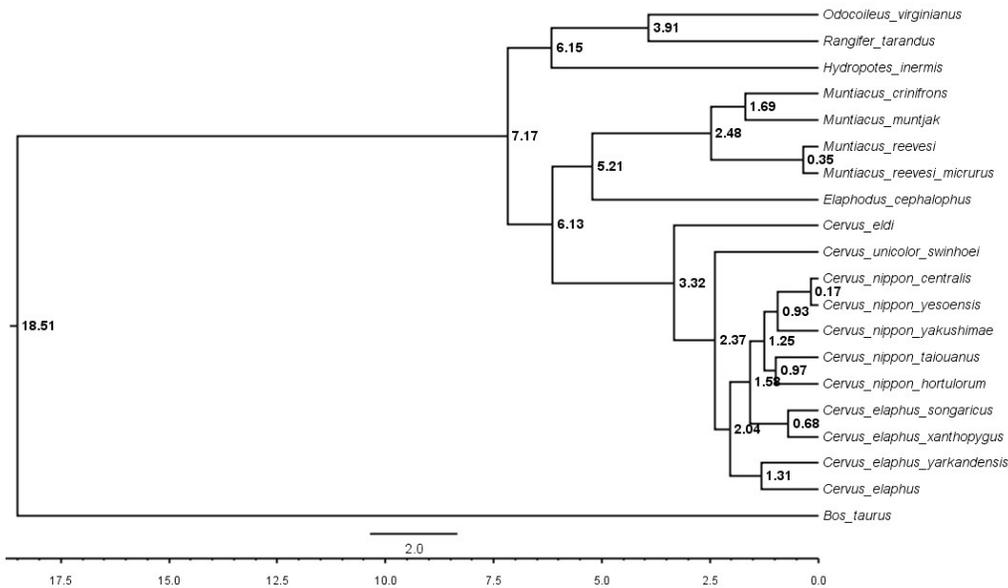
Figure 2. Phylogenetic tree of family Cervidae and outgroup *Bos taurus* based on the whole mitochondria genomes. The divergence times were estimated by Bayesian method, and they were shown on the labels of nodes.

DISCUSSION

Phylogeny and taxonomy of family Cervidae

In our study, we used whole mitochondrial genomes for phylogenetic analysis, which provided a robust phylogenetic resolution spanning the entire evolutionary history of family Cervidae. The phylogenetic trees show that Cervidae is divided into two main clades, named Plesiometacarpalia and Telemetacarpalia in previous research (Pitra et al., 2004; Gilbert et al., 2006). Plesiometacarpalia is also divided into two main clades corresponding to the subfamilies (or tribes) Cervinae and Muntiacinae in the classification of Grubb (1993). This dichotomy was proposed on the basis of molecular data, but our study is the first using all mitochondrial genome sequences, which makes the results more credible. To subfamily Cervinae, *C. nippon*, *C. elaphus*, and the other species of *Cervus sensu* have been added, and the monophyly of subfamily Cervinae is strongly supported by our data and previous data (Polziehn and Strobeck, 2002; Ludt et al., 2004; McDevitt et al., 2009). Muntiacus and Elaphodus have been grouped in the subfamily Muntiacinae, a position weakly confirmed by molecular data; in our phylogenetic trees, this grouping was well supported by Bayesian analysis. However, just three species, *Hydropotes inermis*, *Rangifer tarandus*, and *Odocoileus virginianus*, were sampled from subfamily Telemetacarpalia. Therefore, the monophyly of subfamily Telemetacarpalia is not yet fully classified, but the phylogenetic positions of species with well-known mitochondrial genomes were well determined.

Estimation of divergence times in family Cervidae

In this study, three calibration points in the fossil record were used to determine the divergence times of family Cervidae. The first calibration point for Cervidae was approximately 20 ± 2 mya, which has been used in several studies. The other calibration points were tribe Muntiacinae (8 ± 1 mya) and tribe Odocoileini (5 ± 1 mya; Dong et al., 2004). Combined with the fossil information, molecular analysis had been used to estimate divergence times, but the dates obtained were inconsistent with the paleobiogeographic data. The estimated dates were younger than those suggested by fossil records. The dates obtained from fossil records show that family Cervidae appeared at 20 mya, whereas the divergence time estimated by phylogenetic analysis suggested that it was 7.7-9.6 mya. Therefore, a large gap existed between the fossil record and molecular estimations, and the fossils used as a calibration point for the node Cervidae were not closely related to extant Cervidae. The Cervidae in the fossil record was an ancient deer, which had singular antlers, whereas present deer have paired antlers. Ancient deer was also a ruminant animal, which had been placed in family Cervidae incorrectly, and a large difference existed between it and true Cervidae. Many previous studies have mentioned the limits of using the Cervidae fossil record for estimation of divergence times of family Cervidae, however, and have suggested younger dates for family Cervidae. For example, Gilbert et al. (2006) have suggested that it appeared at 7-9 mya, which is much younger than the date indicated by fossil results. In our estimation, the Cervidae would have appeared at 6-8 mya, which was slightly earlier than the results of Gilbert et al. (2006), and other dates estimated using whole mitochondrial genomes are similarly younger than those previous reported. The whole mitochondrial genome included protein coding regions and no protein coding regions that evolved at different mutation rates. No protein sequences had a high mutation rate, which resulted in younger estimation dates.

In conclusion, based on our phylogenetic results using whole mitochondrial genomes, family Cervidae is divided into two subfamilies, Plesiometacarpalia and Telemetacarpalia, which are also called subfamilies Cervinae and Capreolinae. The two subfamilies contain approximately five tribes, and four of these, Cervinae, Muntiacinae, Hydropotinae, and Odocoileinae, are well determined by the phylogenetic trees of family Cervidae. Finally, the divergence times of the family Cervidae were determined using whole mitochondrial genome analysis, which agreed with the data of other phylogenetic trees. The evolution of Cervidae is complex, and its complete phylogenetic analysis will require additional mitochondrial genome sequences.

ACKNOWLEDGMENTS

REFERENCES

Amato G, Egan M and Schaller GB (2000). Mitochondrial Variation in Muntjac: Evidence for Discovery, Rediscovery, and Phylogenetic Relationships. In: Antelopes, Deer, and Relatives. Fossil Record, Behavioral Ecology, Systematics, and Conservation (Vrba E and Schaller GB, eds.). Yale University Press, New Haven and London, 285-295.
Bouvrain PG, Geraads D and Jehenne Y (1989). Nouvelles données relatives à la classification des Cervidae (Artiodactyla,

Mammalia). *Zool. Anz.* 223: 82-90.

Chan YC, Roos C, Inoue-Murayama M, Inoue E, et al. (2010). Mitochondrial genome sequences effectively reveal the phylogeny of *Hylobates gibbons*. *PLoS One* 5: e14419.

Dong W, Pan Y and Liu J (2004). The earliest *Muntiacus* (Artiodactyla, Mammalia) from the Late Miocene of Yuanmou, southwestern China. *C. R. Palevol.* 3: 379-386.

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

Drummond AJ and Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7: 214.

Flegontov P, Gray MW, Burger G and Lukes J (2011). Gene fragmentation: a key to mitochondrial genome evolution in Euglenozoa? *Curr. Genet.* 57: 225-232.

Gilbert C, Ropiquet A and Hassanin A (2006). Mitochondrial and nuclear phylogenies of Cervidae (Mammalia, Ruminantia): Systematics, morphology, and biogeography. *Mol. Phylogenet. Evol.* 40: 101-117.

Grande C, Templado J and Zardoya R (2008). Evolution of gastropod mitochondrial genome arrangements. *BMC Evol. Biol.* 8: 61.

Grubb P (1993). Mammal Species of the World: A Taxonomic and Geographic Reference (Wilson DE and Reeder DM, eds.). Smithsonian Institution Press, Washington and London, 384-392.

Hassanin A and Douzery EJ (2003). Molecular and morphological phylogenies of ruminantia and the alternative position of the moschidae. *Syst. Biol.* 52: 206-228.

Hundertmark KJ and Bowyer RT (2004). Genetics, evolution, and phylogeography of moose. *Alces* 40: 103-122.

Hundertmark KJ, Shields GF, Udina IG, Bowyer RT, et al. (2002). Mitochondrial phylogeography of moose (*Alces alces*): late pleistocene divergence and population expansion. *Mol. Phylogenet. Evol.* 22: 375-387.

Kan XZ, Yang JK, Li XF, Chen L, et al. (2010). Phylogeny of major lineages of galliform birds (Aves: Galliformes) based on complete mitochondrial genomes. *Genet. Mol. Res.* 9: 1625-1633.

Krajewski C, Sipiorski JT and Anderson FE (2010). Complete mitochondrial genome sequences and the phylogeny of cranes (Gruiformes: Gruidae). *Auk* 127: 440-452.

Kuwayama R and Ozawa T (2000). Phylogenetic relationships among european red deer, wapiti, and sika deer inferred from mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 15: 115-123.

Larkin MA, Blackshields G, Brown NP, Chenna R, et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.

Li YP, Song W, Shi SL, Liu YQ, et al. (2010). Mitochondrial genome nucleotide substitution pattern between domesticated silkmoth, *Bombyx mori*, and its wild ancestors, Chinese *Bombyx mandarina* and Japanese *Bombyx mandarina*. *Genet. Mol. Biol.* 33: 186-189.

Ludt CJ, Schroeder W, Rottmann O and Kuehn R (2004). Mitochondrial DNA phylogeography of red deer (*Cervus elaphus*). *Mol. Phylogenet. Evol.* 31: 1064-1083.

McDevitt AD, Edwards CJ, O'Toole P, O'Sullivan P, et al. (2009). Genetic structure of, and hybridisation between, red (*Cervus elaphus*) and sika (*Cervus nippon*) deer in Ireland. *Mamm. Biol. - Zeitschrift für Säugetierkunde* 74: 263-273.

Md-Zain BM, Mohamad M, Ernie-Muneerah MA, Ampeng A, et al. (2010). Phylogenetic relationships of Malaysian monkeys, Cercopithecidae, based on mitochondrial cytochrome c sequences. *Genet. Mol. Res.* 9: 1987-1996.

Morin PA, Archer FI, Foote AD, Vilstrup J, et al. (2010). Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res.* 20: 908-916.

Pitra C, Fickel J, Meijaard E and Groves PC (2004). Evolution and phylogeny of old world deer. *Mol. Phylogenet. Evol.* 33: 880-895.

Polziehn RO and Strobeck C (2002). A phylogenetic comparison of red deer and wapiti using mitochondrial DNA. *Mol. Phylogenet. Evol.* 22: 342-356.

Randi E, Mucci N, Pierpaoli M and Douzery E (1998). New phylogenetic perspectives on the Cervidae (Artiodactyla) are provided by the mitochondrial cytochrome b gene. *Proc. R. Soc. Lond. B* 265: 793-801.

Ronquist F and Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.

Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.