



Limited phylogenetic distribution of a long tandem-repeat cluster in the mitochondrial control region in *Bubo* (Aves, Strigidae) and cluster variation in Blakiston's fish owl (*Bubo blakistoni*)

Keita Omote, Chizuko Nishida, Matthew H. Dick, Ryuichi Masuda*

Department of Natural History Science, Graduate School of Science, Hokkaido University, Kita-ku, Sapporo 060-0810, Japan

ARTICLE INFO

Article history:

Received 10 September 2012

Revised 15 November 2012

Accepted 20 November 2012

Available online 2 December 2012

Keywords:

Blakiston's fish owl

Bubo

mtDNA control region

Tandem repeat

Slipped-strand mispairing

Phylogeny

ABSTRACT

To investigate the phylogenetic position of Blakiston's fish owl (*Bubo blakistoni*), we sequenced the mitochondrial (mt) DNA control region and cytochrome *b* (*cyt b*) for nine *Bubo* species. Maximum-likelihood analyses of combined control region and *cyt b* sequences, and *cyt b* sequences alone, showed that species formerly placed in genus *Ketupa* comprise a monophyletic group. Unexpectedly, we discovered a long cluster of 20–25 tandem repeat units 77 or 78 bp long in the third control region domain in four of the nine *Bubo* species for which the control region was sequenced (*B. blakistoni*, *B. flavipes*, and *B. ketupu* in the *Ketupa* clade; *B. lacteus*), leading to overall control region lengths of 3.0–3.8 kbp estimated from agarose gel electrophoresis. The control region in *B. lacteus* is the longest (3.8 kbp) reported to date in vertebrates. Sequencing of eight repeat units at each end of the cluster in 20 *B. blakistoni* individuals detected several types of repeat units 77 or 78 bp long, and six patterns in the order of unit types. The occurrence of a repeat cluster in all three species examined in the *Ketupa* clade suggests their common ancestor also had a cluster, whereas a maximum parsimony tree showed repeat-unit types grouping by species, rather than by paralog groups, suggesting independent origins of the clusters. We reconcile these results with a turnover model, in which the range in cluster-length variation and unit types at the 5' end are hypothetically functionally constrained by the protein-binding function of the control region, but otherwise there is a continual turnover of units in evolutionary time, with new unit types arising through mutations, proliferating by duplication of single and double repeat blocks, and being lost through deletion. Estimated free energies for reconstructed secondary structures of single and especially pairs of repeat units were higher than for homologous single-unit blocks in species lacking a repeat cluster, supporting slipped-strand mispairing as the mechanism of cluster turnover.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

The control region is the major non-coding element and one of the most variable regions in vertebrate mitochondrial (mt) DNA (Wenink et al., 1994). The size of the mtDNA control region varies considerably in vertebrates, resulting mainly from variation in the occurrence and number of tandem repeats. In birds, the mtDNA control region is usually 1–1.5 kilo base-pairs (kbp) long, located between the tRNA^{Glu} and tRNA^{Phe} genes, and comprises three domains that include several conserved regions (Roukonen and Kvist, 2002). Although much of the variation consists of nucleotide substitutions and insertions and deletions (indels), considerable length variation among species can occur due to tandemly repeated units in the third domain, with units ranging in size from

four to hundreds of base pairs, and in number from two to more than 100 copies (Lunt et al., 1998). As in other vertebrate groups, species with a tandem repeat region also have a longer control region overall, with the longest previously reported in birds being 3290 bp in the short-eared owl (*Asio flammeus*) (Xiao et al., 2006).

Genus *Bubo* in Family Strigidae comprises large owl species distributed in many parts of the world. Wink et al. (2009) reconstructed owl phylogeny based on mitochondrial cytochrome *b* (*cyt b*) and nuclear *RAG-1* sequences and found *Nyctea*, *Scotooelia*, and *Ketupa* to be embedded within a *Bubo* clade, which supported previous conclusions that these three genera should be included in *Bubo* (e.g., König and Weick, 2008). The former genus *Ketupa* included four species: Blakiston's fish owl (*B. blakistoni*), buffy fish owl (*B. ketupu*), tawny fish owl (*B. flavipes*), and brown fish owl (*B. zeylonensis*). Wink et al. (2009) did not include *B. blakistoni* and *B. flavipes* in their phylogenetic analysis. *Bubo blakistoni* is an endangered species (BirdLife International, 2008) endemic to northeast Asia; its population on Hokkaido Island, Japan, has

* Corresponding author. Fax: +81 11 706 3588.

E-mail address: masudary@mail.sci.hokudai.ac.jp (R. Masuda).

decreased in size and has low genetic diversity, as revealed by a nuclear microsatellite analysis (Omote et al., 2012).

In the process of sequencing the mtDNA control region for phylogenetic reconstruction, we discovered that four of the *Bubo* target species have an exceptionally long region of tandem repeats, while the other five species and the outgroup (*Strix uralensis*) lack this repeat region. In this paper, we (1) report limited *Bubo* phylogenies based on combined control region and *cyt b* sequences (nine species) and *cyt b* sequences alone (15 species); (2) compare the organization of the mtDNA control region among nine *Bubo* species; (3) examine the phylogenetic distribution of the long tandem repeat regions; (4) examine variation in the tandem repeat cluster in *B. blakistoni*; and (5) discuss the evolutionary dynamics of the repeat cluster in *Bubo* based on a comparison of secondary structures between repeat and homologous non-repeat sequences.

2. Materials and methods

2.1. Samples and DNA extraction

Total DNA was extracted from blood or cultured fibroblasts from 20 *B. blakistoni* individuals captured on Hokkaido Island, Japan, and from cultured fibroblasts or feather roots from another eight *Bubo* species and an outgroup taxon: *B. ketupu* and *B. flavipes* (formerly included in *Ketupa*), Verreaux's eagle-owl (*B. lacteus*), spot-bellied eagle-owl (*B. nipalensis*), barred eagle-owl (*B. sumatranus*), Eurasian eagle-owl (*Bubo bubo*), great horned owl (*B. virginianus*), snowy owl (*B. scandiacus*), and Ural owl (*Strix uralensis*; outgroup) (Table 1). Blood samples were preserved in ethanol or dried on filter paper, and frozen at -20°C until use. Fibroblasts were cultured from skin tissue and frozen in liquid nitrogen until use. Total DNA was extracted from blood and fibroblasts by using

the DNeasy Blood & Tissue Kit (Qiagen), and from feather roots by using the QIAamp DNA Micro Kit (Qiagen).

2.2. DNA amplification and sequencing

Polymerase chain reaction (PCR) primers for amplification of the mtDNA control region and *cyt b* gene were designed de novo, or were modified from primers previously reported by Sorenson et al. (1999) (Table 2). PCRs were performed by using PrimeSTAR GXL DNA Polymerase (Takara); cycling conditions were 35 cycles of 98°C for 10 s, 60°C for 15 s, and 68°C for 1 min; products were stored at 4°C . PCR products were separated, and their molecular sizes were estimated, by electrophoresis on 2% agarose gels; they were purified with the QIAquick PCR Purification Kit (Qiagen) and used as templates for nucleotide sequencing. Sequencing was performed with an ABI 3130 or 3730 DNA automated sequencer. To detect artifacts, the repeat region from each individual was sequenced at least twice from independent PCR products. Sequences we obtained have been deposited in the DDBJ/GenBank/EMBL databases under accession numbers AB741537–AB741546 and AB743785–AB743800.

2.3. Sequence data analyses

Sequences were aligned by using MEGA 5.0 software (Tamura et al., 2011). Gene locations were determined by comparisons with mtDNA control region and *cyt b* sequences from the chicken (*Gallus gallus*; accession no. NC001323; Valverde et al., 1994), southern boobook (*Ninox novaeseelandiae*; NC005932; Harrison et al., 2004), and spotted owl (*Strix occidentalis*; AY833623; Barrowclough et al., 2005). An alignment of control region sequences was referred to the following sequence blocks conserved in avian mtDNA (Roukonen and Kvist, 2002): extended termination-associated se-

Table 1
Samples utilized for DNA extraction in this study.

Species	English name	<i>n</i>	Tissue	Source
<i>B. blakistoni</i>	Blakiston's fish owl	20	Blood, fibroblasts	Wild-caught on Hokkaido Island
<i>B. ketupu</i>	Buffy fish owl	1	Feather	Matsue Vogel Park
<i>B. flavipes</i>	Tawny fish owl	1	Fibroblasts	Ueno Zoological Gardens
<i>B. lacteus</i>	Verreaux's eagle-owl	1	Feather	Matsue Vogel Park
<i>B. nipalensis</i>	Spot-bellied eagle-owl	1	Feather	Ueno Zoological Gardens
<i>B. sumatranus</i>	Barred eagle-owl	1	Feather	Ueno Zoological Gardens
<i>B. bubo</i>	Eurasian eagle-owl	2	Fibroblasts	Tama Zoological Park
<i>B. virginianus</i>	Great horned owl	2	Fibroblasts	Sapporo Maruyama Zoo
<i>B. scandiacus</i>	Snowy owl	2	Fibroblasts	Sapporo Maruyama Zoo
<i>S. uralensis</i>	Ural owl	2	Fibroblasts	Kushiro City Zoo

Table 2
PCR primers used in this study for amplification and/or sequencing.

Primer	Nucleotide sequence (5'–3') ^a	Locus	Citation
Lcontrol-01-kb	AGCCACGAATGCTCGTTGTAC	Control region	Newly designed this study
Hcontrol-02-kb	ATTCTGAAAATTAACCCAAACTC	Control region	Newly designed this study
Hcontrol-03-kb	TGAAGAGTTATGGTTTAGGTACG	Control region	Newly designed this study
Lcontrol-10-kb	CTTCTCACGTGAAATCAGCAACC	Control region	Newly designed this study
Hcontrol-11-bl	AGATGGTCCTGACCGAGGAAC	Control region	Newly designed this study
H01251	TCTTGGCATCTTCAGTGCCATGC	tRNA ^{Phe}	Sorenson et al. (1999)
H01530	GTGGCTGGCACAGGATTTACC	12S rRNA	Sorenson et al. (1999)
H01858	TGGATTATAGGACAGGCTCCTCTAC	12S rRNA	Sorenson et al. (1999)
L14080	AACCCTATGCCTTCTCAAAGCC	ND5	Sorenson et al. (1999)
L14649	GGCTACTTCAACCCCTAATACACCG	ND5	Sorenson et al. (1999)
L15560	GTGACAAAATCCATTCCACCC	Cytochrome <i>b</i>	Sorenson et al. (1999)
H16064	CTTCAITCTTTGGTTTACAAGACC	tRNA ^{Thr}	Sorenson et al. (1999)
L16728	CCAAGGCTCTGCGGCTTGAAAAG	tRNA ^{Glu}	Sorenson et al. (1999)

^a Underlined nucleotides were modified from sequences reported by Sorenson et al. (1999).

quences 1 and 2 (ETAS-1 and ETAS-2); B-box, C-box, D-box, E-box, F-box; and conserved sequence box 1 (CSB-1). The control region comprises three domains defined by variability and nucleotide composition (Saccone et al., 1991); the second domain was defined to begin with the F-box, and the third domain to begin with CSB-1.

Phylogenetic trees for the control region (1282 bp, excluding indels), complete *cyt b* (1140 bp), and combined control region + *cyt b* data sets, including nine ingroup taxa and *Strix uralensis* as the outgroup, were reconstructed by maximum likelihood implemented in TREEFINDER March 2011 (Jobb et al., 2004). A partition homogeneity test performed in PAUP 4.0 (Swofford, 2002) indicated that the two genes did not contain significantly different phylogenetic signals ($P = 0.68$). Since analyses of the three data sets gave identical tree topologies, with the combined analysis having the highest nodal support values, we present only the combined analysis. The optimal substitution model for the combined data set determined with MEGA 5.0 was HKY+G1. The robustness of nodes was assessed from analyses of 1000 bootstrap pseudoreplicates.

To examine more extensively relationships among species formerly placed in *Ketupa*, a maximum-likelihood analysis was conducted based on 1001 bp of *cyt b* that included an additional six *Bubo* species over the combined analysis above, for which sequences were obtained from GenBank: pharaoh eagle-owl (*B. ascalaphus*; EU348976), cape eagle-owl (*B. capensis*; EU348978), Indian eagle-owl (*B. bengalensis*; AJ003954), spotted eagle-owl (*B. africanus*; AJ003952), lesser horned owl (*B. magellanicus*; AJ003971), and brown fish owl (*B. zeylonensis*; EU348975) (Wink and Heidrich, 1999; Wink et al., 2009). The optimal substitution model determined with MEGA 5.0 for the expanded *cyt b* data set was HKY+G.

To gain insight into the evolution of the tandem repeat region in *Bubo*, a maximum parsimony tree of repeat units from *B. blakistoni* and the three other species that had a repeat cluster was constructed using PAUP 4.0. The robustness of nodes was assessed from analyses of 1000 bootstrap pseudoreplicates. In addition, a maximum parsimony network of the repeat units in *B. blakistoni* was constructed with TCS 1.21 software (Clement et al., 2000). Potential secondary structures of repeated and non-repeat sequences were examined, and their free energies estimated, using MFOLD (Zuker, 2003).

3. Results

3.1. Phylogenetic relationships among *Bubo* species

The maximum likelihood tree for the control region (1282 bp) and complete *cyt b* sequences (1140 bp) combined (Fig. 1), and

the maximum likelihood tree for partial *cyt b* sequences (1001 bp) but including additional species (Fig. 2), both show *Bubo* divided into two main clades (Clades 1 and 2). Nodal support for these clades was generally high, 89–100% for Clade 1 and 99% for Clade 2. The tree in Fig. 2 (*cyt b*; more taxa) differs in topology for the comparable nodes in Fig. 1 (combined control region and *cyt b*; fewer taxa) in that the three main branches in Clade 1 form an unresolved trichotomy, and *B. ketupu* is the sister group to *B. blakistoni* rather than to *B. flavipes*. In both trees, species formerly placed in *Ketupa* form a well-supported (99–100%) ‘*Ketupa* clade’, with relationships among the species in this clade (Fig. 2) weakly supported, except for the sister group relationship between *B. flavipes* and *B. zeylonensis* in Fig. 2. Clade 1 consists of Asian species, with the exception of *B. lacteus* in Africa.

3.2. Organization of the *Bubo* mtDNA control region

We obtained complete control region sequences (1478–1729 bp long) from *B. bubo*, *B. virginianus*, *B. scandiacus*, *B. nipalensis*, *B. sumatranus*, and *S. uralensis*. By contrast, we were unable to completely sequence the central part of the third domain in *B. blakistoni*, *B. ketupu*, *B. flavipes*, and *B. lacteus*, due to an extended region of tandem repeats roughly 3.0–3.8 kbp long (Fig. 3), estimated from gel electrophoresis.

The control region in all species examined contained conserved sequence blocks: ETAS-1 and ETAS-2 in the first domain; B-box, C-box, D-box, E-box and F-box in the second domain; and CSB-1 in the third domain (Fig. 3). In the first domain, there were many indels and incomplete repeated sequences, most of which were 22 bp long and similar to part of ETAS-2. The first domain in *B. lacteus* contained a unique insertion of more than 100 bp. The second domain was more highly conserved in sequence than the other domains. The repeat region toward the 3' end of the third domain in *B. blakistoni*, *B. ketupu*, *B. flavipes*, and *B. lacteus* contained an estimated 20–25 copies of units 77 or 78 bp long (Table 3); the nucleotide sequences of these units were 80.6–98.5% similar among the four species. By contrast, the third domain in *B. nipalensis*, *B. sumatranus*, *B. bubo*, *B. virginianus*, *B. scandiacus*, and *S. uralensis* contained no complete tandem repeats, but did contain at the same locus a sequence motif 72–82 bp long (Table 3) that was similar to the repeated sequences in the four *Bubo* species having tandem repeats. Similarity between tandem repeat units and non-repeated homologous loci in species lacking a cluster was 56.7–79.1%. The control region in *S. uralensis* contained three and four copies of incomplete repeat units in two individuals.

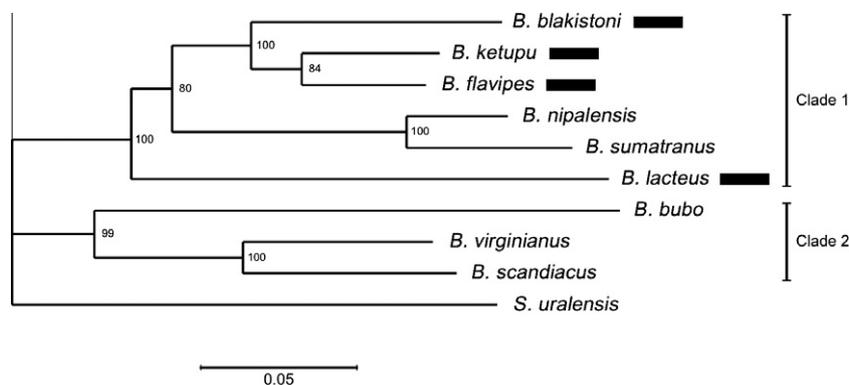


Fig. 1. Maximum-likelihood tree for nine *Bubo* species and the outgroup, *Strix uralensis*, based on a combined analysis of mitochondrial *cyt b* (1140 bp) and control region (1282 bp) sequences. Numbers near nodes indicate bootstrap values derived from 1000 pseudoreplicates. Thick bars indicate species that have a large cluster of tandem repeats in the control region. Accession numbers are as follows (*cyt b* and control region, respectively): *B. blakistoni*, AB741537, AB743785; *B. ketupu*, AB741538, AB743786; *B. flavipes*, AB741539, AB743787; *B. sumatranus*, AB741540, AB743790; *B. nipalensis*, AB741541, AB743789; *B. lacteus*, AB741542, AB743788; *B. bubo*, AB741543, AB743791; *B. virginianus*, AB741544, AB743792; *B. scandiacus*, AB741545, AB743793; *S. uralensis*, AB741546, AB743794. Scale bar at bottom indicates substitutions per site.

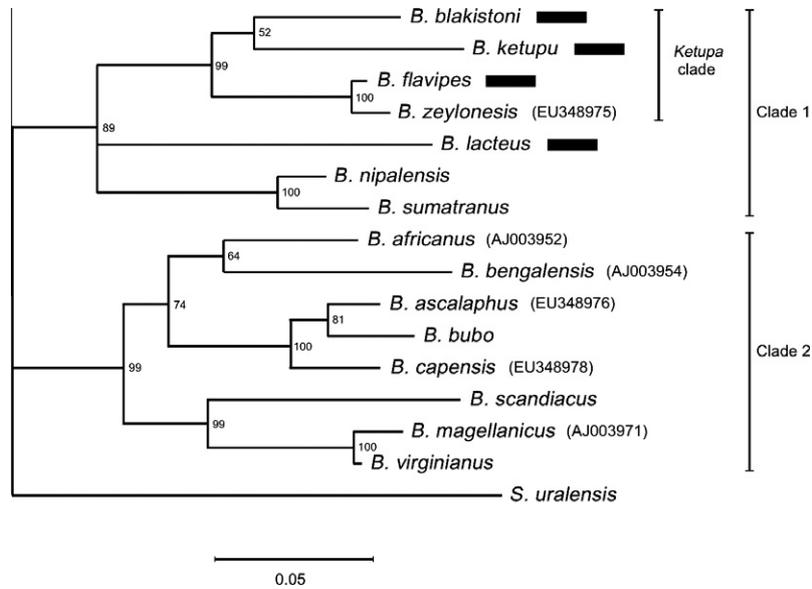


Fig. 2. Maximum-likelihood tree for 15 *Bubo* species and the outgroup, *Strix uralensis*, based on *cyt b* sequences (1001 bp) determined in this study and obtained from GenBank (accession numbers in parentheses). Four species formerly classified in *Ketupa* comprise the *Ketupa* clade within Clade 1. Numbers near nodes indicate bootstrap values derived from 1000 pseudoreplicates. Thick bars indicate species that have a large cluster of tandem repeats in the control region. Scale bar at bottom indicates substitutions per site.

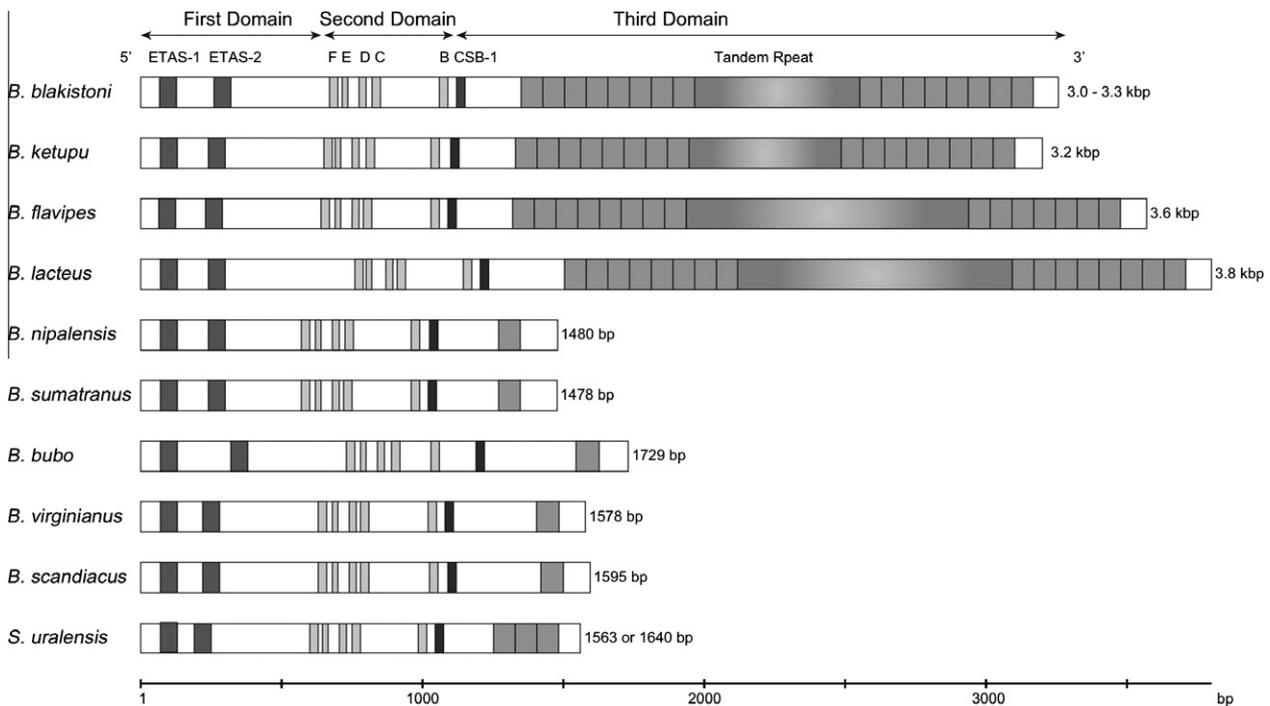


Fig. 3. Molecular organization of the complete mtDNA control region in nine *Bubo* species and *Strix uralensis*. Conserved sequence blocks are shaded and labeled at the top; see the text for details.

Among 20 individuals of *B. blakistoni*, we found four substitution sites in the first domain but no variation in the second domain. The third domain was variable even within single haplotypes, classified according to nucleotide substitutions in the first domain. In addition, there were two size variants among the repeat units: a 77 bp unit of seven types (X_{1-7}) based on five mutation sites, and a 78 bp unit (Y) (Table 4). The average sequence similarity among all pairwise comparisons of repeat unit types was 97.2% (range, 94.8–98.7%). We detected six patterns (Fig. 4) in the order of the unit types, each of which was shared at least by two of the 20 indi-

viduals. Several types of repeat units and complex orders of the types were also found in the third domain of *B. flavipes* and *B. lacteus* (data not shown), but only a single type of repeat unit was observed in *B. ketupu*.

In a maximum-parsimony tree of repeat-unit types (Fig. 5), the units from *B. blakistoni* and *B. lacteus* grouped by species. Units from *B. flavipes* likewise grouped by species, but formed a paraphyletic group that included the single unit type from *B. ketupu*. Unit type X_1 was basal in the *blakistoni* cluster in Fig. 4, suggesting that this type might be the least derived from an ancestral repeat

Table 3
Alignment of nucleotide sequences of representative repeat units and homologous, non-repeated blocks in owl mtDNA control regions.

Species	Sequence (5'-3') ^a
<i>B. blakistoni</i> (X ₁)	CCAGTGAATAATT - TAGAGCAATTGAAA - CGATCAAT - CATCGTATTTCATCATCAAAAATTGA - - - TCGATCAATC - AAA - CAAA
<i>B. ketupu</i>C.....A.....T.....G.....
<i>B. flavipes</i>TG.....G.....
<i>B. lacteus</i>	T.....GG.....T.....C.....A.....T.....TGA.....TC.....G
<i>B. sumatranus</i>C.....AC.....A.....C.....C.....T.....CA.....CA.....C.....T.....GC.....
<i>B. nipalensis</i>T.....AC.....A.....C.....CT.....A.....C.....T.....GT.....G.....CA.....C.....GC.....
<i>B. bubo</i>	TT.TCA..A..CCC..G..C.AC..C.C.C..CC.....C.CC.....C.....CA.CCAAC.A.C..C.....
<i>B. virginianus</i>	.T.TCA..G..CCC.....CCAT..-TC.C.C.C..C.A..CCT.C..C.....CA....-C.C.C..C.....G
<i>B. scandiacus</i>	.T.TCA..G..C.C.....C.AT..-TC.C.T.C..TA...CT.C..C.....G.CA.C--C.C.C..CC.....T...C
<i>S. uralensis</i>	..TGTG.GG..C.....G.....T...A...T...T...C.....A.C--CA..C...ATC.....

^aDashes indicate gaps; dots indicate identity with the *B. blakistoni* X₁ sequence.

Table 4
Alignment of nucleotide sequences of repeat-unit types in *B. blakistoni* mtDNA control regions.

Types	Sequence (5'-3') ^a
X ₁	CCAGTGAATAATTTAGAGCAATTGAAACGATCAATCATCGTATTTCATCATCAAAAATTGATCGATCAATCAAA - CAAA
X ₂T.....
X ₃A.....
X ₄C.....
X ₅C.....
X ₆A.....
X ₇C.....
YTT.....

^aDashes indicate gaps; dots indicate identity with the X₁ sequence.



Fig. 4. Six patterns in the complement and order of repeat-unit types detected among 20 *Bubo blakistoni* individuals. Only eight units at each end of each cluster were sequenced, leaving unit types in the center of the cluster not determined. The unit types are color-coded and aligned by eye, with gaps (-) inserted between units to reflect presumed insertions and deletions of units during cluster evolution. *, one mutational step from X₁; **, two mutational steps from X₁.

sequence, although nodal support was lacking for relationships within this cluster. Similarly, in a maximum parsimony network of repeat-unit types in *B. blakistoni* (Fig. 6), type X₁ was linked to the single unit type found in *B. ketupu*, treated as an outgroup sequence.

3.3. Secondary structures of the repeat sequences

Secondary structures (Fig. 7) reconstructed for single repeat units in the third domain for the 10 owl species consisted of stems and loops, but overall the structure was not well conserved. The units in the four species (*B. blakistoni*, *B. ketupu*, *B. flavipes*, *B. lacteus*) with a tandem repeat cluster formed multiple or extensive stem-and-loop structures (top row, Fig. 7) having comparatively high stability, indicated by high values (-4.6 to -8.9 kcal M⁻¹) for the free energy of DNA folding (Table 5). By contrast, homologous units in the species lacking a tandem repeat cluster formed

one or two generally small stem-and-loop regions (bottom two rows, Fig. 7), with low estimated free energy (-0.3 to -2.2 kcal M⁻¹) (Table 5) and hence secondary structures of low stability. Secondary structures were also reconstructed for all possible two-unit combinations of the unit types in the tandem repeat clusters. The two-unit structures were more complex (Fig. 8) and had higher free-energy values (Table 5) than the single-unit structures (Fig. 7). Some combinations of the repeat units listed in Table 4 and evident in Fig. 4 (X₁-Y, Y-X₄, and X₁-X₄) showed the highest stability (-18.3 kcal M⁻¹).

4. Discussion

4.1. Phylogenetic distribution of the tandem repeat region in *Bubo*

Our phylogenies in Figs. 1 and 2 based on mtDNA control region and *cyt b* sequences showed *Bubo* subdivided into two main clades

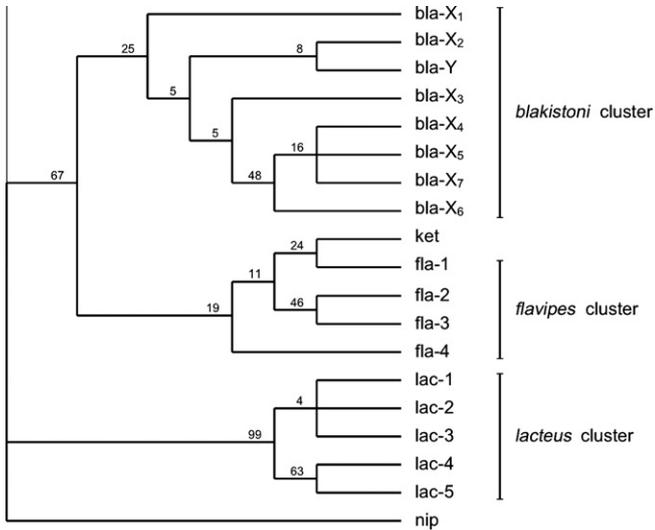


Fig. 5. Maximum-parsimony tree for repeat-unit types (77–79 bp) detected in *B. blakistoni* (bla-), *B. ketupu* (ket), *B. flavipes* (fla-), and *B. lacteus* (lac-).

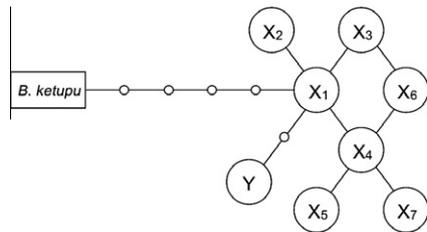


Fig. 6. Maximum-parsimony network for repeat-unit types (X₁–X₇, and Y) found in *B. blakistoni* (circles) and *B. ketupu* (square). Smaller circles indicate intermediate types not detected in this study; each line connecting two circles indicates one mutational step.

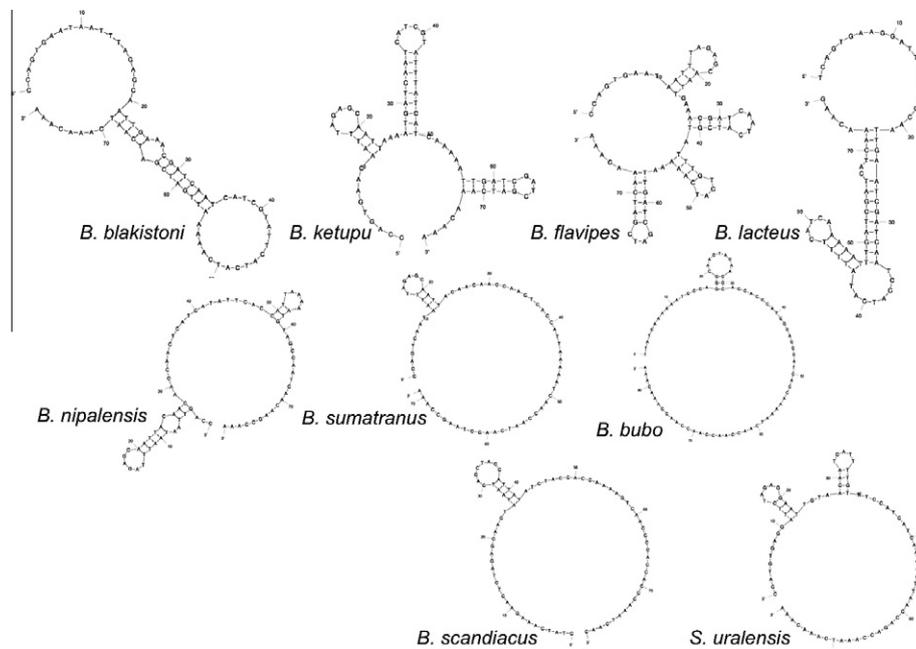


Fig. 7. Single-unit secondary structures deduced with MFOLD for repeated (top row) and homologous, non-repeated sequences (bottom two rows) in the mtDNA control region of eight *Bubo* species and *Strix uralensis*. The type of repeat unit shown for *B. blakistoni* is X₁.

(designated Clades 1 and 2), with species formerly in *Ketupa* included in Clade 1. Our analyses, however, failed to adequately resolve sister-group relationships among members of the well-supported *Ketupa* clade (99% nodal support in the *cyt b* tree, Fig. 2), with low support values at relevant nodes. This suggests that these three Asian species diverged rapidly or relatively recently, and additional markers will be necessary to resolve the relationships among them.

Among the *Bubo* species for which we examined the control region (Fig. 3), we detected a long stretch of tandem repeats only in four species in Clade 1 (Figs. 1 and 2); *B. nipalensis* and *B. sumatranus* in this clade lacked tandem repeats. Three of the species having a tandem repeat region were in the *Ketupa* clade, whereas the fourth, *B. lacteus*, fell outside the *Ketupa* clade in both trees. The control region has not been sequenced in *B. zeylonensis*, but Fig. 2 suggests that this species should also have a long tandem repeat region.

4.2. Characteristics of tandem repeats in the control region in owls

The complete control region was 1.5–1.7 kbp long in five of the *Bubo* species we studied (Fig. 3), similar to or slightly longer than the average in birds, but more than 3 kbp long in *B. blakistoni*, *B. ketupu*, *B. flavipes*, and *B. lacteus* due to the long tandem repeat region; that in *B. lacteus* (about 3.8 kbp) appears to be the longest reported to date in vertebrates, due both to the tandem repeats and a long independent insertion in the first domain. Some previous information on the mtDNA control region existed for owls. Harrison et al. (2004) sequenced the mitochondrial genome of the Tasmanian spotted owl (*Ninox novaeseelandiae*), but did not include the control region. Barrowclough et al. (2005) did not mention tandem repeats in the control region of the California spotted owl (*Strix occidentalis*). Xiao et al. (2006) detected tandem repeats in the third domain of the control region in four species: short-eared owl (*Asio flammeus*), long-eared owl (*Asio otus*), little owl (*Athene noctua*), and tawny owl (*Strix aluco*).

Table 5Characteristics of repeat units, clusters, and homologous loci among nine species of *Bubo* and *Strix uralensis*.

Species	Molecular size of unit (bp)	Number of copies ^a	Free energy (kcal M ⁻¹)	
			One unit	Two units
<i>B. blakistoni</i>	77, 78	20–24	–4.6 to –8.4	–14.9 to –18.3
<i>B. ketupu</i>	77	20	–5.9	–16.5
<i>B. flavipes</i>	77	25	–8.0 to –8.9	–16.3 to –23.4
<i>B. lacteus</i>	77	25	–6.9 to –8.8	–14.5 to –19.1
<i>B. nipalensis</i>	79	1	–2.2	–
<i>B. sumatranus</i>	71	1	–0.3	–
<i>B. bubo</i>	83	1	–1.8	–
<i>B. virginianus</i>	78	1	No pairs	–
<i>B. scandiacus</i>	80	1	–0.4	–
<i>S. uralensis</i>	78	3, 4	–3.6	–

^a The number of copies in the long tandem-repeat clusters was estimated from electrophoresis of PCR products on agarose gels.

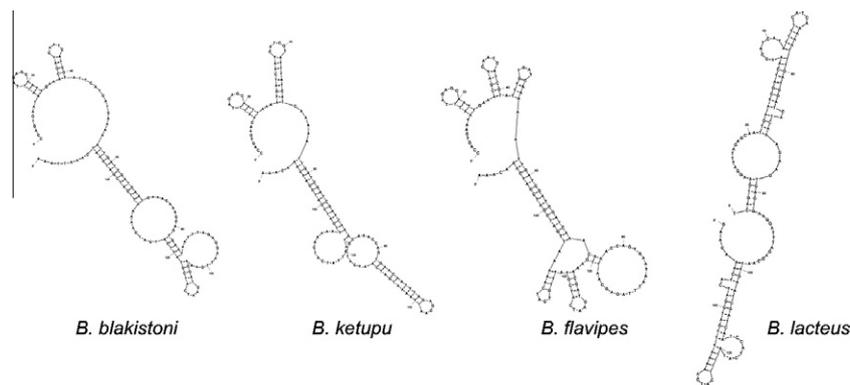


Fig. 8. Two-unit secondary structures deduced with MFOLD for repeated sequences in the mtDNA control region of the four *Bubo* species having a long tandem repeat array. The combination of units shown for *B. blakistoni* is X₁–Y.

Our study is the first to adequately examine the distribution of tandem repeats in any owl genus. The repeat regions in *B. blakistoni*, *B. ketupu*, *B. flavipes*, and *B. lacteus* all consist of more than 20 units, each unit 77 or 78 bp long, with high similarity (80.6–98.5%) between units among the four species. The control region of each of the other five species of *Bubo* includes a single (non-repeated) sequence block 56.7–79.1% similar to the repeat units at the same locus (Fig. 3; Table 3), indicating that the repeat and non-repeat sequences at this locus in the nine *Bubo* species are homologous. In the outgroup, *Strix uralensis*, we detected a pair of non-identical repeat units, with four different unit types between the two individuals examined. The repeat units and their non-repeated homologs in *Bubo* are 56.7–83.6% similar to the 78 bp repeated sequences that Xiao et al. (2006) detected in species of *Asio* and *Strix*, and that we detected in *Strix*, suggesting that the repeat sequences have a common evolutionary origin among species in Strigidae.

In *Bubo blakistoni*, the only owl species for which the tandem repeat region has been at least partly sequenced from many individuals, the number of repeat units and the order of unit types vary among individuals, even within a single haplotype as defined by substitutions in the first domain; *Strix uralensis* also showed individual variation, on the basis of the two individuals we examined.

4.3. Evolution of the large tandem repeat regions in *Bubo*

Mechanisms proposed for the formation of tandem repeats include recombination and transposition (Hasson et al., 1984), unequal crossing over or gene conversion (Hoelzel, 1993), and slipped-strand mispairing (Levinson and Gutman, 1987; Wilkinson and Chapman, 1991). Because avian mtDNA does not extensively recombine (Berlin et al., 2004), slipped-strand mispairing is the

most likely mechanism (Faber and Stepien, 1998). In this mechanism, replication of repeat units occurs when these units misalign upon denaturation of the DNA duplex and one or more units loop out, followed by repair (duplication) of the looped-out region. This mechanism actually allows either the gain or loss of repeat units, and loop formation is facilitated when single strands of the repeat units can form stable secondary structures (Buroker et al., 1990). Our reconstructions of secondary structures with higher stability than those of the non-repeat sequences (Fig. 7; Table 5) are consistent with slipped-strand mispairing as the mechanism responsible for the long tandem repeat regions in *Bubo*.

In a maximum parsimony analysis of all different repeat unit types detected in four species (Fig. 5), the unit types grouped by species (except for *B. ketupu*, with only one unit type), rather than forming paralog groups with units from other species. One interpretation of this pattern is that duplications of an ancestral repeat unit and the expansion of a cluster occurred independently in each species. Arguing against this interpretation is the presence of a repeat cluster in three species (*B. blakistoni*, *B. flavipes*, and *B. ketupu*) comprising a clade within the well-supported *Ketupa* clade (Figs. 7 and 8), the most parsimonious explanation for which is that the common ancestor also had a repeat cluster. An explanation reconciling these apparently contradictory results is that an ancestral cluster was present, but that the repeat units within each species have undergone concerted evolution through gene conversion (Chen et al., 2007), rendering paralogs more similar to one another within a species than between species.

4.4. Dynamics of repeat-cluster evolution in *Bubo blakistoni*

Evidence is accumulating that repeat clusters in the vertebrate mtDNA control region are evolutionarily dynamic, with observed

patterns produced by several distinct processes. The clusters can vary in length (number of repeats) among closely related species, among individuals within species, and even as heteroplasmy within individuals (Lunt et al., 1998; Kvist et al., 2003). The only feature common to all six patterns in the tandem repeat cluster among 20 individuals of *B. blakestoni* (Fig. 4) is the occurrence of two or three X_1 units at the 5' end. We speculate that three X_1 or X_1 -like units at the 5' end are necessary for DNA binding to the control region and are conserved by purifying selection. Fig. 4 further suggests that the six tandem repeat patterns observed in *B. blakestoni* were produced by independent point mutations, and deletions or duplications of one or two repeat units, in different mitochondrial lineages. More than half of the unit types (X_2 , X_3 , X_5 , X_6 , and X_7) differing from X_1 by one or two point mutations show no evidence of having duplicated at all. Unit types X_3 , X_5 , and X_6 are each common to four or five of the patterns, indicating the retention within mitochondrial lineages of repeat units having particular point mutations. On the other hand, pairs or odd-numbered series of adjacent X_4 units in patterns 2 and 5, and Y–Y pairs or triplets adjacent to an X_4 unit in patterns 1, 2, 3, and 5, indicate single duplications of X_4 and Y, respectively. Similarly, the Y– X_4 motif interspersed two or three times in each of patterns 1, 2, 3, and 5, suggest repeated duplications of this doublet occurred. Duplications of doublets may be more frequent than single duplications, as two-unit secondary structures (Fig. 8) are more stable (Table 5) than those of single units (Fig. 7). It is impossible to reconstruct the actual steps leading to variant cluster patterns among individuals, because a particular pattern can result from multiple evolutionary pathways.

We consider repeat-unit type X_1 at the 5' end of the cluster in *B. blakestoni* to be most similar in sequence to the ancestral unit type in the common ancestor, as it is basal in the *blakestoni* cluster in Fig. 5, and is connected to the outgroup (the single type in *B. ketupu*) in the haplotype network in Fig. 6. We speculate that two or three X_1 units at the 5' end of the cluster in *B. blakestoni* were conserved in number and sequence, perhaps as a DNA-binding site; no similar constraint is evident at the 3' end, where X_4 , X_7 (both 77 bp), or Y (78 bp) occur. The overall length of the repeat sequence cluster also appears to have been constrained within a range of variation of 1–4 units, possibly likewise for functional reasons related to DNA binding (Wilkinson and Chapman, 1991), or due to a balance between length mutation and selection against extreme lengths (Wilkinson et al. 1997). We speculate that between the 5' X_1 triad and the 3' end, repeat units freely accumulated point mutations and underwent duplications and deletions of one or two units, with no constraints on the order or complement of units, as long as cluster length did not fall outside the optimal range; presumably, mitochondria whose repeat cluster fell outside this range were selected against at the mitochondrial, cellular, or individual levels. The rate of duplication must have been equal to or greater than the rate of deletion, or the repeat cluster would have been lost (Buroker et al., 1990).

We hypothesize that, in evolutionary time, there was a turnover of repeats within clusters, with new repeat types arising in mitochondrial lineages through point mutations and increasing in frequency through duplications, and with the possibility of some types being lost through deletions and genetic drift (Rand and Harrison, 1989). This sort of turnover has previously been proposed to account for high levels of variation in tandem repeat arrays within and among individuals and between species (e.g., Hoelzel et al., 1993). The turnover model explains how all species in a clade can have inherited a repeat cluster from a common ancestor, and yet have all the repeat units within each species more similar to one another than to any units in related species.

Our initial aim in sequencing the mtDNA control region for multiple individuals of *B. blakestoni* was to detect markers useful for

analyses of geographical population structure. The discovery of tandem repeat arrays in *Bubo* was an unexpected but fortuitous result. Several levels of variation in the array in *B. blakestoni* will provide abundant markers for population analyses (e.g., Lunt et al., 1998; Hernández et al., 2010), and cluster variation among and within species in the *Ketupa* clade in *Bubo* will provide a useful model system for studying how repeat clusters evolve in mitochondrial genomes.

Acknowledgments

We thank Dr. Takeshi Takenaka, Ueno Zoological Gardens (Eri Okumura), Matsue Vogel Park (Hideki Hashimoto), Tama Zoological Park, Sapporo Maruyama Zoo, and Kushiro City Zoo for providing owl specimens. We obtained specimens of *B. blakestoni* through the Blakiston's Fish Owl Conservation Action Program of the Ministry of the Environment, Japan. This study was supported in part by a Grant-in-Aid for Scientific Research (No. 23310164) from the Japan Society for the Promotion of Science; a grant (No. D-1201) from the Environment Research and Technology Development Fund of the Ministry of the Environment, Japan; and a grant from the Pro Natura Fund 2011.

References

- Barrowclough, G.F., Groth, J.G., Mertz, L.A., Gutierrez, R.J., 2005. Genetic structure, introgression, and a narrow hybrid zone between northern and California spotted owls (*Strix occidentalis*). *Mol. Ecol.* 14, 1109–1120.
- Berlin, S., Smith, N.G.C., Ellegren, H., 2004. Do avian mitochondria recombine? *J. Mol. Evol.* 28, 163–167.
- BirdLife International, 2008. *Ketupa blakestoni*. In: IUCN Red List of Threatened Species, Version 2011.2. <www.iucnredlist.org> (accessed on 15.06.12).
- Buroker, N.E., Brown, J.R., Gilbert, T.A., O'Hara, P.J., Beckenbach, A.T., Thomas, W.K., Smith, M.J., 1990. Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. *Genetics* 124, 157–163.
- Chen, J.-M., Cooper, D.N., Chuzhanova, N., Férec, C., Patrinos, G.P., 2007. Gene conversion: mechanisms, evolution, and human disease. *Nat. Rev. Genet.* 8, 762–775.
- Clement, M., Posada, D., Crandall, K., 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657–1660.
- Faber, J.E., Stepien, C.A., 1998. Tandemly repeated sequences in the mitochondrial DNA control region and phylogeography of the pike-perches *Stizostedion*. *Mol. Phylogenet. Evol.* 10, 310–322.
- Harrison, G.L., McLenachan, P.A., Phillips, M.J., Slack, K.E., Cooper, A., Penny, D., 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the late Cretaceous. *Mol. Biol. Evol.* 21, 974–983.
- Hasson, J.-F., Mougkiau, E., Cuzin, F., Yaniv, M., 1984. Simian virus 40 illegitimate recombination occurs near short direct repeats. *J. Mol. Biol.* 177, 53–68.
- Hernández, M.A., Campos, F., Padilla, D.P., 2010. Tandem repeats in the mtDNA control region of the southern grey shrike endemic to the Canary Islands. *Ardeola* 57, 423–427.
- Hoelzel, A.R., 1993. Evolution by DNA turnover in the control region of vertebrate mitochondrial DNA. *Curr. Opin. Genet. Dev.* 3, 891–895.
- Hoelzel, A.R., Hancock, J.M., Dover, G.A., 1993. Generation of VNTRs and heteroplasmy by sequence turnover in the mitochondrial control region of two elephant seal species. *J. Mol. Evol.* 37, 190–197.
- Jobb, G., von Haeseler, A., Strimmer, K., 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol. Biol.* 4, 18.
- König, C., Weick, W., 2008. *Owls. A Guide to the Owls of the World*, second ed. Christopher Helm, London.
- Kvist, L., Martens, J., Nazarenko, A.A., Orell, M., 2003. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Mol. Biol. Evol.* 20, 243–247.
- Levinson, G., Gutman, G.A., 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4, 203–221.
- Lunt, D.H., Whipple, L.E., Hyman, B.C., 1998. Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol. Ecol.* 7, 1441–1455.
- Omote, K., Nishida, C., Takenaka, T., Masuda, R., 2012. Temporal changes of genetic population structure and diversity in the endangered Blakiston's fish owl (*Bubo blakestoni*) on Hokkaido Island, Japan, revealed by microsatellite analysis. *Zool. Sci.* 29, 299–304.
- Rand, D.M., Harrison, R.G., 1989. Molecular population genetics of mtDNA size variation in crickets. *Genetics* 121, 551–569.
- Roukoniemi, M., Kvist, L., 2002. Structure and evolution of the avian mitochondrial genome. *Mol. Phylogenet. Evol.* 23, 422–432.
- Saccone, C., Pesole, G., Sbis, E., 1991. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *J. Mol. Evol.* 33, 83–91.

- Sorenson, M.D., Ast, J.C., Dimcheff, D.E., Tamaki, Y., Mindell, D.P., 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol. Phylogenet. Evol.* 12, 105–114.
- Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods) Version 4. Sinauer Associates, Sunderland.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739.
- Valverde, J.R., Marco, R., Garesse, R., 1994. A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. *Proc. Natl. Acad. Sci. USA* 91, 5368–5371.
- Wenink, P.W., Baker, A.J., Tilanus, M.G.J., 1994. Mitochondrial control-region sequences in two shorebird species, the turnstone and the dunlin, and their utility in population genetic studies. *Mol. Biol. Evol.* 11, 22–31.
- Wilkinson, G.S., Chapman, A.M., 1991. Length and sequence variation in evening bat D-loop mtDNA. *Genetics* 128, 607–617.
- Wilkinson, G.S., Mayer, F., Kerth, G., Petri, B., 1997. Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. *Genetics* 146, 1035–1048.
- Wink, M., El-Sayed, A.-A., Sauer-Gürth, H., Gonzalez, J., 2009. Molecular phylogeny of owls (Strigiformes) inferred from DNA sequences of the mitochondrial *cyt b* and the nuclear *RAG-1* gene. *Ardea* 97, 581–591.
- Wink, M., Heidrich, P., 1999. Molecular evolution and systematics of owls (Strigiformes). In: König, C., Weick, F., Becking, J.-H. (Eds.), *Owls: A Guide to Owls of the World*. Yale University Press, New Haven, pp. 39–57.
- Xiao, B., Ma, F., Sun, Y., Li, Q.-W., 2006. Comparative analysis of complete mitochondrial DNA control region of four species of Strigiformes. *Acta Genet. Sinica* 33, 965–974.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415.