J Mol Evol (1998) 47:449–462

JOURNAL OF MOLECULAR EVOLUTION

Organization and Evolution of the Mitochondrial DNA Control Region in the Avian Genus *Alectoris*

Ettore Randi, Vittorio Lucchini

Istituto Nazionale per la Fauna Selvatica, via Cà Fornacetta 9, 40064 Ozzano dell'Emilia (BO), Italy

Received: 18 August 1997 / Accepted: 20 February 1998

Abstract. The entire mitochondrial DNA control region (mtDNA D-loop) was sequenced in the seven extant species of *Alectoris* partridges. The D-loop length is very conserved (1155 ± 2 nucleotides), and substitution rates are lower than for the mitochondrial cytochrome *b* gene of the same species, on average. Comparative analyses suggest that these D-loops can be divided into three domains, corresponding to the highly variable peripheral domains I and III and to the central conserved domain II of vertebrates (Baker and Marshall 1997). Nevertheless, the first 161 nucleotides of domain I of the *Alectoris*, immediately flanking the $tRNA^{Glu}$, evolve at an unusually low rate and show motifs similar to the mammalian extended termination-associated sequences [ETAS1 and ETAS2 (Sbisà et al. 1997)], which can form stable secondary structures. The second part of domain I contains a hypervariable region with two divergent copies of a tandemly repeated sequence described previously in other species of anseriforms and galliforms (Quinn and Wilson 1993; Fumihito et al. 1995). Some of the conserved sequence blocks of mammals can be mapped in the central domain of *Alectoris.* Domain III is highly variable and has sequences similar to mammalian CSB1. The bidirectional transcription promoter HSP/LSP box of the chicken is partially conserved among the *Alectoris.* This structural organization can be found in the anseriform and galliform species studied so far, suggesting that strong functional constraints might have controlled the evolution of the D-loop since the origin of Galloanserae. Their conserved organization and slow molecular evolution make D-loops of galliforms appropriate for phylogenetic studies, although homoplasy can be be generated at a few hypervariable sites and at some sites which probably have mutated by strand slippage during DNA replication. Phylogenetic analyses of D-loops of *Alectoris* are concordant with previously published cytochrome *b* and allozyme phylogenies (Randi 1996). *Alectoris* is monophyletic and includes three major clades: (1) basal *barbara* and *melanocephala;* (2) intermediate *rufa* and *graeca;* and (3) recent *philbyi*, *magna*, and *chukar*. Comparative description of the organization and substitution patterns of the mitochondrial control region can aid in mapping hypervariable sites and avoid some sources of homoplasy in data sets which are to be used in phylogenetic analyses.

Key words: Mitochondrial DNA — Control region — *Alectoris* — Galliformes — Molecular phylogeny — Avian extended termination-associated sequences — Homoplasy

Introduction

The mitochondrial genome (mtDNA) of vertebrates is a small, circular molecule of 16,000–21,000 nucleotides (nt), with a compact and conserved organization in most, but not all, of the studied species (Wolstenholme 1992; Janke et al. 1994; Quinn 1997). It contains the genes for 2 ribosomal RNAs (rRNAs), 22 transfer RNAs (tRNAs), 13 enzymes, and the control region (D-loop), which regulates replication of the H strand and transcription of

Correspondence to: E. Randi; *e-mail:* MET0217@IPERBOLE.BOLOGNA.IT

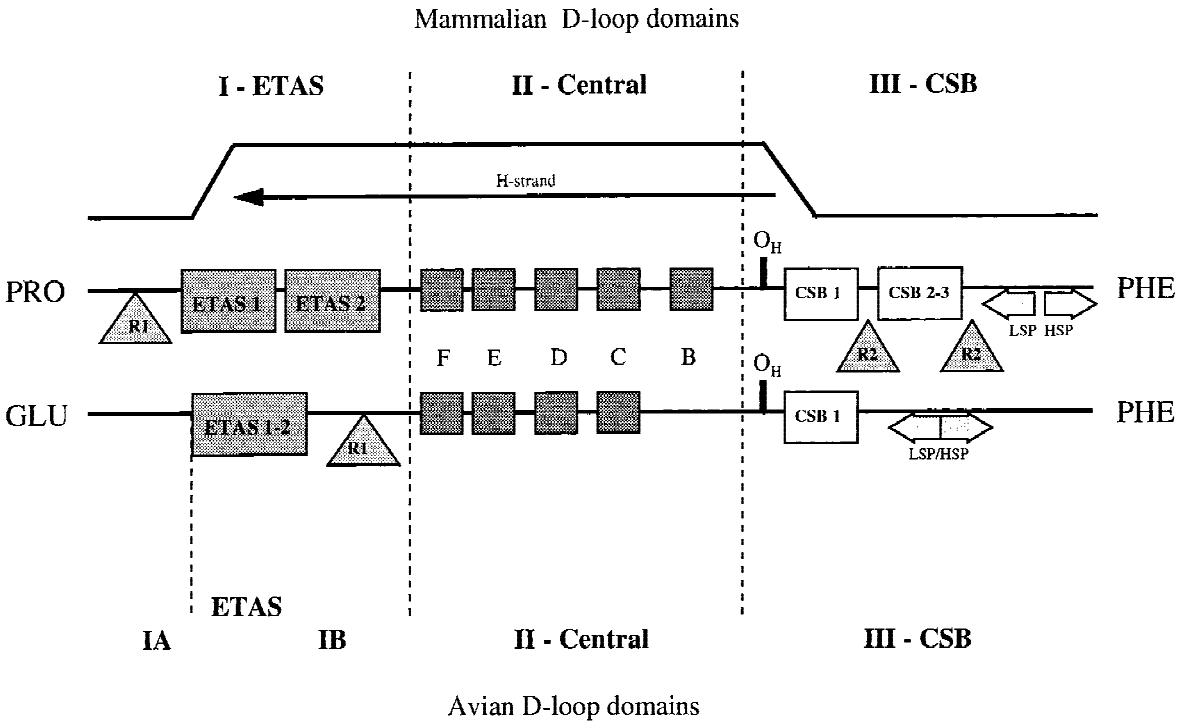

Fig. 1. Schematic comparison of mtDNA D-loop organization in mammals and Galliformes, with particular reference to *Alectoris.* ETAS—extended termination-associated sequence domain and mammalian consensus sequences (ETAS1 and ETAS2) (Sbisà et al. 1997); R1 and R2—tandem repeats in the ETAS and CSB peripheral domains; B, C, D, E, F—conserved sequence blocks of the central domain; O_H—origin of H-strand replication; CSB1, -2, and -3—conserved sequence blocks of the CSB peripheral domain; LSP—light-strand transcription promoter; HSP—heavy-strand transcription promoter.

all mtDNA genes (Clayton 1992). A novel gene order of avian mtDNAs was described in chickens (Desjardins and Morais 1990) and other species of birds (Desjardins et al. 1990; Desjardins and Morais 1991; Quinn and Wilson 1993; Marshall and Baker 1997). Quinn and Wilson (1993) proposed a mechanism which could explain the generation of the avian mtDNA gene order by H-strand displacement during replication, gene duplication, and subsequent partial deletion. In consequence, the avian D-loop is flanked by the genes for $tRNA^{Glu}$ and $tRNA^{Phe}$, and not by $tRNA^{Pro}$ and $tRNA^{Phe}$ as is usual for vertebrates (Fig. 1).

D-Loops are usually considered to be the most variable parts of mtDNAs, in terms of nucleotide substitutions, short insertions/deletions (indels), and dynamics of variable-number tandem repeats (VNTRs). However, site mutability and structural rearrangements are not distributed randomly across the entire region, but affect particular hypervariable sites and domains (Lyrholm et al. 1996; Yang 1996a). Therefore, mammalian D-loops have been partitioned into three domains (Fig. 1): (1) the $tRNA^{Glu}$-adjacent domain I, which contains sequences associated with termination of H-strand replication (TAS) (Doda et al. 1981) and often includes VNTRs (R1 repeats) (Fumagalli et al. 1996), responsible mainly for interspecific length variation and intraspecific heteroplasmy (Hoelzel et al. 1994; Zardoya et al. 1995; Fumagalli et al. 1996); (2) the central conserved domain II (Saccone et al. 1987); and (3) the $tRNA^{Phe}$-adjacent domain III, which is highly variable due to nucleotide substitutions, indels, and VNTRs (R2 repeats) (Fumagalli et al. 1996; Hoelzel et al. 1994; Dufresne et al. 1996) and contains conserved sequences of functional importance—the light- and heavy-strand promoters of transcription (LSP and HSP) (Clayton 1992), the origin of the H strand (O_H), and short conserved sequence blocks (CSB1, CSB2, and CSB3) (Walberg and Clayton 1981), which have critical functions in priming mtDNA replication and controlling RNA/DNA transition (Chang and Clayton 1987; MacKay et al. 1993; Sbisà et al. 1997). These sequences are involved in potential secondary structures (Brown et al. 1986).

Sbisà et al. (1997) recently described the comparative structural organization of 26 complete D-loops from 10 mammalian orders. They were able to identify two conserved extended termination-associated sequences (ETAS1 and ETAS2, which included the shorter TAS sequences) in domain I, discussed presumptive roles of CSB sequences in domain III, and suggested adopting the following nomenclature for the three D-loop domains: the ETAS domain I (adjacent to $tRNA^{Pro}$ in mammals and $tRNA^{Glu}$ in birds), the central conserved domain II, and the CSB domain III (adjacent to $tRNA^{Phe}$).

The organization and evolutionary dynamics of avian D-loops have recently been reviewed by Baker and Marshall (1997). Peripheral domains have hypervariable

blocks (Quinn 1992; Edwards 1993; Wenink et al. 1996) and contain tandem repeats in *Gallus* (Desjardins and Morais 1991; Fumihito et al. 1995), the lesser snow goose (Quinn and Wilson 1993), Ciconiiformes (Wenink et al. 1994; Berg et al. 1995), and passerines (Mundy et al. 1997). Some of these repeats include sequences similar to CSB1 (Desjardins and Morais 1991; Ramirez et al. 1993; Quinn and Wilson 1993). Sequences corresponding to TASs, CSB1, and LSP/HSP were identified in chicken and quail (L'Abbé et al. 1991; Desjardins and Morais 1991; Ramirez et al. 1993). The basic organization of avian and mammalian D-loop seems roughly similar, although only a very limited number of bird species have been studied so far. Rates and patterns of molecular evolution of avian D-loops are not known, with the exception of a few studies at the population level (Quinn 1992; Edwards 1993; Wenink et al. 1996). The phylogenetic content of avian D-loops has been explored only for some phasianid and passerine species (Fumihito et al. 1995; Marshall and Baker 1997).

In this paper we present a study of the organization and evolution of the mtDNA D-loop in partridges of the genus *Alectoris.* The seven living species in the genus *Alectoris* (Aves, Galliformes, Phasianidae) are distributed in the Palearctic, China, the Himalayas, and southern Arabia (see Randi 1996, Fig. 1). Their phylogenetic relationships have been investigated previously using allozyme electrophoresis and nucleotide sequencing of the mitochondrial cytochrome *b* (cyt *b*) gene (Randi et al. 1992; Randi 1996). Here we aim to use D-loop nucleotide sequences to address the following questions: (1) Can comparative analyses of closely related species suggest information on the structural organization and functional importance of avian D-loops? (2) Can the reconstructed phylogenetic relationships among *Alectoris* help in elucidating aspects of the molecular evolution of avian D-loops? and (3) Can D-loops, or portions of D-loops, be used for phylogenetic inference, perhaps in association with other mtDNA genes? Our results suggest that the structural organization and patterns of molecular evolution of the D-loop of *Alectoris* (and probably of other species of Anseriformes and Galliformes as well) are similar to the mammalian homologous region. In particular, our data allow us to define, for the first time in birds, two regions with sequence similarity to the two mammalian ETAS, which were recently described by Sbisà et al. (1997).

Materials and Methods

DNA Samples. Total DNA was extracted, using guanidinium thiocyanate (GuSCN) and diatomaceous silica particles (Sigma D5384) (a procedure modified after Gerloff et al. 1995), from 95% ethanol-preserved livers of the following samples (locality; EMBL accession number of the D-loop sequences): (1) two Barbary partridges—*A. barbara*-1 (Morocco, Salè; AJ222726) and *A. barbara*-8 (Italy, Sardinia; AJ222727); (2) three Arabian red-legged partridges—*A. melanocephala*-1, -4, and -B (Saudi Arabia, Taif; AJ222734, AJ222735, and AJ222736); (3) two red-legged partridges—*A. rufa*-11 (Spain; AJ222740) and *A. rufa*-291 (Portugal; AJ222739); (4) two rock partridges—*A. graeca*-2 (Italian Alps; AJ222730) and *A. graeca*-3 (Italy, Sicily; AJ222731); (5) two chukar partridges—*A. chukar*-2 (Israel, Negev; AJ222728) and *A. chukar*-8 (Kirghizistan, Tianshan; AJ222729); (6) two Philby's rock partridges—*A. philbyi*-1 and -2 (Saudi Arabia, Taif; AJAJ222737 and AJ222738); and (7) two Przhevalski rock partridges—*A. magna*-1 and -2 (Central China, Lazhou; AJ222732 and AJ222733).

PCR Amplification and Sequencing. The entire mtDNA D-loops were PCR amplified using the primers PHDL (5′-AGG ACT ACG GCT TGA AAA GC-3′) and PHDH (5′-CAT CTT GGC ATC TTC AGT GCC-3′) (Fumihito et al. 1995), whose 5′ terminal nucleotides bind to positions 16,750 and 1255 of the chicken mitochondrial genome, respectively (Desjardins and Morais 1990). These primers consistently yield a single and very clean PCR product of the expected size. The mitochondrial origin of these amplifications was checked by comparing nucleotide sequences obtained from PCR fragments of total and mitochondrial DNA isolated by an alkaline lysis procedure (Palva and Palva 1985). Amplifications were performed with AmpliTaq DNA polymerase (Perkin Elmer), 1.5 m*M* $MgCl_2$ in the reaction buffer, and using the following thermal cycle in a 9600 Perkin Elmer machine: 94°C for 2 min, 94°C for 15 s, 55°C for 15 s, 72°C for 1 min (30 cycles), and 72°C for 10 min. PCR products were purified in low-melting agarose gels. All sequences were obtained by double-stranded DNA cycle sequencing with the ABI Prism Dye Terminator procedure in an ABI 373 automatic sequencer, using the external primers PHDL and PHDH and the internal primers PH-L400, 5′-ATT TAT TGA TCG TCC ACC TCA CG-3′; PH-L818, 5′-GGA ATG ATC TTG ACA CTG ATG CAC T-3′; and PH-H521, 5′-TTA TGT GCT TGA CCG AGG AAC CAG-3′ (E.A. Scott, personal communication, Manchester Metropolitan University, Manchester, UK). Additional sequences were obtained from published literature: chicken (Desjardins and Morais 1990), Japanese quail (Desjardins and Morais 1991), lesser snow goose (Quinn and Wilson 1993), Pecking duck (Ramirez et al. 1993), 2 species of *Gallus* (Fumihito et al. 1994), and 14 species of phasianids (Fumihito et al. 1995).

Sequence Analysis. The D-loop sequences were aligned with CLUSTAL W (Thompson et al. 1994). The programs HAIRPIN and RNAFOLD (in PG/GENE; IntelliGenetics) and LOOPDLOOP (Gilbert 1992) were used to identify and draw potential secondary structures in the single-stranded D-loops. Nucleotide composition and genetic distances were computed using MEGA 1.01 (Kumar et al. 1993), with pairwise exclusion of gaps. A maximum-likelihood (ML) test for heterogeneity of substitution rates among domains was performed using 2LRHET (Gaut and Weir 1994). Compositional stationarity and applicability of substitution models were tested with MODELS (Rzhetsky and Nei 1995). Phylogenetic reconstructions were obtained by the ML method, the best suited for modeling the extremely heterogeneous substitution pattern of the D-loops, as implemented in the packages PUZZLE 3.1 (Strimmer and von Haeseler 1996), MOPLHY 2.3 (Adachi and Hasegawa 1996), and PHYLIP 3.5 (programs DNAML and DNAMLK, Felsenstein 1993). Additional phylogenetic trees were computed by neighbor-joining (NJ) (Saitou and Nei 1987) clustering of the estimated percentage sequence divergence, using MEGA. Most-parsimonious (MP) trees were searched using PAUP 3.1.1 (Swofford 1993) with unordered character states, 10 replicates of random addition of terminal sequences, and the branch-and-bound option.

Robustness of the phylogenies was assessed by two approaches: (1) the bootstrap (Felsenstein 1985), with 1000 resamplings followed by a NJ (MEGA) or a maximum-parsimony reconstruction (bootstrap option in PAUP), and (2) the reliability percentages (RP), e.g., the number of times the group appears after 1000 ML puzzling steps (PUZZLE) (Strimmer and von Haeseler 1996).

Results

Mitochondrial Control Region Sequences of Alectoris

It is well-known that nuclear sequences with homology to the D-loop exist in some avian and mammalian species (Quinn 1997, and references therein). D-Loop sequences of *A. barbara, rufa, graeca,* and *chukar* (one liver sample from each species) obtained from mitochondrial DNA purifications were identical to the sequences amplified from the corresponding total DNAs. Therefore, although the alkaline lysis procedure does not exclude possible nuclear DNA contaminations, we suppose that these sequences are authentic mtDNA. The alignment of D-loops of *Alectoris* was straightforward: sequence blocks were conserved and did not require extensive addition of indels (Fig. 2). D-Loops were 1153 nt long in *philbyi,* 1154 in *barbara* and *chukar,* 1155 in *melanocephala, graeca* and *rufa,* and 1157 in *magna,* with short intraspecific length variation among individuals of *rufa, graeca,* and *chukar* (nt 192 was deleted in *rufa*-1; nt 994 and 1109 were deleted in *graeca*-3 and -2, respectively; and nt 1023 and 1024 were deleted in *chukar*-2). The two specimens of *magna* differ for one A-to-G transition (nt 212), and the two *philbyi* differ for one A-to-G transition (nt 518). Sequences were identical between two specimens of *barbara* and three *melanocephala.* Short indels were introduced to optimize the alignment exclusively in the two peripheral domains (Fig. 3). Nucleotide substitutions also occurred more frequently in peripheral domains, particularly at positions 153–316 and 978–1157 of the alignment (Fig. 2).

The distribution of variability, by visual inspection and by plotting the number of variable sites along the alignment (Figs. 2 and 3), suggests that the D-loops of *Alectoris* can be descriptively divided into three domains: a variable domain I (nt 1–316), with 22% variable sites (including indels) among *Alectoris;* a central conserved domain II (nt 317–784), with only 5% variable sites among *Alectoris;* and a variable domain III, spanning the origin of replication O_H to the end of the D-loop (nt 785–1160), with 14% variable sites. Average substitution rates for the three domains were 0.12:0.02:0.08, corresponding to relative proportions of 6:1:4, respectively. These rates were significantly different ($p < 0.001$, as estimated using 2LRHET). Substitution rates among sites were also heterogeneous, with the following values of the gamma parameters: $\alpha = 0.16–0.27$, and $\epsilon = 0.86–0.79$ (as estimated using PUZZLE).

The observed differences in substitution rates among D-loop domains of *Alectoris* correspond to differences among the correspondent D-loop domains in other species of birds (Baker and Marshall 1997), on average. Nevertheless, the distribution of the variable sites within domain I of *Alectoris* was markedly nonrandom (Figs. 2 and 3). The first 161 nucleotides, adjacent to the $tRNA^{Glu}$, apparently evolved at unusually low rates among species (with only 5% variable sites), while the second part of this domain was hypervariable (with 39% variable sites). Average substitution rates for the first (A) and second (B) parts of domain I, central domain II, and domain III were 0.02, 0.20, 0.02, and 0.08, respectively. The presence of a conserved part within domain I, which evolves at the same rate as the conserved central block, suggests a peculiar organization of D-loops of *Alectoris* (schematic mammalian and avian D-loops are compared in Fig. 1).

An extensive comparative analysis of 26 entire mammalian D-loops allowed Sbisà et al. (1997) to identify two long conserved sequence blocks in domain I. These blocks include the formerly described TAS elements (Doda et al. 1981; Foran et al. 1988) and have been denominated extended TAS (ETAS1 and ETAS2) (Sbisà et al. 1997). The remarkable sequence conservation of the ETAS blocks across mammalian species which have evolved during the last 150 million years suggested that ETASs might have important functional roles in controlling the termination of H-strand synthesis (Sbisà et al. 1997). Two partially overlapping sequences with 67 and 45% similarity to the consensus mammalian ETAS1 and ETAS2 (see Sbisà et al. 1997, Fig. 3) can be identified and mapped in the central part of domain I of *Alectoris* (Fig. 2) and other phasianids. The presence and sequence conservation of avian ETASs remain to be confirmed when additional avian D-loops are available. Nevertheless, their presence in galliforms suggests close structural and functional similarity among mammalian and avian D-loop domain I.

Mapping Conserved Motifs of the Avian D-Loop

Comparative analysis of the alignment allows the identification of various conserved motifs in the three domains of D-loops of *Alectoris* (Fig. 2).

Part A of ETAS Domain I (nt 1–161 in Fig. 2). There are three blocks of conserved sequences which are similar to motifs previously identified in other avian and mammalian species. The first block is perfectly conserved among *Alectoris* and has sequence similarity to the ''goose hairpin'' as described in *Anas caerulescens* by Quinn and Wilson (1993, Fig. 2). The secondary structure of this hairpin is determined by a stem of seven complementary Cs/Gs and by a loop containing a TCCC motif also present in mammalian D-loops (Douzery and Randi 1997), which was associated experimentally with termination of H strands (Dufresne et al. 1996). There are two copies of TCCC, mapping at nt 21–24 and 182–185, which flank the cloverleaf putative secondary structure formed by single-strand domain I (Fig. 5). The ''goose hairpin'' is present in the D-loops of 16 species of phasianids (Fumihito et al. 1994; 1995), and in *Anas platyrhyncos* (Ramirez et al. 1993) (Fig. 4a).

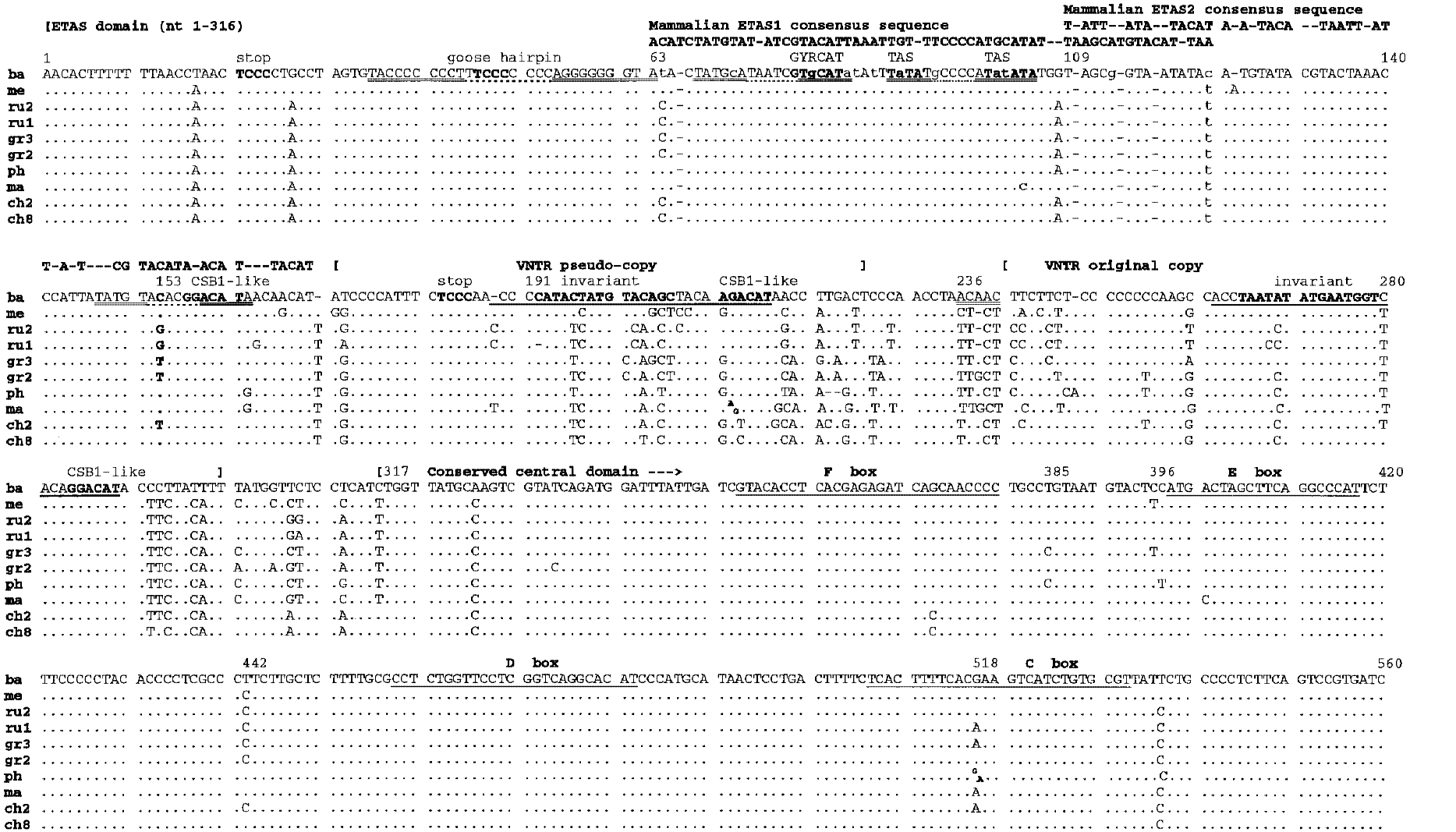

Fig. 2. Aligned nucleotide sequences of the control region of *Alectoris*. *Dots* indicate identity of nucleotides to the reference sequence (*A. barbara*); *dashes* indicate indels. Conserved blocks of putative functional importance, as discussed in the text, are *highlighted*. The putative ends of H-strand DNA synthesis at TCCC motifs are indicated as ''stop'' (in *boldface*). Sequences corresponding to the ''goose hairpin'' (Quinn and Wilson 1993) include a TCCC stop motif (in *boldface*). The stem and loop of this hairpin are underlined with *double lines* and *dots*, respectively. Sequences similar to the mammalian ETAS1 and ETAS2 consensus sequences are indicated at nt 63–121 and 109–169. Nucleotides of the *Alectoris* D-loops which do not correspond to the consensus ETASs are in *lowercase letters*. *Dashes* in the ETAS region were inserted to maximize the similarity of the alignments. TAS sequences and the GYRCAT motif (in *boldface*) can produce two hairpins (stems and loops are underlined with *double lines* and *dots*, respectively). The CSB-1-like motifs are in *boldface*. ''Pseudo'' and ''original'' copies of the repeats of anseriforms and galliforms are delimited by *square brackets* and include the ''invariant'' (Fumihito et al. 1995) and CSB1-like (Quinn and Wilson 1993) motifs. In the central conserved domain there are sequences similar to conserved blocks F, E, D, and C (*underlined*). The putative origin of replication of the H strand (O_H) is *underlined* at nt 785–794. CSB1 is *underlined* at nt 807–832. The presence of CSB2/CSB3 is dubious. A sequence corresponding to the avian LSP/HSP (L'Abbé et al. 1991) is *underlined*, with the conserved TATA box in *boldface*. ba, *A. barbara;* me, *melanocephala;* ru, *rufa;* gr, *graeca;* ph, *philbyi;* ma, *magna;* ch, *chukar*.

```
                                                                                                   647                                                                                700
ba   GCGGCATCTT CTCTCTTCTT TGCTGTTGGT TCCTTTTCTT TTTGGGGCTT CTTCACAGGT TGCCCTTCAC AGTGCGGGTG CGGAGTACTA TTCAAGTGAA GCCAGGACTA CTCCTGCGTT GCGTCCTATC CTAGATCTCT
me   .......... A......... .......... ....C..T.. .C........ .......... .......... .......... ......G... .......... .......... .......... .......... ..........
ru2  .......... A......... .......... ....C..T.. .C........ .......... .......... .......... ......G... .......... .......... .......... .........T .....C....
ru1  .......... A......... .......... ....C..T.. .C........ .......... .......... .......... ......G... .......... .......... .......... .........T .....C....
gr3  .......... ..T....... .......... ....C..T.. .C........ .......... .......... .......... .......... .......... .......... .......... .........T .....C....
gr2  .......... ..T....... .......... ....C..T.. .C........ .......... .......... .......... ......G... .......... .......... .......... .......... ..........
ph   .......... ..T....... .......... ....C..T.. .C........ .......... .......... .......... ......G... .......... ...T...... .......... .......... .........C
ma   .......... ..T....... .......... ....C..T.. .C........ .......... .......... .......... ......G... .......... .......... .......... .......... ..........
ch2  .......... ..T....... .......... ....C..T.. .C........ .......... .......... .......T.. ......G... .......... .......... .......... .......... ..........
ch8  .......... ..T....... .......... ....C..T.. .C........ .......... .......... .......T.. ......G... .......... .......... .......... .......... ..........

                                                               <--- Conserved central domain ] [CSB domain (nt 785-1160)
                                                                                              785   OH 794           807          CSB1           832   840
ba   CGTGTCCCTC AATGAGACGG TTTGCGTGTA TGGGGAATCA TTTTGACACT GATGCACTTT GGATCGCATT TGGTTATGGT TGTTCCACCC CCCCGGTAAA TGGGGCTATA TAGTGAATGC TTGTCGGACA TATTTTTACC
me   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
ru2  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
ru1  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
gr3  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
gr2  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
ph   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
ma   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
ch2  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
ch8  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........

                                  873   CSB2/CSB3    891                                 922                                                                980
ba   ATTTTTCACT TCCTCAATTT TTCCCACAAA ACTAGGGAAT TTCCCACAAA TTTTTTTTTG TTTTTTGTTT AATTTTTTTT TATTTTTTAA AAACATTTTT TAAAAAACAA AATAACATAT AAACTACCGC ATAAATACCC
me   .......... .......... .......... .......... .......... AA........ .......... .......... .G........ .......... .......... .........C .......... .......T..
ru2  .......... .....T.... .......... ..C....... .......... AA........ .......... .......... .......... .......... .......... .......... .......... .......T..
ru1  .......... .....T.... .......... ..C....... .......... AA........ .......... .......... .......... .......... .......... .......... .......... .......T..
gr3  .......... .......... .......... ..C....... .........T .......... .......... .......... .......... .......... .......... .......... .......... .......T..
gr2  .......... .......... .......... ..C....... .......... .......... .......... .......... .TA....... .......... .......... .......... .......... .......T..
ph   .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... .......... .......... .......T..
ma   .......... .......... .......... ..C....... .......... .......... .......... .......... .G........ .......... .......... .........C .....G.... .......T..
ch2  .......... .......... .......... ..C....... .......... .......... .......... .......... .......... .......... .......... .......... .....G.... .......T..
ch8  .......... .......... .......... ..C....... .......... .......... .......... .......... .......... .......... .......... .......... .....G.... .......T..

                994        LSP/HSP         hairpin                                                1073       1084                  1109       1120
ba   CCAAACTATA AAAACGTTTA TGTTTAGTAT ATATACATTG TTGTATTTAA CAACACTATT AGAGAAACTC CACTGCCTAA TCAAACATCA AA-TTAAACA AATTCCCACC AGCACAAAAC CCCACAAACA -AATACTATT
me   T......... .......... ..C....... .......... .......... .......... .......... .......C.. CT......T. ..A..T.... .........T ..T....... .......... ..........
ru2  T.......C. .......... ..CG...... .......... .......... .......... .......... ....A..C.. C.......T. ..AA...... ....A..... .......... .......... ..........
ru1  T.......C. .......... ..CG...... .......... .......... .......... .......... ....A..C.. C.......T. ..AA...... ....A..... .......... .......... ..........
gr3  T.......C. ...-.A.... ..CG...... .......... .......... .......... .......... .......... CT........ ..AA...... ....A..... ........-. ...C...AC A.........
gr2  T.......C. .....A..C. ..CG...... .......... .......... .......... .......... .......... CT........ ..AA...... ....A..... ........-. ........-C A.........
ph   T.......C. .....A.... ..CG...... .......... .--....... .......... .......... .......C.. C.......T. ..AA...... ....A..... ..T.....-. .......... T.........
ma   T....A..C. .....A.... ..CA...... .......... .......... .......... .......... ....A..C.. C......... ..AA...... ....A..... ....A...-. .......... A.........
ch2  T....T..C. .....A.... ..CG...... .......... ..--...... .......... .......... .......C.. C-........ ..AA...... ...AA..... .C......-. .......... ..........
ch8  T....T..C. .....A.... ..CG...... .......... .......... .......... .......... .......C.. C-........ ..AA...... ...AA..... .C......-. .......... ..........

                                          1160
ba   TATATTGTTA GTTACAAACA ATGCATCTTC CACTGCCAAA
me   .......... A......... ..CA.CT... ....A.....
ru2  .......... A......... ..T..CA... ......T...
ru1  .......... A......... ..T..CA... ......T...
gr3  .......... A......... ..T..C.... ...C..T...
gr2  .......... A......... ..T..C.... ...C..T...
ph   .......... .......... ..T....... ...C..T...
ma   .......... .......... ..T..C.... ...C..T...
ch2  .......... A......... ..C..C.... ...C..T...
ch8  .......... A......... ..C..C.... ...C..T...
```

Fig. 2. Continued.

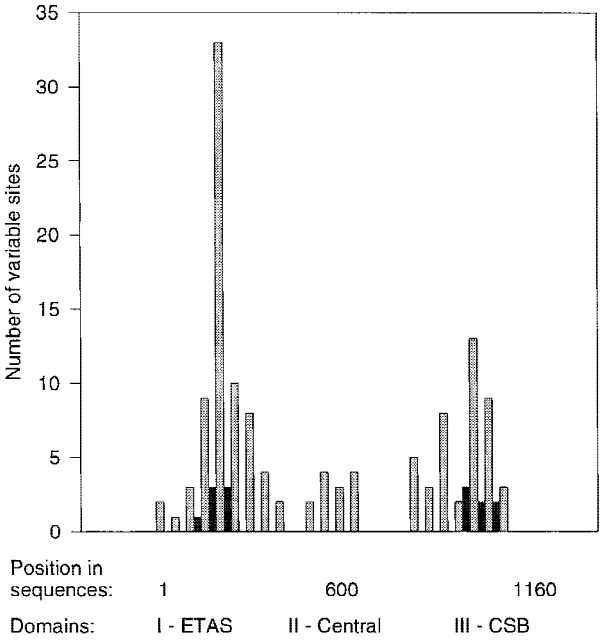

Fig. 3. Plot of the number of variable sites, in a sliding window of 50 nucleotides, in the control region of *Alectoris.* The number of single-nucleotide indels are indicated by *black bars.*

The second block, perfectly conserved among *Alectoris,* has sequence similarity to mammalian TASs, including the highly conserved motif GTGCAT, which is present in all phasianids and anseriforms sequenced so far (Fig. 4b). A putative TAS sequence has been identified in domain I of chaffinches and greenfinch (Marshall and Baker 1997). The motif GYRCAT (Y = C/T; R = A/G) is widespread in domain I of a range of mammalian D-loops (Douzery and Randi 1997), and apparently it has been duplicated in the R1 repeats of many species, including the opossum (Gemmel et al. 1996), several rodents (Stewart and Baker 1994; Fumagalli et al. 1996), and ungulates (Douzery and Randi 1997). Its functional importance is suggested by both comparative (Douzery and Randi 1997) and experimental (Dufresne et al. 1996) findings. These GYRCAT and TAS motifs are included in the putative ETAS1 sequence block of *Alectoris* (nt 63–121; Fig. 2), which has 67% similarity to the mammalian consensus ETAS1 sequence (Sbisà et al. 1997).

The third block has sequence homology to avian CSB1 (cf. the motif GGACAT at positions 156–161 and 826–831). This block is very conserved among galliforms, *Anas* and *Anser* (Fig. 4c), but includes the hypervariable site 153. The CSB1 motif is included in the putative avian ETAS2 block (nt 109–169; Fig. 2), which has 45% similarity to the mammalian consensus ETAS2 sequence (Sbisà et al. 1997). Single-stranded domain I can form potential secondary cloverleaf structures. The configurations of cloverleaves can slightly vary in different species of *Alectoris* (Fig. 5), although (1) four stable hairpins are always possible at the three conserved blocks (underlined in Fig. 2), (2) the free energies of the cloverleaves are very similar, and (3) stem–loop structures of the cloverleaves always include the ''goose hairpin'' and the central part of ETAS1 and ETAS2.

Part B of ETAS Domain I (nt 162–316 in Fig. 2). The second part of the ETAS domain maps immediately upstream to the CSB1 motif and can be divided into two blocks, each one containing a sequence of about 60 nt with similarity to the repeat of anseriforms and galliforms (Desjardins and Morais 1991; Quinn and Wilson 1993; Ramirez et al. 1993; Fumihito et al. 1994, 1995). These repeats have a core sequence 27–29 nt long similar CSB1, as described in chicken and goose (Desjardins and Morais 1991; Quinn and Wilson 1993), which partially corresponds to the ''invariant'' sequence as defined by Fumihito et al. (1994). The first copy of this sequence is very variable among *Alectoris* and has a low similarity to the second copy [i.e., the ''original'' copy of Fumihito et al. (1994)], and therefore we call it the ''pseudo-copy'' (Fig. 2). Tandem duplications of the ''original'' copy were found in red junglefowl (*Gallus gallus*) and green junglefowl (*G. varius*) and can be inserted at approximately position 236–240 in the alignment of *Alectoris,* i.e., between the ''pseudo'' and the ''original'' copies.

Central Conserved Domain II (nt 317–784 in Fig. 2). The central domain of *Alectoris* spans 468 very conserved nucleotides, from the hypervariable site 317 at the end of the ETAS domain to the beginning of the putative O_H sequence block at nucleotide 785 (Fig. 2). Several conserved blocks of the central domain are similar to the F, E, D, and C boxes of vertebrates and other avian species (Quinn and Wilson 1993; Marshall and Baker 1997; Sbisà et al. 1997).

CSB Domain III (nt 785–1160 in Fig. 2). A poly(C) sequence (nt 785–794), similar to the O_H of mammals, maps just a few nucleotides downstream from the putative CSB1 (nt 807–832). CSB1 includes a TATA box, identical to the active site of LSP/HSP of chicken (L'Abbé et al. 1991), and the GGACAT motif, which was apparently duplicated in the ETAS domain. These motifs are also present and highly conserved in CSB1 of chaffinches and greenfinch (Marshall and Baker 1997). In the D-loops of *Alectoris* it is difficult to identify sequences corresponding to mammalian CSB2 and CSB3 (but see Ramirez et al. 1993; Härlid et al. 1997). The bidirectional LSP/HSP promoters (L'Abbé et al. 1991) are almost perfectly conserved. A stable hairpin is possible immediately upstream from the promoters (1019–1035), and farther upstream is the most variable part of the CSB domain, which is rich in poly(T) and poly(A) strings. The CSB domain of *Alectoris* has no tracks of the repeats described in Ciconiiformes (Wenink et al. 1994; Berg et al. 1995).

(a) Alignment of the "goose hairpins"

```
Anser caerulescens     50 ACCCCCCCCCCCCTCCCCCCCCGGGCGGGGT
Anas platyrhyncos      30 ACCCCCCCCCCCCTTCCCCCCCCAGGGGTTGCGGGGTA
Perdix perdix          35 GTACCCCCCCCTTTCCCCCCCCAGGGAGGGTAT
Alectoris*             34 GTACCCCCCCCTTTCCCCCCCCAGGGGGGGTAT
Bambusicola thoracica  36 GTACCCCCCCCTTTCCCCCCCCAGGGGAGGTAT
Chrysolophus pictus    35 GTACCCCCCCCCTTCCCCCCCCAGGGGGGGTAT
Phasianus versicolor   35 GCACCCCCCCCTTCCCCCCCCAGGGGGGGTAT
```

(b) Alignment of the putative TAS sequences and GYRCAT motifs (underlined)

```
Anser caerulescens          87 TTATGCATATTCGTGCATAGATTTATAT
Anas platyrhyncos           88 TTATGCATAT-CGTGCATACATTTATAT
Perdix perdix               67 CTATGTACAG-CGTGCATATCTTTGTGT
Alectoris*                  66 CTATGCATAATCGTGCATATATTTATAT
Coturnix sinensis**         66 CTATGTATAATCGTGCATACATTTATAT
Argusianus argus            67 CTATGTATAATTGTACATACATTTATAT

Consensus mammalian TAS        -------T90A100C100A100T100-------
```

(c) Aligment of CSB1 and CSB1-like sequences

```
CSB1 Anser caerulescens        TATTGGATGAATGCTCGTTGGACATAG
CSB1 Anas platyrhyncos         TATTTAGTGAATGCTCGATGGACATAC
CSB1 Coturnix coturnix         TATATAGTGAATGCTTGCCGGACATAT
CSB1 Gallus gallus         867 TATTTAGTGAATGCTTGTCGGACATAT
CSB1 Alectoris             806 TATATAGTGAATGCTTGTCGGACATAT

CSB1-like Alectoris        145 TATAT-GT-A*A------CGGACAT
                           211                    GGACAT
                           284                    GGACAT
```

Fig. 4. **a** Aligned ''goose hairpins.'' The *underlined* nucleotides indicate complementary positions in the stem of hairpins. *This sequence is shared by all seven species of *Alectoris* and the phasianids *Coturnix sinensis, C. coturnix, Polyplectron bicalcaratum, Argusianus argus, Pavo cristatus, P. muticus, Chrysolophus pictus, Lophura nycthemera, Syrmaticus humiae, Gallus lafayettei, G. sonneratii, G. varius,* and *G. gallus.* **b** Aligned GYRCAT motifs within the putative TAS sequences of *Alectoris* and other galliforms and anseriforms. GYRCAT motifs are *underlined.* *This sequence is shared by all seven species of *Alectoris* and the phasianids *Coturnix coturnix, Pavo cristatus, P. muticus, Syrmaticus humiae, Chrysolophus pictus, Gallus lafayettei, G. varius,* and *G. gallus.* **The sequence of *Coturnix sinensis* is shared with *Bambusicola thoracia, Polyplectron bicalcaratum, Lophura nycthemera, Phasianus versicolor,* and *G. sonneratii.* The consensus mammalian TAS sequence is from Foran et al. (1988). **c** Alignment of CSB1 and CSB1-like sequences. Numbers indicate 5′ nucleotide positions in the alignment of *Alectoris* (Fig. 2) or in the original alignments (see Materials and Methods). For sequence sources, see Materials and Methods.

Sequence Divergence and Phylogenetic Relationships Among Alectoris *Species*

The average nucleotide composition of the D-loop of *Alectoris* was A = 26.3%, T = 32.5%, C = 26.8%, and G = 14.4%, with a bias against G, which is usual for the mtDNA sense strand of vertebrates (Wolstenholme 1992). The G + C biases of the D-loop of *Alectoris* (41.2%) and finches (42.5%) are very similar and not as extreme as reported for other vertebrates [e.g., G + C is only 30% in the D-loops of rainbow fish (Zhu et al. 1994)]. Nucleotide frequencies were not significantly different among species, and thus the Tamura and Nei (1993) model is an appropriate estimator of genetic distances. Genetic distances were 0.0–2.0% within species, 3.0–6.0% between species, and about 20% between *Alectoris* and the phasianid outgroups. There was a trend toward saturation of transitions (Ti) at genetic distances greater than 2.0%. Ti bias declined from 9.0 (within species) to 1.3 (among species) in *Alectoris* and was always <1.0 among the *Alectoris* and the outgroups (Fig. 6a). D-loops of *Alectoris* expressed a significant phylogenetic signal [$g_1 = -1.17$, $p < 0.01$ (Hillis and Huelsenbeck 1992)]. ML phylogenetic analyses (Strimmer and von Haeseler 1996), with Tamura and Nei (1993) distances and the discrete gamma model (Yang 1996b) with four to eight rate categories, showed (Fig. 7) that *A. barbara* and *melanocephala* were basal in the phylogenetic tree, *rufa* and *graeca* were intermediate, and then there was a cluster of derived species including *chukar, magna,* and *philbyi.* The ML tree was fully resolved; only 15 (7.1%) of the 210 analyzed quartets were unresolved, although RP values were lower for the internodes joining *chukar, philbyi,* and *magna.* The Tamura and Nei model with a uniform rate of substitutions among sites produced a ML tree with a similar topology, but the relationships among *chukar, magna,* and *philbyi* were totally unresolved. The ML ratio test of gamma vs uniform rate was highly significant ($\Delta l = 147.25$, $\frac{1}{2}\chi^2 = 9.24$ at the 1% significance level with df = 7). As variable sites are concentrated in short parts of the sequences where they are flanked by other variable sites, we have used the autocorrelated model of Schoeniger and von Haeseler (1994), which produced a topology and RP values identical to the tree shown in Fig. 7, although its log-likelihood was significantly lower than the discrete gamma model ($\Delta l = 44.44$). The same topology was obtained by MOLPHY, using the Hasegawa et al. (1985) model, with Ti/Tv = 4.0 (estimated from the data using PUZZLE) and local rearrangement search. DNAML and DNAMLK

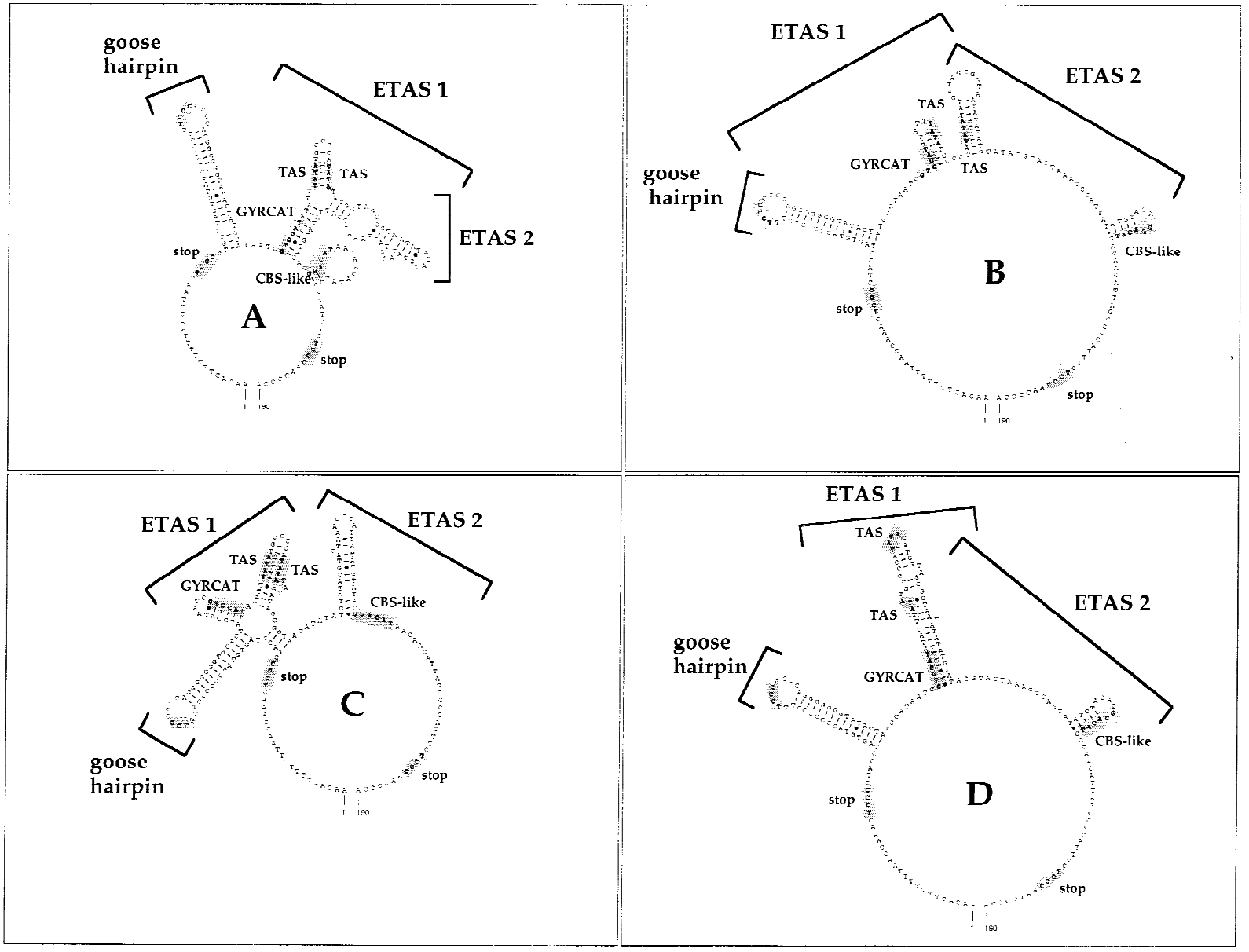

Fig. 5. Cloverleaf potential secondary structures obtained by folding the first 190 nt of the ETAS domain of *Alectoris.* **A** *A. barbara* and *melanocephala* ($\Delta g = -38.5$ kcal); **B** *A. philbyi* and *rufa* ($\Delta g = -35.6$ kcal); **C** *A. chukar* and *graeca* ($\Delta g = -41.1$ kcal); **D** *A. magna* ($\Delta g = -37.4$ kcal). The motifs and secondary structure elements in common in the different species are *highlighted.*

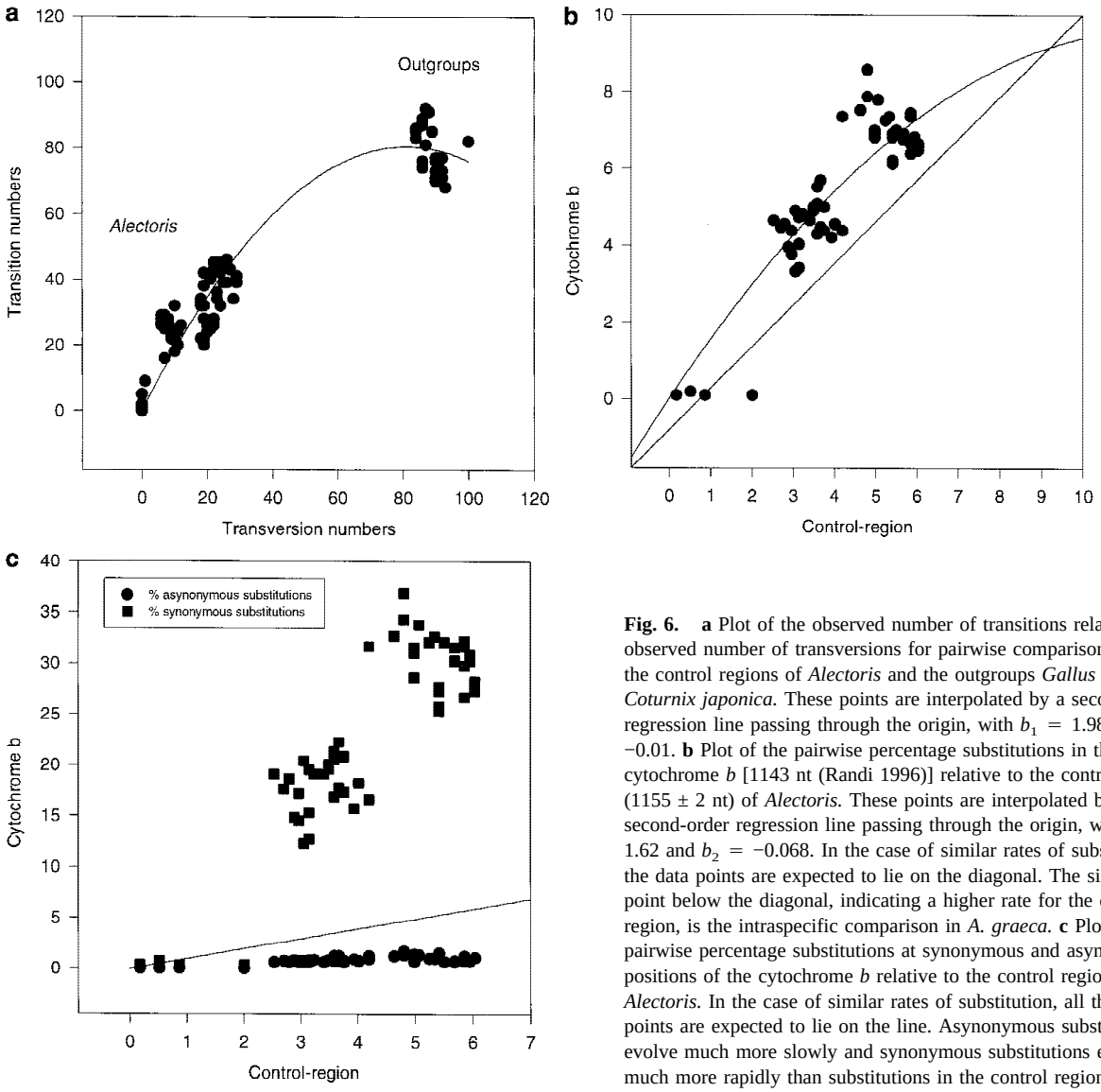

Fig. 6. **a** Plot of the observed number of transitions relative to the observed number of transversions for pairwise comparisons between the control regions of *Alectoris* and the outgroups *Gallus gallus* and *Coturnix japonica.* These points are interpolated by a second-order regression line passing through the origin, with $b_1 = 1.98$ and $b_2 = -0.01$. **b** Plot of the pairwise percentage substitutions in the entire cytochrome *b* [1143 nt (Randi 1996)] relative to the control region (1155 ± 2 nt) of *Alectoris.* These points are interpolated by a second-order regression line passing through the origin, with $b_1 = 1.62$ and $b_2 = -0.068$. In the case of similar rates of substitution, the data points are expected to lie on the diagonal. The single data point below the diagonal, indicating a higher rate for the control region, is the intraspecific comparison in *A. graeca.* **c** Plot of the pairwise percentage substitutions at synonymous and asynonymous positions of the cytochrome *b* relative to the control region of *Alectoris.* In the case of similar rates of substitution, all the data points are expected to lie on the line. Asynonymous substitutions evolve much more slowly and synonymous substitutions evolve much more rapidly than substitutions in the control region.

(in PHYLIP), with empirical base frequencies, Ti/Tv = 4.0, and a uniform rate of substitution across sites and regions, produced two identical trees, corresponding to the tree in Fig. 7, except for a reorganization of the cluster of recent *philbyi, magna,* and *chukar* (i.e., *philbyi* was basal to derived sister *magna* and *chukar*). The ML ratio test of DNAMLK (ML method with the molecular clock) vs DNAML (ML method without the molecular clock) topologies was not significant ($\Delta l = 4.32$, ½ $\chi^2 = 9.24$ at the 1% significance level with df = 7; comparisons among *Alectoris* excluding the outgroups), suggesting that D-loop sequences evolved with fairly constant rates among lineages of *Alectoris.* Discrete gamma models allowed estimation of the shape parameter $\alpha = 0.16$–0.27, which can be used to compute pairwise Tamura and Nei gamma distances. The resulting NJ trees had topologies similar to that of the tree in Fig. 7, but bootstrap values among *chukar, magna,* and *philbyi* were always lower than 50%. The single MP tree (length $L = 327$, using informative characters only; consistency index CI = 0.624; retention index RI = 0.757) was identical to the ML tree, but bootstrap values among *chukar, magna,* and *philbyi* were lower than 50% and the polytomy was unresolved (not shown).

Discussion

Organization and Evolutionary Dynamics of the Control Region of Alectoris

Comparative analyses of the organization and evolution of avian D-loops can help to identify conserved sequences of putative functional importance, to describe the organization of domains and the dynamics of nucleotide substitutions, VNTRs, and indels at different levels of evolutionary divergence. The three domains of the

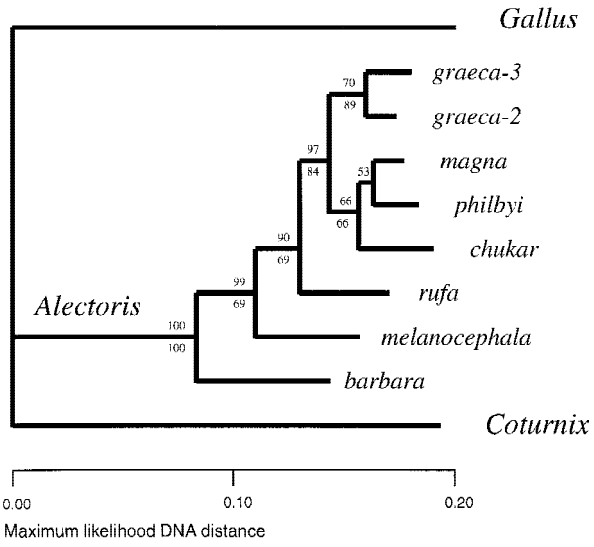

Fig. 7. Maximum-likelihood phylogenetic tree among control-region sequences of *Alectoris.* This tree was obtained using PUZZLE 3.1 (Strimmer and von Haeseler 1996), with 10 sequences (1 for each species and 2 for *A. graeca*), the Tamura and Nei (1993) substitution model with discrete gamma-distributed rates (4 categories), and 1000 puzzling steps. The number of quartets analyzed was 210; there were 15 (7.1%) unresolved quartets. The number of constant sites was 850 (68.7%). The estimated transition/transversion ratio was 1.62 (SE, 0.20). The estimated gamma rate heterogeneity parameter ϵ was 0.86. The log-likelihood of the tree was −3940.07. Reliability percentages (RP > 50%), e.g., the number of times the group appears after 1000 ML puzzling steps, are reported *above each internode.* Bootstrap values (>50%), obtained by PAUP after 500 replicates of the parsimony tree, are reported *below each internode.* Branch lengths are proportional to the maximum-likelihood DNA distances estimated by PUZZLE through the neighbor-joining procedure. *Gallus* and *Coturnix* were used as outgroups.

D-loop of *Alectoris* evolve at highly different rates. The conserved part A of the ETAS domain has sequences similar to the TAS motifs described in mammalian and other avian species; it includes repeated TCCC stop signals and short sequences similar to CSB1. These motifs of putative functional importance are associated with potential cloverleaf or hairpin secondary structures. During the cycles of mtDNA replication and transcription, most of the nascent H strands are abortive in mammals and terminate within domain I at ETASs (Clayton 1992; Sbisà et al. 1997). The role of TASs and secondary structures in blocking the extension of H strands during mtDNA replication has been demonstrated repeatedly (Doda et al. 1981). It is therefore plausible that ETAS domain A contains the critical sequences for regulation of H-strand termination and that its very slow rate of molecular change among *Alectoris* species, and perhaps among galliforms in general, is the consequence of strong selective constraints.

The hypervariable part B of ETAS domain has two blocks of sequences similar to the tandem repeats described in anseriforms and in *Gallus.* The ''original'' copy (Fumihito et al. 1994) of these repeats is conserved among *Alectoris* (72% similarity on average) and among other galliforms. The ''pseudo'' copy has only 58% similarity among *Alectoris* species and is highly divergent from the ''original.'' The second and third tandem duplications of this repeat in *Gallus* (Fumihito et al. 1994) can be inserted between the ''pseudo'' and the ''original'' copies of *Alectoris* (Figs. 1 and 2). As pointed out by Quinn and Wilson (1993) and Fumihito et al. (1994), the second and third duplications of *Gallus* are the most divergent, while the first is almost identical to the ''original'' copy. This pattern of divergence among repeats conforms to a model of unidirectional replication slippage (Fumagalli et al. 1996). However, comparative analyses of D-loops suggest that this mechanism has been activated at least two times during the evolution of anseriforms and galliforms, which, according to Sibley and Ahlquist (1990), are sister clades joined in the superorder Galloanserae. The first duplication probably occurred before their evolutionary splitting, about 66 million years ago (mya), and produced the ''pseudo'' copy of the ''original.'' The first duplicated copy has lost its original functions, if any, and started to evolve faster, just as is expected for the evolution of a pseudogene. This ''pseudo'' copy and the 20 nucleotides flanking its 3′ side constitute the region of hypervariability of D-loops of *Alectoris.* A string of only 50 nt at positions 201–250, corresponding to 2.8% of the total length of the D-loop of *Alectoris,* has 33 substitutions, corresponding to 27% of the total 128 observed variable positions (Fig. 3). Apparently, the mechanism of duplication has been silenced during the early evolution of the Galliformes, but it was reactivated, or retained, in *Gallus*, where it produced conserved copies in the red junglefowl and heteroplasmy in the green junglefowl (Fumihito et al. 1994). It is noteworth that recently published D-loop sequences of South American platyrrhine primates, the marmosets (genus *Callithrix*), have an ''insertion'' sequence, shared with some Old World primates (human, gibbons, and chimps), which is flanked by two direct repeats with TAS-like sequences and is the most variable part of domain I (Tagliaro et al. 1997). This ''insertion'' sequence is located between ETAS1 and ETAS2 (Sbisà et al. 1997). Both mammalian and avian conserved ETAS sequences are interspersed by short variable indels; they are separated by variable numbers of nucleotides in mammals, but they are partially overlapping in *Alectoris.* The presence, organization, and sequence conservation of ETASs in birds will be analyzed more carefully when other D-loop sequences are available.

The central domain of the D-loop is very conserved among the *Alectoris* species and the other galliforms. Nevertheless, there are some sites that might have mutated several times and that are polymorphic within species. For example, according to the reconstructed phylogeny of *Alectoris,* nt 316 had a C-to-T transition from *barbara* to all the other species but reverted to C in

derived *chukar*. Also, site 442 reverted to C in *philbyi*, *magna*, and *chukar*. Sites 334, 385, 396, 442, 518, and 647 are polymorphic in *rufa*, *philbyi*, and *chukar*. Therefore, the CCB of *Alectoris* has sequences which include predominantly invariable sites plus a few interspersed hypervariable sites.

This pattern of heterogeneous substitution rate among sites is well-known for the D-loops of other vertebrate species, including humans (Lyrholm et al 1996; Yang 1996a). Variable rates among sites and domains generate rapid saturation of Ti at the few hypervariable positions and produce lower Ti biases, as observed in comparisons of D-loops and protein-coding mtDNA genes (Zhu et al. 1994). The low bias toward Ti has been related to the lower G+C content of D-loops (Zhu et al. 1994), but most probably it is due to the extreme heterogeneity of substitution rates among sites (Lyrholm et al. 1996). Homoplasy might also derive from nucleotide substitutions generated by strand slippage at DNA replication (Vogler et al. 1997). In the variable part of the CSB domain of *Alectoris* there are some strings of poly(T) and poly(A) (Fig. 2), which can generate strand mispairing and slippage. This domain has an unexpectedly high frequency of transversions at adjacent nucleotides: 22 of 44 variable sites (50%, excluding indels) have transversions, vs only 22 of 84 (26%, excluding indels) sites in the other domains. For example, downstream from a long poly(T) string, nt 890–892 show one or two unusual T-to-A transversions among *barbara*, *graeca*-3, *rufa*, and *melanocephala*. The same is true for sites 922, 923, 1073–1076, 1084–1085, and 1109–1111 (Fig. 2).

Different Organization of Mammalian and Avian Control Regions

Most of the tandemly repeated sequences of mammalian D-loops have two traits in common: (1) they are located upstream from the origin (R2 repeats), or downstream from the termination (R1 repeats), of the H strands; and (2) they contain duplicated TAS sequences and, in some cases, CSB1-like motifs. Therefore, sequences associated with termination of the H strands map upstream from the VNTR regions in mammals (Fig. 1). On the contrary, the D-loops of *Alectoris* and all the other galliforms and anseriforms studied so far have the highly conserved part of the ETAS domain with TAS sequences and other putative signals of termination of the nascent H strands, which always map downstream from the VNTR region (Fig. 1). An important consequence of the peculiar organization of the D-loop of galliforms is that the hypervariable part of the ETAS domain, with its tandemly duplicated copies of the repeat, should always be replicated within the nascent H strands. D-Loop tandem repeats are known to be transcribed during replication of mammalian mtDNA (MacKay et al. 1993).

It has been suggested that mitochondrial VNTRs might be randomly generated by strand slippage during DNA replication and can be quickly eliminated by drift or stabilizing selection for optimal length of the mtDNA molecules (Rand 1993). However, the position of VNTRs near the origins of replication or at sites of termination of the H strands, the presence of copies of regulatory sequences within the repeats, and the potential to fold in secondary structures suggest that VNTRs can provide multiple sites for starting and stopping DNA replication (Dufresne et al. 1996; Douzery and Randi 1997). Therefore, some classes of VNTRs could confer replicative advantages to the mtDNAs (Kuzminov 1996) and persist for millions of years in particular lineages (Biju-Duval et al. 1991; Stewart and Baker 1994). However, the VNTRs confined upstream from the functional signals of termination and the very limited presence of VNTRs among anseriforms and galliforms suggest that functional constraints on the generation of VNTRs and on the control of mtDNA terminations could be stronger in avian than in mammalian D-loops. Peripheral domain I of chaffinches and greenfinch had a TAS-like element associated with a cloverleaf secondary structure, the central domain II had sequences similar to the F, D, and C boxes, and the peripheral domain III had a CSB1 associated with secondary structures (Marshall and Baker 1997). As in *Alectoris*, the length of the D-loop was conserved among three species of chaffinches and the greenfinch (average length, 1230 ± 7 nt). Therefore, contrasting with the widespread presence of VNTRs in mammals, the length of the D-loop is very conserved in the studied species of galliforms and passerines, suggesting strong stabilizing selection for optimal size of the avian mtDNA.

The CSB domain has no obvious tracks of CSB2 and CSB3 in *Alectoris* and passerines (Marshall and Baker 1997), but these blocks are also absent in some mammals, and their functional importance has been questioned (Saccone et al 1991; Marshall and Baker 1977; Sbisà et al. 1997; Douzery and Randi 1997). Moreover, the finches did not show sequences similar to the transcription promoters mapped in duck, chicken, quail, and *Alectoris* species (Marshall and Baker 1997). Thus, comparative analyses suggest that, although blocks of sequences similar to regulatory sites are always present in peripheral domains of avian and mammalian D-loops, their functions are not related strictly to conserved sequence motifs or secondary structures, but they are relatively free to evolve or to coevolve with the nuclear genes coding for the protein factors essential for mtDNA replication and transcription (Clayton 1992; Sbisà et al. 1997).

Evolutionary Rates and Phylogenetic Utility of the Avian D-Loop

Most of the sites that are free to vary are restricted to short regions within the ETAS and CSB domains of the

D-loop, where the few hypervariable sites have experienced repeated substitutions, probably due in part to strand slippage at DNA replication. Highly skewed distributions of mutability among site and domain are expected to produce rapid saturation and introduce homoplasy in the data set (Yang 1996a). Nevertheless, the average slow rate of molecular change preserves the significant phylogenetic signal among the D-loops of *Alectoris*, which evolve slightly more slowly than cytochrome *b* (cyt *b*), on average (Fig. 6b), but more rapidly than asynonymous substitutions (Fig. 6c), and produce phylogenetic trees concordant with cyt *b* trees [compare Fig. 7 in this paper with Figs. 5 and 6 of Randi (1996)]. Allozyme and cyt *b* analyses split the *Alectoris* into three groups: (1) Eurasian species of recent origins (*chukar*, *magna*, and *philbyi*); (2) intermediate West European species (*rufa* and *graeca*); and (3) ancient and more distantly related North African and Arabian species (*barbara* and *melanocephala*) (Randi et al. 1992; Randi 1996). Calibrated rates of molecular evolution suggested that these three groups could have speciated about 1.8–2.0, 2.4–3.8, and 5.9–6.4 million years ago (mya), respectively (Randi 1996). Although the evolutionary framework was quite well established, detailed patterns of nested relationships among the recent species of *Alectoris* were not resolved by allozyme or cyt *b* data (Randi 1996). Phylogenetic trees of D-loops confirm the nested patterns of relationships among *Alectoris* and produced a better resolution of relationships among the recent species (Fig. 7): accordingly, *A. chukar* should be basal to the sister species *philbyi* and *magna*. The resolution of this cluster is improved by the use of ML procedures with gamma-distributed models of the substitution rates among sites. The consequences of multiple substitutions at hypervariable sites and of sequence variation generated by slippage on phylogenetic inference, as well as the analysis of comparative rates of evolution of mtDNA coding vs D-loop sequences among *Alectoris*, will be discussed in a forthcoming paper (Randi and Lucchini, in preparation).

Acknowledgments. This work was supported by the Italian Institute of Wildlife Biology (INFS). We thank D. Dias, A. Meriggi, M. Paganin, I. Artuso, D. Febbo, P.U. Alkon, S. Kark, Y. Alaoui, M. Pellegrini, A. Greth, M. Vassart, P. Paillat, L. Granjeon, N. Liu, and C. Qiang for kindly providing us with tissue samples. E.A. Scott, Manchester Metropolitan University, suggested primer sequences and details on laboratory protocols. C. Saccone, A.J. Baker, and two anonymous reviewers provided helpful comments which improved an early draft of this paper.

References

Adachi J, Hasegawa M. (1996) MOPLHY: programs for molecular phylogenetics, Version 2.3. Institute of Statistical Mathematics, Tokyo

Baker AJ, Marshall HD (1997) Mitochondrial control-region sequences as tools for understanding the evolution of avian taxa. In: Mindell DP (ed) Avian molecular systematics and evolution. Academic Press, New York, pp 49–80

Berg T, Moum T, Johansen S (1995) Variable numbers of simple tandem repeats make birds of the orders Ciconiiformes heteroplasmic in their mitochondrial genomes. Curr Genet 27:257–262

Biju-Duval C, Ennafaa H, Dennebouy N, Monnerot M, Mignotte M, Soriguer RC, El Gaaïed A, El Hili A, Mounolou J-C (1991) Mitochondrial DNA evolution in lagomorphs: origin of systematic heteroplasmy and organization of diversity in European rabbits. J Mol Evol 33:92–102

Brown GG, Gadaleta G, Pepe G, Saccone C, Sbisà E (1986) Structural conservation and variation in the D-loop containing region of the vertebrate mitochondrial DNA. J Mol Biol 192:503–511

Chang D, Clayton D (1987) A novel endonuclease cleaves at a priming site of mouse mitochondrial DNA replication. EMBO J 6:409–417

Clayton DA (1992) Transcription and replication of animal mitochondrial DNAs. In: Wolstenholme DR, Jeon KW (eds) International review of cytology, Vol 141. Mitochondrial genomes. Academic Press, San Diego, pp 217–323

Desjardins P, Morais R (1990) Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. J Mol Biol 212:599–634

Desjardins P, Morais R (1991) Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. J Mol Evol 32:153–161

Desjardins P, Ramirez V, Morais R (1990) Gene organization of the Pecking duck mitochondrial genome. Curr Genet 17:515–518

Doda JN, Wright CT, Clayton DA (1981) Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc Natl Acad Sci USA 78: 6116–6120

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

Dufresne C, Mignotte F, Guéride M (1996) The presence of tandem repeats and the initiation of replication in rabbit mitochondrial DNA. Eur J Biochem 235:593–600

Edwards SV (1993) Mitochondrial gene genealogy and gene flow among island and mainland populations of a sedentary songbird, the grey-crowned babbler (*Pomatostomus temporalis*). Evolution 47: 1118–1137

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:83–791

Felsenstein J (1993) PHYLIP: phylogenetic inference package, Version 3.5. University of Washington, Seattle.

Foran DR, Hixson JE, Brown WM (1988) Comparison of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res 16:5841–5861

Fumagalli L, Taberlet P, Favre L, Hausser J (1996) Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. Mol Biol Evol 13:31–46

Fumihito A, Miyake T, Sumi S-I, Takada M, Ohno S, Kondo N (1994) One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds. Proc Natl Acad Sci USA 91:12505–12509

Fumihito A, Miyake T, Takada M, Ohno S, Kondo N (1995) The genetic link between the Chinese bamboo partridge (*Bambusicola thoracica*) and the chicken and junglefowls of the genus *Gallus*. Proc Natl Acad Sci USA 92:11053–11056

Gaut BS, Weir B (1994) Detecting substitution rate heterogeneity among regions of nucleotide sequence. Mol Biol Evol 11:620–629

Gemmell NJ, Western PS, Watson JM, Marshall-Graves JA (1996) Evolution of the mammalian mitochondrial control region—comparisons of control region sequences between monotreme and therian mammals. Mol Biol Evol 13:798–808

Gerloff U, Schlotterer C, Rassmann K, Rambold I, Hohmann G, Frutth B, Tautz D (1995) Amplification of hypervariable simple sequence

repeats (microsatellites) from excremental DNA of wild living Bonobos (*Pan paniscus*). Mol Ecol 4:515–518

Gilbert DG (1992) LOOPDLOOP, a Macintosh program for visualizing RNA secondary structure. Published electronically on the Internet, available via anonymous ftp to ftp.bio.indiana.edu

Härlid A, Janke A, Arnason U (1997) The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. Mol Biol Evol 14:754–761

Hasegawa M, Kishino H, Yano T (1985) Dating of the human–ape splitting by molecular clock of mitochondrial DNA. J Mol Evol 22:160–174

Hillis DM, Huelsenbeck JP (1992) Signal, noise, and reliability in molecular phylogenetic analyses. J Hered 83:189–195

Hoelzel RA, Lopez JV, Dover G, O'Brien SJ (1994) Rapid evolution of a heteroplasmic repetitive sequences in the mitochondrial DNA control region of carnivores. J Mol Evol 39:191–199

Janke A, Feldmaier-Fuchs G, Thomas WK, Von Haeseler A, Pääbo S (1994) The marsupial mitochondrial genome and the evolution of placental mammals. Genetics 137:243–256

Kumar S, Tamura K, Nei M (1993) MEGA: molecular evolutionary genetic analysis, Version 1.01. Pennsylvania State University, University Park

Kuzminov A (1996) Mutant fixation via plasmid dimerization and its relation to human diseases. Trends Genet 12:246–249

L'Abbé DL, Duhaime J-F, Lang BF, Morais R (1991) The transcription of DNA in chicken mitochondria initiates from one major bidirectional promoter. J Biol Chem 266:10844–10850

Lyrholm T, Leimar O, Gyllensten U (1996) Low diversity and biased substitution patterns in the mitochondrial DNA control region of sperm whales: implications for the estimates of time since common ancestry. Mol Biol Evol 13:1318–1326

MacKay SLD, Ghivizzani SC, Madsen CS, Tanhauser SM, Laipis PJ, Hauswirth WW (1993) Transcribed heteroplasmic repeated sequences in the porcine mitochondrial DNA D-loop region. J Mol Evol 37:36–47

Marshall HD, Baker AJ (1997) Structural conservation and variation in the mitochondrial control region of fringilline finches (*Fringilla* spp.) and the greenfinch (*Carduelis chloris*). Mol Biol Evol 14: 173–184

Mundy NI, Winchell CS, Woodruff DS (1997) Genetic differences between the endangered San Clemente Island loggerhead shrike *Lanius ludovicianus mearnsi* and two neighbouring subspecies demonstrated by mtDNA control region and cytochrome *b* sequence variation. Mol Ecol 6:29–37

Palva TK, Palva ET (1985) Rapid isolation of animal mitochondrial DNA by alkaline extraction. FEBS Lett 192:267–270

Quinn TW (1992) The genetic legacy of Mother Goose—phylogeographic patterns of lesser snow goose *Chen caerulescens caerulescens* maternal lineages. Mol Ecol 1:105–117

Quinn TW (1997) Molecular evolution of the mitochondrial genome. In: Mindell DP (ed) Avian molecular systematics and evolution. Academic Press, New York, pp 3–28

Quinn TW, Wilson AC (1993) Sequence evolution in and around the mitochondrial control region in birds. J Mol Evol 37:417–425

Ramirez V, Savoie P, Morais R (1993) Molecular characterization and evolution of a duck mitochondrial genome. J Mol Evol 37:296–310

Rand DM (1993) Endotherms, ectotherms, and mitochondrial genome-size variation. J Mol Evol 37:281–295

Randi E (1996) A mitochondrial cytochrome *b* phylogeny of the *Alectoris* partridges. Mol Phylogenet Evol 6:214–227

Randi E, Meriggi A, Lorenzini R, Fusco G, Alkon PU (1992) Biochemical analysis of relationships of Mediterranean *Alectoris* partridges. Auk 102:358–367

Rzhetsky A, Nei M (1995) Tests of applicability of several substitution models for DNA sequence data. Mol Biol Evol 12:131–151

Saccone C, Attimonelli M, Sbisà E (1987) Structural elements highly preserved during the evolution of the D-loop containing region in vertebrate mitochondrial DNA. J Mol Evol 26:205–211

Saccone C, Pesole G, Sbisà E (1991) The main regulatory region of mammalian mitochondrial DNA: structure–function model and evolutionary pattern. J Mol Evol 33:83–91

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205:125–140

Schoeniger M, von Haeseler A (1994) A stochastic model for the evolution of autocorrelated DNA sequences. Mol Phylogenet Evol 3:240–247

Sibley CG, Ahlquist JE (1990) Phylogeny and classification of birds. A study in molecular evolution. Yale University Press, New Haven; CT

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

Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, Version 3.1. Illinois Natural History Survey, Champaign

Tagliaro CH, Schneider MPC, Schneider H, Sampaio I, Stanhope MJ (1997) Marmoset phylogenetics, conservation perspectives, and evolution of the mtDNA control region. Mol Biol Evol 14:674–684

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Vogler AP, Welsh A, Hancock JM (1997) Phylogenetic analysis of slippage-like sequence variation in the V4 rRNA expansion segment in tiger beetles (Cicindelidae). Mol Biol Evol 14:6–19

Walberg MW, Clayton DA (1981) Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res 9:5411–5421

Wenink PW, Baker A, Tilanus MGJ (1994) Hypervariable 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

Wenink PW, Baker A, Rosner HU, Tilanus MGJ (1996) Global mitochondrial DNA phylogeography of Holarctic breeding dunlins (*Calidris alpina*). Evolution 50:318–330

Wolstenholme DR (1992) Animal mitochondrial DNA: Structure and evolution. In: Wolstenholme DR, Jeon KW (eds) International review of cytology, Vol 141. Mitochondrial genomes. Academic Press, San Diego, pp 173–216

Yang Z (1996a) Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol Evol 11:367–372

Yang Z (1996b) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39:306–314

Zardoya R, Villalta M, Lòpez-Pérez MJ, Garrido-Pertierra A, Montoya J, Bautista JM (1995) Nucleotide sequences of the sheep mitochondrial DNA D-loop and its flanking tRNA genes. Curr Genet 28: 94–96

Zhu D, Jamieson BGM, Hugall A, Moritz C (1994) Sequence evolution and phylogenetic signal in D-loop and cytochrome *b* sequences of rainbow fishes (Melanotaeniidae). Mol Biol Evol 11:672–683