The Complete Sequence of the Mitochondrial Genome of *Buteo buteo* (Aves, Accipitridae) Indicates an Early Split in the Phylogeny of Raptors

*Elisabeth Haring,* *Luise Kruckenhauser,* *Anita Gamauf,* *Martin J. Riesing,* and *Wilhelm Pinsker*†

*Zoological Department, Museum of Natural History, Vienna, Austria; and †Institute of Medical Biology, University of Vienna, Austria

The complete sequence of the mitochondrial (mt) genome of *Buteo buteo* was determined. Its gene content and nucleotide composition are typical for avian genomes. Due to expanded noncoding sequences, *Buteo* possesses the longest mt genome sequenced so far (18,674 bp). The gene order comprising the control region and neighboring genes is identical to that of *Falco peregrinus*, suggesting that the corresponding rearrangement occurred before the falconid/accipitrid split. Phylogenetic analyses performed with the mt sequence of *Buteo* and nine other mt genomes suggest that for investigations at higher taxonomic levels (e.g., avian orders), concatenated rRNA and tRNA gene sequences are more informative than protein gene sequences with respect to resolution and bootstrap support. Phylogenetic analyses indicate an early split between Accipitridae and Falconidae, which, according to molecular dating of other avian divergence times, can be assumed to have taken place in the late Cretaceous 65–83 MYA.

Introduction

Among birds of prey, buzzards and hawks of the genus *Buteo* are an extremely successful group that is widely distributed, being absent only in Australia, Antarctica, and most parts of the oriental region. The genus currently comprises between 25 (Brown and Amadon 1968) and 27 species (del Hoyo, Elliott, and Sargatal 1994), thus representing about 7% of the species in the family Accipitridae. In the present study, we focus on the common buzzard *Buteo buteo* (Linnaeus, 1758), which is the most abundant raptor species in many parts of Europe (Bijlsma 1997).

The two largest families of birds of prey are the Accipitridae and the Falconidae. The accipitrids, known colloquially as hawks, kites, harriers, vultures, and eagles, are rather similar in their basic morphological structures, although they show great diversity in size, shape, flying ability, ecology, and predatory habits. The falcons resemble the accipitrids in some characteristics, such as a powerful hooked bill, a fleshy cere straddling the bill, heavy bony brow ridges, and a crop (to store freshly eaten food). The differences between the two families have been summarized by Olsen (1995) in a survey of 25 anatomical and behavioral traits.

According to traditional morphological classifications (e.g., Brown and Amadon 1968; Storer 1971; Stresemann and Amadon 1979; Cracraft 1981), Accipitridae and Falconidae belong to the order Falconiformes. However, based on detailed morphological studies of several families, Jollie (1976, 1977) concluded that falconids and accipitrids are not closely related. According to his interpretation, the order Falconiformes is polyphyletic, especially with respect to the inclusion of New World vultures (Cathartidae); that view is supported by studies of several behavioral traits (König 1982) and, more recently, by molecular analyses (e.g., Sibley and Ahlquist 1990; Seibold and Helbig 1995; Wink et al. 1998). Sibley and Ahlquist (1990) estimated the overall genomic similarity by DNA-DNA hybridization and proposed a new classification of birds in which the New World vultures appear as close relatives of the storks (Ciconiidae). In their classification, the falconiform taxa are placed within an expanded order Ciconiiformes, in which they include the infraorders Falconides (including Falconidae and Accipitridae) and Ciconiides (including the family Ciconiidae with the subfamilies Cathartinae and Ciconiinae). Sequence analyses of the *cytochrome b* gene (*cyt b*) (Avise, Nelson, and Sibley 1994; Seibold and Helbig 1995; Wink et al. 1998) support this taxonomic position for New World vultures, but the relationships of the Ciconiidae with respect to Accipitridae and Cathartidae have not been resolved unambiguously. So far, no other genetic markers (mitochondrial or nuclear) have been employed to elucidate the phylogenetic relationships of these families.

Analyses of complete mitochondrial (mt) genomes provide not only sequence data for phylogenetic studies, but also information about structural genomic rearrangements which may serve as additional markers. Sequencing of the first complete vertebrate mt genomes suggested that gene content and gene order are highly conserved, but subsequent sequence data have demonstrated that the gene order in vertebrates is not uniform (for review, see Quinn 1997). A major rearrangement within the mt genome of chickens and other galliform birds has been described by Desjardins and Morais (1990). It comprises the *cyt b* gene, the NADH dehydrogenase subunit 6 gene (*nd6*), and several tRNA genes. Subsequently, this particular gene order has been found in several other avian species, suggesting that this rearrangement could have occurred at the base of the avian branch and thus might be shared by all recent bird species. However, the hypothesis of a universal gene order characteristic for all birds was refuted by the discovery of yet another rearrangement of the mt genes in *Falco peregrinus*, as well as in birds of four additional orders (Mindell, Sorenson, and Dimcheff 1998). In an inves-
tigation of 137 species, representing 13 orders, Mindell, Sorenson, and Dimcheff (1998) hypothesize that this novel arrangement, which includes the control region (CR) and surrounding sections, must have originated independently four times in avian evolution. In addition to the CR, these species possess a second noncoding (nc) region, probably generated through a duplication process. The same arrangement was recently detected in warblers of the genus Phylloscopus (Bensch and Härild 2000). Partial sequence analysis of the B. buteo mt genome (Haring et al. 1999) revealed the existence of a noncoding section corresponding to the nc region of F. peregrinus, which was designated a pseudo control region (ΨCR). Although the position of the functional CR in B. buteo was not determined, this finding suggested that F. peregrinus and B. buteo might share the same gene order.

In this paper, we report the complete sequence of the mt genome of B. buteo. For sequence comparisons, we used the previously published complete mt genomes of nine other bird species, along with that of Alligator mississippiensis as an outgroup, to address the following questions: (1) Do Falconidae and Accipitridae share the same rearrangement within their mt genomes? (2) What are the phylogenetic positions of the genera Buteo, Ciconia, and Falco within the avian tree? (3) Are subsections as useful as complete mt genomes for resolving phylogenies? (Up to now, the majority of avian molecular phylogenies have been based on sequence data of the mitochondrial cyt b gene.) (4) What is the degree of divergence between Buteo and Falco, and can it be related to other splits in the phylogeny of birds to estimate their approximate divergence time?

### Materials and Methods

#### DNA Extraction

DNA from B. buteo was extracted from the blood (stored in EDTA buffer) of a single specimen (specimen b.but-2, common buzzard, B. buteo buteo; Haring et al. 1999) by overnight incubation at 37°C in extraction buffer (10 mM Tris-HCl [pH 8.0], 10 mM EDTA, 50 mM NaCl, 40 mM dithiothreitol, 1% SDS, 0.5 mg/ml proteinase K). DNA was purified by two phenol/chloroform/isoamylalcohol (25:24:1) extractions and one chloroform/isoamylalcohol (1:1) extraction, followed by precipitation with 1/10 volumes 3 M NaAc, 3 × volumes EtOH.

#### PCR Amplification

PCR was carried out with an Eppendorf Thermocycler in a volume of 25 μl containing 1 unit Dynazyme DNA polymerase (Finnzymes OY), 1 μM of each primer, 0.2 mM of each dNTP, and 50–200 ng template DNA. The solutions were heated to 95°C (2 min) and then put through 30 reaction cycles: 95°C (10 s), annealing temperature (10 s), and 72°C (1 min/1 kb expected length), followed by a final extension at 72°C (10 min). Primers used for PCR amplification are listed in table 1. The 11 PCR fragments that were subsequently cloned and sequenced are depicted in figure 1. These overlapping fragments cover the whole mt genome of B. buteo except the region comprising a part of the nd6 gene, tRNAAsp, ψCR, tRNAPro, and a part of the 12S rRNA gene (primer pair nd6-1+/12S-1−) determined in a previous study from the same individual of B. buteo (GenBank accession number AF202186). The 11 fragments sequenced in the present study were obtained us-

### Table 1

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’−3’)</th>
<th>Binding Sitea</th>
<th>Referenceb</th>
</tr>
</thead>
<tbody>
<tr>
<td>12S-3+</td>
<td>CGCAACCAGGACCAAGCCTAGGGCC</td>
<td>305</td>
<td>ps</td>
</tr>
<tr>
<td>16S-3−</td>
<td>GTGTCATCTCAGCCTATACTGAG</td>
<td>1,214</td>
<td>ps</td>
</tr>
<tr>
<td>12S-2+</td>
<td>AAAGCTGAAACAGGTAATTG</td>
<td>998</td>
<td>ps</td>
</tr>
<tr>
<td>16S-2−</td>
<td>ATCCCTGGGTAGCTGTTGCC</td>
<td>2,407</td>
<td>ps</td>
</tr>
<tr>
<td>16S-1+</td>
<td>CCGAGAAACCCCTAGCTGACCCCT</td>
<td>2,172</td>
<td>p.</td>
</tr>
<tr>
<td>nd2−2</td>
<td>TGATGAGAAGGGCTAGGATTTTGGC</td>
<td>4,543</td>
<td>S, H5766</td>
</tr>
<tr>
<td>nd2+1</td>
<td>GGATGATGAGGACCAACCAAGAC</td>
<td>4,490</td>
<td>S, H5758</td>
</tr>
<tr>
<td>cox2−2</td>
<td>TCTCAAGCTCCTATATCTGGTGC</td>
<td>7,404</td>
<td>S, H8628</td>
</tr>
<tr>
<td>atp6−2</td>
<td>ATGTTTCTTGGTGGATAGG</td>
<td>8,673</td>
<td></td>
</tr>
<tr>
<td>atp6+1</td>
<td>AGCTTCCTGCTCTCTAATAG</td>
<td>8,631</td>
<td></td>
</tr>
<tr>
<td>nd4−4</td>
<td>GTTCTTAGGCTAGTATGAGCG</td>
<td>10,060</td>
<td>ps</td>
</tr>
<tr>
<td>nd3−1+</td>
<td>CAGGAGAACCTAGGATGACAC</td>
<td>9,854</td>
<td>ps</td>
</tr>
<tr>
<td>nd4+5</td>
<td>ATGGTACTGCCCACTTAGG</td>
<td>11,387</td>
<td>ps</td>
</tr>
<tr>
<td>nd4+3</td>
<td>ACCAATTACGCGGGCAGACACAG</td>
<td>11,239</td>
<td>ps</td>
</tr>
<tr>
<td>nd5−2</td>
<td>ATGATTCACCCTCTCTCAGCC</td>
<td>12,285</td>
<td>ps</td>
</tr>
<tr>
<td>nd5−3+</td>
<td>AATTTGCAACATGATACATAGC</td>
<td>12,141</td>
<td>ps</td>
</tr>
<tr>
<td>cyt−3</td>
<td>TGACTTGTTCGCGCATGTGGGCC</td>
<td>13,887</td>
<td></td>
</tr>
<tr>
<td>cyt+4</td>
<td>TAAGGAGGGGAACTAGTGG</td>
<td>13,844</td>
<td>ps</td>
</tr>
<tr>
<td>cyt−1</td>
<td>ACCATTCATCATCATGCCC</td>
<td>14,739</td>
<td></td>
</tr>
<tr>
<td>nd6−3</td>
<td>CGTCTTATATGCGGTATG</td>
<td>16,770</td>
<td>ps</td>
</tr>
</tbody>
</table>

a The binding sites (5’ position) within the mitochondrial genome of Buteo are given.
b K = Knight and Mindell (1993); S = Sorenson et al. (1999); H = Haring et al. (1999); ps = present study.
FIG. 1.—Mitochondrial genome of *Buteo*. Overlapping PCR fragments are depicted with their corresponding primers. Arrows at the inner circle indicate orientation of genes. For designation of tRNAs, the three-letter code for amino acids is used. Position 0 of the complete sequence (table 2) is indicated.

Cloning and Sequencing

PCR products were extracted from agarose gels using the QIAquick Gel Extraction Kit (QIAGEN) and cloned (TOPO TA Cloning Kit, Invitrogen). Sequencing of the clones (both directions, M13 primers) was performed by MWG-Biotech (Ebersberg, Germany) with a Li-Cor Sequencer. Due to overlapping clones and overlap of regions sequenced with internal primers, the whole mitochondrial genome was read at least twice (some sections four times). Thus, the complete sequence was determined without any ambiguities.

Sequence Analysis

Alignments of DNA and amino acid sequences were produced with the program CLUSTAL X (Thompson et al. 1997) and improved manually. Distance-based (the neighbor-joining [NJ] algorithm; Saitou and Nei 1987) and maximum-parsimony (MP) methods were used to infer the phylogenetic relationships. All dendrograms were calculated with the software package PAUP (test version 4.0b3a; Swofford 1998) using a heuristic search with the tree bisection-reconnection (TBR) algorithm and a random-taxon-addition sequence. Gaps were treated as "missing," and all characters were...
Complete Mitochondrial Genome of Buteo buteo

The complete mitochondrial genome of Buteo is 18,674 bp in size. In comparison with other complete avian mt genomes, which range from 16,591 bp (Struthio) to 18,068 bp (Falco), Buteo has the largest mt genome sequenced so far. This considerable difference in size is due to an expansion of noncoding sequences. The three bird species with the largest portions of noncoding mt DNA (Buteo, 17.1%; Falco, 14.1%; Smithornis, 10.7%) possess, in addition to the CR, a second nc region (ΨCR). The gene content of the Buteo mt genome is typical of vertebrates (fig. 1 and table 2), consisting of genes for 13 proteins (12 H-strand-encoded, 1 L-strand-encoded), 22 tRNAs, and 2 rRNAs.

The nucleotide composition of the Buteo mt genome (H-strand) is similar to that of other avian species (A = 31.1%, C = 31.7%, G = 13.2%, T = 24.0%). The A+T content of 55.1% is within the range for avian mt genomes (51.6%–55.7%). The usage of translation initiation and termination signals in comparison with other bird species is given in table 3. The most common start codon is ATG. In Buteo, nonstandard start codons are found in the cox1 and nad5 genes (GTG) and in the nd3 gene (ATC). The unusual ATC start codon in the nd3 gene is also found in Smithornis. As in the mt genomes of the other birds and Alligator, TAA is the most frequent stop codon in Buteo; TAG and AGG are used twice. In cox3 and nd4, a terminal T probably serves as the stop signal after it is completed to UAA by post-transcriptional polyadenylation (Ojala, Montoya, and Attardi 1981). Buteo and Falco differ in two start codons and in three stop codons. With respect to the length of intergenic spacers and overlaps, Buteo (67-bp spacers) has a rather compact genome compared with Falco (101-bp spacers) (table 4).

Gene Order and Noncoding Regions

In contrast to the standard avian gene order (e.g., Gallus), the CR of Buteo is located between tRNA^Thr^ and tRNA^Glu^ . Between tRNA^Glu^ and tRNA^Phe^ , at the position of the CR in the majority of the birds analyzed so far, another nc section is found in Buteo which was designated a ΨCR by Haring et al. (1999). It consists of a short nonrepetitive part (23.2% of its length) followed by 48-bp tandem repeats (76.8%). A similar array of nc sections (designated CR and nc) has been described for the mt genome of F. peregrinus (Mindell, Sorenson, and Dimcheff 1998). Thus, gene order and content are identical in Buteo and Falco but differ from the standard gene order of many other birds (including Ciconia, Gallus, Aythya, Rhea, Struthio, Vidua, and Corvus) with respect to CR and ΨCR (table 5). The gene order of Buteo and Falco is similar to that of Smithornis (Mindell, Sorenson, and Dimcheff 1998) and Phylloscopus (Bensch and Härlid 2000), but the ΨCR regions of these species lack any repetitive parts. An appreciable degree of sequence identity is found between Buteo and Falco in the nonrepetitive sections (fig. 2) of the CR (71.4%, 691 bp; uncorrected, gaps = 5 excluded) and the ΨCR (58.2%, 189 bp; gaps = 5 excluded). No sequence similarity was detectable in an intragenomic comparison between CR and ΨCR of Buteo. Some structural features of the nc sequences and their tandem-repetitive sections are compared in table 6. The majority of birds listed in this table have tandem repeats within their CRs. The CR of Buteo contains seven different types of repeats; that of Falco contains only two. The mt genome of Buteo has the largest percentage of tandem repetitive sequences (9.6%). We could find no sequence relationships be-
Table 2
Organization of the Mitochondrial Genome of *Buteo buteo*

<table>
<thead>
<tr>
<th>Gene/Region</th>
<th>Start Position</th>
<th>End Position</th>
<th>Start Codon</th>
<th>End Codon</th>
<th>bp</th>
<th>aa</th>
<th>Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA-Phe</td>
<td>1</td>
<td>70</td>
<td>—</td>
<td>—</td>
<td>70</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>t2SrRNA</td>
<td>71</td>
<td>1,042</td>
<td>—</td>
<td>—</td>
<td>972</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Val</td>
<td>1,043</td>
<td>1,113</td>
<td>—</td>
<td>—</td>
<td>71</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>t6SrRNA</td>
<td>1,114</td>
<td>2,711</td>
<td>—</td>
<td>—</td>
<td>1,598</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Leu-UUR</td>
<td>2,712</td>
<td>2,785</td>
<td>—</td>
<td>—</td>
<td>74</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>nd1</td>
<td>2,795</td>
<td>3,772</td>
<td>ATG</td>
<td>AGG</td>
<td>978</td>
<td>325</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Ile</td>
<td>3,771</td>
<td>3,842</td>
<td>—</td>
<td>—</td>
<td>72</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Gln</td>
<td>3,926</td>
<td>3,856</td>
<td>—</td>
<td>—</td>
<td>71</td>
<td>—</td>
<td>L</td>
</tr>
<tr>
<td>tRNA-Met</td>
<td>3,926</td>
<td>3,994</td>
<td>—</td>
<td>—</td>
<td>69</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>nd2</td>
<td>3,995</td>
<td>5,035</td>
<td>ATG</td>
<td>TAG</td>
<td>1,041</td>
<td>346</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Trp</td>
<td>5,034</td>
<td>5,105</td>
<td>—</td>
<td>—</td>
<td>72</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Ala</td>
<td>5,175</td>
<td>5,107</td>
<td>—</td>
<td>—</td>
<td>69</td>
<td>—</td>
<td>L</td>
</tr>
<tr>
<td>tRNA-Asn</td>
<td>5,250</td>
<td>5,250</td>
<td>—</td>
<td>—</td>
<td>73</td>
<td>—</td>
<td>L</td>
</tr>
<tr>
<td>tRNA-Cys</td>
<td>5,250</td>
<td>5,253</td>
<td>—</td>
<td>—</td>
<td>67</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Tyr</td>
<td>5,319</td>
<td>5,319</td>
<td>—</td>
<td>—</td>
<td>71</td>
<td>—</td>
<td>L</td>
</tr>
<tr>
<td>cox1</td>
<td>5,391</td>
<td>6,941</td>
<td>GTG</td>
<td>AGG</td>
<td>1,551</td>
<td>516</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Ser-UCN</td>
<td>7,006</td>
<td>6,933</td>
<td>—</td>
<td>—</td>
<td>74</td>
<td>—</td>
<td>L</td>
</tr>
<tr>
<td>tRNA-Asp</td>
<td>7,011</td>
<td>7,079</td>
<td>—</td>
<td>—</td>
<td>69</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>cox2</td>
<td>7,682</td>
<td>7,765</td>
<td>ATG</td>
<td>TAA</td>
<td>684</td>
<td>227</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Lys</td>
<td>7,767</td>
<td>7,834</td>
<td>—</td>
<td>—</td>
<td>68</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>atp8</td>
<td>7,836</td>
<td>8,003</td>
<td>ATG</td>
<td>TAA</td>
<td>168</td>
<td>55</td>
<td>H</td>
</tr>
<tr>
<td>atp6</td>
<td>7,994</td>
<td>8,677</td>
<td>ATG</td>
<td>TAA</td>
<td>684</td>
<td>227</td>
<td>H</td>
</tr>
<tr>
<td>cox3</td>
<td>8,577</td>
<td>9,460</td>
<td>ATG</td>
<td>TAA</td>
<td>784</td>
<td>261</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Gly</td>
<td>9,461</td>
<td>9,529</td>
<td>—</td>
<td>—</td>
<td>69</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>nd3</td>
<td>9,530</td>
<td>9,880</td>
<td>ATC</td>
<td>TAG</td>
<td>351</td>
<td>116</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Arg</td>
<td>9,881</td>
<td>9,954</td>
<td>—</td>
<td>—</td>
<td>70</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>nd4</td>
<td>9,956</td>
<td>10,252</td>
<td>ATG</td>
<td>TAA</td>
<td>297</td>
<td>98</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-His</td>
<td>10,246</td>
<td>11,623</td>
<td>ATG</td>
<td>T++</td>
<td>1,378</td>
<td>459</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Ser-AGY</td>
<td>11,569</td>
<td>11,759</td>
<td>—</td>
<td>—</td>
<td>70</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Leu-CUN</td>
<td>11,760</td>
<td>11,830</td>
<td>—</td>
<td>—</td>
<td>66</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>nd5</td>
<td>11,831</td>
<td>13,648</td>
<td>GTG</td>
<td>TAA</td>
<td>1,818</td>
<td>605</td>
<td>H</td>
</tr>
<tr>
<td>cytB</td>
<td>13,664</td>
<td>14,806</td>
<td>ATG</td>
<td>TAA</td>
<td>1,143</td>
<td>380</td>
<td>H</td>
</tr>
<tr>
<td>tRNA-Thr</td>
<td>14,809</td>
<td>14,878</td>
<td>—</td>
<td>—</td>
<td>70</td>
<td>—</td>
<td>H</td>
</tr>
<tr>
<td>CR</td>
<td>14,879</td>
<td>16,550</td>
<td>—</td>
<td>—</td>
<td>1,672</td>
<td>—</td>
<td>L</td>
</tr>
<tr>
<td>tRNA-Pro</td>
<td>16,620</td>
<td>16,551</td>
<td>—</td>
<td>—</td>
<td>70</td>
<td>—</td>
<td>L</td>
</tr>
<tr>
<td>nd6</td>
<td>17,145</td>
<td>16,627</td>
<td>ATG</td>
<td>TAG</td>
<td>519</td>
<td>172</td>
<td>L</td>
</tr>
<tr>
<td>tRNA-Glu</td>
<td>17,149</td>
<td>17,219</td>
<td>—</td>
<td>—</td>
<td>71</td>
<td>—</td>
<td>L</td>
</tr>
<tr>
<td>ΨCR</td>
<td>17,220</td>
<td>18,674</td>
<td>—</td>
<td>—</td>
<td>1,455</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

a H = heavy strand; L = light strand.

Table 3
Usage of Start and Stop Codons

<table>
<thead>
<tr>
<th></th>
<th>But</th>
<th>Fal</th>
<th>Ayt</th>
<th>Smi</th>
<th>Vid</th>
<th>Cor</th>
<th>Rhe</th>
<th>Str</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATG</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>GTG</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ATC</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>ATA</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>ATT</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Stop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAA</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAG</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>AGG</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>AGA</td>
<td>2</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Note.—All = Alligator mississippiensis; Ayt = Aythya americana; But = Buteo buteo; Cie = Ciconia ciconia; Cor = Corvus frugilegus; Fal = Falco peregrinus; Gal = Gallus gallus; Rhe = Rhea americana; Smi = Smithornis sharpei; Str = Struthio camelus; Vid = Vidua chalybeata.

twix between the repeat units in the CR and the ΨCR within or between the mt genomes of *Buteo* and *Falco*. In both species, the largest repetitive block is located in the ΨCR (table 6), but the repeat units differ considerably in length (48 bp vs. 27 bp). We sequenced the ΨCRs of two other accipitrids (*Aquila chrysaetos* and *Haliaeetus albicilla*) which had the same gene order as *Buteo* and *Falco*. Although the repeat units in these species resemble that of *Buteo* in length (*Aquila*, 49 bp; *Haliaeetus*, 48 bp), there is no apparent relatedness among them at the sequence level.

An alignment of the nonrepetitive parts of the CRs of *Buteo*, *Falco*, and *Ciconia* (not shown) revealed several conserved sections (fig. 2). Some of them show similarities to conserved sequence blocks (CSBs) of possible functional importance (CSB1, E box, D box) which have previously been described by Walberg and Clayton (1981), Baker and Marshall (1997), Randi and Lucchini (1998), and Clayton (1991). The section designated “C stretch” is located close to the 5′ end and corresponds to the “goose hairpin” in other avian species (Quinn and Wilson 1993; Randi and Lucchini 1998), which con-
sists of a stem of seven complementary C’s/G’s and a loop containing a TCCG motif that may be involved in H-strand termination (Dufresne, Mignotte, and Guéride 1996). Yet, in the CRs of *Buteo*, *Falco*, and *Ciconia*, the C stretch is not followed by a G stretch, and thus the motif lacks the ability to form a hairpin structure. The predicted TCCG stop motif within the C stretch is found in *Buteo* and *Falco*, whereas in *Ciconia* the corresponding motif is ACCC. Downstream of the C stretch, a 41-bp section corresponding to the consensus sequence of the mammalian ETAS1 (extended termination-associated sequence) described by Sbisà et al. (1997) is found in all three species, but there are no sequences capable of forming hairpin structures. The *Buteo* ETAS1 contains two termination-associated sequence (TAS) motifs, TACAT and TATAT, whereas in *Falco* and *Ciconia* such motifs are absent. Five additional CSBs are found in the three species: (1) a 34-bp section that includes a sequence similar to the E box, although this motif is less conserved than the rest of the section; (2) a 25-bp stretch with high similarity to the D box; (3) a 40-bp section, designated CSB-a, which has no similarity to other described CSBs; (4) a 31-bp section, designated CSB-b, with no similarity to other CSBs (in *Buteo*, the section surrounding CSB-b is duplicated [97-bp repeat; table 6], and therefore this motif

### Table 4
Spacer and Overlaps

<table>
<thead>
<tr>
<th>Genes</th>
<th>But</th>
<th>Fal</th>
<th>Gal</th>
<th>Ayt</th>
<th>Smi</th>
<th>Vid</th>
<th>Cor</th>
<th>Rhe</th>
<th>Str</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA-Leu-UUR</td>
<td>9</td>
<td>15</td>
<td>9</td>
<td>4</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>11</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>nd1</td>
<td>−2</td>
<td>16</td>
<td>−2</td>
<td>−2</td>
<td>21</td>
<td>6</td>
<td>8</td>
<td>14</td>
<td>−2</td>
<td>−1</td>
</tr>
<tr>
<td>tRNA-Ile</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>5</td>
<td>5</td>
<td>14</td>
<td>11</td>
<td>−2</td>
</tr>
<tr>
<td>tRNA-Gln</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>12</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
</tr>
<tr>
<td>tRNA-Met</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>nd2</td>
<td>−2</td>
<td>−1</td>
<td>−2</td>
<td>−2</td>
<td>−1</td>
<td>2</td>
<td>−2</td>
<td>−2</td>
<td>−2</td>
<td>−4</td>
</tr>
<tr>
<td>tRNA-Trp</td>
<td>1</td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>tRNA-Ala</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>tRNA-Ser</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>−6</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>−4</td>
<td>−4</td>
<td>4</td>
</tr>
<tr>
<td>tRNA-Lys</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>atp8</td>
<td>−10</td>
<td>−10</td>
<td>−10</td>
<td>−10</td>
<td>−10</td>
<td>−10</td>
<td>−10</td>
<td>−10</td>
<td>−10</td>
<td>−10</td>
</tr>
<tr>
<td>atp6</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>9</td>
<td>5</td>
<td>7</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
</tr>
<tr>
<td>cox3</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>tRNA-Asp</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>11</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>cox2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>tRNA-Cys</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
</tr>
<tr>
<td>tRNA-Tyr</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>cox1</td>
<td>−9</td>
<td>−9</td>
<td>−9</td>
<td>−9</td>
<td>−9</td>
<td>−9</td>
<td>−9</td>
<td>−9</td>
<td>−9</td>
<td>−5</td>
</tr>
<tr>
<td>tRNA-Ser-UCN</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>tRNA-Gly</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>11</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>nd3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>tRNA-Arg</td>
<td>1</td>
<td>1</td>
<td>−6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>nd4L</td>
<td>−7</td>
<td>−7</td>
<td>−7</td>
<td>−7</td>
<td>−7</td>
<td>−7</td>
<td>−7</td>
<td>−7</td>
<td>−7</td>
<td>−7</td>
</tr>
<tr>
<td>tRNA-Asp</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>11</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>tRNA-His</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>tRNA-Ser-AGY</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
<td>−1</td>
</tr>
<tr>
<td>tRNA-Leu-CUN</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>nd5</td>
<td>15</td>
<td>5</td>
<td>4</td>
<td>−1</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>−5</td>
<td>−5</td>
</tr>
<tr>
<td>tRNA-Pro</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>tRNA-Asp</td>
<td>2</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>26</td>
<td>9</td>
<td>15</td>
<td>10</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>nd6</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>−3</td>
<td>−2</td>
<td>−1</td>
<td>−2</td>
<td>−3</td>
<td>−2</td>
<td>−2</td>
</tr>
<tr>
<td>Total of spacers</td>
<td>67</td>
<td>101</td>
<td>46</td>
<td>36</td>
<td>109</td>
<td>103</td>
<td>84</td>
<td>82</td>
<td>57</td>
<td>135</td>
</tr>
<tr>
<td>Total of overlaps</td>
<td>−33</td>
<td>−31</td>
<td>−33</td>
<td>−35</td>
<td>−28</td>
<td>−29</td>
<td>−29</td>
<td>−32</td>
<td>−34</td>
<td>−26</td>
</tr>
</tbody>
</table>

**Note.**—The lengths of spacers between adjacent tRNA or protein-coding genes are given in base pairs. The numbers relate to the spacers at the 3' ends of the respective genes. Negative numbers represent overlaps; † indicates exact fit. Spacers flanked by tRNA genes or noncoding regions were excluded because their limits could not be determined unambiguously. All = Alligator mississippiensis; Ayt = Aythya americana; But = Buteo buteo; Cie = Ciconia ciconia; Cor = Corvus frugilegus; Fal = Falco peregrinus; Gal = Gallus gallus; Rhe = Rhea americana; Smi = Smithornis sharpei; Str = Struthio camelus; Vid = Vidua chalybea.

## Table 5
Different Gene Orders in Bird and the Alligator

<table>
<thead>
<tr>
<th></th>
<th>But</th>
<th>His</th>
<th>Ser</th>
<th>Leu</th>
<th>ND5</th>
<th>Cyb</th>
<th>Thr</th>
<th>CR</th>
<th>Pro</th>
<th>ND6</th>
<th>Glu</th>
<th>ΨCR</th>
<th>Phe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>But</td>
<td>His</td>
<td>Ser</td>
<td>Leu</td>
<td>ND5</td>
<td>Cyb</td>
<td>Thr</td>
<td>CR</td>
<td>Pro</td>
<td>ND6</td>
<td>Glu</td>
<td>nc</td>
<td>CR</td>
</tr>
<tr>
<td>Fal</td>
<td>Fal</td>
<td>His</td>
<td>Ser</td>
<td>Leu</td>
<td>ND5</td>
<td>Cyb</td>
<td>Thr</td>
<td>CR</td>
<td>Pro</td>
<td>ND6</td>
<td>Glu</td>
<td>nc</td>
<td>Phe</td>
</tr>
<tr>
<td>Smi</td>
<td>Smi</td>
<td>His</td>
<td>Ser</td>
<td>Leu</td>
<td>ND5</td>
<td>Cyb</td>
<td>Thr</td>
<td>CR</td>
<td>Pro</td>
<td>ND6</td>
<td>Glu</td>
<td>nc</td>
<td>Phe</td>
</tr>
<tr>
<td>Gal</td>
<td>Gal</td>
<td>His</td>
<td>Ser</td>
<td>Leu</td>
<td>ND5</td>
<td>Cyb</td>
<td>Thr</td>
<td>—</td>
<td>Pro</td>
<td>ND6</td>
<td>Glu</td>
<td>CR</td>
<td>Phe</td>
</tr>
<tr>
<td>All</td>
<td>All</td>
<td>Ser</td>
<td>His</td>
<td>Leu</td>
<td>ND5</td>
<td>ND6</td>
<td>Glu</td>
<td>Cyb</td>
<td>Thr</td>
<td>Pro</td>
<td>—</td>
<td>Phe</td>
<td>CR</td>
</tr>
</tbody>
</table>

**Note.**—Only the rearranged section is depicted. Genes coding for tRNAs are named according to their corresponding amino acids. All = Alligator mississippiensis; But = Buteo buteo; Fal = Falco peregrinus; Gal = Gallus gallus; Smi = Smithornis sharpei.
The phylogenetic analysis was carried out using the coding regions of Buteo and the nine bird species listed above and Alligator. Each gene was aligned separately after excluding nonmatching positions at the length-variable 3' ends. Although the Alligator sequence was considerably diverged from the avian sequences, a reasonable alignment could be achieved. In a distance matrix (not shown) calculated from the concatenated coding sequences (15,742 bp), the distances (HKY85; Hasegawa, Kishino, and Yano 1985) between all ingroup taxa and Alligator, respectively, are rather homogenous, indicating no major differences in substitution rates. The distances among ingroup taxa range from 18.1% to 27.2%. The distance between Buteo and Falco (21.3%) is on the same order of magnitude as that between the two struthioniform species Rhea and Struthio (19.6%) and between the two galloanserine species Gallus and Aythya (22.3%). Concerning the phylogenetic relationships of Buteo, it is remarkable that with 11 out of 14 genes compared, Buteo appears more closely related to Ciconia than to Falco (table 7), an affiliation that does not conform with classical taxonomy. The divergence of the total coding sequence is more extensive between Buteo and Falco than between Buteo and Ciconia ($\chi^2 = 34.86$, df = 1, $P < 0.001$).

Table 7 shows the uncorrected distances (DNA and proteins) between pairs of related species calculated for different sections of the mt genome. In general, for each region, the values observed for the six species pairs are in the same range. In the Buteo/Falco comparison, the most conserved regions are the tRNAs (with respect to 16S rRNA: $\chi^2 = 7.55$, df = 1, $P < 0.01$; with respect to cox1: $\chi^2 = 6.72$, df = 1, $P < 0.01$). Among the protein-coding genes, cox(1-3) + cyt b are more conserved than nd(1-6) + ATP(8+6) ($\chi^2 = 46.63$, df = 1, $P < 0.001$). Among amino acid sequences, COX(1-3) + CYTb are also more conserved than ND(1-6) + ATP(8+6) ($\chi^2 = 93.18$, df = 1, $P < 0.001$).

An MP dendrogram based on the complete coding sequence (15,742 bp) is depicted in figure 3. Three species pairs are supported by high bootstrap values: Gallus/Aythya, Rhea/Struthio, and Corvus/Vidua. In the cluster of the three ciconiiform species, Buteo and Falco appear as sister taxa, although with weak bootstrap support. The passeriform split is at the base of the tree, yet Passeriformes do not form a monophyletic group, since Smithornis branches off as the most basal bird taxon. The corresponding NJ dendrogram (not shown) resembles the MP dendrogram with two exceptions: Buteo clusters with Ciconia (bootstrap value = 97), and Smithornis is not the sister group to all other ingroup taxa. Instead, there is a trichotomy of the three lineages (Smithornis/Corvus/Vidua) remaining avian taxa). In both the MP and the NJ dendrograms, the passeriform taxa are placed at the base of the birds, and the two pairs Gallus/Aythya and Rhea/Struthio are grouped as one clade that is the sister group of the cluster Buteo/Falco/Ciconia. Dendrograms were also calculated from deduced protein sequences and from DNA sequences using transversions only (data not shown). However, in general, better resolution and stronger bootstrap support were obtained with DNA sequences that included all substitutions, so those were used for subsequent phylogenetic analysis.

To estimate in detail which clusterings were supported by different sections of the mt genome, MP analyses were performed for each gene separately, as well as
## Table 6

<table>
<thead>
<tr>
<th>Species</th>
<th>Total bp</th>
<th>NC bp</th>
<th>PCR bp</th>
<th>CR bp</th>
<th>SPACER bp</th>
<th>Tandem Repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buteo</td>
<td>18,674</td>
<td>171</td>
<td>1,650</td>
<td>1,455</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
</tr>
<tr>
<td>Falco</td>
<td>18,068</td>
<td>141</td>
<td>1,570</td>
<td>1,481</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
</tr>
<tr>
<td>Ciconia</td>
<td>17,347</td>
<td>101</td>
<td>1,510</td>
<td>1,441</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
</tr>
<tr>
<td>Aythya</td>
<td>16,714</td>
<td>67</td>
<td>1,570</td>
<td>1,441</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
</tr>
<tr>
<td>Smithornis</td>
<td>16,484</td>
<td>67</td>
<td>1,570</td>
<td>1,441</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
</tr>
<tr>
<td>Vidua</td>
<td>16,846</td>
<td>67</td>
<td>1,570</td>
<td>1,441</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
<td>13 repeats of 11 bp (one incomplete repeat of 8 bp)</td>
</tr>
</tbody>
</table>

**Discussion**

**Gene Order and Noncoding Regions**

The sequence data revealed that *Buteo* and *Falco* share the same mt gene order, one which is different from that of most other bird species. This difference could have arisen in the common ancestor of Falconidae and Accipitridae prior to their split, or it could have arisen independently, as has been postulated for other
Table 7  
Sequence Divergences of Mitochondrial Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>But/Fal</th>
<th>But/Cic</th>
<th>Cor/Vid</th>
<th>Rhe/Str</th>
<th>Gal/Ayt</th>
<th>Birds/All</th>
<th>bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>12sRNA</td>
<td>16.6</td>
<td>12.9</td>
<td>11.1</td>
<td>11.3</td>
<td>14.6</td>
<td>29.9</td>
<td>1,036</td>
</tr>
<tr>
<td>16sRNA</td>
<td>15.7</td>
<td>15.2</td>
<td>8.0</td>
<td>13.0</td>
<td>17.4</td>
<td>28.1</td>
<td>1,717</td>
</tr>
<tr>
<td>tRNAs</td>
<td>12.4</td>
<td>10.8</td>
<td>8.7</td>
<td>8.4</td>
<td>11.7</td>
<td>28.8</td>
<td>1,597</td>
</tr>
<tr>
<td>atp(8+6)</td>
<td>22.4</td>
<td>21.4</td>
<td>19.8</td>
<td>18.9</td>
<td>21.2</td>
<td>35.5</td>
<td>842</td>
</tr>
<tr>
<td>cox1</td>
<td>14.6</td>
<td>12.2</td>
<td>13.1</td>
<td>16.6</td>
<td>16.3</td>
<td>22.1</td>
<td>1,551</td>
</tr>
<tr>
<td>cox2</td>
<td>16.8</td>
<td>13.9</td>
<td>19.7</td>
<td>16.7</td>
<td>16.7</td>
<td>32.9</td>
<td>690</td>
</tr>
<tr>
<td>cox3</td>
<td>16.4</td>
<td>13.1</td>
<td>17.1</td>
<td>18.2</td>
<td>14.5</td>
<td>25.2</td>
<td>764</td>
</tr>
<tr>
<td>cytb</td>
<td>17.0</td>
<td>15.5</td>
<td>17.3</td>
<td>16.2</td>
<td>17.1</td>
<td>31.9</td>
<td>1,146</td>
</tr>
<tr>
<td>nd1</td>
<td>18.4</td>
<td>19.4</td>
<td>19.3</td>
<td>20.9</td>
<td>23.2</td>
<td>32.4</td>
<td>983</td>
</tr>
<tr>
<td>nd2</td>
<td>24.6</td>
<td>21.2</td>
<td>22.5</td>
<td>22.0</td>
<td>25.1</td>
<td>45.3</td>
<td>1,041</td>
</tr>
<tr>
<td>nd3</td>
<td>14.8</td>
<td>14.8</td>
<td>20.2</td>
<td>16.5</td>
<td>22.8</td>
<td>33.0</td>
<td>352</td>
</tr>
<tr>
<td>nd(4L+4)</td>
<td>22.2</td>
<td>15.9</td>
<td>18.2</td>
<td>20.6</td>
<td>22.4</td>
<td>39.1</td>
<td>1,671</td>
</tr>
<tr>
<td>nd5</td>
<td>21.3</td>
<td>16.9</td>
<td>19.0</td>
<td>19.0</td>
<td>23.9</td>
<td>40.2</td>
<td>1,818</td>
</tr>
<tr>
<td>nd6</td>
<td>24.5</td>
<td>25.2</td>
<td>25.9</td>
<td>27.1</td>
<td>23.4</td>
<td>51.5</td>
<td>528</td>
</tr>
<tr>
<td>Total coding</td>
<td>18.3</td>
<td>15.8</td>
<td>16.8</td>
<td>17.8</td>
<td>19.1</td>
<td>33.3</td>
<td>15,742</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein</th>
<th>But/Fal</th>
<th>But/Cic</th>
<th>Cor/Vid</th>
<th>Rhe/Str</th>
<th>Gal/Ayt</th>
<th>Bird/All</th>
<th>bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP(8+6)</td>
<td>19.1</td>
<td>17.4</td>
<td>14.9</td>
<td>11.0</td>
<td>14.9</td>
<td>38.5</td>
<td>282</td>
</tr>
<tr>
<td>COX(1–3)</td>
<td>6.8</td>
<td>3.7</td>
<td>6.7</td>
<td>2.9</td>
<td>4.9</td>
<td>20.7</td>
<td>1,007</td>
</tr>
<tr>
<td>CYTb</td>
<td>8.2</td>
<td>7.4</td>
<td>12.4</td>
<td>6.3</td>
<td>8.4</td>
<td>32.4</td>
<td>380</td>
</tr>
<tr>
<td>ND(1–6)</td>
<td>18.5</td>
<td>14.9</td>
<td>15.0</td>
<td>13.0</td>
<td>19.0</td>
<td>46.8</td>
<td>2,127</td>
</tr>
<tr>
<td>All proteins</td>
<td>14.4</td>
<td>11.4</td>
<td>12.5</td>
<td>9.5</td>
<td>13.9</td>
<td>37.8</td>
<td>3,796</td>
</tr>
</tbody>
</table>

**Fig. 3.**—Single most-parsimonious tree (tree length = 21,916, consistency index = 0.588) for complete mt coding sequences of 10 avian taxa (15,742 bp) generated with a heuristic search using the tree bisection-reconnection (TBR) algorithm and a random-taxon-addition sequence. Gaps were treated as “missing,” and all characters were weighted equally. Alligator was used as the outgroup. Bootstrap values (%, 100 replicates) are given at the nodes. Branch lengths are proportional to nucleotide substitutions. The bar indicates 500 substitutions.

**Table 7**  
Sequence Divergences of Mitochondrial Genes

As a possible mechanism for the rearrangement in *Falco*, a tandem duplication of the entire region including the CR and tRNA\(^{\text{Glu}}\) with subsequent deletion of duplicated sequences except parts of the CR has been assumed (Mindell, Sorenson, and Dimcheff 1998). While in *Falco* the sequence similarity between CR and \(\Psi\)CR corroborates this hypothesis, in *Buteo* no apparent sequence homology can be detected between the CR and the \(\Psi\)CR. The CRs of *Buteo* and *Falco* can be well aligned, at least in the conserved parts, whereas no reasonable alignment could be achieved between the \(\Psi\)CR of *Buteo* and the corresponding \(\Psi\)CR region of *Falco*. Therefore, it cannot be decided from sequence divergence alone whether the similar rearrangements in *Buteo* and *Falco* are a shared derived character, nor can we trace unequivocally the events (duplication or transposition) that caused the rearrangements. Nevertheless, a strong argument in favor of a common origin comes from structural similarities between the \(\Psi\)CRs of *Buteo* and *Falco*. Both sequences are composed of a short non-repetitive section followed by a long stretch of tandem repeats. The same structure has been found within the \(\Psi\)CRs of two other accipitrids, *A. chrysaetos* (Masuda et al. 1998) and *H. albicipilla* (present study). In contrast, *Smithornis* (Mindell, Sorenson, and Dimcheff 1998) and *Phylloscopus* (Bensch and Härld 2000), two passeriform species that have presumably acquired the same gene order independently, do not possess any repetitive sequences in the region corresponding to the \(\Psi\)CR in *Buteo*. The sequences of the \(\Psi\)CR sections of other species (belonging to the Cuculiformes, Piciformes, and Passeriformes) with rearranged gene orders (described by Mindell, Sorenson, and Dimcheff 1998) have not yet been published. Thus, an array of tandem repeats in the \(\Psi\)CR section of the mt genome has been detected so far only in falconid and accipitrid species.

**Relationships Among Buteo, Falco, and Ciconia**

Although we included all avian mt genomes available, our main interest in this study was the group *Buteo/Falco/Ciconia*. Nevertheless, despite the huge amount of DNA sequence, the relationships within this triad were not clearly resolved. Whereas in the distance-based dendrogram of the complete coding sequence *Buteo* clusters with *Ciconia*, the MP analysis yielded the
Table 8
Comparison of Bootstrap Values

<table>
<thead>
<tr>
<th>Gene</th>
<th>But/Fal</th>
<th>But/Cic</th>
<th>Cor/Vid</th>
<th>Rhe/Str</th>
<th>Gal/Ayt</th>
<th>bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>12SrRNA</td>
<td>—</td>
<td>54</td>
<td>100</td>
<td>92</td>
<td>—</td>
<td>1,036</td>
</tr>
<tr>
<td>16SrRNA</td>
<td>92</td>
<td>—</td>
<td>100</td>
<td>—</td>
<td>52</td>
<td>1,717</td>
</tr>
<tr>
<td>tRNAs</td>
<td>88</td>
<td>—</td>
<td>99</td>
<td>98</td>
<td>79</td>
<td>1,597</td>
</tr>
<tr>
<td>atp8 + 6</td>
<td>—</td>
<td>—</td>
<td>65</td>
<td>95</td>
<td>—</td>
<td>842</td>
</tr>
<tr>
<td>cox1</td>
<td>—</td>
<td>—</td>
<td>93</td>
<td>54</td>
<td>—</td>
<td>1,551</td>
</tr>
<tr>
<td>cox2</td>
<td>—</td>
<td>—</td>
<td>82</td>
<td>76</td>
<td>—</td>
<td>690</td>
</tr>
<tr>
<td>cox3</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>784</td>
</tr>
<tr>
<td>cytb</td>
<td>—</td>
<td>—</td>
<td>68</td>
<td>68</td>
<td>—</td>
<td>1,146</td>
</tr>
<tr>
<td>nd1</td>
<td>—</td>
<td>92</td>
<td>60</td>
<td>—</td>
<td>—</td>
<td>983</td>
</tr>
<tr>
<td>nd2</td>
<td>—</td>
<td>—</td>
<td>97</td>
<td>90</td>
<td>—</td>
<td>1,041</td>
</tr>
<tr>
<td>nd3</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>94</td>
<td>—</td>
<td>352</td>
</tr>
<tr>
<td>nd4L + 4</td>
<td>—</td>
<td>—</td>
<td>100</td>
<td>97</td>
<td>—</td>
<td>1,671</td>
</tr>
<tr>
<td>nd5</td>
<td>—</td>
<td>—</td>
<td>79</td>
<td>97</td>
<td>98</td>
<td>1,818</td>
</tr>
<tr>
<td>nd6</td>
<td>—</td>
<td>—</td>
<td>95</td>
<td>—</td>
<td>—</td>
<td>528</td>
</tr>
<tr>
<td>RNAS(r + t)</td>
<td>96</td>
<td>—</td>
<td>100</td>
<td>99</td>
<td>68</td>
<td>4,350</td>
</tr>
<tr>
<td>cox-(1–3)</td>
<td>—</td>
<td>—</td>
<td>99</td>
<td>99</td>
<td>79</td>
<td>3,025</td>
</tr>
<tr>
<td>nd-(1–6)</td>
<td>—</td>
<td>—</td>
<td>100</td>
<td>100</td>
<td>—</td>
<td>6,408</td>
</tr>
<tr>
<td>Total coding</td>
<td>66</td>
<td>—</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>15,742</td>
</tr>
</tbody>
</table>

NOTE.—The bootstrap values (100 replicates) are shown for nodes joining species pairs that are supposed to cluster according to traditional taxonomy (Corvus/ Vidua, Rheia/Struthio, Gallus/Aythya), as well as the pairs Buteo/Falco and Buteo/Ciconia. The maximum-parsimony dendrograms were calculated for single genes, groups of genes, and the total coding sequence. Total coding = all RNA and protein-coding sequences. All = Alligator mississippiensis; Ayth = Aythya americana; But = Buteo buteo; Cic = Ciconia ciconia; Cor = Corvus frugilegus; Fal = Falco peregrinus; Gal = Gallus gallus; Rhe = Rheia americana; Smi = Smithornis sharpei; Str = Struthio camelus; Vid = Vidua chalybeata.

expected group *Buteo/Falco*. Saturation effects are not the only explanation for the poor resolution, because the topology of the rest of the tree (with the exception of the position of *Smithornis*) seems to be unambiguous. It is more likely that unequal substitution rates among *Buteo*, *Falco*, and *Ciconia* are the reason for the conflicting topologies. Another possibility might be that the radiation of the lineages leading to *Buteo*, *Falco*, and *Ciconia* occurred within a relatively short time frame. The fossil evidence for the first appearance of the three lineages is rather vague: the earliest storklike fossil, *Palaeoepiphippia/rhynchus*, stems from the early Oligocene (37–26 MYA) of Fayum, Egypt (Olson 1985). The oldest accipitrid fossils are also from early Oligocene deposits in France. These are thought to be *Buteo*-like (Newton and Olsen 1990), but according to Feduccia (1996), these remains are in need of revision. Falconids have been reported from 55 MYA (del Hoyo, Elliott, and Sargatal 1994) and 48 MYA (Peters 1991), respectively, but from fragmentary fossils. The first well-documented falconids, however, were described from the Eo-Oligocene in France and England (Peters 1991). Therefore, from the fossil record, it is not possible to decide which of the three lineages split first. With respect to anatomical traits, storks differ from accipitrids and falconids mainly in numerous characteristics of skeletal bone structure, skull formation, and the arrangement of some muscles (Rea 1983). One exceptional behavioral trait which storks share with cathartids, but not with accipitrids and falconids, is that storks keep cool by squirting their legs with urine. Although the distance-based algorithms favor a topology with *Buteo* and *Ciconia* as sister taxa, a closer relationship between *Buteo* and *Falco* is suggested by the following arguments: In the CR, the most variable part of the mt genome, sequence similarity is stronger between *Buteo* and *Falco* (71.4%, 691 bp; uncorrected, gaps ≥5 excluded) than between either of them and *Ciconia* (Buteo/Ciconia: 66.8%; Falco/Ciconia: 69.0%, 691 bp; uncorrected, gaps ≥5 excluded). Furthermore, *Buteo* and *Falco* share, in contrast to *Ciconia*, the same derived rearrangement of the CR. Thus, the topology of the ciconiform clade in the MP dendrogram is not only in accordance with the classical taxonomic view, but is also corroborated by structural features of the mt genome.

Usefulness of Markers

Investigations of mt genes of various vertebrates by Russo, Takezaki, and Nei (1996) indicated that amino acid sequences were more informative than nucleotide sequences for reconstructing reliable trees. Our results do not conform with this, since resolution, as well as bootstrap support, decreases when amino acid sequences are used. Moreover, *nd5*, *cyt b*, and *nd4* do not appear to be the most appropriate genes in our analyses. Nevertheless, the two studies differ with respect to taxonomic level: Russo, Takezaki, and Nei (1996) analyzed vertebrate phylogeny, whereas we focused on avian evolution only. Moreover, we also included RNA sequences for our comparisons.

With respect to usefulness of marker genes, the results of our MP analyses of nucleotide sequences can be interpreted as follows: In general, dendrograms derived from single mt genes have low resolution and bootstrap support for expected species pairs. Whereas most genes support the species pairs *Corvus/Vidua* and *Rheia/Struthio*, the clades *Gallus/Aythya* and *Buteo/Falco* are found in the 16S and tRNA trees only. Therefore, none of the protein-coding genes can be recommended as reliable markers for phylogenetic studies at this taxonomic level. Among the concatenated protein-coding sequenc-
es, cox(1–3) (although only half as long) appears better than nd(1–6). Concatenated rRNAs plus tRNAs (4,350 bp) resolve all four expected species pairs and are thus as good as the complete coding sequence (15,742 bp). In a dendrogram calculated from the complete coding sequence except nd5 and 12S, the two genes supporting the Buteo/Ciconia clustering, the bootstrap value of the Buteo/Falco clade rose to 92, while the rest of the tree remained unchanged. A tree based on only 16S and tRNAs yielded high bootstrap support in the Buteo/Falco/Ciconia cluster but lower values for the other nodes. To summarize, the concatenated RNA genes (rRNAs, tRNAs) appear to be the favorable combination, although bootstrap support is lower than that for the total sequence.

Basal Relationships and Dating of Splits

The phylogeny presented in this study, which is based on MP analyses of mt genomes, is in accordance with several other studies based on whole mt genomes (Härlid, Janke, and Arnason 1997, 1998; Härlid and Arnason 1999; Mindell et al. 1999; Waddell et al. 1999) and nuclear sequences (Stapel et al. 1984; Caspers et al. 1997). It is also compatible with the results of Griffiths (1994), based on syringeal morphology. In comparison to the dendrogram of Mindell et al. (1999), three additional taxa were included in the present study: Buteo, Ciconia, and Corvus. Whereas in the maximum-likelihood tree of Mindell et al. (1999) Falco clusters with Smithornis, in our trees Smithornis never clusters with ciconiform species. Instead, the Ciconiiformes appear as a stable monophyletic group that is the sister group of the clad Galloanserinae/Ratitae. The Passeriformes split at the base of the avian tree, although in some of the dendrograms they do not appear to be a monophyletic group. Instead, Smithornis (a representative of the suboscines, which are considered the most basal passeriform group) splits off as the most basal lineage of the dendrogram. Our mt-based phylogeny contradicts those derived from other studies of mt as well as nuclear genes, in which ratites appear at the base of the avian tree (Groth and Barrowclough 1999; van Tuinen, Sibley, and Hedges 2000). This incongruity is not necessarily due to marker selection (nuclear/mitochondrial). One reason for it might be that the studies differ in taxon composition. For example, in the dendrogram presented by Groth and Barrowclough (1999) based on the nuclear gene RAG-1, no ciconiform and no suboscine species are represented. On the other hand, sequences of the mt genomes of birds included in the RAG-1 study (cranes, loons, penguins, hemipodes, shorebirds, rollers) have not been published so far. Another reason for the incongruity might be the different lengths of the marker sequences (e.g., 3-kb RAG vs. 15,742-bp mt genes). In our MP dendrogram, the branches of the passeriform taxa appear shorter. This may be due either to the fact that the Alligator sequence is not a suitable outgroup to root the tree or to a slower substitution rate in the Passeriformes. Thus, we cannot rule out the possibility that the basal position of Passeriformes might be caused by long-branch attraction of the ciconiiform and galloanserine clusters. Altogether, a reasonable comparison between nuclear and mt-derived phylogenies will be possible only when (1) representatives of more avian groups (with both mt and nuclear genes) have been analyzed and (2) more nuclear sequence data (different genes) are available.

Unfortunately, the fossil record allows neither corroboration nor falsification of our data concerning the position of passeriformes. When the reports about the first appearances of avian groups are compared, no clear conclusions about the succession of splits or divergence times of the various lineages can be drawn. For example, according to Feduccia (1996, p. 166), the oldest fossils of Ratitae are from the Paleocene (65–53 MYA), and according to Houde and Haubold (1987), the oldest ostrich fossils are from the early Eocene (53–37 MYA), the epoch from which the oldest putative passeriform fossils (Boles 1995) as well as falconids (see above) are also dated. As the oldest neognathous fossils are from the Mesozoic (Olson 1992), there is at least no support from the fossil record for the hypothesis that the Pa
eognathae represent the most basal lineage.

Faced with the incomplete fossil record for birds of prey, the dating of divergences has to rely on a molecular approach. The following molecular datings are available. For the Rheal/Gallus split, Härlid, Janke, and Arnason (1997) calculated 80–90 MYA, and Waddell et al. (1999) calculated 92 MYA. For the divergence of Aythyg/Gallus, 68 MYA has been estimated (Waddell et al. 1999). According to our HKY85 distances (22.3% for Aythya/Gallus, 23.7% for Galloanserinae/Struthioniformes), these two splits should be closer. From the two different reference points, the divergence of Buteo/Falco (21.3%) can be estimated to have occurred in the late Cretaceous, either at 72–83 MYA or at 65 MYA.

Supplementary Material

The complete sequence of the mt genome of B. buteo is registered under GenBank accession number AF380305. For the analysis of VCR, the following partial sequences have been determined: VCR of H. albicilla, AY034150; repeat units of VCR of A. chrysaetos, AY034151. Alignments of avian mt sequences can be viewed under “Sequences” at our web site at http://www.nhm-wien.ac.at/NHM/1Zoo/first_zoological_department/web/chemsyst/cuhp_24e.html.

Acknowledgments

We are grateful to Werner Mayer for critical discussions and comments on the manuscript and to two anonymous referees for numerous valuable suggestions. This work was supported by the Austrian Science Foundation (FWF) (project number P14069-BIO).

LITERATURE CITED


Kumar, S. 1996. PHYLTEST: a program for testing phylogenetic hypotheses. Version 2.0. Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, University Park.


Elizabeth Kellogg, reviewing editor

Accepted June 12, 2001