Molecular Phylogenetics and Evolution 61 (2011) 760–771

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenetic utility and evolution of indels: A study in neognathous birds

Łukasz Paśko [a], Per G.P. Ericson [b], Andrzej Elzanowski [c,*]

[a] Institute of Zoology, University of Wrocław, 21 Sienkiewicz Street, PL-50-335 Wrocław, Poland
[b] Swedish Museum of Natural History, P.O. Box 50007, SE-10405 Stockholm, Sweden
[c] Museum and Institute of Zoology, Polish Academy of Sciences, 64 Wilcza Street, PL-00-679 Warszawa, Poland

ARTICLE INFO

Article history:
Received 10 January 2011
Revised 28 July 2011
Accepted 30 July 2011
Available online 6 August 2011

Keywords:
Bird
Indel
Insertion
Deletion
Phylogeny
Genomics

ABSTRACT

Indels are increasingly used in phylogenetics and play a major role in genome size evolution, and yet both the phylogenetic information content of indels and their evolutionary significance remain to be better assessed. Using three presumably independently evolving nuclear gene fragments (28S rDNA, β-fibrinogen, ornithine decarboxylase) from 29 families of neognathous birds, we have obtained a topology that is in general agreement with the current molecular consensus tree, supports the monophyly of Metaves, and provides evidence for the unresolved relationships within the Charadriiformes. Based on the retrieved topology, we assess the relative impact of indels and nucleotide substitutions and demonstrate that the superposition of the two kinds of data yields a topology that could not be obtained from either data set alone. Although only two out of three gene fragments reveal the deletion bias, the combined nucleotide insertion-to-deletion ratio is 0.22, indicating a rapid decrease of intron length. The average indel fixation rate in the neognaths is 2.5 times faster than that in therian (placental) mammals of similar geologic age. As in mammals, there is a considerable variation of indel fixation rate that is 1.5 times higher in Galloanseres compared to Neoaves, and 2.4 times higher in the Rallidae compared to the average for Neoaves (8.2 times higher compared to the related Gruidae). Our results add to the evidence that indel fixation rates correlate with lineage-specific evolutionary rates.

1. Introduction

Insertions and deletions (indels) attract increasing interest because they prove to be useful in phylogenetic reconstruction and play a major role in genomic evolution. Indels have been credited with relatively low levels of homoplasy compared to nucleotide substitutions (Simmons et al., 2001), but this may be true only for the best studied indels in noncoding sequences, where most indels occur, e.g., in birds the incidence of indels in the coding sequences is only 10% of that in the noncoding DNA (Brandström and Ellegren, 2007). Coding sequence indels that affect protein structure are obviously subject to functional constraints (Wolf et al., 2007). However, at least some non-coding sequences may also be subject to selective pressures (Bird et al., 2006) and thus their indels may not be immune to homoplasy.

Different kinds of indels have been thought to differ in their sensitivity to homoplasy. First, large indels have been expected to carry more phylogenetic information (Fain and Houde, 2004), although this has earlier been disputed by Simmons et al. (2001). Second, the deletion bias alone should increase the probability of homoplasious deletions as compared to insertions by the very fact that deletions are more frequent than are insertions, and thus the latter should be given more phylogenetic weight (Edwards et al., 2005). However, only limited empirical support has been provided for the first hypothesis and none for the second.

Major progress has been achieved in resolving the relationships among extant birds (Ericson et al., 2006; Hackett et al., 2008), in particular within the Neoaves that are subdivided into a well-supported clade Coronaves and weakly supported clade Metaves. Indels proved to be important phylogenetic markers in a wide range of organisms including birds (Fain and Houde, 2004; Edwards et al., 2005; Chojnowski et al., 2008). Indels provided strong support for the monophyly of non-struthioniform paleognaths (Harshman et al., 2008), galloanserines (Fain and Houde, 2004), and Neoaves (Groth and Barrowclough, 1999; Fain and Houde, 2004). Indels in β-fibrinogen intron 7 showed a very high consistency index (CI = 0.91) in the phylogeny of Columbidae (Johnson, 2004) and four of them increased bootstrap support for the subdivison of Neoaves into Coronaves and Metaves but their segregation between these two proposed clades was not fully consistent, suggesting some homoplasy (Fain and Houde, 2004). Well-established monophyly of eight ordinal-level neoavian clades, including [kagu, sunbittern] and [swifts, hummingbirds], is supported by intron indels in the clathrin heavy chain genes (Chojnowski et al., 2008) and the monophyly of Apodiformes (Johansson et al., 2001) with

* Corresponding author.
E-mail address: elzanowski@miiz.waw.pl (A. Elzanowski).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.07.021

Aegothelidae as their basal branch (Fidler et al., 2004; Barrowclough et al., 2006) was confirmed by indels in three genes. Ericson et al. (2003) found in a myoglobin intron one indel synapomorphic for all Lari and one separating the majority of Lari from the Glareolidae. Also, indels helped establish relationships within the uniform orders of Psittaciformes (de Kloet and de Kloet, 2005) and Passeriformes (Ericson et al., 2000; Spicer and Dunipace, 2004) and even more uniform passeriform clades such as the sylvioid warblers (Alström et al., 2006).

Avian genomes have long been known to be smaller than those of most other vertebrates (Gregory, 2004a) except for the other two groups of flying amniotes, bats and pterosaurs (Organ and Shedlock, 2009; Smith and Gregory, 2009). The size of avian genome is only 30–40% of that of mammals (Brandström and Ellegren, 2007). The reduction of genome size is attributed primarily to the loss of interspersed repetitive elements and probably preceded the origin of flight in the evolution of archosaurs (Organ et al., 2007). Indels (up to 400 bp) have been implicated in the evolution of genome size (Oliver et al., 2007) but deletion bias is deemed too weak to be a primary determinant of genome size variation (Gregory, 2004b) although its impact on the avian genome has yet to be quantified. Whatever their effect on genome size, indels clearly influence gene length, primarily by shrinking or expanding introns. The net impact of indels on gene size ultimately depends on the nucleotide insertion-to-deletion (NITD) ratio that is a resultant of the frequency and the length of insertions and deletions, and thus a deletion bias alone does not necessarily lead to gene contraction. For example, there is a strong deletion bias of 3.0 in the introns of three mammalian (eutherian) nuclear genes, and yet the NITD ratio of 1.23 is much greater than unity as the insertions are on the average ca 3.7 times longer than are deletions (Matthee et al., 2007).

We sequenced three nuclear gene fragments and, using both substitution and indel data, confirmed the relationships within a major subset of the neognathous clades including the majority of landbirds, waterbirds and raptors, as recovered by Ericson et al. (2006) and Hackett et al. (2008). Subsequently, we studied the distribution of indels on a well-supported tree in order to obtain new information on their utility for phylogenetic reconstruction and their significance in genomic evolution.

2. Materials and methods

2.1. Material

We sequenced fragments of three genes: 28S rDNA, ornithine decarboxylase (ODC), and β-fibrinogen (β-FIB) in 42 species representing 29 families of neognathous birds, and one palaeognathous species as an outgroup (Table 1). Whenever possible, we sampled at least two terminal taxa from a family level group to minimize long branch attraction between distant families.

In the 28S gene we sequenced segments 13–15 that constitute its most upstream part, corresponding to nucleotides 3429 through 4200 of the all rDNA array in the map for *Mus musculus* (Hillis and Dixon, 1991; Palumbi, 1996). For amplifications we used the universal primers, 28v-5′ (28SVI)-forward and 28Sjj-3′ (28SXI)-reverse.

In the β-FIB gene we sequenced the entire (last) intron 7 with minor fragments of exons 7 and 8. For amplifications we used the primers FIB-BI7U-forward and FIB-BI7L-reverse (Prychitko and Moore, 1997). Our 926-nucleotide pre-alignment sequence of *Gallus gallus bankiva* corresponds to nucleotides 21,396,564 through 21,397,489 of *Gallus gallus* chromosome 4 (NCBI Gallus_gallus-2.1 genome map, NCBI Reference Sequence: NC_006091). In *G. gallus bankiva,* intron 7 comprises 893 nucleotides (96.4% of the obtained sequence) and the flanking exons 7 and 8 are represented only by 33 nucleotides (3.6% of the obtained sequence). The trimmed sequences of β-FIB that we used for analyses represent effectively intron 7 only.

For ODC amplifications we used the primers OD6-forward and OD8-reverse and, in many cases, an additional pair of internal primers OD6int-forward and OD8int-reverse (Allen and Omland, 2003). Our 706-nucleotide pre-alignment sequence of *Tragopan temminckii* corresponds to nucleotides 99,664,489 through 99,665,196 of *G. gallus* chromosome 3 (NCBI Gallus_gallus-2.1 genome map, NCBI Reference Sequence: NC_006090.2). In *G. gallus*, gene ODC comprises 11 exons and 10 introns (NCBI Reference Sequence: XM_419949.2). As mapped using BLAST, our sequence from *T. temminckii* comprises entire introns 5 and 6 (jointly 461 nucleotides), the ends of exons 5 (111 nucleotides) and 7 (52 nucleotides), and the entire exon 6 (82 nucleotides). Exon sequences make up 34.7% and intron sequences 65.3% of our sequenced fragment.

There seems to be some variation in the structure of the ODC gene, which confuses the numbering of exons and introns. Allen and Omland (2003) sequenced the ODC gene fragment AF491996.1 in *Icterus spurius* and were first to use it in avian phylogenetics. When mapped on chromosome 3 of the zebra finch *Taeniopygia guttata,* another passerine that has only 10 exons and 9 introns in the ODC gene (NCBI Reference Sequence: XM_002198078.1), the AF491996.1 sequence covers exons 4 through 6 and complete introns 4 and 5. However, when mapped on chromosome 3 of *G. gallus*, the AF491996.1 sequence covers exons 5 through 7 and complete introns 5 and 6. Thus, the exon and intron numbers in *I. spurius* do not correspond to those in either *G. gallus* (with the phase shift of 1) or *T. guttata* (with the phase shift of 2).

Aside from the names, most important for the phylogenetic purposes is that all sequenced fragments are orthologous rather than paralogous. There is no evidence for β-FIB gene paralogues in birds (Morgan-Richards et al., 2008) and there is only one copy of true ODC gene. The only avian ODC paralogue is ODCp/AZIB that belongs to a distant family (Ivanov et al., 2010). Our simulations using ODCp/AZIB sequence of *G. gallus* (NCBI *G. gallus*-2.1 genome map, NCBI Reference Sequence: NW_001471654.1) show, that its sequence differs in about 51% from the true ODC gene and does not bind the primers OD6 and OD8 that we used for amplification. The 28S fragment occurs in multiple copies all of which are subject to concerted evolution (Hillis and Dixon, 1991) that is slow compared to the other two genes and thus the influence of this gene on our final topology is weak. Our matrix is therefore unlikely to include data from any paralogues rather than orthologues and thus we do not expect discrepancies between the phylogenies of species and genes we used. Hence we used as a reference the best species tree (Fig. 1) rather than locus specific phylogenies.

2.2. DNA extraction, amplification and sequencing

All sequences used in this study have been obtained *de novo*. Total genomic DNA was extracted mostly from skeletal muscles or blood samples, and only in a few cases from feathers, using modifications of phenol–chloroform method (Laird et al., 1991) or Qiagen DNAeasy Tissue Kit following the manufacturer protocols. The target DNA fragments were amplified in most cases using ABI GeneAmp 9600 thermocycler and Qiagen Taq PCR Core Kit following manufacturer protocols. All three genes were subjected to (with slight modifications on the case-by-case basis) an initial denaturation (5 min at 95 °C); 35 cycles of denaturation (40 s at 95 °C), annealing (40 s at 52 °C for 28S, 57 °C for β-FIB, 59 °C for ODC), extension (60 s at 72 °C); and a final extension (8 min at 72 °C).

Table 1
Taxa, source collections[a], sequence length[b], and GenBank accession numbers.

Family	Species	28S			ODC			β-FIB		
		Coll	Length	Accession	Coll	Length	Accession	Coll	Length	Accession
Diomedeidae	*Diomedea nigripes*	USNM	652	EF552807	NRM	720	EF552718	NRM	910	EF552760
Procellariidae	*Fulmarus glacialis*	NRM	642	EF552812	NRM	701	EF552723	NRM	933	EF552765
Fregatidae	*Fregata minor*	USNM	644	EF552811	USNM	708	EF552722	USNM	915	EF552764
Sulidae	*Sula bassana*	NRM	677	EF552834	NRM	699	EF552744	NRM	938	EF552786
Anhingidae	*Anhinga anhinga*	NRM	644	EF552793	NRM	686	EF552709	NRM	899	EF552751
Pelecanidae	*Pelecanus crispus*	ZZK	730	EF552822	ZZK	711	EF552733	ZZK	972	EF552775
Balaenicipitidae	*Balaeniceps rex*	SM	637	EF552841	SM	713	EF552711	SM	868	EF552753
Scopidae	*Scopus umbretta*	ZZK	698	EF552831	NRM	712	EF552741	NRM	456	EF552783
Ardeidae	*Ardea cocoi*	NRM	639	EF552794	NRM	718	EF552710	NRM	951	EF552752
	Botaurus stellaris	ZZK	685	EF552796	ZZK	725	EF552713	ZZK	875	EF552755
	Egretta thula	NRM	639	EF552808	NRM	719	EF552719	NRM	721	EF552761
Threskiornithidae	*Eudocimus ruber*	ZZK	700	EF552809	ZZK	723	EF552720	ZZK	918	EF552762
	Harpiprion caerulescens	NRM	641	EF552815	NRM	602	EF552726	NRM	912	EF552768
	Theristicus caudatus	NRM	647	EF552836	NRM	715	EF552745	NRM	946	EF552787
Ciconiidae	*Ciconia ciconia*	ZZK	710	EF552802	ZZK	712	EF552716	ZZK	826	EF552758
	Mycteria americana	NRM	697	EF552820	NRM	709	EF552731	NRM	952	EF552773
	Jabiru mycteria	NRM	639	EF552817	NRM	706	EF552728	NRM	870	EF552770
Cathartidae	*Cathartes aura*	NRM	682	EF552800	NRM	714	EF552715	NRM	931	EF552757
	Coragyps atratus	NRM	638	EF552803	NRM	720	EF552717	NRM	949	EF552759
Accipitridae	*Accipiter nisus*	ZZK	709	EF552791	ZZK	733	EF552707	ZZK	934	EF552749
	Heterospizias meridionalis	NRM	639	EF552816	NRM	720	EF552727	NRM	931	EF552769
	Leptodon cayanensis	NRM	657	EF552819	NRM	707	EF552730	NRM	898	EF552772
Pandionidae	*Pandion haliaetus*	ZZK	663	EF552821	NRM	712	EF552732	NRM	927	EF552774
Falconidae	*Polyborus plancus*	NRM	646	EF552827	NRM	722	EF552738	NRM	855	EF552780
	Falco tinnunculus	NRM	652	EF552814	NRM	721	EF552725	NRM	784	EF552767
Phoenicopteridae	*Phoenicopterus chilensis*	NRM	650	EF552825	NRM	719	EF552736	NRM	980	EF552778
Podicipedidae	*Podiceps cristatus*	ZZK	703	EF552826	NRM	717	EF552737	ZZK	956	EF552779
Gaviidae	*Gavia arctica*	NRM	652	EF552814	NRM	721	EF552725	NRM	784	EF552767
Spheniscidae	*Spheniscus demersus*	ZZK	713	EF552832	ZZK	725	EF552742	ZZK	887	EF552784
Laridae	*Larus ridibundus*	ZZK	710	EF552818	ZZK	731	EF552729	ZZK	698	EF552771
Charadriidae	*Vanellus vanellus*	ZZK	661	EF552840	ZZK	722	EF552748	ZZK	925	EF552790
Scolopacidae	*Scolopax rusticola*	ZZK	703	EF552830	ZZK	725	EF552740	ZZK	746	EF552782
Burhinidae	*Burhinus oedicnemus*	ZZK	662	EF552798	ZZK	723	EF552714	ZZK	849	EF552756
Rallidae	*Rallus aquaticus*	ZZK	683	EF552828	ZZK	701	EF552739	ZZK	948	EF552781
	Fulica atra	ZZK	683	EF552828	ZZK	701	EF552739	ZZK	948	EF552781
Gruidae	*Balearica regulorum gibbericeps*	ZZK	643	EF552795	ZZK	681	EF552712	ZZK	822	EF552754
	Grus japonica	ZZK	643	EF552795	ZZK	681	EF552712	ZZK	822	EF552754
Phaethontidae	*Phaethon rubricauda*	NRM	686	EF552823	NRM	708	EF552734	ZZK	974	EF552776
Eurypygidae	*Eurypyga helias*	ZZK	639	EF552810	ZZK	569	EF552721	ZZK	977	EF552763
Phasianidae	*Tragopan temminckii*	ZZK	620	EF552839	ZZK	706	EF552747	ZZK	687	EF552789
	Gallus gallus bankiva	ZZK	666	EF552813	ZZK	466	EF552724	ZZK	926	EF552766
Anatidae	*Anas platyrhynchos*	ZZK	680	EF552792	ZZK	711	EF552708	ZZK	969	EF552750
Struthionidae	*Struthio camelus*	ZZK	693	EF552833	ZZK	709	EF552743	ZZK	960	EF552785

[a] Collection acronyms: SM – Forschungsinstitut und Museum Senckenberg, Frankfurt/M; ZZK – Laboratory of Vertebrate Zoology, Department of Zoology, University of Wrocław; NRM – Department of Vertebrate Zoology, Swedish Museum of Natural History, Stockholm; USNM – National Museum of Natural History, Washington, DC.
[b] The length of high quality contigs that were used for alignments and phylogeny.

Purifications after PCR were done with Qiagen QIAquick PCR Purification Kit or EURX BlueMatrix PCR/DNA Clean-Up Purification Kit. The PCR products were cycle-sequenced in both directions using the same primers, conditions and thermocyclers, and ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Kit. Total sequencing reaction volumes were typically 20 μl, containing 2 μl of cleaned PCR product, 1 μl dye-terminators, 1 μl of primer, and 4 μl of buffer solution. To remove residual dye terminators, the sequenced products were precipitated with $C_2H_5OH/MgCl_2$, following the protocol of Applied Biosystems for ABI 3100Avant. Cleaned products were re-suspended in 4 μl formamide-EDTA or ddH_20, denatured at 95 °C for 2 min., and run on ABI 3100Avant Genetic Analyzer.

Chromatographs were called for bases using ABI KB basecaller and ABI Sequence Analysis 5.1 software (default basecaller settings: no mixed bases and most probable base in each situation). Chromatographs for each taxon were assembled into contigs and edited using Aligner 3 for MacOSX (CodonCode Corporation). The majority of consensus sequences were assembled from 2 to 8 strands, which makes them very reliable as each strand has been obtained from separate sequencing reaction. Ambiguities and heterozygosities were resolved using appropriate IUPAC IUB ambiguity codes.

2.3. Alignment

The sequences were pre-aligned in Clustal W with interface BioX1.5 and with default settings (gap opening penalty 10, gap extension penalty 5, gap separation distance 8, delay divergent sequences 30%) and without additional settings. The pre-aligned sequences were trimmed and aligned in the SOAP1.1 program for multiple alignments (Löytynoja and Milinkovitch, 2001) using different combinations of gap opening (from 5 to 12 with step 1) and gap extension (from 6 to 9 with step 0.5) penalties. The obtained alignment matrix was used for a strict consensus alignment without unstable nucleotide blocs (ambiguously aligned positions) that leave room for nonhomologous matches. Consensus alignments were viewed and edited in 4SALE (Seibel et al., 2006) and submitted to subsequent analyses as nexus files. Indels were coded as binary characters using SeqState1.22 program (Müller, 2005) and the simple indel coding (sic) algorithm (Simmons and Ochoterena, 2000; Simmons et al., 2007) which is approximately as effective

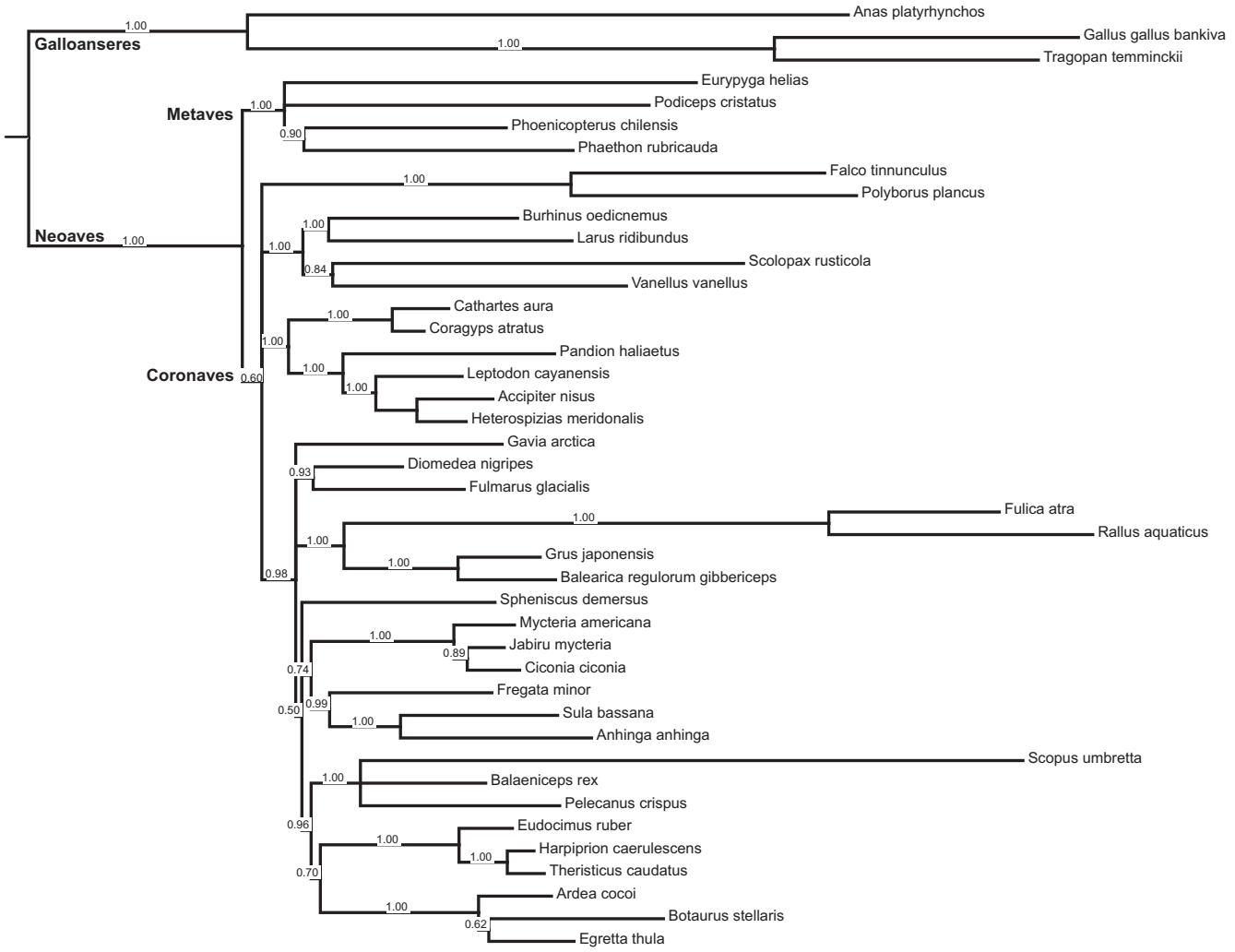

Fig. 1. Bayesian 50% majority rule tree from three substitution partitions and three indel sic-coding partitions (SUBIND). Branch lengths correspond to number of characters state changes (substitutions and indels).

as the modified multiple indel coding (Simmons et al., 2007). The sequence fragments with indels were removed prior to the Bayesian analysis of substitution partitions if a gap was present in over 50% of sequences in the alignment.

2.4. Phylogenetic analyses

We used the ostrich (*Struthio camelus*) as an outgroup. In order to avert long branch attraction, we based our phylogenetic analysis primarily on Bayesian inference (Huelsenbeck et al., 2001, 2002).Six data partitions were defined *a priori*, two for each of the three sequenced gene fragments – one for substitutions and one for indels. We did not use codon position partitioning because protein coding sequences were effectively absent from two genes (β-FIB and 28S) and represented only about one third of the ODC sequence. In order to determine the contribution of indels to the resolution of phylogeny we performed three Bayesian analyses: for all six partitions jointly (Fig. 1: SUBIND tree), for three substitution partitions alone (Fig. 2: SUB tree), and for three indel partitions alone (Fig. 3: IND tree). In addition, we performed three locus specific analyses for substitutions (Fig. S1: β -FIB tree, Fig. S2: ODC tree, Fig. S3: 28S tree).

The models of nucleotide substitutions for Bayesian analyses were selected individually for each partition using the Bayesian Information Criterion (BIC) and two Akaike Information Criterion (AIC) variants implemented in MultiPhyl program (Keane et al., 2007) that can test 56 currently available models of sequence evolution. The BIC test was considered decisive in cases of conflict. The best models for single gene partitions were HKY + I for 28S, TVM + G for β-FIB, and TVM + I + G for ODC. However, as TVM models cannot be routinely executed by any of the programs that are broadly used for Bayesian analysis, they were replaced, respectively, by GTR + G and GTR + I + G models that have one free parameter extra (in the MrBayes script these models are: 28S – nst = 2 rates = propinv; Fibr – nst = 6 rates = gamma and ODC – nst = 6 rates = invgamma). For indel partitions we applied Bayesian analysis using the model implemented in MrBayes for standard discrete data and based on the Mkv model (Lewis, 2001).

Regardless of the model used, all analyses were performed using the same logic and procedures of Bayesian inference as implemented in MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001). To reduce the chance of reaching the apparent stationarity on local optima, two separate runs consisting of four Markov chains for each analysis were performed (in every case three chains were cold and one heated, as a default in MrBayes). Each chain was performed by 20×10^6 generations and was sampled every 100 generations. The assumptions were congruent with the default settings to a random starting tree, priors constraining the same set of branch lengths for each partition, and heating values of Markov chains. Stationarity and convergence of analyses was estimated

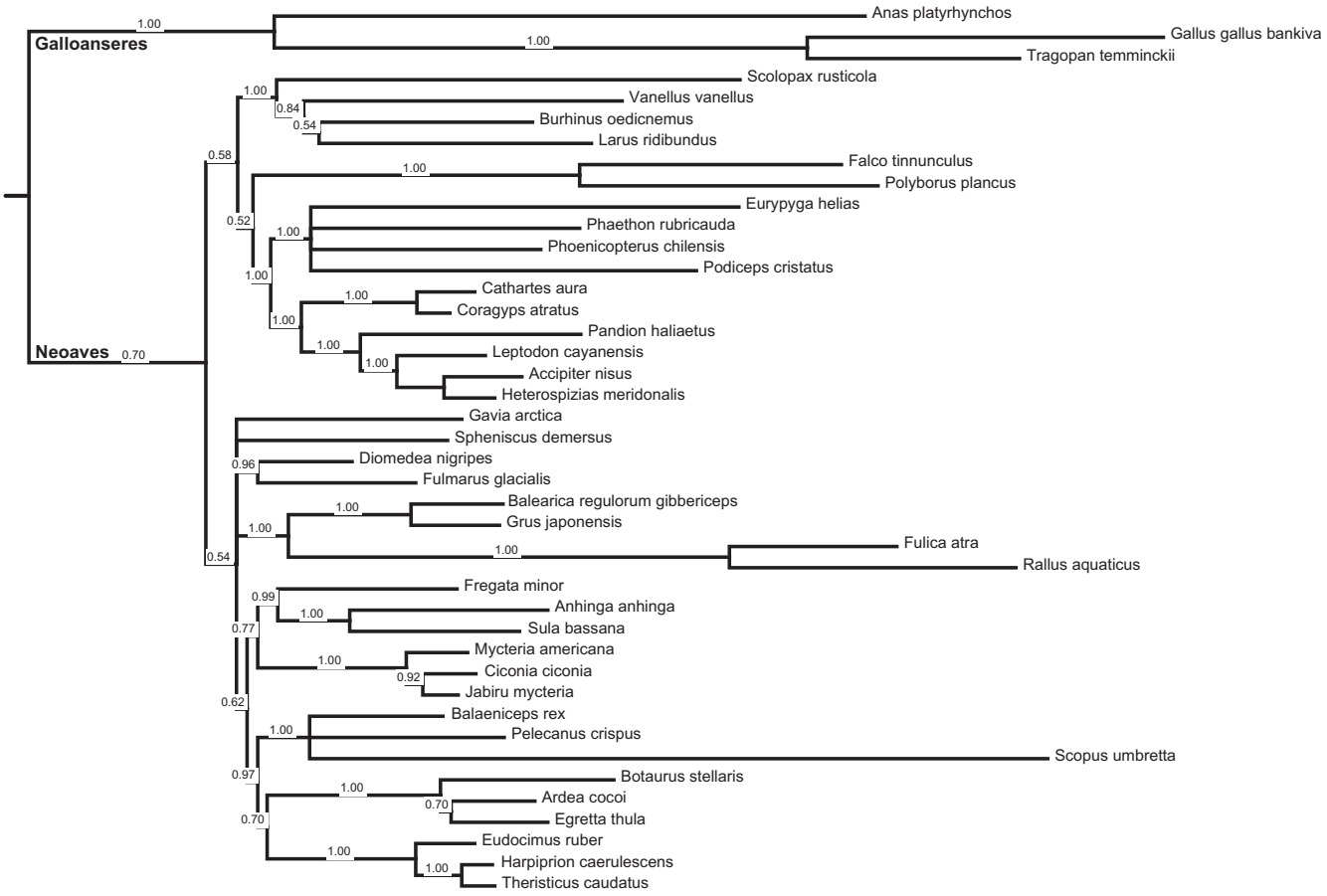

Fig. 2. Bayesian 50% majority rule tree from three substitution partitions (SUB). Branch lengths correspond to number of substitutions.

by default MrBayes statistics and graphically in Tracer (Rambaut and Drummond, 2007). Burn-in trees and parameters were discarded (5000 samples or 0.5×10^6 generations in every case) and the remaining trees and associated parameters saved, with the frequency of clades representing estimation of posterior probabilities. All phylogenetic reconstructions were generated using MrBayes 3.12 parallel version that was run on WCSS supercomputers in the operational environment of ScientificLinux.

2.5. Indel analyses

\In order to determine the significance of indels as both phylogenetic markers and indicators of genomic evolution we compared the SUB, IND, and SUBIND trees and mapped the indels onto the SUBIND tree. The MacClade 4.08 parsimony criterion (Maddison and Maddison, 1992) was used to reconstruct the number of indel events (characters), and to determine their states that is, to distinguish between insertions or deletions. Characters were assumed as unweighted, unordered, and thus reversible (Fitch parsimony). All most parsimonious reconstructions of character evolution were examined for each character. Only unambiguously reconstructed insertions and deletions (206 out of the total of 287 indels) were used in further analyses (a change was counted as unambiguous only when it occurred on the same branch in all most parsimonious reconstructions of character evolution). We determined the number of synapomorphic indels supporting each clade (Fig. 4) and calculated the levels of homoplasy as indicated by standard fit indices, consistency index (CI), retention index (RI), and rescaled consistency index (RC), that were calculated for each indel separately (single-character indices) and for all indels in a gene fragment jointly (ensemble indices). When calculating individual fit indices for polarized indels, we categorized as deletions only those indels whose all state changes were deletions, and as insertions only those indels whose all state changes were insertions. Indels with state changes including any combination of insertion(s) and deletion(s) have been excluded from the analysis. Based on the obtained figures, we determined a correlation of the length of parsimony informative indels with their fit indices (nonparametric Spearman R) and the differences in mean fit indices of such indels between the three genes (Kruskal–Wallis ANOVA or K–W test and median test).

We analyzed the entire set of 287 unpolarized indel characters, which is immune to potential errors of the accepted topology (Fig. 3), and, separately insertions and deletions in that topology. Single insertions and deletions were treated as separate indel characters when analyzing the set of polarized indels, the number of which is a sum of unambiguously polarized indels and the number of extra (more than one) times they occur (as homoplasies) in the accepted topology. We determined: (1) the numbers and length frequencies of unpolarized indels in each gene and in all three genes jointly by testing their agreement with the normal distribution (Kolmogorov–Smirnov–Lilliefors or K–S–L test, Shapiro–Wilk or S–W test); (2) the numbers and length frequencies of insertions and deletions in each gene and in all three genes jointly (K–S–L and S–W tests); (3) differences in unpolarized indel length between the three genes (K–W and median tests); (4) differences in insertion and deletion length separately between the three genes (K–W and median tests); (5) differences in length between insertions and deletions for each gene fragment (K–W and median test). All statistical tests and characterization of indels were performed using the Statistica8 (StatSoft, Inc., 2008) package.

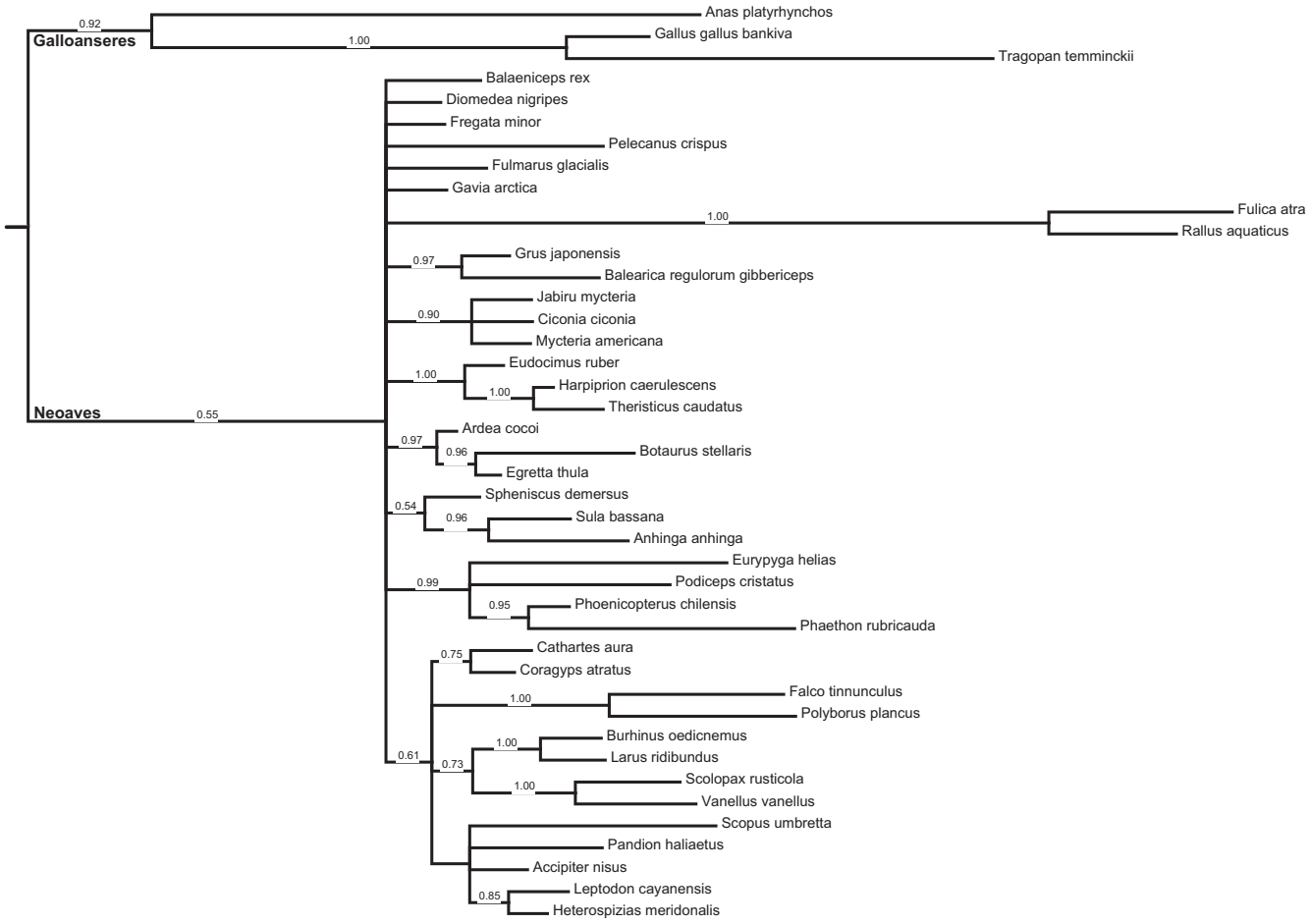

Fig. 3. Bayesian 50% majority rule tree from three indel sic-coding partitions (IND). Branch lengths correspond to number of indels. Note that the tree for Neoaves differs substantially from the best (SUBIND) tree in Fig. 1 and so does the distribution of indels from that in Fig. 4.

In an attempt to determine whether the sequence length of each gene fragment increased or decreased in avian evolution as a result of indel fixation, we compared the mean length values of insertions and deletions in each gene fragment with the confidence intervals for the experimental insertion/deletion ratios. Also, we calculated the indel fixation rates (per one site = nucleotide) in Neoaves and Galloanseres by averaging the numbers of indels from, respectively, 27 neoavian and 2 galloanserine lineages that we defined as leading to separate families. We counted the indels in Fig. 4 from the split Neoaves/Galloanseres through terminal branches. In order to avoid an accidental bias by several families represented by more than one species, we averaged the number of indels from two or three terminal branches, thus reducing their impact on the neoavian or galloanserine average to that of single-species families.

We employed correlation statistics in order to assess whether the distribution of rate heterogeneity was similar among substitutions and indels. First, we compared the branch lengths estimates from the SUB and IND consensus trees by patristic distances approach (903 compared distances in every matrix). We then calculated patristic distance matrices for the SUB and IND trees and tested the correlations between them by Pearson correlation coefficient (r) and Mantel test (100,00,000 permutations) with the help of PATRISTIC v1.0 and ZT v1.1. Second, we tested correlation between patristic distance matrices (as in the case of SUB and IND trees) and branch length correlation for indel and substitution branch length sets using parsimony reconstructed branch lengths for indels and substitutions on constrained topology SUBIND (MacClade 4.08). Because of moderate sample size (77 branches in every case) and the lack of normal distributions of branch lengths (Shapiro–Wilk test), we adopted a nonparametric Spearman rank correlations R test.

3. Results and discussion

Altogether we sequenced three gene fragments from 43 avian species in 30 families (Table 1). The final data matrix used for Bayesian analyses (after coding and exclusion of indel sequences) comprises 2184 nucleotides (647 bp of 28S, 857 bp of ODC, and 680 bp of β-FIB) and 287 indels (15 of 28S, 128 of β-FIB, and 144 of ODC).

3.1. Phylogeny

For each of six Bayesian analyses including the three main analyses (SUB, IND and SUBIND) and the three analyses for substitutions only in each gene fragment separately, we obtained a sample of 195,000 trees from each run, thus a total of 390,000 sampled trees from one analysis. One SUB, one IND, and one SUBIND Bayesian 50% majority rule consensus tree was obtained from each sample. The estimated marginal likelihoods (lnL) in SUB analysis for tree sample are −18760.37 (arithmetic mean)/−18808.78 (harmonic mean) for run 1 and −18758.85/−18822.27 for run 2. The corresponding values for IND analysis are −1520.42/−1591.02 (run1) and −1521.85/−1588.51 (run2) and for SUBIND analysis −20321.017/−20371.66 (run1) and −20320.50/−20373.39 (run2).

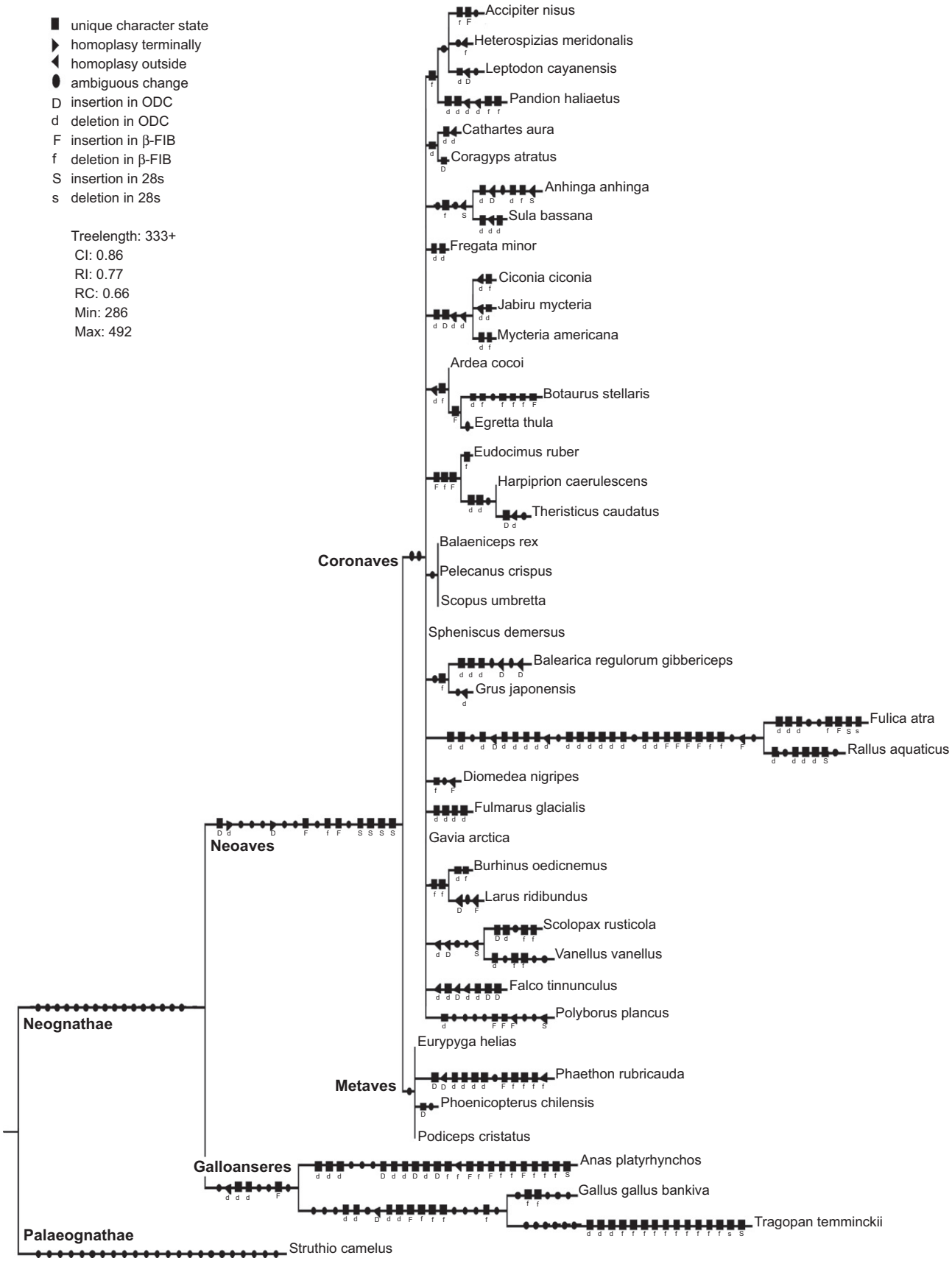

Fig. 4. All indels mapped on SUBIND tree using MacClade 4.08 parsimony. Branch lengths proportional to number of indels, hence the nodes not separated by at least one indel have been collapsed, resulting in polytomies as well as accidental changes in the sequential order of taxa. Two ambiguous changes at the root of Coronaves are from β-FIB and ODC, and the one at the root of Metaves is from β-FIB gene.

Table 2
Basic statistical description of identified indels.

	N	Mean	Median	Mode	Mode frequency	Min. length	Max. length	Std. dev.	Skewness
All indels									
28S	15	4.33	3	1	5	1	23	5.58	3.01
β-FIB	128	6.47	2	1	46	1	168	16.27	8.06
ODC	144	4.78	3	1	38	1	21	4.74	1.79
Total	287	5.51	3	1	89	1	168	11.45	10.47
Insertions									
28S	12	3	2	2	4	1	8	2.17	1.34
β-FIB	24	3.63	2	1	11	1	21	5.34	2.73
ODC	23	3.96	3	1	10	1	20	4.97	2.50
Total	59	3.63	2	1	24	1	21	4.66	2.74
Deletions									
28S	2	12	12	Multiple	1	1	23	15.56	1.34
β-FIB	57	9.16	3	1	14	1	168[a]	23.16	6.07
ODC	88	4.98	3	1	22	1	21	4.90	1.63
Total	147	6.69	3	1	37	1	168	15.04	8.79

[a] The omission of this extremely long deletion (the longest of all indels recorded in this study) barely changes the significance levels in K–W and median tests.

The estimated marginal likelihoods (lnL) for β-FIB analysis are: −10125.12 (arithmetic mean)/−10176.97 (harmonic mean) for run 1 and −10125.43/−10173.16 for run 2. The values for ODC analysis are: −7094.53/−7142.65 (run1) and −7094.59/−7150.15 (run2) and for 28S analysis: −1300.87/−1386.33 (run1) and −1304.92–1392.19 (run2). The convergence diagnostic (PSRF or potential scale reduction factor) of 1.00 for all free parameters of the model, 33 for SUB, 4 for IND, 36 for SUBINF, 12 for β -FIB, 13 for ODC, and 7 for 28S indicates a good sample from the posterior probability distribution.

The best supported, SUBIND tree (Fig. 1) that is based on both substitutions and indels, is in general agreement with the topologies obtained by both Ericson et al. (2006) and Hackett et al. (2008). While the tree recovered from indels alone (Fig. 3) is less structured than the tree based on substitutions (Fig. 2) neither the substitutions (Fig. 2) nor the indels alone (Fig. 3) yield the subdivision of Neoaves into Metaves and Coronaves. This suggests that the combined information from substitutions and indels generates an added value that allowed us to obtain essentially the same topology as in Ericson et al. (2006) although theirs is based on 5007 aligned nucleotides from five genes compared to our 2184 nucleotides and 287 indels from three genes.

However, the subdivision of Neoaves into Coronaves and Metaves remains questionable (Mayr, 2011a) and we found the Coronaves clade poorly supported (Fig. 1: Pp = 0.6) with Metaves nested in Coronaves in 40% of topologies. In the majority of cases (30% of all topologies) the Metaves were nested as a sister group to the (Cathartidae (Pandionidae (Accipitridae))) clade. The clade (Metaves (Cathartidae (Pandionidae (Accipitridae)))) was flanked in 61% of cases by the Falconidae as the sister group. While this second most frequent topology has been obtained with substitutions alone (Fig. 2), the IND tree (Fig. 3) leaves the monophyletic Metaves in a broad polytomy and thus does not preclude the possibility of the Metaves–Coronaves dichotomy.

While the Metaves–Coronaves dichotomy remains highly uncertain, we obtained maximum support for the monophyly of Metaves (Fig. 1) that have also been recovered in both SUB (Fig. 2: Pp = 1) and IND (Fig. 3: Pp = 0.99) tree alone. This support comes mostly from the β-FIB gene as Metaves are recovered only in the trees based on substitutions in this one locus (Figs. S1–S3) and our additional bayesian analysis of single locus indels (results not shown) leads to the same conclusion. However, both substitutions (Fig. S2) and indels in the ODC gene alone support the monophyly of the *Phaethon–Phoenicopterus–Podiceps* clade. This is the first direct support for this clade, although in a limited version, that comes independent of the β-FIB gene.

We also obtained maximum support for the clade (Cathartidae (Pandionidae (Accipitridae))) to the exclusion of Falconidae, and a high support (Pp = 0.98) for the clade of aquatic and semiaquatic birds. However, in our topology the gruiforms are nested within the water birds whereas in Ericson et al. (2006) and Hackett et al. (2008) the gruiforms, cuculiforms, and musophagids are together closest relatives of the water birds but outside their clade. This difference may possibly result from our omission of the cuculiforms and musophagids that may attract the gruiform branch.

Within the Charadriiformes, our topology (Fig. 1) identifies clades [Burhinidae, Laridae] and [Charadriidae, Scolopacidae] with the maximum support (Pp = 1) for both clades (Fig. 3) and thus agrees with the trees obtained by Paton et al. (2003) using RAG-1 gene sequence and Livezey (2010) using phenotypic characters. However, in 12 out of 19 charadriiform published topologies (Livezey, 2010; Mayr, 2011b), the Burhinidae are placed closer to the Charadriidae than to Lari. Fain and Houde (2007) left a polytomy with the position of Burhininidae unresolved, Chu (1995) placed Burhinidae equidistant to Charadriidae and Lari, and one of three Mayr's (2011b: Fig. 3A) topologies is a polytomy of four clades each comprising one of the four discussed taxa. Since the relationships within the Charadriiformes have yet to be settled, our topology may prove correct despite the paucity of sampled families.

Nearly all the ensembled fit indices CI, RI i RC for indels are well over 0.5 and thus much higher than those for substitutions (Table 3). Also Fain and Houde (2004) and Johnson (2004) found in FGB-int7 the consistency indices for indels to be much higher than for the nucleotides. This is also true for all mean indices that are significantly higher for indels (Table 4) than those for substitutions except for 28S indels (the exception being probably due to a small sample size). We suggest that a single indel carries more character weight than a single substitution, although clearly much more work is needed to determine realistic and safe multiplication factors.

No correlation of the fit indices with indel length has been found either as calculated for all indels jointly or for particular gene fragments separately (Table 5). This provides evidence against phylogenetic weighting of indel length, in agreement with Simmons et al. (2001). Fain and Houde (2004: Fig. 2) plotted β-FIB parsimony informative indel length and consistency indices as evidence for a greater weight of longer indels but the correlation of 0.2 (Spearman R) as calculated from their data is rather low although significant (0.05). Analogous correlation calculated from our data for β-FIB parsimony informative indels is even lower (Spearman R = 0.10) and insignificant. This suggests a weak correlation between β-FIB indel length and CI that is insignificant with

Table 3
Ensembled fit indices (CI_inf – consistency index, RI_inf – retention index, and RC_inf – rescaled consistency index) for parsimony informative indels (left figure) and substitutions (right figure) as mapped on the SUBIND tree (Fig. 1).

	CI_inf	RI_inf	RC_inf
28S	0.60/0.38	0.78/0.52	0.47/0.20
β-FIB	0.73/0.55	0.85/0.53	0.62/0.29
ODC	0.66/0.49	0.69/0.47	0.46/0.23
All fragments	0.68/0.52	0.77/0.50	0.53/0.26

Table 4
Differences between insertions and deletions (left figure) and beetwen all indels and substitutions (right figure) in mean fit indices (CI_inf – consistency index, RI_inf – retention index, and RC_inf – rescaled consistency index, all calculated for parsimony informative characters) for every gene fragment.

	CI_inf	RI_inf	RC_inf
28S	–/p = 0.31	–/p = 0.30	–/p = 0.25
β-FIB	p = 0.52/p = 0.00	p = 0.56/p = 0.00	p = 0.56/p = 0.00
ODC	p = 0.01/p = 0.00	p = 0.01/p = 0.00	p = 0.01/p = 0.00
All fragments	p = 0.10/p = 0.00	p = 0.11/p = 0.00	p = 0.12/p = 0.00

Table 5
Correlations (Spearman R, all insignificant at 0.05) between the length of parsimony informative indels and their phylogenetic uitility as measured by CI_inf – consistency index, RI_inf – retention index, and RC_inf – rescaled consistency index.

	CI_inf	RI_inf	RC_inf
All indels	−0.01	0.00	0.00
ODC	−0.09	−0.07	−0.08
β-FIB	0.10	0.08	0.09
28S	0.23	0.23	0.23

Table 6
Differences in mean fit indices (CI_inf – consistency index, RI_inf – retention index, and RC_inf – rescaled consistency index, all calculated for parsimony informative characters) between indels (left figure) and substitutions (right figure) from three different gene fragments.

	CI_inf	RI_inf	RC_inf
K–W test	p = 0.52/p = 0.02	p = 0.5/p = 0.25	p = 0.53/p = 0.32
Median test	p = 1.00/p = 0.04	p = 1.00/p = 0.57	p = 1.00/p = 0.47

our sample of 40 indels but significant with the larger sample of 161 indels in Fain and Houde's study.

The mean values of all individual fit indices CI, RI i RC are higher for deletions than for insertions for all three genes jointly but the difference is barely significant (K–W test, p = <0.1) (Table 4) although it is significant for the ODC fragment alone. Thus our results do not support the hypothesis that deletions are more likely to be homoplasious because of being more frequent than insertions and show that the opposite may be true for some genes.

The phylogenetic utility of indels, as measured by their individual fit indices, does not significantly vary between individual genes (Tables 6 and 7). In contrast, the averaged individual consistency index is significantly higher for substitutions in β-FIB as compared to the ODC fragment (K–W test with post hoc multiple comparisons and median test) although the difference is small (Tables 6 and 7).

3.2. Genomics

There are no differences in indel length between the three genes, either for all indels jointly or insertions and deletions separately (Table S1, K–W and median tests). Aside from the only two deletions in 28S, a normal distribution of indel lengths can be rejected at the significance levels of 0.01 (Fig. S1 and Table S3). In fact, the distributions of both insertions and deletions for each fragment reveal a strong right skewness (Table 2, Fig. S1), suggesting an exponential or Zipfian (Yule's) distribution. The most frequent are short, one- to three-nucleotide indels that contribute over 50% of all indels. Second in frequency are 4–11-nucleotide indels. The least frequent are medium-size indels of 12–20 nucleotides, whereas indels over 20 nucleotides are somewhat more frequent, which results in slightly bimodal distributions (Fig. S1 and Table S3). Such a distribution is consistent with the proposal that different mechanisms are responsible for the generation of short and long indels (Creer, 2007).

The average insertion and deletion lengths in the neognathous birds are, respectively, 3.63 and 6.69 nucleotides (Table 2), the deletion bias is ca 2.5 (147/59), and thus the overall NITD ratio amounts to 0.22, suggesting a rapid decrease in gene size. However, this may not be representative for the entire avian (or at least neognathous) genome, as the deletion bias as calculated by comparing the entire genomes of two phasianids (*Gallus* and *Meleagris*) is much lower (1.42) and limited to macrochromosomes (Brandström and Ellegren, 2007).

Table 7
Basic statistical description of individual fit indices for parsimony informative polarized indels and substitutions.

		Deletions					Insertions					Substitutions			
	N	Mean	Min.	Max.	Std. dev.	Valid *N*	Mean	Min.	Max.	Std. dev.	Valid *N*	Mean	Min.	Max.	Std. dev.
All															
CI_inf	50	0.88	0.25	1	0.23	28	0.78	0.25	1	0.28	807	0.60	0.11	1	0.25
RI_inf	50	0.83	0	1	0.36	28	0.69	0	1	0.41	807	0.45	0.00	1	0.38
RC_inf	50	0.80	0	1	0.38	28	0.64	0	1	0.46	807	0.33	0.00	1	0.38
28S															
CI_inf	0	–	–	–	–	6	0.79	0.25	1	0.33	19	0.62	0.17	1	0.35
RI_inf	0	–	–	–	–	6	0.79	0.25	1	0.33	19	0.60	0.00	1	0.39
RC_inf	0	–	–	–	–	6	0.72	0.06	1	0.44	19	0.49	0.00	1	0.46
β-FIB															
CI_inf	16	0.94	0.5	1	0.17	14	0.89	0.5	1	0.21	481	0.62	0.14	1	0.24
RI_inf	16	0.88	0	1	0.34	14	0.82	0	1	0.37	481	0.45	0.00	1	0.39
RC_inf	16	0.88	0	1	0.34	14	0.80	0	1	0.39	481	0.34	0.00	1	0.39
ODC															
CI_inf	34	0.86	0.25	1	0.25	8	0.58	0.33	1	0.27	307	0.58	0.11	1	0.25
RI_inf	34	0.81	0	1	0.36	8	0.40	0	1	0.42	307	0.44	0.00	1	0.37
RC_inf	34	0.77	0	1	0.40	8	0.31	0	1	0.43	307	0.31	0.00	1	0.37

Table 8
Insertion and deletion frequencies.

	All polarized indels	Deletion numbers	Deletion frequency	0.95 freq. confidence interval	Insertion numbers	Insertion frequency	Deletion to insertion ratio	Insertion to deletion ratio
ODC	111	88	0.79	0.71–0.86	23	0.21	3.83	0.26
β-FIB	81	57	0.70	0.59–0.80	24	0.30	2.38	0.42
28S	14	2	0.14	0.02–0.43	12	0.86	0.17	6

There are no detectable length differences between insertions and deletions in 28S and ODC genes, but such differences are present in the β-FIB intron where deletions are on the average 2.5 times longer, and have the median value 1.5 times higher than insertions (Tables 2 and S2). Accordingly, the distribution of deletions in β-FIB is much more right skewed than that of insertions (Table 2, Fig. S1). In conjunction with the deletion bias of 2.38 (Table 8), the differences in length result in the NITD ratio as low as 0.17 which must lead to a rapid shortening of this intron. Deletion bias has been recorded in β-fibrinogen intron 7 in many avian lineages (Prychitko and Moore, 2003) and turned out to be as high as ca. 6 in the Columbidae with a record low NITD ratio of 0.05 (Johnson, 2004) although this may be heavily exaggerated by microsatellite length variation (Brandström and Ellegren, 2007: p. 1699).

In the ODC fragment, deletions are 3.83 times more frequent than insertions and the difference is statistically significant (Table 8). Without significant differences in length between insertions and deletions (Tables 2 and S2) this must lead to a shortening of this gene with the NITD ratio of 0.21.

In contrast to the other two genes, in the 28S fragment insertions are significantly more frequent (Table 8) and prevail over deletions with the overall NITD ratio of 1.38. This is a slowly evolving gene expanded primarily in the early neoavian evolution with rare deletions occurring only in terminal branches (Fig. 4).

The crown-group neognaths (van Tuinen, 2009) and placentals (therians) (Murphy and Eizirik, 2009) split at about the same time, some 105 mya, and yet the avian (neognathous) indel fixation rate of 0.01385 per site is nearly 2.5 times higher than 0.00559 in placental mammals as determined by Matthee et al. (2007). This difference may possibly be somewhat exaggerated but seems ways too big to be accounted for by differences in methods. Another caveat stems from the fact that mammalian and avian fixation rates were calculated from different nonoverlapping sets of three genes that may differ in indel fixation rates. However, our results provide strong evidence that indel fixation rates are likely to be on the average much higher in neognathous birds than in therian mammals.

We discovered a considerable taxonomic variation in indel fixation rates (Fig. 4). The neoavian rate of 0.01132 indels/site, which is based on a large sample and thus reliable, is nearly 1.5 times lower than the galloanserine rate of 0.01637 indels/site, which is based on a small sample of three species from two modern dominant families (Anatidae and Phasianidae) and thus may be exaggerated. However, even after removing the value for *Tragopan* as a possible outlier, the rate of 0.01465 indels/site, which is identical with either *Gallus* or *Anas*, remains 1.3 times higher than the neoavian average that includes an obvious outlier for the Rallidae.

Among the Neoaves, the rates within most family-level and other infaordinal clades (Accipitridae, Cathartidae, Suloidea, Ciconiidae, Scopoidea, and Charadrii) are fairly uniform, but extremely uneven in the Ardeidae, Threskiornithidae, and Gruidae. The most striking are the accelerated rates in the Rallidae and *Tragopan*. Assuming the split between the rails and cranes at 64.5 mya (Houde, 2009) and that between the neoavians and galloanserines at 105 mya (van Tuinen, 2009), the indel fixation rate of 0.0002626 indel/site/my in the Rallidae has been over 2.4 times higher than the average rate of 0.0001078 indel/site/my for Neoaves (including the Rallidae) and 8.2 times higher than the average for the two Gruidae (Fig. 4). The indel fixation rate of in *Tragopan* lineage within the Phasianinae (pheasants) has been 3.5 times that in the *Gallus* (junglefowl, chicken) lineage. Since the junglefowls are only distantly related to the pheasants (Crowe et al., 2006), there is a possibility that other pheasants may share the fast rate with *Tragopan*. Highly differentiated indel fixation rates have been reported among the placentals, with rodents and afrotherians accumulating indels more than twice as fast as other lineages (Matthee et al., 2007).

There is no obvious explanation for the striking differences of indel fixation rates in either mammals or birds. However, there is good evidence that indel fixation rates correlate with lineage-specific evolutionary rates as determined by other measures. It is known that indels, including those affecting non-coding sequences, may be subject to selective pressures (Bird et al., 2006) and that indels increase mutation rates in the surrounding sequences (Tian et al., 2008). Accordingly, a high indel fixation rate in rodents correlates with elevated rates of change in both nuclear and mitochondrial genes (Matthee et al., 2007).

Our results support a strong association between indel fixation rates and lineage-specific substitution rates as demonstrated by high and significant correlations between SUB tree and IND tree patristic distance matrices (Pearson $r = 0.91$, $p < 0.0001$), between patristic distance matrices for indel and substitution branch length reconstructed on constrained topology SUBIND (Pearson $r = 0.88$, $p < 0.0001$), and for indel and substitution branch length reconstructed on constrained topology SUBIND (Spearman $R = 0.85$, $p < 0.05$).

Highest substitution rates in both Rallidae and Phasianidae have been recorded by Hackett et al. (2008) and ourselves (Fig. 2). In addition, the rails are known for their high adaptability to insular habitats and a remarkably fast evolution of flightlessness (Slikas et al., 2002) and the phasianid galliforms show a much higher recombination rates than the passerines (Ellegren, 2007; Stapley et al., 2008). Also, on a smaller taxonomic scale, the striking difference in the number of indels between the bittern (*Botaurus stellaris*) and two heron species (Fig. 4) correlates with about two times faster substitution rate in the same genes (Fig. 2) and 1.25 times faster evolution of single-copy nuclear DNA (Sheldon et al., 2000) in the bitterns compared to herons. We suggest that the variation of indel fixation rates may be accounted for in terms of lineage-specific evolutionary rates rather than general, phylogeny-independent allometric relationships. However, while all studied lineages with above-average indel fixation rates also show heightened substitution rates, the opposite is not true as shown by heightened substitution rates in *Scopus*, *Scolopax*, the falconids, *Eurypyga*, and *Podiceps* (Figs. 1 and 2), that do not reveal any above-average indel fixation rates (Fig. 4).

Acknowledgments

We thank our technicians, K. Nowak and A. Dąbrowska, for sequencing work, and K. Lipiec-Sidor for managing tissue samples which were kindly provided by J. Dean (National Museum of Natural History, Smithsonian Institution), A. Kruszewicz (The Warsaw Zoo), P. Ćwiertnia (The Poznań Zoo), and T. Grabiński (The Wrocław Zoo). The manuscript greatly benefitted from comments by three

anonymous reviewers as well as M. Wolsan (Museum and Institute of Zoology, Polish Academy of Sciences). The laboratory part of this project was supported by Poland's Ministry of Science and Higher Education grant 3 PO4 C 06925 and the computational part by Wrocław Network-Supercomputer Center (WCSS) grant 106/2008.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.07.021.

References

Allen, E.A., Omland, K.E., 2003. Novel intron phylogeny supports plumage convergence. Appl. Bioinformatics 4, 65–69.
Alström, P., Ericson, P.G.P., Olsson, U., Sundberg, P., 2006. Phylogeny and classification of the avian superfamily Sylvioidea. Mol. Phylogenet. Evol. 38, 381–397.
Brandström, M., Ellegren, H., 2007. The genomic landscape of short insertion and deletion polymorphisms in the chicken (*Gallus gallus*) genome: a high frequency of deletions in tandem duplicates. Genetics 176, 1691–1701.
Barrowclough, G.F., Groth, J.G., Mert, L.A., 2006. The RAG-1 exon in the avian order Caprimulgiformes: phylogeny, heterozygosity, and base composition. Mol. Phylogenet. Evol. 41, 238–248.
Bird, C.P., Stranger, B.E., Dermitzakis, E.T., 2006. Functional variation and evolution of non-coding DNA. Curr. Opin. Genet. Dev. 16, 559–564.
Chojnowski, J.L., Kimball, R.T., Braun, E.L., 2008. Introns outperform exons in analyses of basal avian phylogeny using clathrin heavy chain genes. Gene 410, 89–96.
Chu, P.C., 1995. Phylogenetic re-analysis of Strauch's osteological data set for the Charadriiformes. Condor 97, 174–196.
Creer, S., 2007. Choosing and using introns in molecular phylogenetics. Evol. Bioinform. 3, 99–108.
Crowe, T.M., Bowie, R.C.K., Bloomer, P., Mandiwana, T.G., Hedderson, T.A.J., Randi, E., Pereira, S.L., Wakeling, J., 2006. Phylogenetics, biogeography and classification of and character evolution in, gamebirds (Aves: Galliformes): effects of character exclusion, data partitioning and missing data. Cladistics 22, 495–532.
de Kloet, R.S., de Kloet, S.R., 2005. The evolution of the spindlin gene in birds: sequence analysis of an intron of the spindlin W and Z gene reveals four major divisions of the Psittaciformes. Mol. Phylogenet. Evol. 36, 706–721.
Edwards, S.V., Jennings, W.B., Shedlock, A.M., 2005. Phylogenetics of modern birds in the era of genomics. Proc. Roy. Soc. Lond. B 272, 979–992.
Ellegren, H., 2007. Molecular evolutionary genomics of birds. Cytogenet. Genome Res. 117, 120–130.
Ericson, P.G.P., Andersson, C.L., Britton, T., Elzanowski, A., Johansson, U.S., Källersjö, M., Ohlson, J.I., Parsons, T.J., Zuccon, D., Mayr, G., 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biol. Lett. 2, 543–547.
Ericson, P.G.P., Johansson, U.S., Parsons, T.J., 2000. Major divisions in Oscines revealed by insertions on the nuclear gene *c-myc*: a novel gene in avian phylogenetics. Auk 117, 1069–1078.
Ericson, P.G.P., Envall, I., Irestedt, M., Norman, J.A., 2003. Inter-familial relationships of the shorebirds (Aves: Charadriiformes) based on nuclear DNA sequence data. BMC Evol. Biol. 3, 16.
Fain, M.G., Houde, P., 2004. Parallel radiations in the primary clades of birds. Evolution 58, 2558–2573.
Fain, M.G., Houde, P., 2007. Multilocus perspectives on the monophyly and phylogeny of the order Charadriiformes (Aves). BMC Evol. Biol. 7, 35. doi:10.1186/1471-2148-7-35.
Fidler, A.E., Kuhn, S., Gwinner, E., 2004. Convergent evolution of strigiform and caprimulgiform dark-activity is supported by phylogenetic analysis using the arylalkylamine N-acetyltransferase (Aanat) gene. Mol. Phylogenet. Evol. 33, 908–992.
Gregory, T.R., 2004a. A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves. Evolution 56, 121–130.
Gregory, T.R., 2004b. Insertion–deletion biases and the genome size. Gene 324, 15–34.
Groth, J.G., Barrowclough, G.F., 1999. Basal divergences of birds and the phylogenetic utility of nuclear RAG-1 gene. Mol. Phylogenet. Evol. 10, 377–390.
Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, E.L., Braun, M.J., Chojnowski, J.L., Cox, W.A., Han, K.L., Harshman, J., Huddleston, C.J., Marks, B.D., Miglia, K.J., Moore, W.S., Sheldon, F.H., Steadman, D.W., Witt, C.C., Yuri, T., 2008. A phylogenomic study of birds reveals their evolutionary history. Science 320, 1763–1768.
Harshman, J., Braun, E.L., Braun, M.J., Huddleston, C.J., Bowie, R.C.K., Chojnowski, J.L., Hackett, S.J., Han, K.-L., Kimball, R.T., Marks, B.D., Miglia, K.J., Moore, W.S., Reddy, S., Sheldon, F.H., Steadman, D.W., Steppan, S.J., Witt, C.C., Yuri, T., 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. PNAS 105, 13462–13467.
Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Quart. Rev. Biol. 66, 411–453.
Houde, P., 2009. Cranes, rails, and allies (Gruiformes). In: Hedges, S.B., Kumar, S. (Eds.), The Timetree of Life. Oxford University Press, Oxford, pp. 40–444.
Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.
Huelsenbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Syst. Biol. 51, 673–678.
Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian influence of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.
Ivanov, I.P., Firth, A.E., Atkins, J.F., 2010. Recurrent emergence of catalytically inactive ornithine decarboxylase homologous forms that likely have regulatory function. J. Mol. Evol. 70, 289–302.
Johansson, U.S., Parsons, T.J., Irestedt, M., Ericson, P.G.P., 2001. Clades within the 'higher land birds', evaluated by nuclear DNA sequences. J. Zool. Syst. Evol. Res. 39, 37–51.
Johnson, K.P., 2004. Deletion bias in avian introns over evolutionary timescales. Mol. Biol. Evol. 21, 599–602.
Keane, T.M., Naughton, T.J., McInerney, O., 2007. MultiPhyl: a high-throughput phylogenomics webserver using distributed computing. Nucl. Acids Res. 35, W33–W37.
Laird, P.W., Zijderveld, A., Linders, K., Rudnicki, M.A., Jaenisch, R., Berns, A., 1991. Simplified mammalian DNA isolation procedure. Nucl. Acids Res. 19, 4293.
Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50, 913–925.
Livezey, B.C., 2010. Phylogenetics of modern shorebirds (Charadriiformes) based on phenotypic evidence: analysis and discussion. Zool. J. Linn. Soc. 160, 567–618.
Löytynoja, A., Milinkovitch, M.C., 2001. SOAP, cleaning multiple alignments from unstable blocks. Bioinformatics 17, 573–574.
Matthee, C.A., Eick, G., Willow-Munro, S., Montgelard, C., Pardini, A., Robinson, T.J., 2007. Indel evolution of mammalian introns and the utility of non-coding nuclear markers in eutherian phylogenetics. Mol. Phylogenet. Evol. 42, 827–837.
Maddison, W.P., Maddison, D.R., 1992. MacClade: analysis of phylogeny and character evolution. Sinauer, Sunderland, Massachusetts.
Mayr, G., 2011a. Metaves, Mirandornithes, Strisores and other novelties – a critical review of the higher-level phylogeny of neornithine birds. J. Zool. Syst. Evol. Res. 49, 58–76.
Mayr, G., 2011b. The phylogeny of charadriiform birds (shorebirds and allies) – reassessing the conflict between morphology and molecules. Zool. J. Linn. Soc. 161, 916–934.
Müller, K., 2005. SeqState – primer design and sequence statistics for phylogenetic DNA data sets. Appl. Bioinformatics 4, 65–69.
Morgan-Richards, M., Trewick, S.A., Bartosch-Härlid, A., Kardailsky, O., Phillips, M.J., McLenachan, P.A., Penny, D., 2008. Bird evolution: testing the Metaves clade with six new mitochondrial genomes. BMC Evol. Biol. 23, 8–20.
Murphy, W.J., Eizirik, E., 2009. Placental mammals (Eutheria). In: Hedges, S.B., Kumar, S. (Eds.), The Timetree of Life. Oxford University Press, Oxford, pp. 471–474.
Oliver, M.J., Petrov, D., Ackerly, D., Falkowski, P., Schofield, O.M., 2007. The mode and tempo of genome size evolution in eukaryotes. Genome Res. 17, 594–601.
Organ, C.L., Shedlock, A.M., Meade, A., Pagel, M., Edwards, S.E., 2007. Origin of avian genome size and structure in non-avian dinosaurs. Nature 446, 180–184.
Organ, C.L., Shedlock, A.M., 2009. Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. Biol. Lett. 5, 47–50.
Palumbi, S.R., 1996. Nucleic acids II: the polymerase chain reaction. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics, second ed. Sinauer Assoc., Sunderland, MA, pp. 05–221.
Paton, T.A., Baker, A.J., Groth, J.G., Barrowclough, G.F., 2003. RAG-1 sequences resolve phylogenetic relationships within charadriiform birds. Mol. Phylogenet. Evol. 29, 268–278.
Prychitko, T.M., Moore, W.S., 1997. The utility of DNA sequences of an intron from the β-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). Mol. Phylogenet. Evol. 8, 193–204.
Prychitko, T.M., Moore, W.S., 2003. Alignment and phylogenetic analysis of beta-fibrinogen intron 7 sequences among avian orders reveal conserved regions within the intron. Mol. Biol. Evol. 20, 762–771.
Rambaut, A., Drummond, A.J., 2003–2007. Tracer. MCMC Trace Analysis Tool. Version v1.4.
Seibel, P.N., Müller, T., Dandekar, T., Schultz, J., Wolf, M., 2006. 4SALE – a tool for synchronous RNA sequence and secondary structure alignment and editing. BMC Bioinformatics 7, 498.
Sheldon, F.H., Jones, C.E., McCracken, K.G., 2000. Relative patterns and rates of evolution in heron nuclear and mitochondrial DNA. Mol. Biol. Evol. 17, 437–450.
Simmons, M.P., Mueller, K., Norton, A.P., 2007. The relative performance of indel-coding methods in simulations. Mol. Phylogenet. Evol. 44, 724–740.
Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49, 369–381.
Simmons, M.P., Ochoterena, H., Carr, T.G., 2001. Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. Syst. Biol. 50, 454–462.
Slikas, B., Olson, S.L., Fleischer, R.C., 2002. Rapid, independent evolution of flightlessness in four species of Pacific rails (Rallidae): an analysis based on mitochondrial sequence data. J. Avian Biol. 33, 5–14.
Smith, J.D.L., Gregory, T.R., 2009. The genome sizes of megabats (Chiroptera: Pteropodidae) are remarkably constrained. Biol. Lett. 5, 347–351.

Spicer, G.S., Dunipace, L., 2004. Molecular phylogeny of songbirds (Passeriformes) inferred from mitochondrial 16S ribosomal RNA gene sequences. Mol. Phylogenet. Evol. 30, 325–335.

Stapley, J., Birkhead, T.R., Burke, T., Slate, J., 2008. A linkage map of the zebra finch *Taeniopygia guttata* provides new insights into avian genome evolution. Genetics 179, 651–667.

StatSoft, 2008. STATISTICA (data analysis software system), version 8. StatSoft, Tulsa, OK, USA.

Tian, D., Wang, Q., Zhang, P., Araki, H., Yang, S., Kreitman, M., Nagylaki, T., Hudson, R., Bergelson, J., Chen, J.-Q., 2008. Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. Nature 455, 105–108.

van Tuinen, M., 2009. Birds (Aves). In: Hedges, S.B., Kumar, S. (Eds.), The Timetree of Life. Oxford University Press, Oxford, pp. 09–411.

Wolf, Y., Madej, T., Babenko, V., Shoemaker, B., Panchenko, A.R., 2007. Long-term trends in evolution of indels in protein sequences. BMC Evol. Biol. 7, 19.