New View on the Organization and Evolution of Palaeognathae Mitogenomes Poses The Question on the Ancestral Gene Rearrangement in Aves

Adam Dawid Urantowka
Uniwersytet Przyrodniczy we Wroclawiu
Aleksandra Kroczak
Uniwersytet Wroclawski
Pawel Mackiewicz (pamac@smorfland.uni.wroc.pl)
Uniwersytet Wroclawski https://orcid.org/0000-0003-4855-497X

Research article

Keywords: ancestral state, Aves, duplication, mitochondrial genome, mitogenome, Neognathae, Palaeognathae, phylogeny, rearrangement

Posted Date: September 8th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-41469/v1

Version of Record: A version of this preprint was published on December 7th, 2020. See the published version at https://doi.org/10.1186/s12864-020-07284-5.

Abstract

Background

Bird mitogenomes differ from other vertebrates in gene rearrangement. The most common avian gene order, first identified in *Gallus gallus*, is considered ancestral for all Aves. However, other rearrangements including a duplicated control region and neighboring genes have been reported in many representatives of avian orders. The repeated regions can be easily omitted due to inappropriate DNA amplification or genome sequencing, and may thus be readily overlooked. This raises a question about the actual prevalence of mitogenomic duplications and the validity of the current view on the avian mitogenome evolution. In this context, Palaeognathae is especially interesting because is sister to all other living birds (Neognathae). So far, a unique duplicated region has been found in one palaeognath mitogenome, that of *Eudromia elegans*.

Results

Therefore, we applied an appropriate PCR strategy to look for omitted duplications in other palaeognaths. The analyses revealed the duplicated control regions with adjacent genes in *Crypturellus, Rhea* and *Struthio* as well as ND6 pseudogene in three moa. The copies are very similar and were subjected to concerted evolution. Mapping the presence and absence of duplication onto the Palaeognathae phylogeny indicates that the duplication was an ancestral state for this avian group. This feature was inherited by early diverged lineages and lost two times in others. Comparison of incongruent phylogenetic trees based on mitochondrial and nuclear sequences showed that two variants of mitogenomes could exist in the evolution of palaeognaths. Included data for other avian mitogenomes revealed that the last common ancestor of all birds and early diverging lineages of Neoaves may also possess the mitogenomic duplication.

Conclusions

The duplicated control regions with adjacent genes are more common in avian mitochondrial genomes than it was previously thought. These two regions could increase effectiveness of replication and transcription as well as the number of replicating mitogenomes per organelle. In consequence, energy production by mitochondria may be also more efficient. However, further physiological and molecular analyses are necessary to assesses the potential selective advantages of the mitogenome duplications.

1. Background

Animal mitochondrial genomes are characterized by compact organization and almost invariable gene content, so specific changes in them are especially interesting because they can be associated with major transitions in animal evolution [1, 2]. The first fully sequenced avian mitogenome from *Gallus gallus* [3] turned out to contain single versions of 37 genes and one control region as in most other vertebrates, but organized in a different order (Fig. 1). This rearrangement is believed to have derived from the typical

vertebrate gene order by a single tandem duplication of the fragment located between ND5 and tRNA-Phe genes followed by random losses of one copy of duplicated genes and control regions. Due to the prevalence of the *Gallus gallus* gene order in other birds, this rearrangement type is generally believed to be an ancestral state for all Aves. As a consequence, it is called common, standard or typical.

However, the growing number of avian mitochondrial genomes sequenced in recent years has revealed that other gene orders may be also be present in a frequency higher than it was previously thought. To date, several distinct variations of mitochondrial rearrangements have been reported in many representatives of the following avian orders: Accipitriformes [4, 5], Bucerotiformes [6], Charadriiformes [7], Coraciiformes [8], Cuculiformes [9–11], Falconiformes [4], Gruiformes [12], Passeriformes [13, 14], Pelecaniformes [4, 15, 16], Phoenicopteriformes [17, 18], Piciformes [4, 19], Procellariiformes [20, 21], Psittaciformes [22, 23], Strigiformes [24], Suliformes [15, 20, 25] and Tinamiformes [26]. All these rearrangements include an additional region between ND5 and tRNA-Phe genes, which seems to be particularly susceptible to duplication.

The most fully duplicated avian region (GO-FD; Fig. 1) was found in mitogenomes of all representatives of Gruidae [12] and Suliformes [15, 20, 25, 27], the majority of Pelecaniformes [4, 16] and Procellariiformes [4, 21, 28], as well as some Bucerotiformes and Psittaciformes species [6, 23]. All other avian gene orders containing the duplicated elements result from subsequent degenerations of the rearrangement GO-FD due to pseudogenization or loss of some genes and/or control region [22, 23].

It has been commonly assumed that the mitogenomic duplications are derived states and occurred independently in many avian lineages, for example in Psittaciformes and Passeriformes [13, 22, 29–31]. However, an independent origin of identical gene orders in different avian lineages seems unlikely because of the great number of possible arrangements [32–35]. Accordingly, recent analyses of additional parrot mitogenomes showed that the last common ancestor of Psittaciformes had a duplicated control region with adjacent genes [23]. The same could be true for Accipitriformes [4, 5, 36–38], Falconiformes [4, 39], Gruidae [12] and Pelecaniformes [4, 15, 16] because all or almost all members of these groups contain mitogenomes with the duplicated regions.

The lack of duplication in some fully sequenced mitogenomes may be false and result from omission of identical repeats due to an inappropriate PCR strategy, insufficient sequencing methods or incorrect genome assembly. This problem was already addressed by Gibb, et al. [4], who found the fully duplicated gene order in *Thalassarche melanophris* mitogenome, which had been previously annotated without the duplication [40]. Similarly, two other mitogenomes of *Notiomystis cincta* and *Turdus philomelos* showed a novel duplicated gene order after a re-analysis [41], although previously the single version had been reported [42]. All re-amplified and re-sequenced crane mitogenomes also revealed the existence of duplication [12], which had not been found earlier [43]. Omitted duplications were also found within the mitochondrial genomes of Strigopoidea and Cacatuoidea, demonstrating that the ancestral parrot contained duplication in its mitogenome [23]. Recently, Mackiewicz, et al. [14] showed that even the last common ancestor of a larger monophyletic group of Aves including Psittaciformes, Passeriformes and

Falconiformes could have had a duplication of the control region with adjacent genes in the mitochondrial genome.

The growing number of formerly unidentified duplications implies that many avian mitogenomes published so far without duplication may, in fact, have it. Therefore, a diligent search for potential duplications is crucial to understanding the evolution of the avian mitochondrial genome. Palaeognathae are particularly important to this subject because all comprehensive avian phylogenies have placed them as the sister group to the rest of birds, called Neognathae [44–48]. Palaeognaths comprise 25 genera and 82 species [49, 50], which are currently grouped into three extinct and five extant orders: the flighted Lithornithiformes known from Paleocene and Eocene of North America and Europe, and possibly from the Late Cretaceous; the flighted tinamous (Tinamiformes) of South and Central America; the flightless ratites containing the recently extinct New Zealand moa (Dinornithiformes) and Madagascan elephant birds (Aepyornithiformes) as well as the extant African ostrich (Struthioniformes), South American rheas (Rheiformes), Australian emu and Australasian cassowaries (Casuariiformes), and New Zealand kiwi (Apterygiformes). Phylogenetic relationships between these groups have been controversial. Molecular analyses have revealed that the ratites are paraphyletic and suggested that flightlessness evolved several times among ratites independently [51–60].

So far, a duplicated region (Cytb/tRNA-Thr/tRNA-Pro/CR1/ND6/tRNA-Glu/CR2) has been found only in one representative of palaeognaths, namely *Eudromia elegans* [26]. This rearrangement has not been identified in any other avian species. Other Palaeognathae mitogenomes have a typical single avian gene order or were published as incomplete, especially in the part adjacent to the control region [26]. However, it cannot be ruled out that an inadequate PCR strategy was unable to amplify identical repeats or even prevented the completion of the mitochondrial genome sequencing and assembly in the presence of non-identical repeats [61]. Therefore, we applied another PCR strategy that allows the amplification of the fragment between two control regions including a potentially omitted duplication in representatives of *Struthio*, *Rhea*, *Casuarius*, *Dromaius* and *Crypturellus*. The new data help elucidate the evolution of the Palaeognathae mitogenome in terms of duplication events, and also have implications for mitogenomic evolution in Aves as a whole.

2. Results And Discussion

2.1. Duplicated gene order identified in mitogenomes of analysed Palaeognathae taxa

Using an appropriate PCR strategy (Fig. 2), the diagnostic CR1/CR2 fragments were obtained for *Struthio camelus* (Fig. S1a in Additional file 1), *Rhea pennata* (Fig. S1b in Additional file 1), *Rhea americana* (Fig. S1c in Additional file 1) and *Crypturellus tataupa* (Fig. S1d in Additional file 1). Only two out of 16 or 48 reactions failed in taxa for which species-specific primers were designed based on the previously published sequences of complete mitogenomes (*Struthio camelus* and *Rhea* species) (Table S1 in Additional file 2). In the case of *Crypturellus tataupa*, amplicons were obtained only for six out of 12

tested reactions. This was caused by the fact that primers dedicated for this species were designed based on the *Eudromia elegans* mitogenome [26]. As in the published *Crypturellus tataupa* genomic sequence [62], the control region and adjacent genes were missing. Sequencing and annotation of the obtained amplicons revealed the presence of tRNA-Pro/ND6/tRNA-Glu fragments between two control regions for *Struthio camelus*, *Rhea pennata*, *Rhea americana* and *Crypturellus tataupa* (Fig. 1). The duplicated fragment obtained for *Struthio camelus* differed only in one nucleotide from the homologous region present in the previously published mitogenome (Fig. S2a in Additional file 1). These fragments in rheas showed 100% identity with corresponding homologous regions (Fig. S2b and Fig. S2c in Additional file 1).

Although the high identity strongly indicates a mitochondrial origin of the amplified CR1/CR2 fragments, additional diagnostic reactions were designed to exclude a possibility of nuclear mitochondrial DNA inserts (NUMTs) amplification. Based on the obtained sequences of ND6 genes, appropriate primers were created to amplify ND6-1/ND6-2 regions. Sequencing of the obtained PCR products revealed the ND6/tRNA-Glu/CR/tRNA-Pro/ND6 gene order for all analyzed species. The corresponding CR/tRNA-Pro/ND6 regions overlapped the appropriate CR1/CR2 diagnostic fragments and showed 100% identity. Additional PCR reactions (see Materials and methods and Fig. 2) were run to complete the missing parts of CRs and to reveal the order of genes preceding the first control region. Finally, the complete mitogenomic fragments containing the duplicated regions were obtained by assembling four overlapping fragments (Fig. 2). Their length was: 8,554 bp for *Struthio camelus*, 8,254 for *Rhea Americana*, 8,360 bp for *Rhea paennata* and 7,044 bp for *Crypturellus tataupa* (Table 1; Fig. S3 in Additional file 1). In all cases the same gene order was found (GO-I; Fig. 1e, Table 1, Fig. S3 in Additional file 1), which was previously annotated only for two Passeriformes species, *Notiomystis cincta* and *Turdus philomelos* [41]. This gene order differs from the most complete known avian duplication (GO-FD; Fig. 1d) in the lack of the second copies of cytb and tRNA-Thr genes, expected between CR1 and tRNA-Pro2 gene. The presence of identical copies of tRNA-Glu gene (Fig. S2a-d in Additional file 1) enabled us to position precisely the 5' ends of both control regions. The 3’ ends of the CR2s precede tRNA-Phe genes as in all other avian gene orders including two potentially functional control regions. The number of nucleotides between the tRNA-Glu copies and appropriate poly-C sequences located at the 5' ends of CRs vary from 2 bp (*Rhea americana*, *Rhea pennata* and *Crypturellus tataupa*) to 26 bp for *Struthio camelus* (Table S2 in Additional file 2). The CR2 in *Rhea pennata* and *Crypturellus tataupa* is longer than CR1, which obey the rule observed in 13 crane species [12]. The tandem duplications found in the mitogenomes of *Struthio camelus*, *Rhea americana*, *Rhea pennata* and *Crypturellus tataupa* make them longer compared with their previous genomic versions assuming the typical avian gene order.

Table 1

Palaeognathae species analyzed in this study in terms of duplicated regions as well as gene orders found within their mitogenomic fragments, which were amplified and sequenced. The sequences are presented in Figure S3.

Sources of samples: ZOO WRO – Zoological Garden in Wrocław, Poland; ZOO WAW – Zoological Garden in Warsaw, Poland; ZOO KAT - Zoological Garden in Katowice. L – tRNA for leucine, ND5 - NADH dehydrogenase subunit 5, cytb – cytochrome b, T – tRNA for threonine, P – tRNA for proline, ND6 – NADH dehydrogenase subunit 6, E – tRNA for glutamic acid, CR – control region, F – tRNA for phenylalanine, 12S – 12S rRNA, V – tRNA for valine, 16S – 16S rRNA.

Species	Sample type	Source	Accession number	Length (bp)	Fragment
Struthio camelus	Blood	ZOO WRO	MH264503	8 554	L/ND5/cytb/T/P1/ND6-1/E1/CR1/P2/ND6-2/E2/CR2/F/12S/V/16S
Rhea americana	Blood	ZOO KAT	MK696563	8 254	ND5/cytb/T/P1/ND6-1/E1/CR1/P2/ND6-2/E2/CR2/F/12S/V/16S
Rhea pennata	Blood	ZOO WAW	MK696564	8 306	ND5/cytb/T/P1/ND6-1/E1/CR1/P2/ND6-2/E2/CR2/F/12S/V/16S
Casuarius casuarius	Blood	ZOO WAW	-	-	-
Dromaius novaehollandiae	Blood	ZOO WAW	-	-	-
Crypturellus tataupa	Blood	ZOO WAW	MK696562	7 044	ND5/cytb/T/P1/ND6-1/E1/CR1/P2/ND6-2/E2/CR2/F/12S

2.2. Probable presence of mitochondrial CR1/CR2 fragments in Casuarius casuarius and Dromaius novaehollandiae nuclear genomes

In the case of two other Palaeognathae species, *Casuarius casuarius* and *Dromaius novaehollandiae*, an attempt to amplify the CR1/CR2 fragment was also made. Similar to other taxa, species-specific D-F and D-R primers (Fig. 2; Table S1 in Additional file 2) were designed using the sequences of previously published complete mitogenomes (AF338713.2 and AF338711.1). In contrast to the results obtained for the other Palaeognathae species, most PCR reactions failed to amplify the expected fragments. In *Dromaius novaehollandiae*, amplicons were obtained only for 3 out of 25 tested reactions (Fig. S4a in Additional file 1, Table S1 in Additional file 2). Analogously, PCR products were obtained only for 4 out of 56 reactions for *Casuarius casuarius* (Fig. S4b in Additional file 1, Table S1 in Additional file 2). Moreover, single DNA fragments were not obtained for any of these seven reactions, although different annealing temperatures were applied (Fig. S4 in Additional file 1). Taking into account the heterogeneity of the obtained DNA fragments as well as the fact that most of the tested reactions failed, we can conclude that

the PCR products presented in Figure S4 in Additional file 1 were not amplified on the mitochondrial genome template. The D-F and D-R primers as well as the applied PCR are highly specificity and diagnostic for the presence of control region duplication in parrots [23], cranes [12] as well as black-browed albatross, ivory-billed aracari and osprey [4]. Therefore, the seven positive amplicons most likely represent mitochondrial DNA fragments located in the nuclear genomes, i.e. NUMTs. It means that *Casuarius casuarius* and *Dromaius novaehollandiae* or their ancestors had mitogenomes comprising two control regions, which were then transferred into the nucleus during evolution.

2.3. Reannotation of Eudromia elegans mitochondrial gene order

The GO-I gene order found in this study for four Palaeognathae taxa differs from that of the published mitogenomic sequence of *Eudromia elegans* [26]. This rearrangement appears to be a degenerated form of GO-I gene order as it lacks the first copy of ND6 and tRNA-Glu genes as well as the second copy of tRNA-Pro gene. This fact prompted us to search for a potential tRNA-Pro pseudogene hidden within the last 122 nucleotides of the first control region of *Eudromia elegans* mitogenome. In fact, the comparison of CR1 sequence with the functional tRNA-Pro sequence of this species revealed a very significant similarity (E-value = $1.2 \cdot 10^{-6}$; 81% identity) between these sequences along the 84-bp alignment (Fig. S5a in Additional file 1), which supports the presence of the tRNA-Pro pseudogene in the *Eudromia* mitogenome in the position between 16,272 bp and 16,349 bp. After reannotation of this pseudogene, the length of CR1 reduced to 1,352 bp. The newly annotated *Eudromia* gene order was defined as GO-P1 in Fig. 1e.

2.4 Reannotation of mitochondrial gene order in the mitogenomes of Anomalopteryx didiformis, Emeus crassus and Dinornis giganteus

Our analysis of 5’ spacers, i.e. fragments of control regions located between the tRNA-Glu gene and poly-C motif, revealed that they are much longer in annotated *Anomalopteryx didiformis*, *Emeus crassus* and *Dinornis giganteus* mitogenomes than in other Palaeognathae species. These spacers of the most Palaeognathae taxa are from 2 bp to 33 bp in length (Table S2 in Additional file 2). Surprisingly, the lengths of *Anomalopteryx didiformis*, *Emeus crassus* and *Dinornis giganteus* spacers are 133 bp, 157 bp and 150 bp respectively. Additionally, all three fragments contain a purine-rich insertion (Fig. S5b in Additional file 1) analogous to that in parrot ND6 pseudogenes (Fig. S5c in Additional file 1) [23]. In the Psittaciformes mitogenomes (*Probosciger aterrimus goliath*, *Eolophus roseicapilla* and *Cacatua moluccensis*), this insertion is preceded by a fragment (433–450 bp) almost identical with the first ND6 copy followed by a highly degenerated region. This similar sequence pattern prompted us to search for potential ND6 pseudogenes within the 5’ spacers of control regions in *Anomalopteryx didiformis*, *Emeus crassus* and *Dinornis giganteus*. The comparison of 5’ CR sequences with appropriate ND6 genes of these species revealed a significant similarity between the aligned sequences (Table S3 in Additional file

2). Those from *Anomalopteryx didiformis* were identical in 71% with E-value = 0.13 (Fig. S5d in Additional file 1) and from *Emeus crassus* in 73% with E-value = 0.0015 (Fig. S5e in Additional file 1). The alignment of *Dinornis giganteus* sequences was much more significant with E-value = $5.8 \cdot 10^{-106}$ and the sequences showed 83% identity (Fig. S5f in Additional file 1). The obtained identity and E values are in the range of those obtained for ND6 pseudogenes and their functional copies annotated in other avian species, i.e. 65% – 96% and 0–0.23, respectively (Table S3 in Additional file 2).

Assuming the presence of ND6 pseudogenes in *Anomalopteryx didiformis*, *Dinornis giganteus* and *Emeus crassus* mitogenomes, the length of their CR is reduced to 1,347 bp, 1,360 bp and 1,346 bp, respectively. The CR sequences show 71–81% identity at 5' spacers on the length 165 bp (Fig. S5b in Additional file 1). The new avian gene order present in these reannotated mitogenomes is indicated as GO-P2 in Fig. 1e.

2.5. Comparison of the duplicated regions of Palaeognathae mitogenomes

The GO-I gene order found in four Palaeognathae species (Fig. 1, Table 2) is characterized generally by a high similarity between paralogous sequences, i.e. copies found within the same mitogenome. The second copies of tRNA-Pro, ND6 and tRNA-Glu are identical with the first ones in the case of *Struthio camelus*, *Rhea americana* and *Rhea pennata* (Table 3). The second copy of tRNA-Glu is also identical with the first one in *Crypturellus tataupa* mitogenome. However, the first copies of tRNA-Pro and ND6 genes of this species differ from their paralogous sequences in three nucleotides (Table 3). Two control regions of analyzed species show a slightly greater variation in identity, from 94.4% (*Rhea pennata*) to 97.8% (*Crypturellus tataupa*). The difference is mainly located at their 3' ends, except for *Rhea* taxa, whose control regions differ also at their 5' ends (Fig. S2 in Additional file 1).

Table 2

Avian mitochondrial genomes analyzed in this study. GO-I, GO-P1 and GO-P2 indicate gene orders with the duplicated region. GO-TA means the typical avian gene order without duplication. Incomplete mitogenomes are marked with an asterisk and their unknown gene order is marked with question mark.

Order	Species	Accession	Length [bp]	Gene Order
Struthioniformes	*Struthio camelus*	AF338715.1	16,595	GO-I
Rheiformes	*Rhea americana*	AF090339.1	16,704	GO-I
Rheiformes	*Rhea pennata*	AF338709.2	16,749	GO-I
Casuariiformes	*Casuarius casuarius*	AF338713.2	16,756	GO-TA
Casuariiformes	*Casuarius bennetti*	AY016011.1*	12,348	?
Casuariiformes	*Dromaius novaehollandiae*	AF338711.1	16,711	GO-TA
Aepyornithiformes	*Aepyornis sp.*	KY412176.1	16,688	GO-TA
Aepyornithiformes	*Aepyornis hildebrandti*	KJ749824.1*	15,547	?
Aepyornithiformes	*Mullerornis agilis*	KJ749825.1*	15,731	?
Apterygiformes	*Apteryx mantelli*	KU695537.1	16,694	GO-TA
Apterygiformes	*Apteryx owenii*	GU071052.1	17,020	GO-TA
Apterygiformes	*Apteryx haastii*	AF338708.2	16,980	GO-TA
Tinamiformes	*Crypturellus tataupa*	AY016012.1*	12,205	GO-I
Tinamiformes	*Eudromia elegans*	AF338710.2	18,305	GO-P1
Tinamiformes	*Tinamus guttatus*	KR149454.1	16,750	GO-TA
Tinamiformes	*Tinamus major*	AF338707.3	16,701	GO-TA
Dinornithiformes	*Anomalopteryx didiformis*	AF338714.1*	16,716	?
Dinornithiformes	*Anomalopteryx didiformis*	MK778441.1	17,043	GO-P2
Dinornithiformes	*Emeus crassus*	AF338712.1*	16,662	?
Dinornithiformes	*Emeus crassus*	AY016015.1	17,061	GO-P2
Dinornithiformes	*Dinornis giganteus*	AY016013.1	17,070	GO-P2

Table 3

Comparison of two copies of selected genes as well as control regions (CRs) in mitogenomes from five Palaeognathae taxa.

Species	Copy	Length (bp)	Percent of residues identical between two copies and number of aligned residues (in parentheses)						
tRNA-Pro	**ND6**	**tRNA-Glu**	**CR**	**tRNA-Pro**	**ND6**	**tRNA-Glu**	**CR**		
Struthio camelus	1st	70	522	68	1035	100 (70)	100 (522)	100 (68)	96.9 (1023)
2nd	70	522	68	1036					
Rhea americana	1st	70	525	69	1118	100 (70)	100 (525)	100 (69)	94.4 (1076)
2nd	70	525	69	1118					
Rhea pennata	1st	70	525	69	1103	100 (70)	100 (525)	100 (69)	94.1 (1076)
2nd	70	525	69	1183					
Crypturellus tataupa	1st	70	522	69	1059	95.7 (70)	99.4 (522)	100 (69)	97.8 (1059)
2nd	70	522	69	1196					
Eudromia elegans	1st	73	–	–	1352	80.6 (84)	–	–	98.2 (1252)
2nd	78 (ψ)	522	70	1350					

In contrast to GO-I gene order, the newly defined *Eudromia elegans* GO-P1 rearrangement is characterized by the presence of only one ND6 and tRNA-Glu gene each (Fig. 1). Moreover, the second copy of tRNA-Pro is a pseudogene, which has substantially diverged from its functional version (Fig. S5a in Additional file 1). Therefore, it seems that the GO-P1 rearrangement is a degenerated form of the GO-I gene order, in which two genes were removed and one gene was pseudogenized. Surprisingly, despite the high degree of degeneration in comparison with other analysed Palaeognathae species, two control regions of *Eudromia elegans* maintain the highest sequence identity (Table 3). However, the alignment of these regions clearly shows the presence of several deletions/insertions (Fig. S6 in Additional file 1).

The comparison of paralogous control regions in Palaeognathae taxa revealed that the second control regions (CR2s) are much longer only in two species, i.e. *Rhea pennata* and *Crypturellus tataupa* (Table 3).

Such a difference in the length of CRs seems to be a rule in most avian mitogenomes with a duplicated region [23]. Interestingly, CRs in *Rhea americana* are identical in length, while those in *Struthio camelus* and *Eudromia elegans* differ only in one and two nucleotides, respectively (Table 3).

2.6. Phylogenetic relationships within Palaeognathae based on mitogenomes

The three phylogenetic methods resulted in a consistent topology (Fig. 3). The earliest diverging lineage of Palaeognathae was *Struthio camelus* (representing Struthioniformes) and the next Rheiformes (Rheidae). Dinornithiformes (Dinornithidae + Emeidae) is grouped with Tinamiformes (Tinamidae), whereas Casuariiformes (Dromaiidae + Casuariidae) is sister to Aepyornithiformes (Aepyornithidae) + Apterygiformes (Apterygidae). Almost all nodes are very well supported. The least significant are two nodes: one clustering Casuariiformes, Aepyornithiformes and Apterygiformes, and the other encompassing the palaeognath lineages separated after the divergence of *Struthio* and *Rhea*. Nevertheless, these two nodes obtained the highest posterior probability in MrBayes analysis, i.e. 1.0 and support in the Shimodara-Hasegawa-like approximate likelihood ratio test (SH-aLRT) equal to 93 and 78, respectively.

In order to eliminate a potential artefact related with the compositional heterogeneity in the third codon positions of protein-coding genes, we created phylogenetic trees based on the RY-coding alignment (Fig. 4). The tree topology produced by the three methods was the same as that for the uncoded alignment. The posterior probability of the two controversial nodes was still very high in MrBayes tree, i.e. 0.99 and the SH-aLRT support was 89 and 82, respectively.

Moreover, we performed phylogenetic analyses based on ten alignments, from which we sequentially excluded partitions characterized by the highest substitution rate (Table S4 in Additional file 2). The calculations provided in total 16 topologies out of which 5 are worthy of mention because they were produced by many independent approaches or were obtained also in other studies (Fig. 5). The topology t1 was identical with that based on the alignments including all sites and demonstrated rheas as the sister to all other non-ostrich palaeognaths. Such a tree was produced by MrBayes, PhyloBayes and IQ-TREE using the alignment without sites characterized by the highest substitution rate, as well as by MrBayes and IQ-TREE using the alignment after removing sites with two highest rate categories. The posterior probabilities for the clade including palaeognaths other than ostrich and rheas were very high in MrBayes, i.e. 1 and 0.98, respectively, or moderate, i.e. 0.87 in PhyloBayes. In the topology t2, the *Rhea* clade was grouped with Casuariiformes + Apterygiformes. However, the support of this grouping was very weak and occurred only in MrBayes tree and IQ-TREE consensus bootstrap tree based on the alignments without seven and eight highest rate categories, respectively. A greater Bayesian support (0.95–0.97) was obtained by the node encompassing rheas with Casuariiformes in the topology t3 based on the alignments after removing three, four and five highest rate categories. This topology was also produced in MrBayes using the alignment without eight highest rate categories and in IQ-TREE for the alignments without four, five and six highest rate categories. However, the node support was generally weak. The

topology t4 was produced only by PhyloBayes for the alignments without two, three, four, five, seven and eight highest rate categories. As in the topology t1, the *Rhea* clade was also sister to all other palaeognaths excluding *Struthio*, but Casuariiformes were clustered with the rest non-ostrich palaeognaths, not directly with Aepyornithiformes and Apterygiformes. The posterior probability values of the clade including palaeognaths sister to rheas did not exceed 0.8. The topology t5 differed from the others because *Struthio camelus* was placed within other Palaeognathae and the external position was occupied by Dinornithiformes + Tinamiformes, whereas *Rhea* was grouped with Casuariiformes. This topology was obtained for the alignments without three (IQ-TREE) and six highest rate categories (MrBayes and IQ-TREE). Nevertheless, the controversial nodes were poorly supported.

Removing the sites with the highest substitution rate eliminated the alignment positions that were saturated with substitutions, but the number of parsimony informative sites decreased, too (Fig. 6a). Therefore, the stochastic error could increase for the short alignments and the inferred phylogenetic relationships could be unreliable. After elimination of sites with two highest rate categories, the mean phylogenetic distance in the MrBayes tree decreased abruptly from 0.94 to 0.33 substitutions per site and the maximum distance dropped from 1.99 to 0.69 substitutions per site (Fig. 6a). The sharp decrease was also visible in the number of informative sites, which constituted 56% of those in the original alignment. However, the sisterhood of rheas to other non-ostrich palaeognaths was still present in the trees based on the purged alignments and the latter group was relatively highly supported (Fig. 6b). After removing sites with at least three highest rate categories, the alignment was deprived of more than half of informative sites and alternative topologies were favored, though with smaller support values (Fig. 6b).

Among the applied topology tests, the BIC approximation produced all Bayesian posterior probabilities for the alternative topologies much smaller than 0.05 indicating a strong rejection of the tested alternatives (Table S5 in Additional file 2). Moreover, the topology t4 performed significantly worse than t1 in two bootstrap tests, whereas the bootstrap probabilities for the topology t2 were 0.063, i.e. very close to the 0.05 threshold. Other tests did not reject the alternative topologies. However, Bayes factor was greater than 9 indicating an overwhelming support for the topology t1 because the commonly assumed threshold for such interpretation is 5 [63].

2.7. Comparison of Palaeognathae tree topologies

All the phylogenetic analyses imply that the relationships presented in the topology t1 describe the most probable evolutionary history between the mitochondrial genomes of palaeognaths. Such relationships, but not always on the full taxa set, were also obtained in other studies based on mitochondrial genes [55, 56], selected nuclear genes [48, 54, 57], the joined set of nuclear and mitochondrial genes [46, 52, 58] as well as the concatenated alignments of many nuclear markers [45, 59, 60]. However, the application of a coalescent species tree approach on these markers and the analysis of retroelement distribution indicated a closer relationship between rheas and the clade of Casuariiformes + Apterygiformes [45, 59, 60]. This phylogeny was also generated for selected nuclear genes [53] and in a supertree approach [47]. These relationships are presented in the topology t2 but are, however, insignificant for the mitochondrial gene

set. An alternative, poorly supported topology, in which *Rhea* clustered with Tinamiformes or Dinornithiformes + Tinamiformes, was obtained for some nuclear genes [54, 57].

Two topologies, t1 and t2, aspire to be the real species tree but it is not easy to evaluate which one is true. Although t1 was found in many studies based on concatenated alignments of many markers, it has been criticized as a true species tree in favor of t2, which was obtained in coalescent-based approaches also based on huge data sets [59, 60]. Moreover, the most common gene tree topologies differed from the coalescent species tree. The topology t2 was supported by 229 loci and was only the 7th in the ranking of the most common gene tree topologies, whereas the topology t1 was the 2nd, supported by 280 loci [59]. The largest number of markers, i.e. 357, indicated another topology, in which Rheiformes was grouped with the clade of Dinornithiformes + Tinamiformes. This discrepancy was explained by the existence of an empirical anomaly zone resulting from incomplete lineage sorting (ILS) across short internal branches leading to the last common ancestor of Rheiformes, Apterygiformes and Casuariiformes [59, 60].

Although coalescent species trees account for ILS, simulations showed that species tree methods based on gene tree summation may not provide significantly better performance over concatenation alignment methods, which can perform even better [64]. Prum, et al. [45], who applied extensive avian taxon sampling and loci with slow substitution rates, found no single locus that was able to fully resolve the tree topology. They concluded that this lack of phylogenetic information can challenge the accuracy of a coalescent-based summary approach relative to concatenation. The multispecies coalescent models work under some conditions [65]. For example, they assume that incongruent gene trees are independently generated from a coalescence process occurring along the lineages of the species tree and there is no selection on the studied markers. Although Cloutier, et al. [60] and Sackton, et al. [59] analyzed noncoding nuclear elements, they were conserved or ultraconserved. Therefore, we cannot exclude that they were involved in essential regulatory functions and subjected to selection [66–68], which could influence the model assumptions.

Generally, nuclear markers are more prone to ILS and hidden gene paralogy [69–76] than mitochondrial genes, which are present in a haploid genome and maternally inherited [77]. Thus, the time needed to completely sort the ancestral polymorphism of mitochondrial DNA is on average four times smaller than for nuclear genes [78]. Introgression of mtDNA is another reason for the discrepancy between the gene and species trees. However, this process concerning maternally inherited mtDNA is restricted between heterogametic avian species because female hybrids are characterized by reduced viability [79–85]. In agreement with that, a survey of causes of mtDNA gene tree paraphyly in birds found that 8% of studied species had paraphyly attributable to incorrect taxonomy, and only 2% on account of incomplete lineage sorting, 1% due to introgressive hybridization, and 3% because of incomplete lineage sorting or hybridization [86]. Moreover, the mitochondrial genes are located on one molecule and are inherited together [74, 87], so they should bear a consistent phylogenetic signal. Accordingly, analyses based on complete mitogenomes have provided well-resolved phylogenies of various avian groups [9, 14, 40–42, 88–91]. Nevertheless, it is not inconceivable that the ILS effect can influence mtDNA in rapidly radiating taxa, in which on-going speciation occurs before genetic sorting [92].

2.8. Distribution of mitogenomic rearrangements in phylogenetic trees of Palaeognathae

We considered both t1 and t2 topologies to analyze the presence and absence of the mitogenomic duplication in the phylogenetic context. Using these relationships, we mapped the mitogenomic features onto these topologies and inferred ancestral states for the individual lineages using maximum parsimony (MP) and maximum likelihood (ML) methods (Fig. 7).

Two methods applied to the topology t1 clearly indicate that the last common ancestor of palaeognaths contained a duplicated region in its mitogenome (Fig. 7a and b). This state was inherited by the ostrich and rhea lineages. The ML method provided the probability $P > 0.982$ of this state for the last common ancestors of all palaeognaths and non-ostrich palaeognaths. The last common ancestor of the remaining groups, i.e. Dinornithiformes, Tinamiformes, Casuariiformes, Aepyornithiformes and Apterygiformes, could also contain a duplication with $P = 0.899$. However, the last common ancestor of Casuariiformes, Aepyornithiformes and Apterygiformes lost this duplication ($P = 0.914$). In turn, the last common ancestor of Dinornithiformes and Tinamiformes still had the duplicated region ($P = 0.983$), which was probably lost in *Tinamus*, whereas *Anomalopteryx*, *Dinornis* and *Emeus* maintained only a pseudogenized ND6.

According to the topology t2, the last common ancestors of all palaeognaths and non-ostrich palaeognaths also had the duplication in their mitogenomes with the probability of at least 0.982 (Fig. 7c and d). The duplication was preserved in the last common ancestor of Dinornithiformes and Tinamiformes ($P = 0.996$) as well as the last common ancestor of Rheiformes, Casuariiformes, Aepyornithiformes and Apterygiformes ($P = 0.906$). Among four latter orders, only Rheiformes maintained the duplication, whereas the last common ancestor of the other orders had a mitogenome without the duplicated region ($P = 0.914$).

One could argue that the traces of ND6 pseudogene in *Anomalopteryx*, *Dinornis* and *Emeus* are equivocal. Therefore, we conducted analyses in which we assumed that mitogenomes of these genera had already lost the duplication (Fig. S7 in Additional file 1). However, the general conclusion that the last common ancestor of all Palaeognathae had a mitogenome with a duplicated region did not change independently of the assumed topology. The probability of this state was 0.916 and 0.829 for topology t1 and t2, respectively. The last common ancestor of non-ostrich palaeognaths also had this feature with $P = 0.889$ and 0.697 for these topologies, respectively.

As discussed in the previous section, the topology t1 is highly supported by mitochondrial data, whereas t2 is supported by nuclear markers in coalescent-based approaches and is regarded as the true species tree by some authors [59, 60]. Assuming that the t1 presents the real mitogenomic history, we superimposed the mitogenome phylogeny onto the potential species tree (Fig. 8). In order to reconcile the alternative positions of rheas in these topologies, we should assume that there existed a heteroplasmy, i.e. at least two types of mitochondrial genomes, in the last common ancestor of non-ostrich palaeognaths. Both mitogenomes probably initially contained a duplication, as indicated by the inferred

ancestral states. The lineages of these mitogenomes are marked in Fig. 8 as 1 and 2. Dinornithiformes and Tinamiformes inherited only mitogenome 2, whose duplicated regions had begun to fade and likely disappeared in *Tinamus*. However, the common ancestor of Rheiformes, Casuariiformes, Aepyornithiformes and Apterygiformes preserved two mitogenomes, but genome 2 lost the duplicated region during the course of evolution. Then, the mitochondrial lineages were segregated: mitogenome 1 with the duplication was left in rheas, whereas mitogenome 2 without the duplication was passed to Casuariiformes, Aepyornithiformes and Apterygiformes.

The assumption about the presence of heteroplasmy in some period of Palaeognathae evolution seems probable because such genomic diversity has also been reported in various avian groups: Accipitriformes [5], Bucerotiformes [6], Charadriiformes [7, 93, 94], Ciconiiformes [95], Columbiformes [96], Gruiformes [97], Passeriformes [98, 99], Pelecaniformes [16, 100], Piciformes [4], Procellariiformes [28], Psittaciformes [101], Sphenisciformes [102], Strigiformes [24] and Suliformes [25]. This period was likely very short. According to the results of molecular dating performed by Kimball, et al. [47], the time elapsed since the divergence of the clade Dinornithiformes + Tinamiformes and that including Rheiformes, Casuariiformes, Aepyornithiformes and Apterygiformes to the separation of Rheiformes from the three latter orders was only 3.6 million years.

2.9. Implications for mitogenome evolution in all Aves

The finding that a palaeognath ancestor contained a mitogenomic duplication challenges the common assumption that this feature evolved independently in individual lineages of Neognathae, i.e. the sister group of Palaeognathae [13, 22, 29–31]. Increasing data indicate that mitogenomes with a duplicated region are present in all or a vast majority of representatives of diverse Neognathae lineages, i.e. Accipitriformes [4, 5, 36–38], Falconiformes [4, 39, 103], Gruiformes [12], Pelecaniformes [4, 15, 16], Psittaciformes [23] and Suliformes [25]. It suggests that ancestors of these groups could have also had a mitogenomic duplication. The presence of this state was recently inferred for the last common ancestor of three closely related orders, Falconiformes, Passeriformes and Psittaciformes [14]. Therefore, it is interesting to consider if this feature was present much earlier in the evolution of birds or even in the last common ancestor of all known Aves.

The current data indicate that mitogenomes with duplications are distributed in 25 out of 43 avian orders (Table S6 in Additional file 2). Among 852 species with known mitogenomes, 262 have a duplication, 532 do not contain this feature and 58 are too incomplete to classify them into one of these categories. However, the number of the former group is likely underestimated because of difficulties with the amplification and sequencing of repeated regions [4]. Accordingly, reanalysis of 13 crane mitogenomes, previously annotated without the duplication, showed that all of them contain in fact the duplicated region [12]. Similarly, 15 mitogenomes of parrots from Cacatuidae and Nestoridae also revealed this character after re-examination using appropriate PCR and sequencing methods [23].

Another important example is the order Chardriiformes. So far, only two representatives have been shown to have duplication in mitochondrial genomes: *Calidris pugnax* (GQ255993.1) and *Turnix velox*

(MK453380.1). However, annotation of another mitogenome from *Alca torda* (CM018102.1) revealed the presence of the most fully duplicated avian region (GO-FD) (Fig. 1 and Fig. S8 in Additional file 1). Therefore, we surveyed mitogenomes of 20 additional Charadriiformes annotated without duplication to identify potentially unrecognized duplication of the mitogenomes. Using diagnostic PCR reactions amplifying a fragment between two control regions, we found GO-FD gene order in 17 examined species (data not shown; paper in preparation). The obtained results indicate that underestimation of the mitogenomes with the duplication ranges from 85–100% in the case of Charadriiformes, Gruidae and Psittaciformes. It should be emphasized that the omission of GO-FD gene order can be common in the amplification and sequencing of avian mitogenomes using standard procedures due to the presence of two nearly identical copies. Applying other methods, we obtained previously omitted sequences of the GO-FD rearrangement in mitogenomes of five additional avian orders: Cathartiformes (*Cathartes aura*), Ciconiiformes (*Ciconia nigra*), Gaviiformes (*Gavia arctica*), Podicipediformes (*Podiceps cristatus*) and Sphenisciformes (*Spheniscus demersus*) (Fig. S9 in Additional file 1). Furthermore, *ab initio* annotation of *Calypte anna* (Apodiformes) and *Puffinus lherminieri* (Procellariiformes) mitogenomes deposited in GenBank database revealed the presence of GO-FD (Fig. S8 in Additional file 1). The same gene order has been identified in mitogenomes of *Morus serrator* (Suliformes) as well as *Ketupa blackistoni* and *Ketupa flavipes* (Strigiformes) after re-annotation of their duplicated gene rearrangements (Fig. S8 in Additional file 1).

The gathered data (Table S6 in Additional file 2) were used to reconstruct the evolution of mitogenomic duplications in all birds. We mapped the data onto the phylogenetic tree of Aves obtained recently in a supertree approach by Kimball, et al. [47]. If we assume that underestimation of mitogenomes with the duplication is only 48% (much smaller than that above-mentioned), the total number of such mitogenomes in the avian orders already containing at least one mitogenome of this type would exceed the number of mitogenomes without the duplication. It would strongly suggest that the last common ancestor of these orders also contained the mitogenomic duplication in the past. For the orders without this feature, we assumed that they were ancestrally deprived of it. Using these assumptions, we reconstructed ancestral states in the Aves tree (Fig. 9). Both MP and ML methods indicated that the last common ancestor of all Aves contained a mitogenome with the duplication with the probability of 0.776. This state was passed to the ancestors of the main lineages leading to modern orders. The probability of this feature was still above 0.5 for the last common ancestors of Palaeognathae (0.951), Neornithes (0.621), Neoaves (0.760) and later diverged clades (> 0.935). Within Neornithes, the duplication was lost five times independently in the ancestors of Galloanserae, Columbiformes, Otidiformes, Eurypygiformes + Phaethontiformes and Trogoniformes. A similar conclusion can be drawn for the tree topology obtained by Prum, et al. [45] (Fig. S10 in Additional file 1), under which the last common ancestor of all Aves could have had a mitogenome with duplication (P = 0.685), which was also inherited by ancestors of deeply diverged lineages. This state was lost at least four times in Neornithes: Galloanserae, Columbiformes + Otidiformes, Eurypygiformes + Phaethontiformes and Trogoniformes. Cuculiformes would recover the duplication again under this scenario.

This view about the ancestry of the mitogenomic duplication can be further supported by the distribution of the most fully duplicated avian region (GO-FD). It includes the repetition of Cytb/tRNA-Thr/tRNA-Pro/ND6/tRNA-Glu/CR in which only the cytochrome b gene is pseudogenized (Fig. 1). This rearrangement type occurs in 55 mitogenomes distributed among 14 bird orders (Fig. 9, Table S7 in Additional file 2). The length and complexity of this duplication suggests that it is unlikely that it occurred independently at least 8 times (see Fig. 9) because it would require the same recombination pattern and replication errors [12, 16, 25, 104]. Thus, it seems more probable that this state was inherited from shared ancestors by the avian lineages that contain this rearrangement type. In other bird groups, the duplicated regions were subjected to degenerations and lost selected elements.

3. Conclusions

The obtained results indicate that duplicated control regions with adjacent genes are more common in Palaeognathae mitochondrial genomes than it was previously thought. What is more, this feature was most likely present in the last common ancestor of this avian group. Two duplicated control regions can lead to a more effective initiation of replication or transcription and a greater number of replicating mitogenomes per organelle, which may increase energy production by mitochondria [105–109]. In support of this, it has been found in parrots that keeping two copies of control region is associated with morphological features related to more energetically costly active flight [23]. Following this finding, we can infer that the last common ancestor of palaeognaths was volant because mapping of mitogenomic features onto the phylogenetic trees suggested that the ancestor most likely contained the duplicated regions. This hypothesis corresponds well to recently proposed scenarios for the evolutionary history of Palaeognathae, which may have originated in the Late Cretaceous in the Northern Hemisphere [58]. Around the Cretaceous-Paleogene boundary, they may have migrated to the Southern Hemisphere, where they dispersed widely and diversified. They extended their distribution by long-distance overseas dispersal to the Gondwana-derived landmasses, such as New Zealand and Madagascar [58]. These long-distance dispersals must have been accompanied by very good adaptations to active flight. However, modern palaeognath species are flightless or at most poor flyers (tinamous). Therefore, the maintenance of two control regions in some palaeognaths may be the legacy of their ancestors. Alternatively, these regions can still provide benefits even for flightless species during long-distance walking and running, which require a lot of energy.

Once the duplication occurred in a mitogenome of the ancestral palaeognath, it was preserved during the evolution of Struthioniformes, Rheiformes and some Tinamiformes. The duplicated regions were subjected to concerted evolution as in many other avian groups [4, 6, 14, 23, 25, 28, 30, 101, 103–106, 110–112]. This process resulted in homogenization of some parts of duplicated regions and degeneration of others. Reconciliation of the mitogenome-based phylogenetic tree with a probable species tree based on nuclear markers suggests that the evolution of mitochondrial genomes in Palaeognaths could have involved the presence of different mitogenomic variants in one cell, i.e. heteroplasmy, in a short period of time.

The distribution of mitogenomes with duplications across the avian phylogenetic tree suggests that the last common ancestors of not only Palaeognathae but also major Neognathae groups and even all known Aves could have had a mitogenomic duplication, which was then inherited by many modern lineages. Whether the duplicated region is only a neutral trait or can provide a real selective advantage is an interesting question. Previous studies have suggested that bird species having mitogenomes with the duplicated control region can be characterized by longer life-span [23, 113] as well as a greater metabolic rate and energy production [14, 23]. However, further understanding the potential selective advantages of mitogenome duplications will require direct study at physiological and molecular levels.

4. Materials And Methods

4.1. Samples and DNA extraction

Blood samples from Palaeognathae species were obtained from Polish zoological gardens (Table 1). They were taken as dry blood spots on a fiber filter dedicated to laboratory analyses and were preserved at -20°C in Eppendorf tubes sealed with parafilm to avoid damping. Total DNA was extracted with Sherlock AX Kit (A&A Biotechnology) according to the manufacturer's protocol.

4.2. PCR strategy for tandem duplication survey

To verify the presence or absence of tandem duplication within the mitochondrial genomes of Palaeognathae representatives, we used the strategy proposed by Gibb, et al. [4] and recently successfully applied to parrot mitogenomes [23]. Because two control regions (CRs) are common in the majority of avian duplicated rearrangements [22] and the paralogous CRs are usually nearly identical [12, 101], we designed appropriate primers D-F and D-R that were to anneal to the central parts of CRs (Table S1 in Additional file 2, Fig. 2). This property makes the PCR strategy diagnostic because the expected amplicons occur only when two control regions are present in the genome. Due to the high variability of the control region sequences from the analyzed taxa, it was impossible to use universal primers, which forced us to design primers specific to each genus or even species (Table S1 in Additional file 2). Based on the selected primers, we ran 16 different reactions for *Struthio camelus* and *Rhea americana*, 48 reactions for *Rhea pennata*, 12 reactions for *Crypturellus tataupa*, 25 reactions for *Dromaius novaehollandiae* and 56 reactions for *Casuarius casuarius* (Table S1 in Additional file 2).

4.3. PCR strategy for amplification of the mitogenomic fragments containing the whole duplicated regions

The diagnostic fragment designed for tandem duplication survey comprises incomplete control regions, i.e. the second part of CR1 and the first part of CR2, as well as genes located between the two CRs. Therefore, in the case of taxa for which such a fragment was obtained, appropriate PCR reactions were performed to complete the missing parts of CRs and to reveal the order of genes preceding the CR1. The obtained partial CR1 sequences were used to design species specific CR-R primers for amplification of tRNA-Leu/CR1 fragment (for *Struthio camelus*) or ND5/CR1 fragments (for *Rhea americana*, *Rhea*

pennata, *Crypturellus tataupa*) (Fig. 2). Similarly, the partial CR2 sequences were used to design species specific CR-F primers for amplification of CR2/16S fragments (for *Struthio camelus*, *Rhea americana*, *Rhea pennata*) or CR2/12S fragment (for *Crypturellus tataupa*) (Fig. 2). Appropriate L-F, ND5-F, 12S-R and 16S-R primers were designed based on reference mitogenomic sequences of the analyzed or related taxa deposited in GenBank (AF338715.1, AF090339.1, AF338709.2, AF338710.2). Suitable elongation times were applied to avoid amplification of tRNA-Leu/CR2, ND5/CR2, CR1/12S and CR1/16S fragments, which would contain two copies of some genes and/or control regions. Additional diagnostic ND6-1/ND6-2 fragments were amplified to confirm that the CR1/CR2 sequences were not errors of the PCR reactions or copies present in the nuclear genome, i.e. nuclear mitochondrial DNA inserts (NUMTs). Finally, the whole duplicated mitogenomic regions of *Struthio camelus*, *Rhea americana*, *Rhea pennata* and *Crypturellus tataupa* were amplified in four overlapping fragments (Fig. 2). Both amplicons containing only one control region fragment (CR1 or CR2) were 3 kb to 5 kb in length, which excludes a possibility of NUMTs amplification, whose average size is usually below 1 kb [114]. Despite the length of two diagnostic fragments depends on elements (genes and/or control region) located between CRs or ND6 genes, their length was longer than 1 kb, i.e. 2–3 kb on average. Uncropped and unprocessed agarose gels were presented in Fig. S11 in Additional file 1.

4.4. DNA amplification and sequencing

The PCR amplifications were performed in 25 µl reaction mixture containing 50 ng of the DNA template, 1U DreamTaq Green DNA Polymerase (Thermo Fisher Scientific), 2.5 µl of 10 x buffer, 0.6 µl of 10 mM dNTPs, and 0.6 µl of each primer (10 µM). In the case of diagnostic fragments (CR1/CR2, ND6-1/ND6-2), following program was used: 94°C for 5 min; 94°C for 30 sec, 58°C for 30 sec, 72°C for 120 sec repeated 35 times; and 72°C for 5 min. In the case of all other fragments the reaction conditions were as follows: 94°C for 5 min; 94°C for 30 sec, 58°C for 30 sec, 72°C for 180 sec repeated 35 times; and 72°C for 5 min. For each amplified fragment, the appropriate amount of the PCR reaction mixtures was cleaned with the use of Clean-up Kit (A&A Biotechnology) to obtain the final volume of 100 µl with the concentration of at least 50 ng/µl. The two DNA strands of the cleaned PCR products were sequenced using the Primer Walking method (Wyzer Biosciences Inc., Cambridge, MA). Overlaps between four fragments amplified for each species were sufficient to assemble the whole mitogenomic regions containing duplicated elements with the use of appropriate reference mitogenomes of *Notiomystis cincta* and *Turdus philomelos*. The annotation of genes was performed in MITOS [115].

Long control regions were searched for potential pseudogenes using the optimal global:local algorithm (glsearch) form FASTA package version 36.3.8 g [116]. We assumed 1 million of shuffles, match/mismatch scores 5/-4 and gap cost opening/extension − 10/-1.

4.5. Phylogenetic analyses

Phylogenetic relationships between Palaeognathae were inferred based on all available 19 complete or almost complete mitochondrial genomes (Table 2). Five representatives of Neognathae were used as an outgroup: *Anas platyrhynchos* (NC_009684.1), *Anseranas semipalmata* (NC_005933.1), *Crax daubentoni*

(NC_024617.1), *Gallus gallus* (NC_001323.1) and *Numida meleagris* (NC_034374.1). The sequence records were downloaded from the GenBank database. The analyses were based on nucleotide sequences of 13 protein coding genes, 12S and 16S rRNAs, as well as 22 tRNAs. Sequences of the control region (CR) were excluded due to their high variation. The sequences were aligned in MAFFT using a slow and accurate algorithm L-INS-i with 1,000 cycles of iterative refinement [117]. The resulted alignments were edited manually in JalView [118] and sites suitable for phylogenetic study were selected in GBlocks [119]. The concatenated alignment of mitochondrial genes consisted of 15,351 bp.

We applied three phylogenetic approaches in the phylogenetic analyses: the maximum likelihood method in IQ-TREE [120], as well as Bayesian analyses in MrBayes [121] and PhyloBayes [122]. We considered 63 potential partitions, i.e. three codon positions for each individual protein coding gene and separate partitions for each of the RNA genes to test the necessity of using separate substitution models.

The ModelFinder program associated with IQ-TREE [123, 124], proposed 12 partitions with individual substitution models (Table S8 in Additional file 2). In the tree search, we used the more thorough and slower NNI search. In IQ-TREE, we applied Shimodara-Hasegawa-like approximate likelihood ratio test (SH-aLRT) assuming 10,000 replicates and non-parametric bootstrap with 1,000 replicates.

In MrBayes, we assumed 11 substitution models for the appropriate partitions as proposed by PartitionFinder [125]. However, we implemented mixed models rather than fixed ones to specify appropriate substitution models across the large parameter space [126], but the assumptions on the heterogeneity rate across sites were adopted from PartitionFinder results (Table S4 in Additional file 2). Two independent runs each using 8 Markov chains were applied and trees were sampled every 100 generations for 20,000,000 generations. In the final analysis, we selected trees from the last 12,866,000 generations that reached the stationary phase and convergence, i.e. when the standard deviation of split frequencies (SD) stabilized and was below 0.00004.

In PhyloBayes, we applied the CAT-GTR + Γ5 model with the number of components, weights and profiles of the model inferred from the data. Two independent Markov chains were run for 100,000 generations with one tree sampled for each generation. The last 50,000 trees from each chain were collected to compute posterior consensus trees after reaching convergence, when the largest discrepancy observed across all bipartitions (maxdiff) was 0.0055.

To reduce a potential compositional heterogeneity in sequences related with AT or GC bias, we recoded respective nucleotides for purines (R) and pyrimidines (Y) in the third codon positions of protein-coding genes. In MrBayes and PhyloBayes, we adopted the assumptions as described above, whereas in IQ-TREE, we applied 11 partitions with individual substitution models proposed by ModelFinder (Table S9 in Additional file 2). The posterior consensus was calculated for trees from the last 5,845,000 generations in MrBayes for SD < 0.00013 and 90,000 generations in PhyloBayes for maxdiff = 0.007.

Besides the study based on the full alignment, we also performed phylogenetic analyses using ten alignments with sequentially removed partitions characterized by the highest substitution rate. The

partition-specific rate was taken from MrBayes estimations. The settings for this study were the same as above. In the case of IQ-TREE analyses, the best substitution models were calculated for each set every time.

The best tree produced in IQ-Tree was compared in Consel [127] with alternative topologies using available tests. We applied 1,000,000 replicates and site-wise log-likelihoods for the compared trees calculated in IQ-Tree under the best fitted substitution models. Moreover, we analyzed competitive topologies using Bayes Factor in MrBayes based on the stepping-stone method estimating the mean marginal likelihood using 50 steps of the sampling algorithm and other parameters as described above.

The data about the presence and absence of duplication in the palaeognath mitogenomes were mapped on the phylogenetic tree using Mesquite [128]. The lack of data about the duplication was coded as missing data. We applied maximum parsimony and maximum likelihood reconstruction methods. Depending on the AIC criterion, we used the better fit model for the data: either the Mk1 model (Markov k-state 1 parameter model) or the AsymmMk model (asymmetrical Markov k-state 2 parameter model).

Abbreviations

CR
control region
CR1
the first copy of control region
CR2
the second copy of control region
GO
gene order
GO-FD
fully duplicated gene order
GO-P
Palaeognathae gene order
ILS
incomplete lineage sorting
ML
maximum likelihood
MP
maximum parsimony
NUMTs
nuclear mitochondrial DNA inserts
SH-aLRT
Shimodara-Hasegawa-like approximate likelihood ratio test

Declarations

Ethics approval and consent to participate

No permissions were required for the analyzed specimens because the blood samples were not collected specifically for the purpose of this study. They were obtained by veterinarians during standard administering veterinary care and periodic health examinations of birds kept in Polish Zoological Gardens (Gdańsk, Katowice, Warsaw, Wrocław) and Animal Rehabilitation Centre (Kątna).

Consent for publication

Not applicable.

Availability of data and materials section

All the data generated or analyzed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Science Centre Poland (Narodowe Centrum Nauki, Polska) under Grant no. 2015/17/B/NZ8/02402 and 2017/25/N/NZ8/01197. Some computations were carried out at the Wrocław Center for Networking and Supercomputing under the grant no. 307.

Authors' contribution

Conceived and designed the study: ADU, PM. Performed molecular experiments: ADU, AK. Performed bioinformatic analyses: PM, AK. Analyzed the data: ADU, PM, AK. All authors wrote the paper. All authors participated in the improvement of the manuscript and approved the final version.

Acknowledgements

We are very grateful to Daniel Field and Albert Chen for thorough English editing, which significantly improved the manuscript. We are grateful to: Andrzej Kruszewicz, Marcin Chrapowicki and Ryszard Topola (Zoological Garden in Warsaw); Radosław Ratajszczak, Mirosław Piasecki and Krzysztof Kałużny (Zoological Garden in Wrocław); Michał Targowski and Michał Krause (Zoological Garden in Gdańsk); Jolanta Kopiec and Michał Załuski (Zoological Garden in Katowice); Tomasz Grabiński (Animal Rehabilitation Center, Kątna) and Grzegorz Zaniewicz (University of Gdańsk, Department of Vertebrate Ecology and Zoology, Avian Ecophysiology Unit) for providing samples of examined species. We are also grateful to Mr. Jan Kieleczawa (Wyzer Biosciences Inc., Cambridge, MA) for patience and commitment in sequencing of PCR products rich in G and/or C repeats.

References

1. Lavrov DV. Key transitions in animal evolution: a mitochondrial DNA perspective. Integrative comparative biology. 2007;47(5):734–43.
2. Lavrov DV: **Key Transitions in Animal Evolution**: **a Mitochondrial DNA Perspective**. In: *Key Transitions in Animal Evolution.* Edited by Desalle R, Schierwater B. Boca Raton: CRC Press; 2011: 34–53.
3. Desjardins P, Morais R. Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. Journal of molecular biology. 1990;212(4):599–634.
4. Gibb GC, Kardailsky O, Kimball RT, Braun EL, Penny D. Mitochondrial genomes and avian phylogeny: Complex characters and resolvability without explosive radiations. Molecular biology evolution. 2007;24(1):269–80.
5. Roques S, Godoy JA, Negro JJ, Hiraldo F. Organization and variation of the mitochondrial control region in two vulture species, Gypaetus barbatus and Neophron percnopterus. J Hered. 2004;95(4):332–7.
6. Sammler S, Bleidorn C, Tiedemann R. Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. BMC Genomics. 2011;12:35.
7. Verkuil YI, Piersma T, Baker AJ. A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes). Mol Phylogenet Evol. 2010;57(1):411–6.
8. Huang R, Zhou YY, Yao YF, Zhao B, Zhang YL, Xu HL. Complete mitochondrial genome and phylogenetic relationship analysis of Garrulax affinis (Passeriformes, Timaliidae). Mitochondrial DNA A. 2016;27(5):3502–3.
9. Pacheco MA, Battistuzzi FU, Lentino M, Aguilar RF, Kumar S, Escalante AA. Evolution of Modern Birds Revealed by Mitogenomics: Timing the Radiation and Origin of Major Orders. Molecular biology evolution. 2011;28(6):1927–42.
10. Pratt RC, Gibb GC, Morgan-Richards M, Phillips MJ, Hendy MD, Penny D. Toward resolving deep neoaves phylogeny: data, signal enhancement, and priors. Molecular biology evolution. 2009;26(2):313–26.
11. Wang N, Liang B, Huo J, Liang W. Complete mitochondrial genome and the phylogenetic position of the Lesser Cuckoo, Cuculus poliocephalus (Ayes: Cuculiformes). Mitochondrial DNA A. 2016;27(6):4409–10.
12. Akiyama T, Nishida C, Momose K, Onuma M, Takami K, Masuda R. Gene duplication and concerted evolution of mitochondrial DNA in crane species. Mol Phylogenet Evol. 2017;106:158–63.
13. Caparroz R, Rocha AV, Cabanne GS, Tubaro P, Aleixo A, Lemmon EM, Lemmon AR: **Mitogenomes of two neotropical bird species and the multiple independent origin of mitochondrial gene orders in Passeriformes**. *Molecular biology reports* 2018.
14. Mackiewicz P, Urantowka AD, Kroczak A, Mackiewicz D. Resolving Phylogenetic Relationships within Passeriformes Based on Mitochondrial Genes and Inferring the Evolution of Their Mitogenomes in

Terms of Duplications. Genome Biol Evol. 2019;11(10):2824–49.

15. Rodrigues P, A'lvarez P, Verdugo C. Complete mitochondrial genome of the Neotropic cormorant (Phalacrocorax brasilianus). Mitochondrial DNA A. 2017;28(3):401–2.
16. Zhou X, Lin Q, Fang W, Chen X. The complete mitochondrial genomes of sixteen ardeid birds revealing the evolutionary process of the gene rearrangements. BMC Genomics. 2014;15:573.
17. Luo X, Kang XB, Zhang DY. Complete mitochondrial genome of the American flamingo, Phoenicopterus ruber (Phoenicopteriformes, Phoenicopteridae). Mitochondrial DNA A. 2016;27(5):3519–20.
18. Morgan-Richards M, Trewick SA, Bartosch-Haerlid A, Kardailsky O, Phillips MJ, McLenachan PA, Penny D. Bird evolution: testing the Metaves clade with six new mitochondrial genomes. BMC evolutionary biology. 2008;8:20.
19. Fuchs J, Pons JM, Pasquet E, Bonillo C. Complete mitochondrial genomes of the white-browed piculet (Sasia ochracea, Picidae) and pale-billed woodpecker (Campephilus guatemalensis, Picidae). Mitochondrial DNA A. 2016;27(5):3640–1.
20. Gibb GC, Kennedy M, Penny D. Beyond phylogeny: pelecaniform and ciconiiform birds, and long-term niche stability. Mol Phylogenet Evol. 2013;68(2):229–38.
21. Lounsberry ZT, Brown SK, Collins PW, Henry RW, Newsome SD, Sacks BN. Next-generation sequencing workflow for assembly of nonmodel mitogenomes exemplified with North Pacific albatrosses (Phoebastria spp.). Mol Ecol Resour. 2015;15(4):893–902.
22. Eberhard JR, Wright TF. Rearrangement and evolution of mitochondrial genomes in parrots. Mol Phylogenet Evol. 2016;94(Pt A):34–46.
23. Urantowka AD, Kroczak A, Silva T, Padron RZ, Gallardo NF, Blanch J, Blanch B, Mackiewicz P. New insight into parrots' mitogenomes indicates that their ancestor contained a duplicated region. Molecular biology evolution. 2018;35(12):2989–3009.
24. Hanna ZR, Henderson JB, Sellas AB, Fuchs J, Bowie RCK, Dumbacher JP. Complete mitochondrial genome sequences of the northern spotted owl (Strix occidentalis caurina) and the barred owl (Strix varia; Aves: Strigiformes: Strigidae) confirm the presence of a duplicated control region. PeerJ. 2017;5:e3901.
25. Morris-Pocock JA, Taylor SA, Birt TP, Friesen VL. Concerted evolution of duplicated mitochondrial control regions in three related seabird species. BMC evolutionary biology. 2010;10:14.
26. Haddrath O, Baker AJ: **Complete mitochondrial DNA genome sequences of extinct birds**: **ratite phylogenetics and the vicariance biogeography hypothesis**. *Proceedings Biological sciences / The Royal Society* 2001, **268**(1470):939–945.
27. Zhang LT, Zhang M, He SY. The complete mitochondrial genome of great cormorant, Phalacrocorax carbo (Phalacrocorax, Phalacrocoracidae). Mitochondrial DNA A. 2017;28(1–2):1–2.
28. Abbott CL, Double MC, Trueman JW, Robinson A, Cockburn A. An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in Thalassarche albatrosses. Molecular ecology. 2005;14(11):3605–13.

29. Mindell DP, Sorenson MD, Dimcheff DE. Multiple independent origins of mitochondrial gene order in birds. Proc Natl Acad Sci USA. 1998;95(18):10693–7.
30. Schirtzinger EE, Tavares ES, Gonzales LA, Eberhard JR, Miyaki CY, Sanchez JJ, Hernandez A, Mueller H, Graves GR, Fleischer RC, et al. Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. Mol Phylogenet Evol. 2012;64(2):342–56.
31. Alstrom P, Ericson PG, Olsson U, Sundberg P. Phylogeny and classification of the avian superfamily Sylvioidea. Mol Phylogenet Evol. 2006;38(2):381–97.
32. Boore JL. The use of genome-level characters for phylogenetic reconstruction. Trends Ecol Evol. 2006;21(8):439–46.
33. Boore JL, Brown WM. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr Opin Genet Dev. 1998;8(6):668–74.
34. Boore JL, Fuerstenberg SI. Beyond linear sequence comparisons: the use of genome-level characters for phylogenetic reconstruction. Philos T R Soc B. 2008;363(1496):1445–51.
35. Rokas A, Holland PW. Rare genomic changes as a tool for phylogenetics. Trends Ecol Evol. 2000;15(11):454–9.
36. Qin XM, Guan QX, Shi JP, Hou LX, Qin PS. Complete mitochondrial genome of the Spilornis cheela (Falconiformes, Accipitridae): Comparison of S. cheela and Spizaetus alboniger. Mitochondrial DNA. 2013;24(3):255–6.
37. Jiang L, Chen J, Wang P, Ren QQ, Yuan J, Qian CJ, Hua XH, Guo ZC, Zhang L, Yang JK, et al: **The Mitochondrial Genomes of Aquila fasciata and Buteo lagopus** (**Aves**, **Accipitriformes**): **Sequence**, **Structure and Phylogenetic Analyses** (**vol 10**, **e0136297**, 2015). *PloS one* 2015, **10**(10):e0141037.
38. Liu G, Li C, Du Y, Liu X. The complete mitochondrial genome of Japanese sparrowhawk (Accipiter gularis) and the phylogenetic relationships among some predatory birds. Biochem Syst Ecol. 2017;70:116–25.
39. Sveinsdottir M, Guomundsdottir L, Magnusson KP. Complete mitochondrial genome of the gyrfalcon Falco rusticolus (Aves, Falconiformes, Falconidae). Mitochondrial DNA A. 2017;28(3):370–1.
40. Slack KE, Delsuc F, McLenachan PA, Arnason U, Penny D. Resolving the root of the avian mitogenomic tree by breaking up long branches. Mol Phylogenet Evol. 2007;42(1):1–13.
41. Gibb GC, England R, Hartig G, McLenachan PA, Smith BLT, McComish BJ, Cooper A, Penny D. New Zealand Passerines Help Clarify the Diversification of Major Songbird Lineages during the Oligocene. Genome Biol Evol. 2015;7(11):2983–95.
42. Barker FK. Mitogenomic data resolve basal relationships among passeriform and passeridan birds. Mol Phylogenet Evol. 2014;79:313–24.
43. Krajewski C, Sipiorski JT, Anderson FE. Complete Mitochondrial Genome Sequences and the Phylogeny of Cranes (Gruiformes: Gruidae). Auk. 2010;127(2):440–52.
44. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science.

2014;346(6215):1320–31.

45. Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature. 2015;526(7574):569–73.
46. Burleigh JG, Kimball RT, Braun EL. Building the avian tree of life using a large-scale, sparse supermatrix. Mol Phylogenet Evol. 2015;84:53–63.
47. Kimball RT, Oliveros CH, Wang N, White ND, Barker FK, Field DJ, Ksepka DT, Chesser RT, Moyle RG, Braun MJ, et al. A Phylogenomic Supertree of Birds. Diversity. 2019;11(7):109.
48. Hackett SJ, Kimball RT, Reddy S, Bowie RC, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han KL, Harshman J, et al. A phylogenomic study of birds reveals their evolutionary history. Science. 2008;320(5884):1763–8.
49. Gill F, Donsker D. **IOC World Bird List** (**v9.2**). https://wwwworldbirdnamesorg/ 2019.
50. Clements JF, Schulenberg TS, Iliff MJ, Roberson D, Fredericks TA, Sullivan BL, Wood CL: **The eBird/Clements checklist of birds of the world**: **v**2018. http://wwwbirdscornelledu/clementschecklist/download/ 2018.
51. Baker AJ, Haddrath O, McPherson JD, Cloutier A. Genomic support for a moa-tinamou clade and adaptive morphological convergence in flightless ratites. Molecular biology evolution. 2014;31(7):1686–96.
52. Grealy A, Phillips M, Miller G, Gilbert MTP, Rouillard JM, Lambert D, Bunce M, Haile J. Eggshell palaeogenomics: Palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (Aepyornis sp.) eggshell. Mol Phylogenet Evol. 2017;109:151–63.
53. Haddrath O, Baker AJ: **Multiple nuclear genes and retroposons support vicariance and dispersal of the palaeognaths**, **and an Early Cretaceous origin of modern birds**. *Proceedings Biological sciences / The Royal Society* 2012, **279**(1747):4617–4625.
54. Harshman J, Braun EL, Braun MJ, Huddleston CJ, Bowie RCK, Chojnowski JL, Hackett SJ, Han KL, Kimball RT, Marks BD, et al. Phylogenomic evidence for multiple losses of flight in ratite birds. Proc Natl Acad Sci USA. 2008;105(36):13462–7.
55. Mitchell KJ, Llamas B, Soubrier J, Rawlence NJ, Worthy TH, Wood J, Lee MS, Cooper A. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. Science. 2014;344(6186):898–900.
56. Phillips MJ, Gibb GC, Crimp EA, Penny D. Tinamous and Moa Flock Together: Mitochondrial Genome Sequence Analysis Reveals Independent Losses of Flight among Ratites. Syst Biol. 2010;59(1):90–107.
57. Smith JV, Braun EL, Kimball RT. Ratite nonmonophyly: independent evidence from 40 novel Loci. Syst Biol. 2013;62(1):35–49.
58. Yonezawa T, Segawa T, Mori H, Campos PF, Hongoh Y, Endo H, Akiyoshi A, Kohno N, Nishida S, Wu J, et al. Phylogenomics and Morphology of Extinct Paleognaths Reveal the Origin and Evolution of the

Ratites. Current biology: CB. 2017;27(1):68–77.

59. Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, et al. Convergent regulatory evolution and loss of flight in paleognathous birds. Science. 2019;364(6435):74–8.
60. Cloutier A, Sackton TB, Grayson P, Clamp M, Baker AJ, Edwards SV. Whole-genome analyses resolve the phylogeny of flightless birds (Palaeognathae) in the presence of an empirical anomaly zone. Syst Biol. 2019;68(6):937–55.
61. Harrison GL, McLenachan PA, Phillips MJ, Slack KE, Cooper A, Penny D. Four new avian mitochondrial genomes help get to basic evolutionary questions in the late Cretaceous. Molecular biology evolution. 2004;21(6):974–83.
62. Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature. 2001;409(6821):704–7.
63. Kass RE, Raftery AE. Bayes Factors. J Am Stat Assoc. 1995;90(430):773–95.
64. Tonini J, Moore A, Stern D, Shcheglovitova M, Orti G. **Concatenation and Species Tree Methods Exhibit Statistically Indistinguishable Accuracy under a Range of Simulated Conditions**. *PLoS currents* 2015, 7.
65. Liu L, Wu SY, Yu LL. Coalescent methods for estimating species trees from phylogenomic data. Journal of Systematics Evolution. 2015;53(5):380–90.
66. Chiang CW, Liu CT, Lettre G, Lange LA, Jorgensen NW, Keating BJ, Vedantam S, Nock NL, Franceschini N, Reiner AP, et al. Ultraconserved elements in the human genome: association and transmission analyses of highly constrained single-nucleotide polymorphisms. Genetics. 2012;192(1):253–66.
67. Katzman S, Kern AD, Bejerano G, Fewell G, Fulton L, Wilson RK, Salama SR, Haussler D. Human genome ultraconserved elements are ultraselected. Science. 2007;317(5840):915.
68. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 2006;444(7118):499–502.
69. Kuraku S. Palaeophylogenomics of the Vertebrate Ancestor-Impact of Hidden Paralogy on Hagfish and Lamprey Gene Phylogeny. Integrative comparative biology. 2010;50(1):124–9.
70. Feiner N, Begemann G, Renz AJ, Meyer A, Kuraku S. The origin of bmp16, a novel Bmp2/4 relative, retained in teleost fish genomes. BMC evolutionary biology. 2009;9:277.
71. Kuraku S. Impact of asymmetric gene repertoire between cyclostomes and gnathostomes. Semin Cell Dev Biol. 2013;24(2):119–27.
72. Gribaldo S, Philippe H. Ancient phylogenetic relationships. Theor Popul Biol. 2002;61(4):391–408.
73. Maddison WP. Gene trees in species trees. Syst Biol. 1997;46(3):523–36.
74. Moore WS. Inferring phylogenies from mtDNA variation - mitochondrial-gene trees versus nuclear-gene trees. Evolution. 1995;49(4):718–26.

75. Page RDM. Extracting species trees from complex gene trees: Reconciled trees and vertebrate phylogeny. Mol Phylogenet Evol. 2000;14(1):89–106.
76. Martin AP, Burg TM. Perils of paralogy: Using HSP70 genes for inferring organismal phylogenies. Syst Biol. 2002;51(4):570–87.
77. Hudson RR, Turelli M. Stochasticity overrules the "three-times rule": Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. Evolution. 2003;57(1):182–90.
78. Pamilo P, Nei M. Relationships between Gene Trees and Species Trees. Molecular biology evolution. 1988;5(5):568–83.
79. Brumfield RT, Jernigan RW, McDonald DB, Braun MJ. Evolutionary implications of divergent clines in an avian (Manacus: Aves) hybrid zone. Evolution. 2001;55(10):2070–87.
80. Carling MD, Brumfield RT. Haldane's Rule in an Avian System: Using Cline Theory and Divergence Population Genetics to Test for Differential Introgression of Mitochondrial, Autosomal, and Sex-Linked Loci across the Passerina Bunting Hybrid Zone. Evolution. 2008;62(10):2600–15.
81. Rheindt FE, Edwards SV. Genetic Introgression: An Integral but Neglected Component of Speciation in Birds. Auk. 2011;128(4):620–32.
82. Saetre GP, Borge T, Lindell J, Moum T, Primmer CR, Sheldon BC, Haavie J, Johnsen A, Ellegren H. Speciation, introgressive hybridization and nonlinear rate of molecular evolution in flycatchers. Molecular ecology. 2001;10(3):737–49.
83. Saetre GP, Borge T, Lindroos K, Haavie J, Sheldon BC, Primmer C, Syvanen AC. Sex chromosome evolution and speciation in Ficedula flycatchers. P Roy Soc B-Biol Sci. 2003;270(1510):53–9.
84. Tegelstrom H, Gelter HP. Haldane Rule and Sex Biased Gene Flow between 2 Hybridizing Flycatcher Species (Ficedula-Albicollis and F-Hypoleuca, Aves, Muscicapidae). Evolution. 1990;44(8):2012–21.
85. Turelli M, Orr HA. The dominance theory of Haldane's rule. Genetics. 1995;140(1):389–402.
86. Mckay BD, Zink RM. The causes of mitochondrial DNA gene tree paraphyly in birds. Mol Phylogenet Evol. 2010;54(2):647–50.
87. Berlin S, Ellegren H. Evolutionary genetics. Clonal inheritance of avian mitochondrial DNA. Nature. 2001;413(6851):37–8.
88. Urantowka AD, Kroczak A, Mackiewicz P. The influence of molecular markers and methods on inferring the phylogenetic relationships between the representatives of the Arini (parrots, Psittaciformes), determined on the basis of their complete mitochondrial genomes. BMC evolutionary biology. 2017;17:166.
89. Tamashiro RA, White ND, Braun MJ, Faircloth BC, Braun EL, Kimball RT. What are the roles of taxon sampling and model fit in tests of cyto-nuclear discordance using avian mitogenomic data? Mol Phylogenet Evol. 2019;130:132–42.
90. Powell AF, Barker FK, Lanyon SM. Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: an avian case study. Mol Phylogenet Evol. 2013;66(1):69–79.

91. Meikejohn KA, Danielson MJ, Faircloth BC, Glenn TC, Braun EL, Kimball RT. Incongruence among different mitochondrial regions: A case study using complete mitogenomes. Mol Phylogenet Evol. 2014;78:314–23.
92. Funk DJ, Omland KE. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol S. 2003;34:397–423.
93. Moum T, Bakke I. Mitochondrial control region structure and single site heteroplasmy in the razorbill (Alca torda; Aves). Current genetics. 2001;39(3):198–203.
94. He XL, Ding CQ, Han JL. Lack of Structural Variation but Extensive Length Polymorphisms and Heteroplasmic Length Variations in the Mitochondrial DNA Control Region of Highly Inbred Crested Ibis, Nipponia nippon. PloS one. 2013;8(6):e66324.
95. Berg T, Moum T, Johansen S. Variable Numbers of Simple Tandem Repeats Make Birds of the Order Ciconiiformes Heteroplasmic in Their Mitochondrial Genomes. Current genetics. 1995;27(3):257–62.
96. Ball RM, Avise JC. Mitochondrial-DNA Phylogeographic Differentiation among Avian Populations and the Evolutionary Significance of Subspecies. Auk. 1992;109(3):626–36.
97. Avise JC, Zink RM. Molecular Genetic-Divergence between Avian Sibling Species - King and Clapper Rails, Long-Billed and Short-Billed Dowitchers, Boat-Tailed and Great-Tailed Grackles, and Tufted and Black-Crested Titmice. Auk. 1988;105(3):516–28.
98. Mundy NI, Winchell CS, Woodruff DS. Tandem repeats and heteroplasmy in the mitochondrial DNA control region of the loggerhead shrike (Lanius ludovicianus). J Hered. 1996;87(1):21–6.
99. Qian CJ, Wang YX, Guo ZC, Yang JK, Kan XZ. Complete mitochondrial genome of Skylark, Alauda arvensis (Aves: Passeriformes): The first representative of the family Alaudidae with two extensive heteroplasmic control regions. Mitochondrial DNA. 2013;24(3):246–8.
100. Cho HJ, Eda M, Nishida S, Yasukochi Y, Chong JR, Koike H. Tandem duplication of mitochondrial DNA in the black-faced spoonbill, Platalea minor. Genes genetic systems. 2009;84(4):297–305.
101. Eberhard JR, Wright TF, Bermingham E. Duplication and concerted evolution of the mitochondrial control region in the parrot genus Amazona. Molecular biology evolution. 2001;18(7):1330–42.
102. Slack KE, Janke A, Penny D, Arnason U. Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene. 2003;302(1–2):43–52.
103. Cadahia L, Pinsker W, Negro JJ, Pavlicev M, Urios V, Haring E. Repeated Sequence Homogenization Between the Control and Pseudo-Control Regions in the Mitochondrial Genomes of the Subfamily Aquilinae. J Exp Zool Part B. 2009;312B(3):171–85.
104. Shao R, Barker SC, Mitani H, Aoki Y, Fukunaga M. Evolution of duplicate control regions in the mitochondrial genomes of metazoa: a case study with Australasian Ixodes ticks. Molecular biology evolution. 2005;22(3):620–9.
105. Kumazawa Y, Ota H, Nishida M, Ozawa T. Gene rearrangements in snake mitochondrial genomes: Highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. Molecular biology evolution. 1996;13(9):1242–54.

106. Arndt A, Smith MJ. Mitochondrial gene rearrangement in the sea cucumber genus Cucumaria. Molecular biology evolution. 1998;15(8):1009–16.
107. Umeda S, Tang YY, Okamoto M, Hamasaki N, Schon EA, Kang DC. Both heavy strand replication origins are active in partially duplicated human mitochondrial DNAs. Biochem Biophys Res Commun. 2001;286(4):681–7.
108. Tang Y, Manfredi G, Hirano M, Schon EA. Maintenance of human rearranged mitochondrial DNAs in long-term cultured transmitochondrial cell lines. Molecular biology of the cell. 2000;11(7):2349–58.
109. Jiang ZJ, Castoe TA, Austin CC, Burbrink FT, Herron MD, McGuire JA, Parkinson CL, Pollock DD. Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. BMC evolutionary biology. 2007;7:123.
110. Kumazawa Y, Ota H, Nishida M, Ozawa T. The complete nucleotide sequence of a snake (Dinodon semicarinatus) mitochondrial genome with two identical control regions. Genetics. 1998;150(1):313–29.
111. Eda M, Kuro-o M, Higuchi H, Hasegawa H, Koike H. Mosaic gene conversion after a tandem duplication of mtDNA sequence in Diomedeidae (albatrosses). Genes genetic systems. 2010;85(2):129–39.
112. Kurabayashi A, Sumida M, Yonekawa H, Glaw F, Vences M, Hasegawa M. Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in mantellid frogs from Madagascar. Molecular biology evolution. 2008;25(5):874–91.
113. Skujina I, McMahon R, Lenis VPE, Gkoutos GV, Hegarty M. Duplication of the mitochondrial control region is associated with increased longevity in birds. Aging-Us. 2016;8(8):1781–9.
114. Richly E, Leister D. NUMTs in sequenced eukaryotic genomes. Molecular biology evolution. 2004;21(6):1081–4.
115. Bernt M, Braband A, Schierwater B, Stadler PF. Genetic aspects of mitochondrial genome evolution. Mol Phylogenet Evol. 2013;69(2):328–38.
116. Pearson WR, Wood T, Zhang Z, Miller W. Comparison of DNA sequences with protein sequences. Genomics. 1997;46(1):24–36.
117. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology evolution. 2013;30(4):772–80.
118. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91.
119. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56(4):564–77.
120. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology evolution. 2015;32(1):268–74.
121. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a

large model space. Syst Biol. 2012;61(3):539–42.
122. Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Molecular biology evolution. 2004;21(6):1095–109.
123. Chernomor O, von Haeseler A, Minh BQ. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. Syst Biol. 2016;65(6):997–1008.
124. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods. 2017;14(6):587–9.
125. Lanfear R, Calcott B, Ho SY, Guindon S. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular biology evolution. 2012;29(6):1695–701.
126. Huelsenbeck JP, Larget B, Alfaro ME. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. Molecular biology evolution. 2004;21(6):1123–33.
127. Shimodaira H, Hasegawa M. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics. 2001;17(12):1246–7.
128. Maddison WP, Maddison DR: **Mesquite**: **a modular system for evolutionary analysis**.. In., Version 3.31 edn; 2017.

Figures

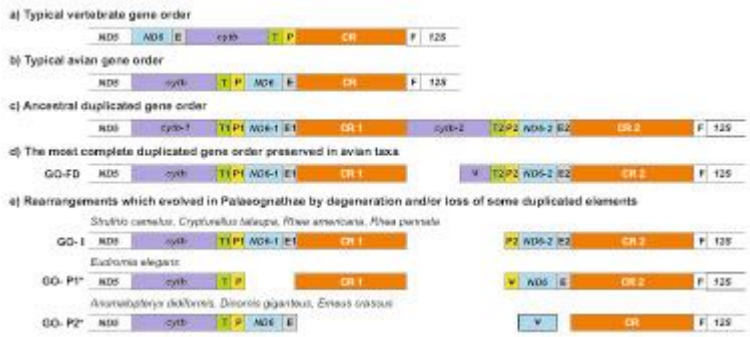

Figure 1

The comparison of various mitochondrial gene orders between ND5 and 12S rRNA: a typical vertebrate gene order (a), a typical avian gene order (b), an ancestral duplicated gene order assuming the tandem duplication of the cytb to CR segment (c), the most fully duplicated avian gene order, which has been found in representatives of Bucerotiformes, Gruiformes, Procellariiformes, Psittaciformes and Suliformes (d), rearrangements that evolved by degeneration and/or loss of some duplicated elements in Palaeognathae and some Passeriformes: Notiomystis cincta and Turdus philomelos (e). ND5 - NADH dehydrogenase subunit 5 gene; cytb – cytochrome b gene; T – tRNA gene for threonine; P – tRNA gene for proline; ND6 – NADH dehydrogenase subunit 6; E – tRNA gene for glutamic acid; CR – control region; F – tRNA gene for phenylalanine; 12S – 12S rRNA gene. Pseudogenes are marked by Ψ and coloured

correspondingly to their functional gene copy. Gene orders reannotated in this study are marked with an asterisk.

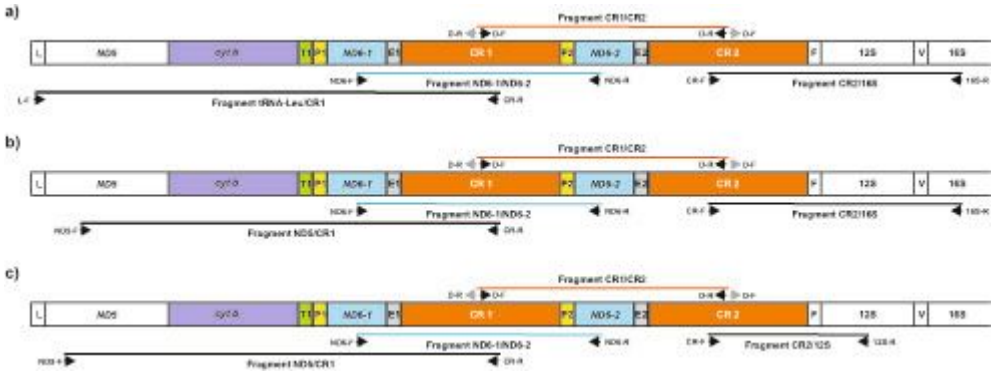

Figure 2

Strategy used in this study for identification of gene orders within duplicated regions in palaeognaths: Struthio camelus (a), Rhea americana and Rhea pennata (b) and Crypturellus tataupa (c) mitogenomes. L – tRNA for leucine, ND5 - NADH dehydrogenase subunit 5, cytb – cytochrome b, T – tRNA for threonine, P – tRNA for proline, ND6 – NADH dehydrogenase subunit 6, E – tRNA for glutamic acid, CR – control region, F – tRNA for phenylalanine, 12S – 12S rRNA, V – tRNA for valine, 16S – 16S rRNA. L-F, ND5-F, CR-R, ND6-F, ND6-R, D-F, D-R, CR-F, 12S-R, 16S-R: primers that were used for amplification of four overlapping mitogenomic fragments.

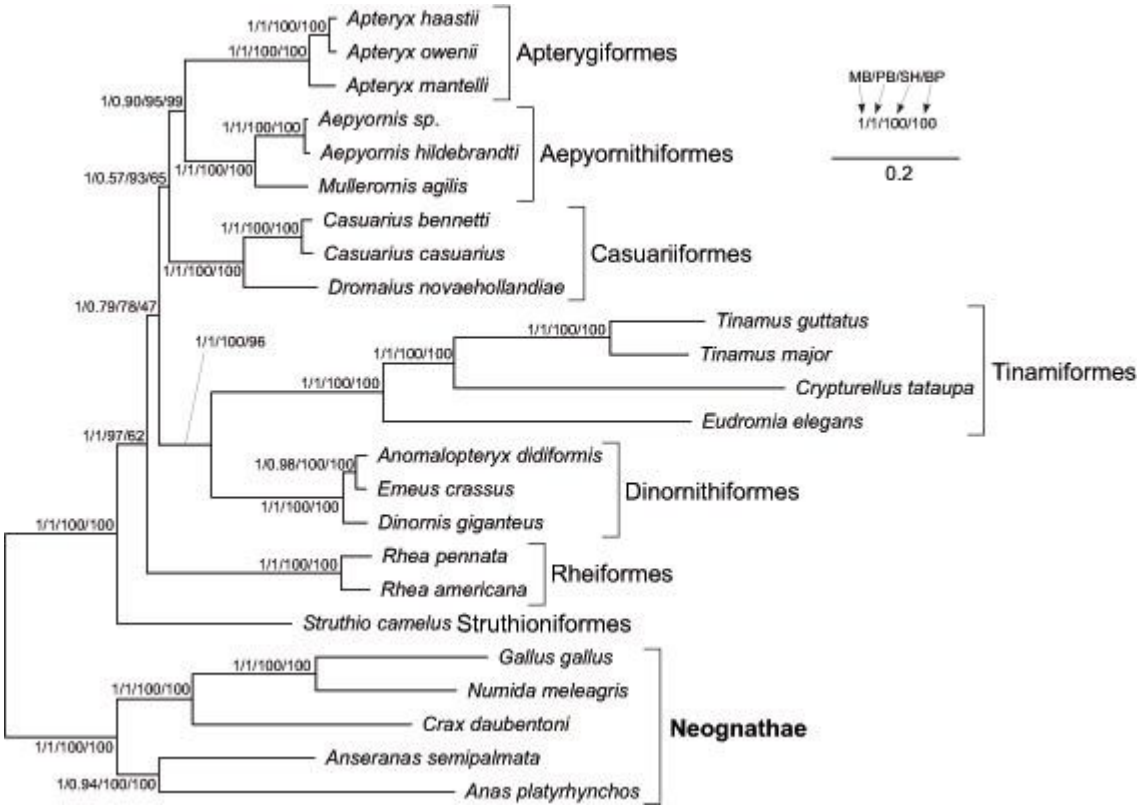

Figure 3

The phylogram obtained in MrBayes based on nucleotide sequences of mitochondrial genes. The values at nodes, in the following order MB/PB/SH/BP, indicate: posterior probabilities found in MrBayes (MB) and PhyloBayes (PB) as well as SH-aLRT (SH) and non-parametric bootstrap (BP) percentages calculated in IQ-TREE.

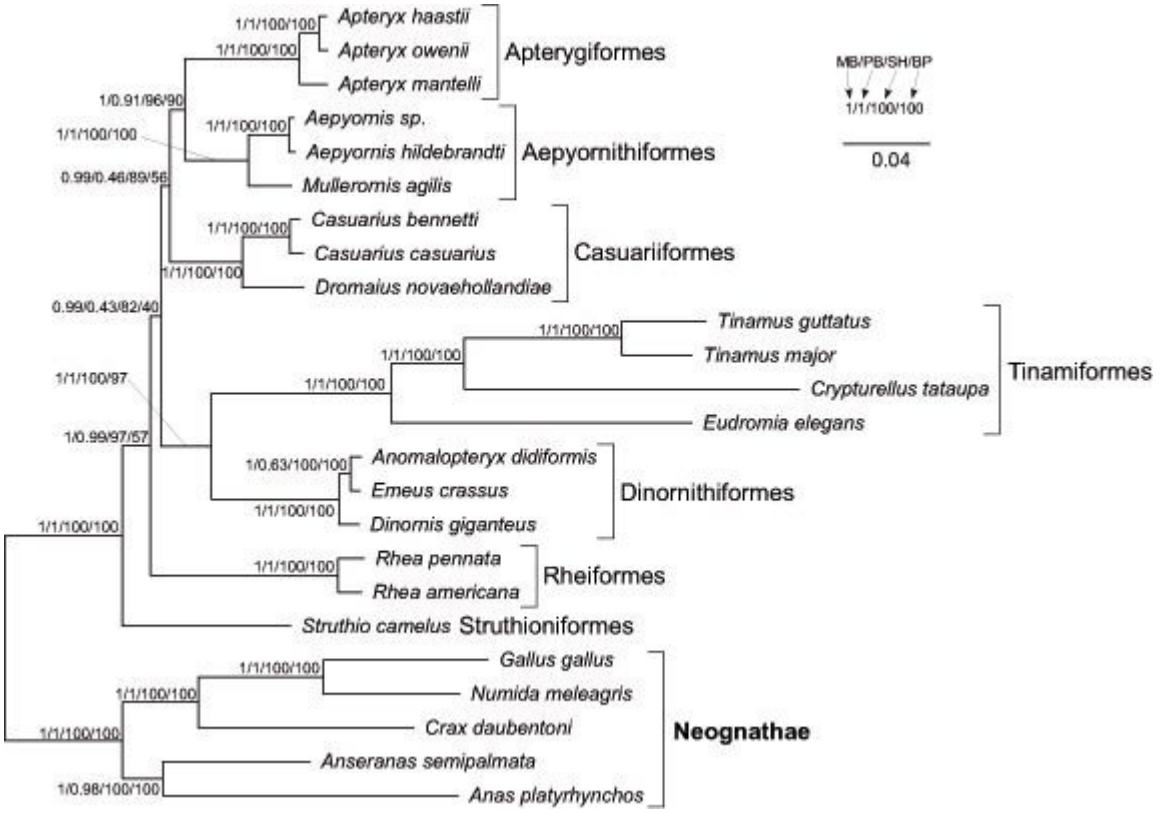

Figure 4

The phylogram obtained in MrBayes based on RY-recoded sequences of mitochondrial genes. See Figure 3 for further explanations.

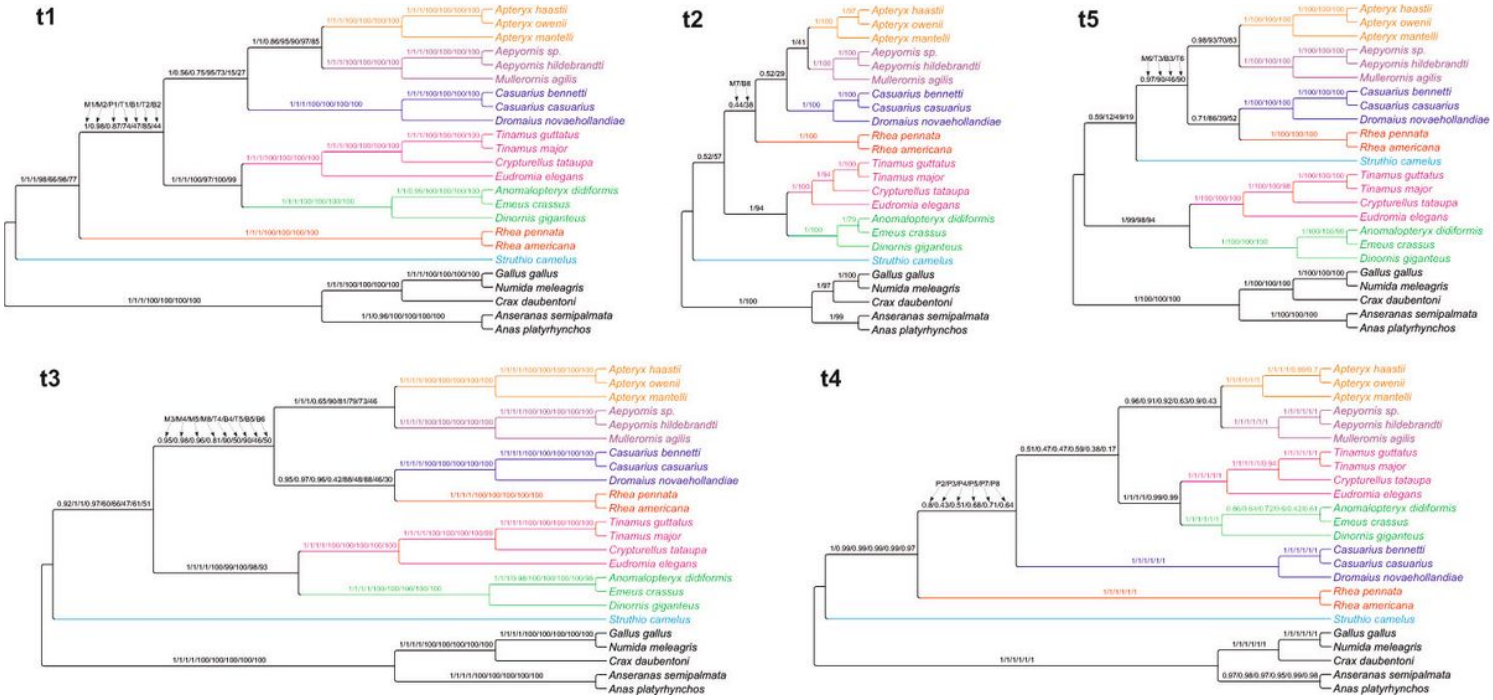

Figure 5

The most frequent tree topologies obtained in the phylogenetic analyses of mitochondrial gene alignments. Partitions characterized by the highest substitution rate were sequentially excluded from the alignment. The values at nodes indicate support values received for various partitions in different approaches. The approaches' names were marked with the letter: MrBayes with M, PhyloBayes with P, SH-aLRT in IQ-TREE with T and non-parametric bootstrap in IQ-TREE with B. The digits after the letters indicate the number of the highest rate partitions removed from the analysis.

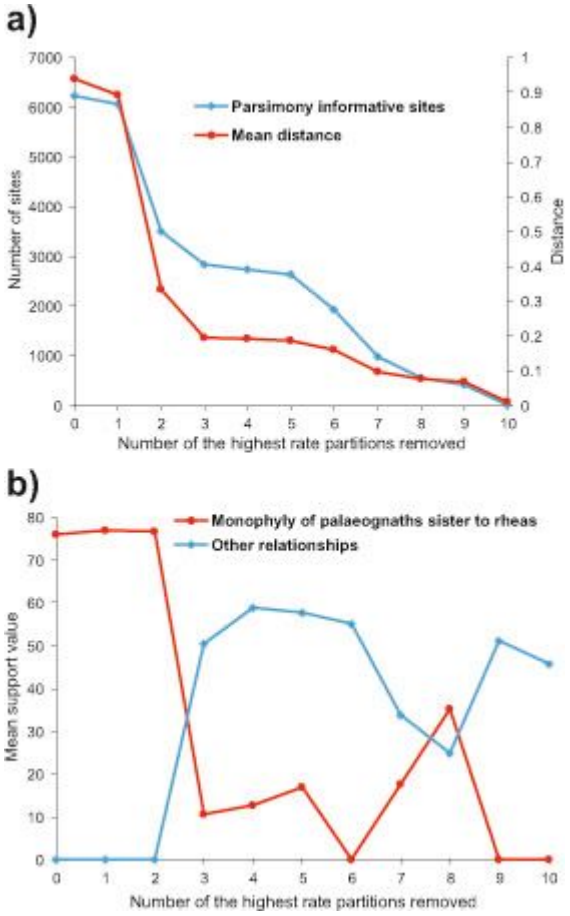

Figure 6

Influence of removing partitions with the highest substitution rate on alignment and tree parameters: the number of parsimony informative sites and mean distance (a) as well as mean support values (b). The mean phylogenetic distance was obtained from MrBayes tree. The mean support values were calculated from posterior probabilities received in MrBayes and PhyloBayes as well as SH-aLRT and non-parametric bootstrap percentages obtained in IQ-TREE. The posterior probabilities were scaled to 100%.

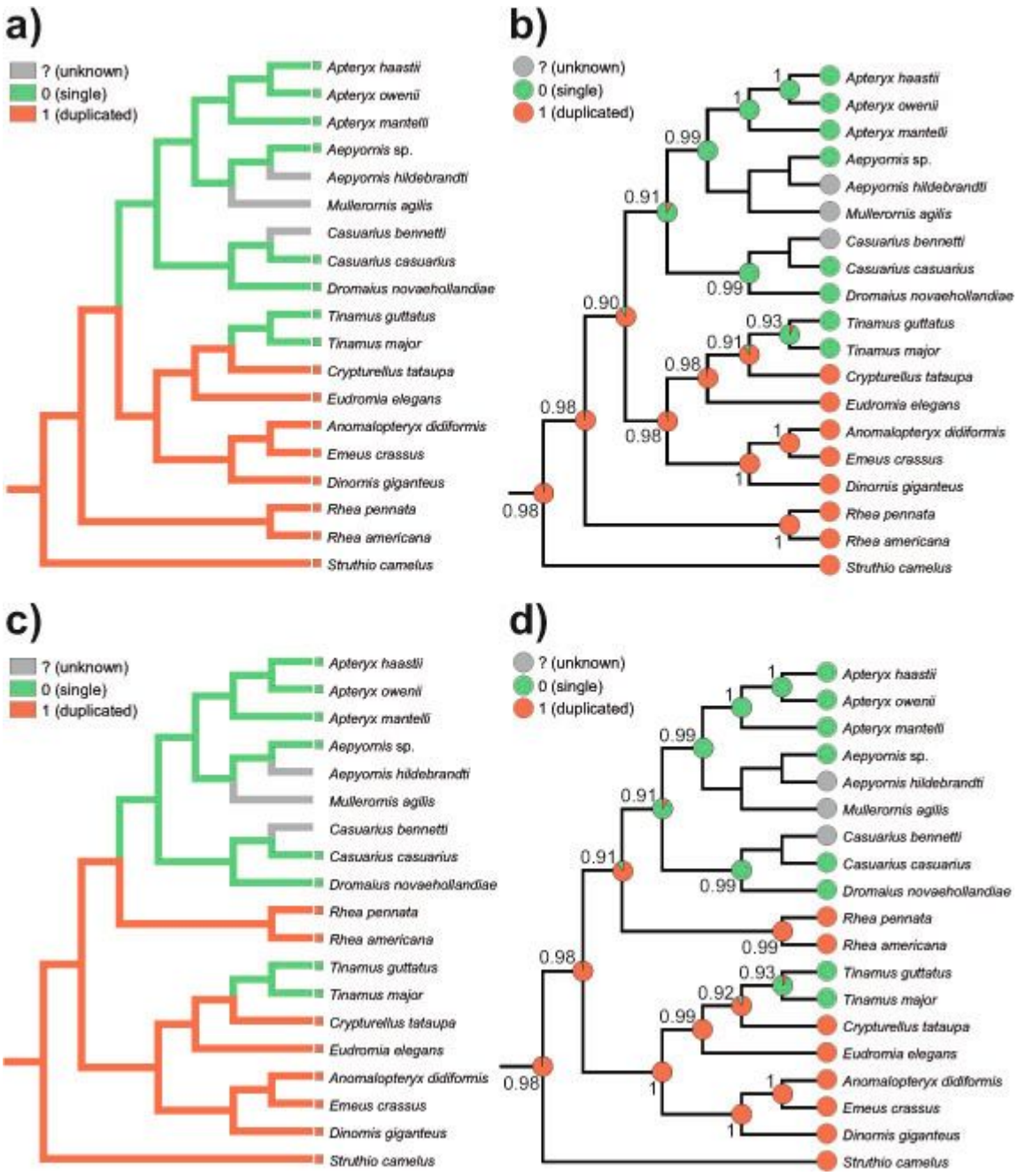

Figure 7

Reconstruction of ancestral states and mapping of mitogenomic duplications onto the Palaeognathae trees. Two methods, maximum parsimony (a, c) and maximum likelihood (b, d) were applied for two tree topologies: t1 based on sequences of mitochondrial genes (a, b) and t2 supported by nuclear markers in coalescent-based approaches (c, d). The area of colors at nodes in b and d corresponds to the probability of the given state, single or duplicated region. The probability value for a more probable state was also given at these nodes. Mk1 model was applied for ML approach.

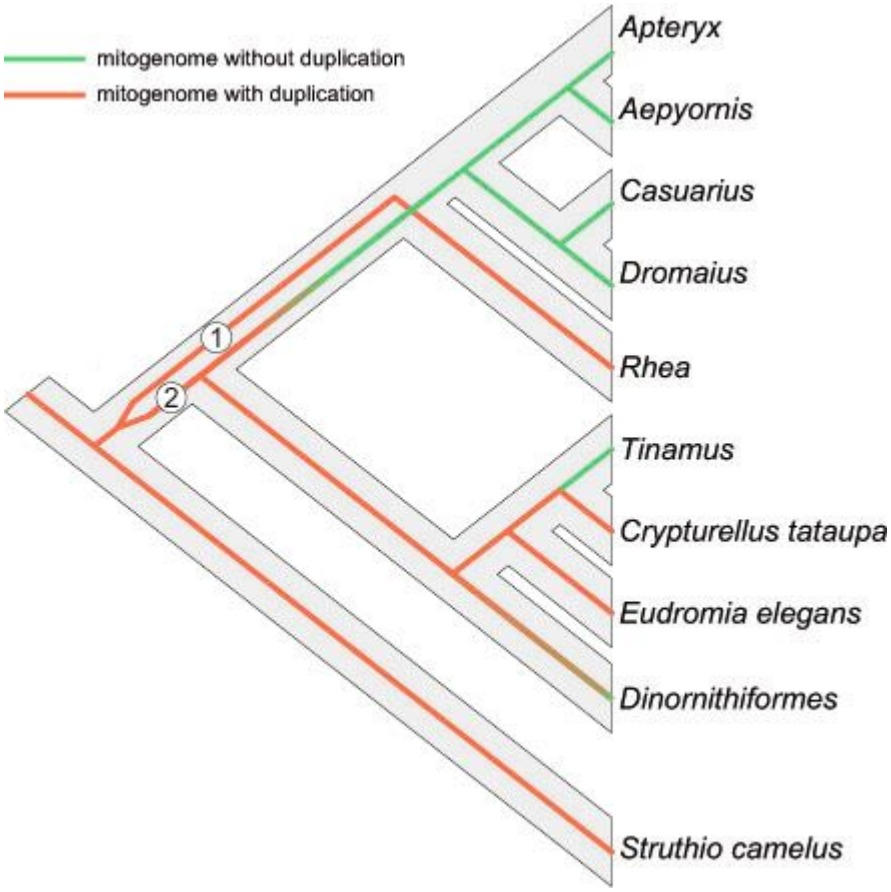

Figure 8

Superposition of the mitogenome phylogeny (thin colored lines) onto the potential species tree of Palaeognathae (grey thick branches). Two mitochondrial lineages were labelled as 1 and 2. Lineages with and without duplication were indicated in different colors. The tonal transition from orange to green indicates gradual disappearing of the duplication.

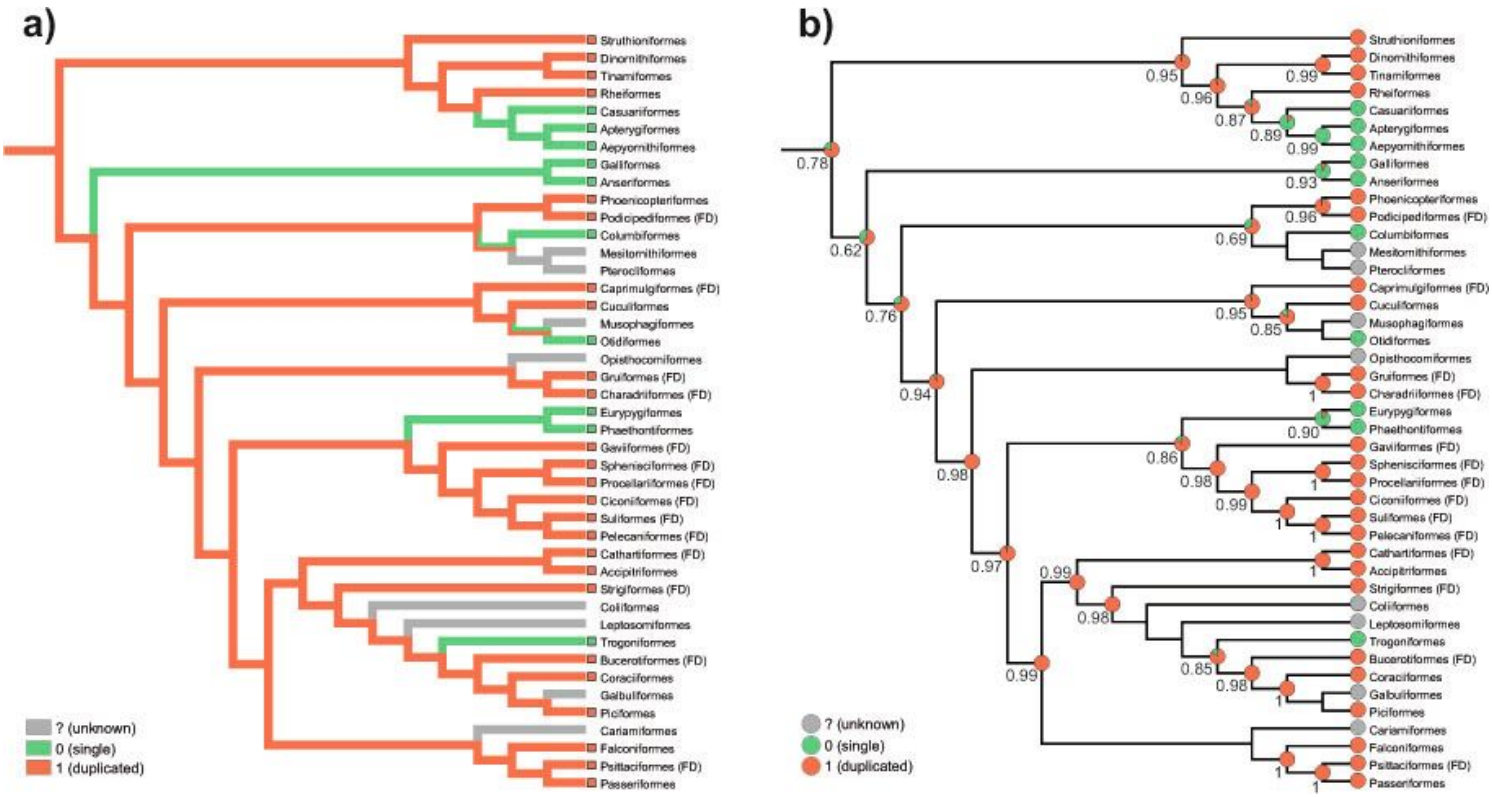

Figure 9

Maximum parsimony (a) and maximum likelihood (b) reconstruction of ancestral states and mapping of mitogenomic duplications onto the Aves tree topology obtained by Kimball, et al. [47]. The area of colors at nodes in b corresponds to the probability of the given state, single or duplicated region. The probability value for a more probable state was also given at these nodes. FD indicates the presence of the most fully duplicated avian region. Mk1 model was applied for ML approach.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.pdf
- Additionalfile2.xlsx