Electronic Supplementary Material for:

Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: A Comment on Ericson et al.

Joseph W. Brown*, Robert B. Payne, and David P. Mindell

* To whom correspondence should be addressed. Email: josephwb@umich.edu

Sequence data

We used the same alignment as was analyzed in Ericson et al. (2006), which was made freely available by those authors (http://www.nrm.se/inenglish/researchandcollections/zoology/vertebratezoology/birds.4.4e32c81078a8d9249800014590.html). From the combined total alignment of 5299 bp, 1536 were deemed ambiguously aligned and so were excluded, yielding a final combined data set of 3763 bp (Table ESM-1). Although data matrices differ notably in size across our two studies, we find no influence of our relatively conservative site exclusion approach on inferred date estimates (see below). No attempt was made at phylogenetic reconstruction; rather, we used their inferred tree (their Figure ESM-9) for dating purposes to make the results directly comparable.

Divergence time estimation

Divergence times were estimated using the MULTIDISTRIBUTE package (Thorne, 2003). This is a Bayesian approach to modelling rate heterogeneity across a tree in an ancestor-descendent fashion (Thorne & Kishino 2002; Thorne et al. 1998). Estimates of the transition/transversion rate ratio κ and the gamma site class-specific rates under the F84+G model were calculated for each gene individually in the baseml program of the PAML 3.15 package (Yang 1997). The output from baseml was used as the input for the MULTIDISTRIBUTE program estbranches, which produces ML estimates of branch lengths and their approximate variance–covariance matrix for each gene. Finally, substitution rates and divergence times were estimated through MCMC approximation in Multidivtime. Here, the logarithm of the substitution rate at the end of a branch is modelled with a normal distribution, the mean of which has an expected value equal to the rate at the beginning of the branch. While rates are implicitly assumed to be autocorrelated from ancestor to descendent nodes, this autocorrelation may decay with increasing branch lengths.

Multidivtime prunes the outgroup for this final analysis, so the taxa involved in dating are all representatives of Neognathae. We defined a diffuse prior for the age of the root of this tree: mean (rttm) = 100, standard deviation (rttmsd) = 40. To investigate the influence of this prior we also ran analyses assuming a much younger divergence: rttm = 80, rttmsd = 50. Additional priors, determined from an average across genes, are given in Table ESM-2. Bigtime, the maximum age allowed for the root, was set at 160 MY, as this is well beyond molecular estimates of the age of the divergence between Galloanseres and Neoaves. The 22 internal fossil constraints used are given in Table ESM-3. The program was run without the assumption of correlated changes in substitution rates across genes. Following a burnin of 10^5 samples, 10^4 samples were taken at a sampling interval of 10^2. Analyses were repeated with different initial conditions to check for convergence of the MCMC chain. Results from analyses assuming our alternative prior distributions (Table ESM-2) were indistinguishable. Chronograms were constructed using FigTree v1.0 (Rambaut 2006) and TS-Creator (http://www.stratigraphy.org).

Influence of site exclusions, fossil constraints, and dating method

Given the numerous concerns we identify with the study of Ericson et al. (2006), it is interesting to examine which aspects of our reanalysis contribute to the discordance in inferred dates between our two studies. To this end, we also reanalyze the data of Ericson et al. (2006) in PATHd8 using our alignment (Table ESM-1) and different fossil complements. To make results comparable, we use the same topology as above. As PATHd8 cannot accommodate multiple genes, we estimated maximum likelihood branch lengths on the concatenated data matrix using the optimal DNA substitution model (TIM+I+G) as inferred using AIC in Modeltest (Posada and Crandall 1998) and PAUP* (Swofford 2003). We will refer to this tree hereafter as T_{ML}.

To determine whether our relatively conservative site exclusion approach contributed to the older dates inferred here, we analyzed T_{ML} in PATHd8 using the same settings as in Ericson et al. (2006). Specifically, we placed an upper limit of 95 MY on the age of Neoaves and fixed the age of divergence between Trochilidae (hummingbirds) and Apodidae (swifts) at 47.5 MY. All other fossil calibrations (Table ESM-3) were treated as minimum age constraints. We exclude the fossil representing stem Strigiformes (owls; constraint F in Ericson et al. 2006) from all analyses as it was found to be superseded by the fossil from stem Coliiformes (mousebirds; constraint E in Ericson et al. 2006). The resulting chronogram (Figure ESM-1, green dashed lines) is in very close agreement with the published chronogram of Ericson et al. (2006). In fact, regarding the K-T boundary, analysis of our alignment yielded slightly younger dates with only two neoavian divergence point estimates lying in the Cretaceous (versus five in Ericson et al. 2006). We can therefore be confident that our conservative alignment is not responsible for the older dates presented in Figure 1. We note, however, that analysis of our alignment yielded reasonable age estimates for the divergence between Paleognathae and Neognathae and between paleognath families Rheidae and Apterygidae, contra to the results of Ericson et al. (2006). This may have come about because the less conservative data exclusion of Ericson et al. (2006) included ambiguously aligned sites involving these taxa.

We next analyzed T_{ML} in PATHd8 using the calibration recommendations outlined in our manuscript. Specifically, we set a liberal upper limit on the age of Neoaves at 120 MY, and for the required fixed calibration used the fossil from stem Sphenisciformes (penguins) at 62 MY (Slack et al. 2006). Additional fossils were again treated as minimum age constraints. Replaced fossil constraints are denoted by a prime (´) symbol in Table ESM-3. The resulting chronogram (Figure ESM-1, solid black lines) illustrates the strong systematic influence of our calibration changes on the inferred age estimates. In contrast to the estimates inferred using the original constraints of Ericson et al. (2006), substantial diversification is inferred to have occurred in the Cretaceous using our calibrations. For example, whereas we infer the initial divergence within Neoaves to be ~ 70 MY using the original calibrations, the estimate increases to ~ 94 MY using our revised calibrations. Furthermore, ages for all neoavian divergences are on average 18.95 MY older using our calibration scheme (range 0 – 41.5 MY). Most striking, however, are the differences in age estimates for the required fixed nodes. While Ericson et al. (2006) fix the divergence between Trochilidae and Apodidae at 47.5 MY, we estimate this divergence at 86.5 MY. Our older estimate of 86.5 MY is similar to the 75.7 ± 7.8 MY estimate reported by van Tuinen & Hedges (2001). In our PATHd8 analysis we fixed the origin of stem Sphenisciformes at 62 MY, whereas Ericson et al. (2006) estimate this event at 55 MY (their minimum age constraint for that node), which is 7 MY younger than the minimum age known from the fossil record. Again, we find much younger dates for nodes associated with the paleognath outgroups than was found in Ericson et al. (2006).

Finally, we can compare age estimates generated in Multidivtime (Figure 1) and PATHd8 (Figure ESM-1) using the same revised fossil constraints. Two important differences can be seen between these two figures. First, the initial divergence within Neoaves is estimated as slightly older at ~ 100 MY using the Bayesian method of Multidivtime (versus ~ 94 MY in PATHd8). We note that this finding was precluded by the maximum age constraint of 95 MY set by Ericson et al. (2006). Second, the results from Multidivtime indicate a more gradual diversification of Neoaves than those from PATHd8. This is likely due to the ability of Multidivtime to accommodate information from individual genes, as these genes are likely informative in different parts of the tree. While Multidivtime tends to produce older age estimates than PATHd8 for many nodes, the difference is much slighter and insignificant when the large posterior credible intervals are included.

References

Clarke, J. A., Tambussi, C. P., Noriega , J. I., Erikson, G. M., & Ketcham, R. A. 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. *Nature* 433, 305-308.

Ericson, P. G. P., Anderson, C. L., Britton, T., Elzanowski, A., Johansson, U. S., Källersjö, M., Ohlson, J. I., Parsons, T. J., Zuccon, D., & Mayr, G. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. *Biol. Let.* 4, 543-547.

Mayr, G. 2004. Old world fossil record of modern-type hummingbirds. *Science* 304, 861-864.

Posada, D., & K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817-818.

Rambaut, A. 2006. FigTree v1.0. Available from the author (http://evolve.zoo.ox.ac.uk/software.html?id=figtree).

Slack, K. E., Jones, C. M., Ando, T., Harrison, G. L., Fordyce, R. E. & Penny, D. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Mol. Biol. Evol*. 23, 1144-1155.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). version 4 (beta 10). Sinauer Associates, Sunderland, Massachusetts.

Thorne, J. L. 2003. MULTIDISTRIBUTE. Available from the author (http://statgen.ncsu.edu/thorne/multidivtime.html).

Thorne, J. L. & Kishino, H. 2002. Divergence time and evolutionary rate estimation. *Syst. Biol.* 51, 689-702.

Thorne, J. L, Kishino, H., Painter, I. S. 1998. Estimating the rate of evolution of the rate of evolution. *Mol. Biol. Evol.* 15, 1647-1657.

van Tuinen, M., & Hedges, S. B. 2001. Calibration of avian molecular clocks. *Mol. Biol. Evol.* 18, 206-213.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in BioSciences* 13, 555-556.

Table legends

Table ESM-1. Aligned nuclear DNA fragment lengths. The alignment is taken directly from Ericson et al. (2006). ‘Original’ refers to aligned sequence lengths prior to excluding ambiguously aligned nucleotide sites.

Table ESM-2. Priors used in the Multidivtime analyses. All priors were determined from an average across the five genes. Those listed under ‘Preferred Prior’ were used to generate the results presented in Figure 1. To examine the influence of prior probabilities on inferred age estimates Multidivtime was also ran with deliberately young age priors (listed under ‘Young Prior’).

Table ESM-3. Fossil calibrations used in this study. For fossils used in Ericson et al. (2006), see that paper for the original fossil references. All fossils are treated as minimum age constraints, and are placed in the tree exactly as in Ericson et al. (2006). Replaced fossil calibrations are denoted by a prime (´) symbol. Fossil calibration F (stem group Strigiformes) from Ericson et al. (2006) was not used as it was found to be redundant with the more derived fossil calibration E (stem group Coliiformes).

Figure ESM-1. Chronograms generated using PATHd8 on the concatenated data matrix using the fossil constraints of Ericson et al. (2006; green dashed chronogram) and our reanalysis (black solid chronogram). The dashed vertical red line marks the K-T boundary.

Table ESM-1. Aligned nuclear DNA fragment lengths. The alignment is taken directly from Ericson et al. (2006). 'Original' refers to aligned sequence lengths prior to excluding ambiguously aligned nucleotide sites.

Gene	Original	This study	Excluded sites
β-fibrinogen (intron 7)	1705	944	24-110 121-123 126 127 168-170 176 199 200 214 232 240 241 253 256 257-275 304-315 362 363 366-375 377 378 395 401 402 420 429 432 440 451-453 468 493-514 532 541 561 596 609 612 613 619 660-675 679-684 690 720 746-748 750 770 771 774 777 778-784 793 794 797 812 813 826-829 843-1123 1144 1150 1151 1190-1192 1195-1203 1222 1224 1227 1228 1229 1231 1247-1251 1265 1271-1278 1294 1296 1307 1335-1340 1346 1350-1353 1359 1360 1389 1394 1396-1408 1455 1468-1472 1491-1509 1513-1529 1532-1540 1551-1572 1579-1656 1660 1663 1673-1685 1692-1701
c-*myc* (exon 3)	510	498	64-75
myoglobin (intron 2)	1061	720	24-30 34 42 54 77-88 91 109 110 118 133 134 137-140 160 172 182-339 347-349 362-364 373 374 385 392 393 408-418 455 485 548 565-568 591 597-600 604 605 618-621 633-637 663 675 697 701 721-723 735-751 759-801 832 839-846 874 875 983 1004-1023 1027 1036-1039
ornithine decarboxylase (ODC; introns 6,7; exon 7)	1093	671	80-82 86-178 184 192 193 198 208 213 215-276 287 289 292 299 300 307 333 339 346 354-356 364 373-386 391-399 404 411-427 438 448 450 451 545 546 558-563 588-648 658 670-683 690-692 702-707 712-726 734 735 742 753 762-770 781 787 790 791 799 808 812-822 828 832 833 839-860 866 869 877-879 892 913 918 926 927 943 948 952 953 956 974 975 984-988 994 999 1005-1008 1013 1014 1031-1034 1036 1042 1056
RAG-1	930	930	-
Total	5299	3763	1536

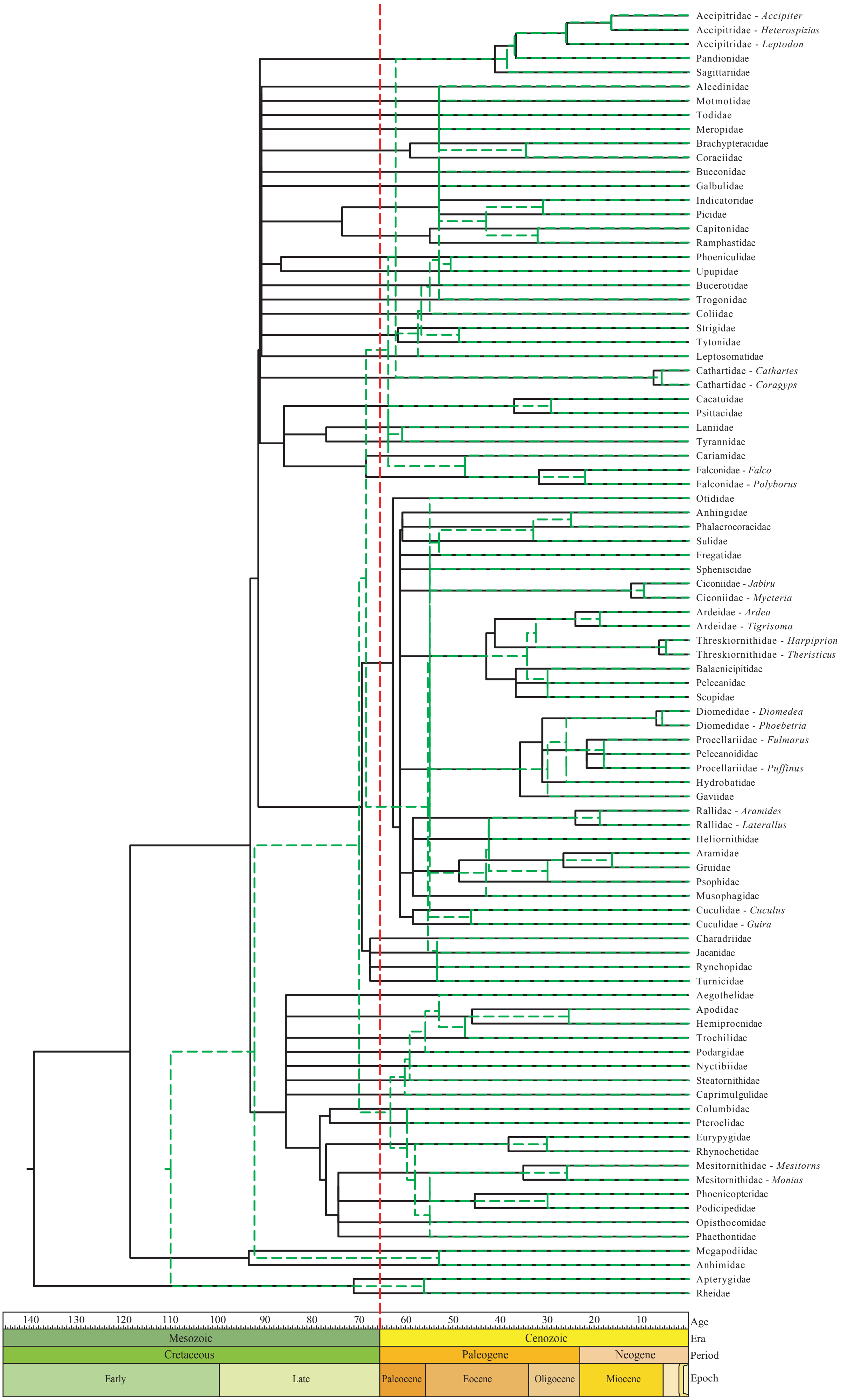
Table ESM-2. Priors used in the Multidivtime analyses. All priors were determined from an average across the five genes. Those listed under ‘Preferred Prior’ were used to generate the results presented in Figure 1. To examine the influence of prior probabilities on inferred age estimates Multidivtime was also ran with deliberately young age priors (listed under ‘Young Prior’).

Parameter	Preferred Prior	Young Prior
rttm (mean time separating root and present)	100	80
rttmsd (standard deviation of rttm)	40	50
rtrate (mean rate at root node)	0.0016	0.0013
rtratesd (standard deviation of rtrate)	0.0016	0.0013
brownmean (mean of Brownian motion constant)	0.01875	0.015
brownmeansd (standard deviation of brownmean)	0.01875	0.015
bigtime (oldest age allowed for root node)	160	160

Table ESM-3. Fossil calibrations used in this study. For fossils used in Ericson et al. (2006), see that paper for the original fossil references. All fossils are treated as minimum age constraints, and are placed in the tree exactly as in Ericson et al. (2006). Replaced fossil calibrations used in our reanalyses are denoted by a prime (´) symbol. Fossil calibration F (stem group Strigiformes) from Ericson et al. (2006) was not used as it was found to be redundant with the more derived fossil calibration E (stem group Coliiformes).

Symbol	Fossil Calibration	Age (MY)	Source
A	crown Pici	30	Ericson et al. 2006
B	stem Upupidae + Phoeniculidae	47.5	Ericson et al. 2006
C	stem Coraciidae + Brachypteraciidae	47.5	Ericson et al. 2006
D	stem Trogoniformes	53	Ericson et al. 2006
E	stem Coliiformes	55	Ericson et al. 2006
G	crown Pandionidae	37	Ericson et al. 2006
H	stem Cariamidae	47.5	Ericson et al. 2006
I	stem Phalacrocoracidae	25	Ericson et al. 2006
J	crown Sulidae	33	Ericson et al. 2006
K	stem Fregatidae	53	Ericson et al. 2006
L	stem Sphenisciformes	55	Ericson et al. 2006
L´	stem Sphenisciformes	62	Slack et al. 2006
M	crown Balaenicipitidae	30	Ericson et al. 2006
N	crown Heliornithidae	14	Ericson et al. 2006
O	stem Jacanidae	30	Ericson et al. 2006
P	stem Apodiformes	53	Ericson et al. 2006
Q	stem Trochilidae	47.5	Ericson et al. 2006
Q´	stem Trochilidae	30	Mayr 2004
R	crown Pteroclididae	30	Ericson et al. 2006
S	stem Phoenicopteriformes	30	Ericson et al. 2006
T	stem Phaethontidae	55	Ericson et al. 2006
V	stem Gruidae + Aramidae	30	Ericson et al. 2006
U	stem Galliformes	54	Ericson et al. 2006
U´	stem Anatidae	66	Clarke et al. 2005
X	Stem Gaviiformes	30	Ericson et al. 2006

Figure ESM-1


Accipitridae - *Accipiter*
Accipitridae - *Heterospizias*
Accipitridae - *Leptodon*
Pandionidae
Sagittariidae
Alcedinidae
Motmotidae
Todidae
Meropidae
Brachypteracidae
Coraciidae
Bucconidae
Galbulidae
Indicatoridae
Picidae
Capitonidae
Ramphastidae
Phoeniculidae
Upupidae
Bucerotidae
Trogonidae
Coliidae
Strigidae
Tytonidae
Leptosomatidae
Cathartidae - *Cathartes*
Cathartidae - *Coragyps*
Cacatuidae
Psittacidae
Laniidae
Tyrannidae
Cariamidae
Falconidae - *Falco*
Falconidae - *Polyborus*
Otididae
Anhingidae
Phalacrocoracidae
Sulidae
Fregatidae
Spheniscidae
Ciconiidae - *Jabiru*
Ciconiidae - *Mycteria*
Ardeidae - *Ardea*
Ardeidae - *Tigrisoma*
Threskiornithidae - *Harpiprion*
Threskiornithidae - *Theristicus*
Balaenicipitidae
Pelecanidae
Scopidae
Diomedidae - *Diomedea*
Diomedidae - *Phoebetria*
Procellariidae - *Fulmarus*
Pelecanoididae
Procellariidae - *Puffinus*
Hydrobatidae
Gaviidae
Rallidae - *Aramides*
Rallidae - *Laterallus*
Heliornithidae
Aramidae
Gruidae
Psophidae
Musophagidae
Cuculidae - *Cuculus*
Cuculidae - *Guira*
Charadriidae
Jacanidae
Rynchopidae
Turnicidae
Aegothelidae
Apodidae
Hemiprocnidae
Trochilidae
Podargidae
Nyctibiidae
Steatornithidae
Caprimulgulidae
Columbidae
Pteroclidae
Eurypygidae
Rhynochetidae
Mesitornithidae - *Mesitorns*
Mesitornithidae - *Monias*
Phoenicopteridae
Podicipedidae
Opisthocomidae
Phaethontidae
Megapodiidae
Anhimidae
Apterygidae
Rheidae
140
130
120
110
100
90
80
70
60
50
40
30
20
10
Age
Mesozoic
Cenozoic
Era
Cretaceous
Paleogene
Neogene
Period
Early
Late
Paleocene
Eocene
Oligocene
Miocene
Epoch