Avian phylogeny and divergence times based on mitogenomic sequences

Kerryn Elizabeth Slack

2012

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Genetics

Institute of Molecular BioSciences
Massey University
Palmerston North,
New Zealand

Abstract

Despite decades of research using a variety of data sources (such as morphological, paleontological, immunological, DNA hybridization and short DNA sequences) both the relationships between modern bird orders and their times of origin remain uncertain. Complete mitochondrial (mt) genomes have been extensively used to study mammalian and fish evolution. However, at the very beginning of my study only the chicken mt sequence was available for birds, though seven more avian mt genomes were published soon after. In order to address these issues, I sequenced eight new bird mt genomes: four (penguin, albatross, petrel and loon) from previously unrepresented orders and four (goose, brush-turkey, gull and lyrebird) to provide improved taxon sampling. Adding these taxa to the avian mt genome dataset aids in resolving deep bird phylogeny and confirms the traditional placement of the root of the avian tree (between paleognaths and neognaths). In addition to the mt genomes, in a collaboration between paleontologists and molecular biologists, the oldest known penguin fossils (which date from 61- 62 million years ago) are described. These fossils are from the Waipara Greensand, North Canterbury, New Zealand, and establish an excellent calibration point for estimating avian divergence times. Bayesian analysis of the DNA sequence data, using the penguin calibration point plus two others, indicates a substantial radiation of modern bird orders in the Late Cretaceous (80 - 65 million years ago). Biotic interactions between modern birds and declining groups such as pterosaurs and early bird groups (e.g. *Hesperornis* and *Ichthyornis*) may thus have played an important role during this time.

Acknowledgements

This thesis is for my family – Mum, Dad, Jude, Stu and Buzz: THANK YOU for your support, your encouragement and, above all, your patience!

Words cannot express my gratitude to my supervisors: David Penny (Professor of Theoretical Biology and former Co-Director, Allan Wilson Center for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand) – more than just a supervisor: a true friend, colleague and mentor; Mike Hendy (Professor of Mathematical Biology and founding Executive Director, Allan Wilson Center for Molecular Ecology and Evolution, Massey University) – I always appreciate your dry sense of humour Mike!; Ulfur Arnason (Professor of Evolutionary Molecular Systematics, Department of Cell and Organism Biology, University of Lund, Lund, Sweden) – I'll never forget the 'rock party' Ulfur!; and Bengt-Olle Bengtsson (Professor of Genetics, Department of Cell and Organism Biology, University of Lund) – always patient, always kind.

Thanks also to my unofficial supervisor: Pete Lockhart (Institute of Molecular BioSciences, Massey University, Palmerston North, NZ**)** – the archetypal 'absent-minded professor' ☺.

A big thanks to my other co-authors:
Axel Janke (Professor of Ecological Genomics, Goethe University, Senckenberg, Frankfurt, Germany),
G. L. (Abby) Harrison (University of Oxford, Oxford, U.K.),
P. A. (Trish) McLenachan (Institute of Molecular BioSciences, Massey University, NZ),
Matt J. Phillips (University of Queensland, Brisbane, Australia),
Alan Cooper (Professor, University of Adelaide, Adelaide, Australia),
Craig M. Jones (Institute of Geological and Nuclear Sciences, Lower Hutt, NZ),
Tatsuro Ando (Department of Geology, University of Otago, Dunedin, NZ),
R. Ewan Fordyce (Professor, Department of Geology, University of Otago, Dunedin, NZ), and
Frederic Delsuc (Laboratoire de Paléontologie, Phylogénie et Paléobiologie, Institut des Sciences de l'Evolution, Université Montpellier II, Montpellier, France).

My deepest appreciation to the following: Anette Gullberg (University of Lund) and Trish McLenachan, Abby Harrison, Alicia Gore, Gillian Gibb and Charmaine Carlisle (all of/formerly of Massey University) for technical assistance; Matt ('font of vertebrate information') Phillips and Fred Delsuc for assistance with analysis; and Tim White and Warwick Allen (both of Massey University) for their assistance with computer programs, for running analyses on the Helix supercomputer and for help with other computer matters. I would also like to thank all those who helped out, by sharing information, advice and access, in the preparation of the manuscripts, Te Papa for the use of facilities, Massey Interloans for sterling service and the anonymous reviewers who commented on the early manuscript drafts.

A huge thanks to my NZ friends and colleagues, especially Trish McLenachan, Abby Harrison, Alicia Gore, Gillian Gibb, Charmaine Carlisle, Matt Phillips, Pete Lockhart and Tim White, plus Mike Charleston, Ant Poole, Mort Piripi, Simon Hills, Olga Kardailsky, Andrew Clarke and Richard Carter, for friendship, fun, support and fascinating conversations. Thank you also to my Swedish friends and colleagues, including Anette Gullberg (again), Xiufeng Xu, Tina Ledje, Anna Härlid, Ann-Sofie Rasmussen, Björn Ursing, Christian Elmerot, Martin Johansson and all of the others within the Lund Department of Cell and Organism Biology who helped me out with various things. I would also like to thank all those I met at assorted conferences and seminars for reminding me that there is more to life than bird mitochondrial genomes and that science is both interesting and stimulating!

I would like to thank Alan Tennyson and Sandy Bartle of Te Papa (Museum of NZ) for supplying the penguin and albatross samples and Rod Hitchmough (Department of Conservation, Wellington) for his hatchet job cutting them up (much appreciated Rod), Anette Gullberg (University of Lund) for obtaining the goose sample, Dick Gill (NZ Department of Conservation, Waikanae office) for the petrel and the gull, Einar Arnason (University of Iceland) and Thorvaldur Björnsson (Museum of Natural History, Reykjavik, Iceland) for the loon, Darryl Jones (Australian School of Environmental Studies, Griffith University, Brisbane) and Ian Owens (University of Queensland) for the brush-turkey and Cathy Nock (Centre for Animal Conservation Genetics, Southern Cross University [Lismore campus], New South Wales) for the lyrebird.

The work carried out in this study was supported at different times by a Swedish Institute Guest Scholarship, a Massey University Doctoral Scholarship, a University of Lund (Department of Genetics) Summer Stipendium, a Sven and Lilly Lawski Stipendium, a University of Lund (Department of Genetics) Doktorandtjänst, the Jörgen Lindströms Stipendiefond, the Swedish Science Council, the Nilsson-Ehle Foundation, the Allan Wilson Center for Molecular Ecology and Evolution (a NZ Centre for Research Excellence) and the NZ Marsden Fund.

Chapter 1. Introduction

This introduction provides background to some of the key issues in avian phylogeny up to the time my first paper on bird evolution was published - Slack et al. (2003; Chapter 2). The date of this publication defines the scope of the introduction, thus references after 2003 are generally not included in this chapter, although there are exceptions, such as Clements (2005) which is used for nomenclature, and Penny and Phillips (2004) which a gives a range of hypotheses to be tested. In my summary (Chapter 5) I give an updating of the current understanding of avian evolution.

Class Aves consists of all extant birds (Neornithes) together with some extinct groups such as enantiornithines ('opposite birds'), hesperornithids (toothed waterbirds) and icthyornithids (toothed seabirds). Birds are usually considered as one of the five main groups of vertebrates, the others being fish, amphibians, reptiles and mammals (although 'reptiles' is not a strictly monophyletic group). Avians are descended from non-mammalian reptiles (almost certainly theropod dinosaurs; see Cracraft 1988 and references therein) and can easily be distinguished from other vertebrates through morphological characteristics such as feathers, wings and beaks (among others). There are almost 10,000 species of extant modern birds and they are globally distributed, being found from the Arctic to the Antarctic, on every continent and most if not all islands, and in a huge variety of environments.

Assuming that Neornithes are a monophyletic group, for which there is substantial evidence (see, for example, Cracraft 2001 and references therein), three main questions about neornithine evolution remain:

1. When and where did neornithines originate?
2. How are the main lineages of modern birds interrelated?
3. What has been their temporal and spatial pattern of diversification?

Although all aspects of avian phylogeny are important in their own right, this study focuses on aspects of questions 2 and 3:

1. The avian tree. This is split into two subprojects:
   a) finding the root of the avian tree, and
   b) resolving some deep avian relationships.
2. Dating avian divergences using well-supported early fossils (i.e. how many lineages of modern birds existed before the Cretaceous-Tertiary boundary?).

1.1 A brief overview of avian phylogeny and taxonomy

Despite centuries of study, relationships between modern birds remain unclear. Two relationships which are now well-accepted are:

1. That the earliest division within modern birds is between Palaeognathae (ratites and tinamous) and Neognathae (all other modern birds) and

2. That Galliformes (landfowl e.g. chickens, quail, megapodes, cracids) and Anseriformes (waterfowl: screamers, ducks, geese and swans) are early offshoots within Neognathae.

While there was some earlier debate about whether or not Galliformes and Anseriformes were most closely related to each other relative to other groups, there is increasing evidence that they are indeed sister groups (see for example Cracraft 1988, and references therein; Sibley and Ahlquist 1990, and Groth and Barrowclough 1999). This therefore divides Neognathae into Galloanserae (Galliformes and Anseriformes) and Neoaves (all other neognaths).

Relationships within Neoaves are far less clear. The number of avian orders is not agreed upon, let alone the number of families, genera, and species (new species are steadily being recognised, old ones are being reduced to synonymy or otherwise eliminated, and so on). Nevertheless, as illustrated in the following figures, there is reasonable agreement on the main groups. The figures are given in approximate chronological order and start with Figure 1, which is from Cracraft (1988). This was one of the more recent morphological trees before 1990, when the DNA/DNA hybridization studies of Sibley and Ahlquist were published (see Figure 2).

Figure 1 shows three extinct lineages of birds which are sister to the extant Aves; these are not discussed further because they are outside the scope of molecular data. Then there is the standard split between Paleognathae (ratites and tinamou) and Neognathae (all other birds). This split has been recognised for over 100 years, and is accepted by virtually all evolutionary schemes for birds; two apparent exceptions for molecular data will be discussed later. More interesting perhaps is the union of Galliformes and Anseriformes. Cracraft (1988) was one of the first to strongly advocate 'Galloanseriformes' (Galloanserae) as a group, and all recent work continues to support this. The remaining birds form the Neoaves, which is by far the most numerous avian group, comprising about 95% of bird species (see Table 1) and Cracraft made an important start in subdividing Neoaves into six supra-ordinal groups (Figure 1). Conversely, it is equally interesting which groups are not placed together – the Charadriiformes (shore birds, including sea gulls) are not grouped with the other marine birds, and this is still supported (GC Gibb, unpublished results).

The DNA/DNA hybridization results of Sibley and Ahlquist (1990) are shown in Figure 2, and this represents the first attempt at a comprehensive avian tree from molecular data alone. The three main divisions of Paleognathae, Galloanserae and the Neoaves are there, but there are some anomalies in that the root now appears between the Eoaves (Paleognathae and Galloanserae) and the Neoaves. However, this is just a question of the placement of the root of the tree; the unrooted tree is not changed. Apart from the lack of a secure placing of the buttonquail (*Turnix* spp.) the Neoavian part of the tree is quite well defined. However, because it was generated from DNA-DNA hybridisation data, the tree was (necessarily) generated using only distance methods and so cannot use as much information as methods that use primary sequence data (Penny 1982).

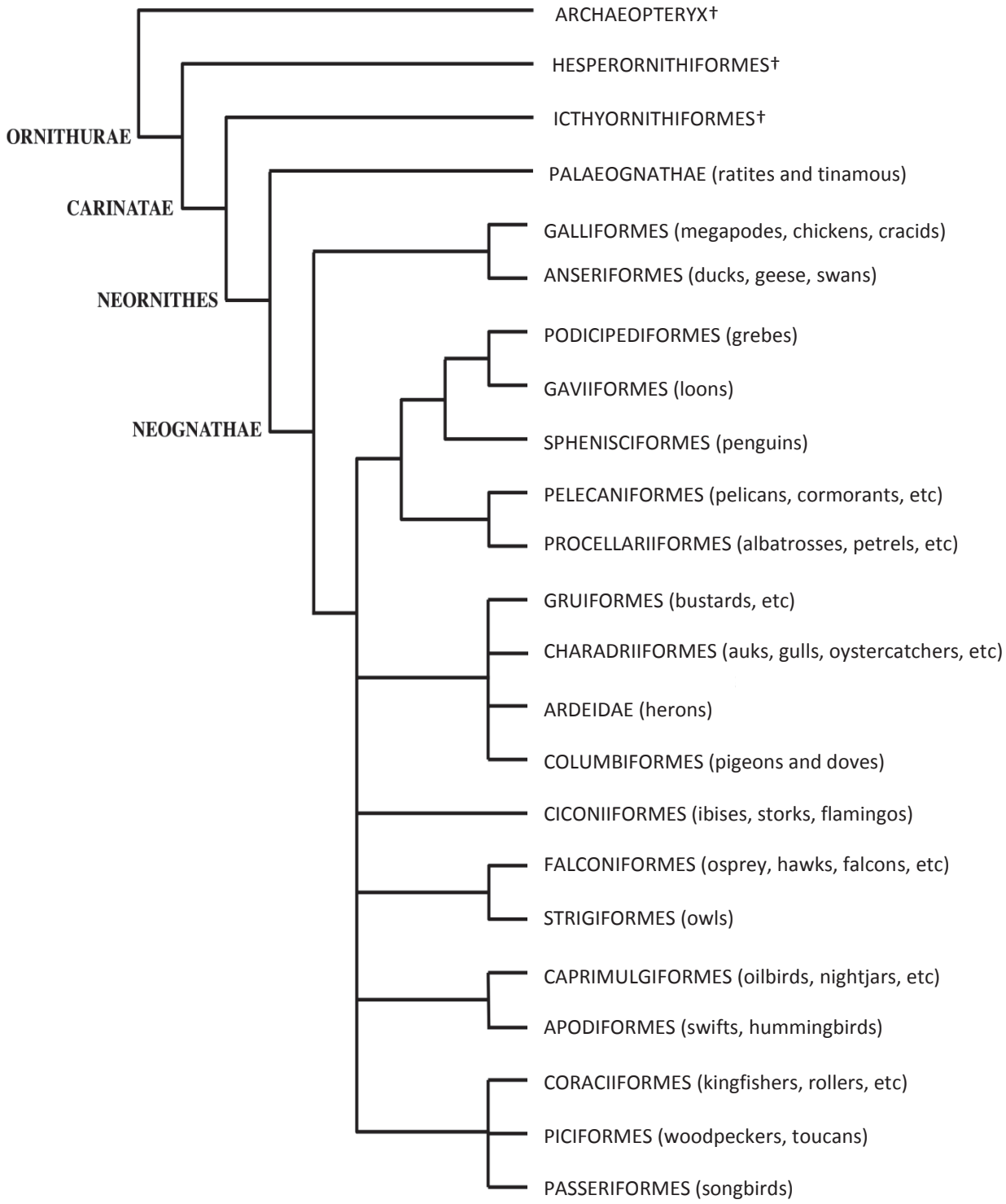


Figure 1. An important start by Cracraft (1988, Figure 9.1) to define the major clades of birds (Aves) using an explicit tree, postulated derived morphological characters and, where appropriate, comparisons to molecular-based hypotheses († marks extinct taxa). Common names (in brackets) have been added for clarity. Not placed: Hamerkop (Scopidae; probably part of Ciconiiformes), shoebill stork (Balaenicipitidae; traditionally placed in Ciconiiformes), secretary-bird (Sagittaridae; probably a falconiform) and cathartid vultures (Cathartidae; usually placed in Falconiformes).

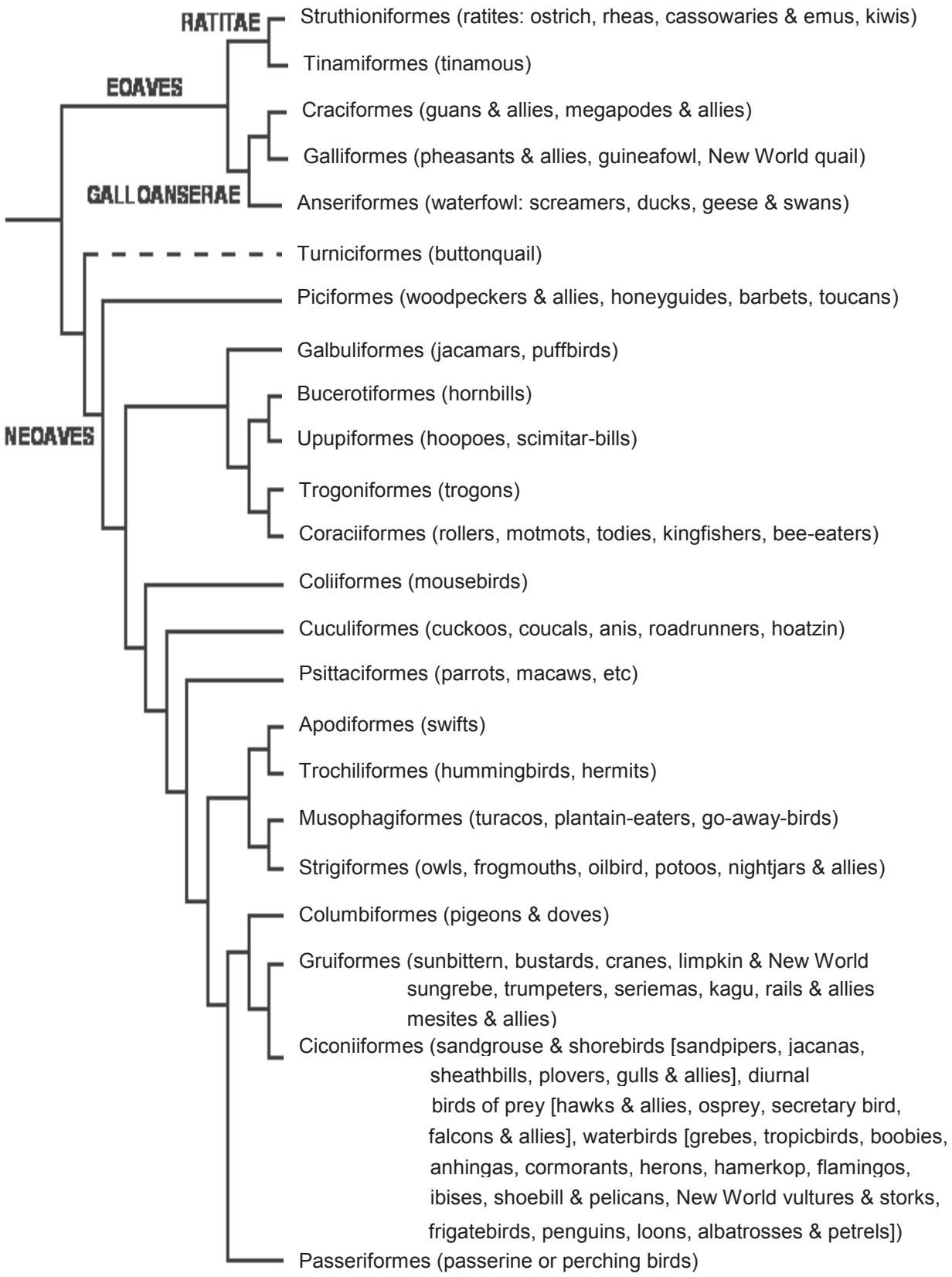


Figure 2. Phylogeny of modern birds (Neornithes) from Sibley and Ahlquist (1990, Figure 353, Chapter 16) and based on DNA/DNA hybridization. Common names have been added for clarity. One surprise is that Galloanserae is grouped with Ratitae to form Eoaves – this is now recognised as a problem of the placement of the root of the avian tree.

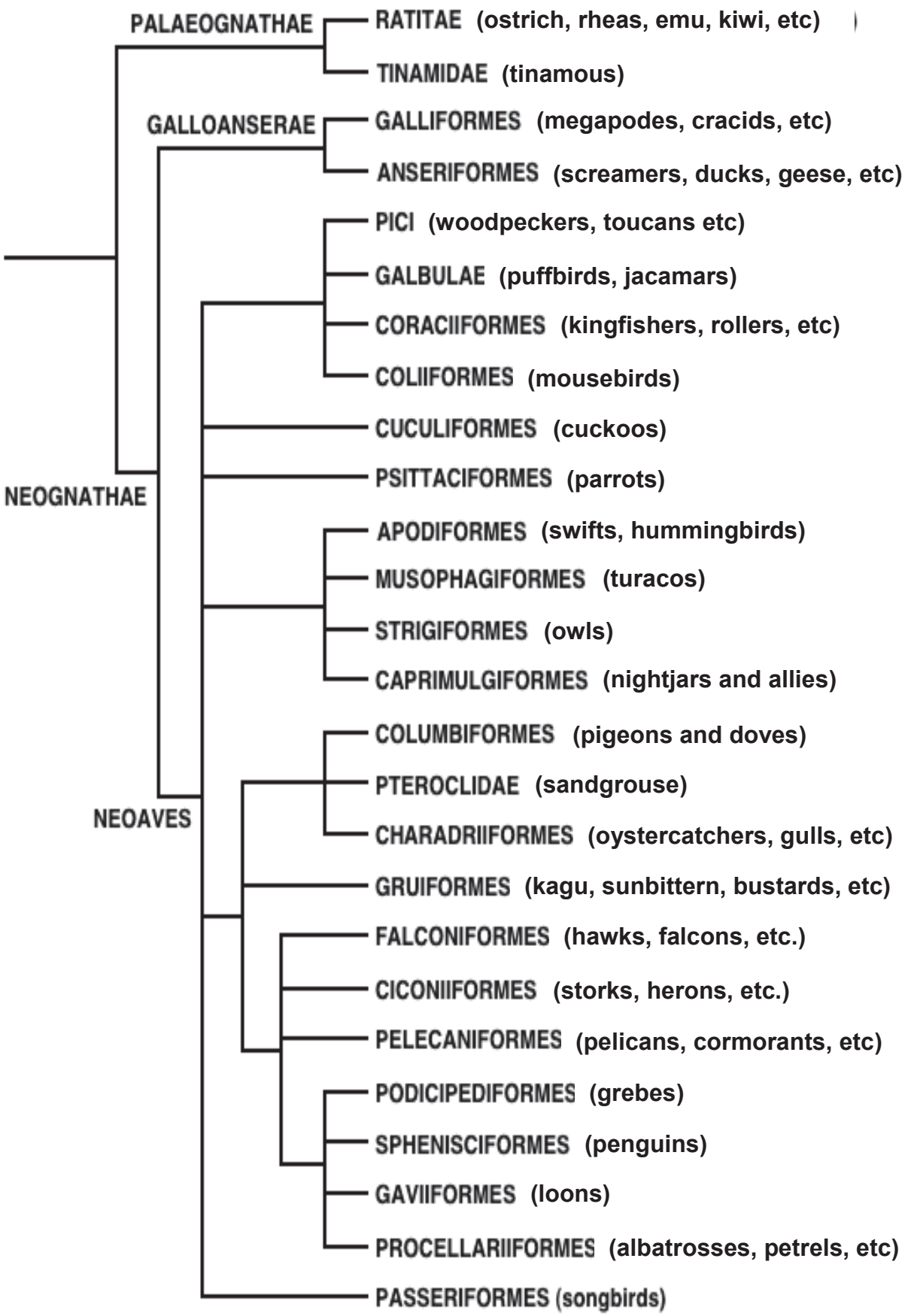


Figure 3. Phylogeny of modern birds (Neornithes) from Cracraft (2001, Figure 6). Common names have been added for clarity.

Figure 3 is from Cracraft (2001) and provides a comprehensive overview of phylogenetic relationships based now on both morphological and molecular data. The molecular data include immunological distances, amino acid sequences, DNA hybridisation data, and nuclear and mt DNA sequences. In addition, the tree includes information from biogeography i.e. biogeographical mapping of phylogenetic relationships, current geographical distribution, inferred historical geographical distribution, and historical reconstruction of continental geography. In other words, after the tree was inferred Cracraft looked at geographical distribution of lineages, age and geographical location of well-supported fossils, and mapped the tree onto a historical reconstruction of continental geography.

Comparing Figures 1 and 3, the same basic Paleognathae/Neognathae and Neoaves/ Galloanserae divisions are found, and there is still a 6-way division of Neoaves. However, there are several changes in these subdivisions. For example, in the more recent paper the Passeriformes are a separate group on their own; the hawks etc (Falconiformes) are separated from the owls (Strigiformes), and the group containing grebes/loons/penguins/albatrosses/ pelicans etc. is expanded to include Ciconiiformes and Falconiformes.

Several taxonomic schemes exist. For taxonomic names this thesis follows Clements (2005; see Table 1). Previous schemes were based on books that were bulky, cost-prohibitive and quickly out-of-date. For example, before the first edition of Clements's 'Birds of the world: A Checklist' was published in 1974, the standard reference was the 16-volume 'Checklist of the Birds of the World' by J. L. Peters and others, the first volume of which was published in 1934 (revised in 1979), and the last volume in 1987.

Clements (2005) was selected because it is:

- a single volume,
- readily available and affordable,
- regularly up-to-dated (including via the internet),
- based on conservative taxonomy,
- the official world checklist of the American Birding Association.

In addition:

- appropriate citations are included (e.g. prominent ornithological journals),
- it covers extant taxa only,
- almost all species names (scientific and English) follow the American Ornithologists' Union (AOU) nomenclature (British Ornithologists' Union names are used in some cases where there is a conflict in the AOU area with a long-standing name in the Palearctic region),
- for species that occur outside the AOU region of North America, the most widely currently accepted nomenclature is used - any exceptions are explained and referenced,

Table 1. Avian higher order taxonomy (Clements 2005).

Order	English Names	Number of Families	Number of Species
Struthioniformes	ostrich; rheas; cassowaries; emu; kiwis	5	12
Tinamiformes	tinamous	1	46
Sphenisciformes	penguins	1	17
Gaviiformes	loons	1	5
Podicipediformes	grebes	1	19
Procellariiformes	albatrosses; shearwaters and petrels; storm-petrels; diving-petrels	4	114
Pelecaniformes	tropicbirds; pelicans; gannets and boobies; cormorants; anhinga and darter; frigatebirds	6	67
Ciconiiformes	herons, egrets and bitterns; hamerkop; storks; shoebill; ibises and spoonbills	5	117
Phoenicopteriformes	flamingos	1	6
Anseriformes	screamers; ducks, geese and swans	2	162
Falconiformes	New World vultures; osprey; hawks, eagles and kites; secretary-bird; falcons and caracaras	5	313
Galliformes	megapodes; guans, chachalacas and curassows; turkeys; grouse and allies; New World quail; pheasants and partridges; guineafowl	7	283
Opisthocomiformes	hoatzin	1	1
Gruiformes	mesites; buttonquail; cranes; limpkin; trumpeters; rails, gallinules and coots; sungrebe and finfoots; kagu; sunbittern; seriemas; bustards	11	204
Charadriiformes	jacanas; painted-snipes; crab plover; oystercatchers; ibisbill; avocets and stilts; thick-knees; pratincoles;coursers; plovers and lapwings; Magellanic plover; sandpipers and allies; skimmers; seedsnipes; plains-wanderer;sheathbills;skuas and jaegers; gulls; terns; auks, murres and puffins	19	355
Pterocliformes	sandgrouse	1	16
Columbiformes	pigeons and doves	1	309
Psittaciformes	cockatoos and allies; parrots, macaws and allies	2	364
Cuculiformes	turacos; cuckoos	2	164
Strigiformes	barn-owls; typical owls	2	215
Caprimulgiformes	oilbird; owlet-nightjars; frogmouths; potoos; nightjars and allies	5	120
Apodiformes	swifts; treeswifts; hummingbirds	3	442
Coliiformes	mousebirds	1	6
Trogoniformes	trogons and quetzals	1	40
Coraciiformes	kingfishers; todies; motmots; bee-eaters; typical rollers; ground-rollers; cuckoo-roller; hoopoes; woodhoopoes and scimitar-bills; hornbills	10	219
Piciformes	jacamars; puffbirds; barbets; toucans; honeyguides; woodpeckers and allies	6	411
Passeriformes	perching birds	98	5892
		202	9919

the taxonomic sequence of families follows that of the yet-to-be completed series of the 'Handbook of the Birds of the World' (del Hoyo, J., et al. editors; Lynx Edicions; volume 1 was published in 1992; the series will be complete when volume 16 is published in late 2011).

There are many important similarities between Figure 3 (Cracraft 2001) and Table 1 (Clements 2000), but also some significant differences. For example, Cracraft (2001) has 26 main groups with 21 orders and 5 subordinal groups. By contrast, Clements (2005) has 27 orders. Cracraft (2001) separates Pici (woodpeckers and allies) and Galbulae (puffbirds and jacamars), whereas Clements (2005) unites them as Piciformes. Similarly, Cracraft (2001) separates cuckoos and turacos, but Clements joins them together in Cuculiformes. In addition, Clements has flamingos, hoatzin and trogons as separate orders.

In summary, and without going further into all the details, comparing Figures 1 - 3 and Table 1 gives an indication of the difficulty of resolving the main divisions within Neoaves. The intent here is not to review in detail the history of avian classification but only to indicate how many interesting hypotheses remain to be tested.

1.2 Fossils and the ages of modern bird lineages

A large part of the problem in dating the evolution of birds is that, in general, they have a relatively poor fossil record; some of the reasons for this are discussed in more detail in Chapter 3. To summarize, the oldest birds were non-marine, often small, and had light bones that were usually hollow (hollow bones reduce weight for flying whilst maintaining overall strength). Taken together, these features reduce the long-term survival of good fossils. In addition, Cracraft (2001) suggested that much of the evolution of early birds occurred in the southern hemisphere, but there was formerly an understandable bias towards northern hemisphere fossils being discovered. Van Tuinen et al. (2000) also suggested a bias in the avian fossil record against non-marine taxa (especially with regards to Cretaceous fossils). Indeed, the two oldest fossils attributable to modern neornithine birds are both aquatic species: *Vegavis* (Anseriformes) from Antarctica (Clarke et al. 2005) and a fossil penguin (*Waimanu*) from North Canterbury, New Zealand (Chapter 3, Slack et al. 2006).

Overall, very few Cretaceous fossils have been identified as neornithine birds. Perhaps the two most important are *Presbyornis* (Noriega and Tambussi 1995) and *Vegavis* (Clarke et al. 2005), both members of the order Anseriformes and both from the Late Cretaceous of Antarctica. *Vegavis* appears to be a deeply diverging member of the waterfowl lineage, with cladistic studies placing it as a sister group to Anatidae (ducks, geese and swans - excluding the magpie goose [Anseranatidae]) (Livezey 1997; Ericson 1997). *Presbyornis* is considered to be reasonably closely related to *Vegavis* and thus also deeply diverging.

Another important pair of fossils is the loons (Gaviiformes) from the Late Cretaceous (Chatterjee 1989, Olson 1992). Professor Judd Case (an avian paleontologist, Eastern Washington University, Spokane, WA) in a personal communication suggests that there are two loon fossils - one is definitely a loon but may be an early Tertiary intrusion into Cretaceous strata, while the other is definitely Cretaceous - but may not be a crown group (modern) loon. It is certainly a priority to clarify both the classification and the age of these two fossils.

Other even earlier fossils consist of a considerable numbers of fossil footprints of shorebird-like species dating from around 75 MYA (million years ago; Lockley 1998). It has been suggested that these were created by relatives of modern gulls (Charadriiformes; see for example, Olson and Parris 1987, Case and Tambussi 1999). However, obviously foot-prints (by themselves) do not give sufficient information for a firm identification. Nevertheless, molecular dating (e.g. Cooper and Penny 1997) does suggest that Charadriiformes had diverged by this time. This question is discussed in more detail in Chapter 3.

There is also one report of a fossil bone from a parrot (Psittaciformes) from the Late Cretaceous of North America (Stidham 1998). However, the identification was made on a single dentary bone, and its identity has been disputed by some authors (see Dyke and Mayr 1999).

Although avian phylogeny is interesting in its own right, there other reasons for needing to know the basics of both avian phylogeny and avian divergence times. There are a variety of hypotheses for early avian evolution (see Penny and Phillips 2004). At one end of the spectrum there are what may be called the ‘fossil literalist’ hypotheses. These (for example, Olson 1989; Feduccia 1995; 1999) take the fossil data literally and, without considering evolutionary mechanisms, suggest that there was a rapid early Tertiary radiation following a supposed ‘mass extinction’ caused by the extraterrestrial impact that marks the K/T (Cretaceous/Tertiary) boundary. Supposedly, one (or a few) ‘lucky’ Late Cretaceous neornithines (the transitional ‘shorebirds’) survived this K/T event, and all other pre-neornithine lineages went extinct. Under this model, almost all modern birds (possibly excluding paleognaths) subsequently originated and diversified in the northern super-continent of Laurasia.

At the other end of the spectrum are those that take the results of molecular clock studies fairly literally (for example, Hedges et al. 1996, Cooper and Penny 1997, Rambaut and Bromham 1998). These studies propose that most modern avian orders originated in the Cretaceous, with the deepest divergence dating around 100 MYA, or more. These hypotheses usually emphasize the limitation of evolutionary rates imposed by known mechanisms of genetic change that can lead to major morphological and ecological divergences.

By obtaining further agreement on both avian phylogeny and divergence times, this thesis aims to evaluate where in this spectrum of hypotheses avian evolution is most likely to lie.

1.3 Complete mitochondrial genome data

When this project began it was becoming routine to use DNA sequences for phylogenetics – but usually from just a single gene or a small number of genes. My concern was that using such short sequences can result in a significant sampling error that is difficult to evaluate, especially because this was before the availability of modern Bayesian methods that might allow for some uncertainty. The use of longer sequences, such as mitochondrial (mt) genomes, promised to reduce sampling error. A chicken mt genome was already available, and groups such as Anna Härlid and Ulfur Arnason of Lund University (Sweden), and David Mindell's group (then at the University of Michigan, USA) were leading the way for avian phylogenetics in using whole avian mt genomes to evaluate evolutionary relationships (see later).

Mt DNA seemed to be advantageous for the following types of reasons (see, for example, Kvist 2000 and references therein):

1. As mentioned above, the use of complete mt genomes is expected to reduce sampling effects ('stochastic error') associated with short datasets. There are, of course, still problems with identifying the best model of sequence evolution.
2. There are certain technical advantages in using mt genomes. They are compact, small and easier to isolate compared to nuclear DNA regions.
3. Mt genomes provide another independent dataset that is complementary to nuclear and morphological data.
4. There is essentially no orthology/paralogy problem within birds - there are no duplicate copies of the same gene (mt genes are equivalent to each other between species).
5. Mt genomes are maternally inherited and have no recombination, thus there is only one tree (no paralogy problem). For deeper avian divergences there should be good agreement with the nuclear tree, though it would not be surprising if this differed from the morphological tree (because of problems such as convergence of morphological features).
6. On average, mt DNA has a higher mutation rate than nuclear DNA (that is, there are more changes per site for mt DNA). Thus much longer nuclear sequences are expected to be required to provide the same amount of information as shorter mt DNA sequences. (The corollary to this, of course, is that saturation will eventually become a problem when mt DNA is used to study very deep divergences.)

For these reasons I decided that it was worth focussing on whole mt genomes of birds in order to investigate questions in avian evolution. It was fortunate that technologies continued to improve, and soon long range PCR was also available (see individual chapters for more details). There were still potential problems with nucleotide composition bias (Lockhart et al.

1994), and close attention was also paid to selection of taxa in order to reduce 'long branch attraction' effects (Hendy and Penny 1989; Holland et al. 2003).

1.4 Overview of earlier studies using complete mt genomes

At the time I started my study of avian phylogeny, some of the first studies using complete avian mt genomes were being published. Some interesting but unexpected results were found in these early studies. For example, as shown in Figures 4 and 5, both Härlid and Arnason (1999) and Mindell et al. (1999) reconstructed an unexpected placement of the root of the avian tree.

The unrooted tree (that is, excluding the outgroup) agreed with previous hypotheses in that Paleognathae (Ratites), Galloanserae (Galloanseriformes), and Neoaves formed distinct groups. However, when the outgroup was added in, the rooted tree (Figures 4 and 5) was distinctly different in that Passeriformes were both

paraphyletic, and

the earliest diverging group among extant (modern) birds.

While the 'neoteny' hypothesis of Härlid and Arnason (1999) was well thought out, another possible explanation was that of 'long branch attraction' (e.g. Hendy and Penny 1989; Holland et al. 2003) between the fast-evolving alligator and passerine lineages. Under long branch attraction, an unrooted subtree might be correct but, when rooted using a distant outgroup, the outgroup may attach to an incorrect long branch in the ingroup. This effect could possibly account for the observation that the unrooted tree was in agreement with a long-term understanding of avian relationships, but the rooted tree was definitely not. If long branch attraction was a problem, then it was important to add additional taxa because breaking up long branches is known to reduce the problem (Hendy and Penny 1989).

1.5 Selection of taxa

In this section I give some of my reasoning for the selection of taxa to be sequenced. Clearly, availability of taxa played a role, but the two main criteria were:

1. Calibration points, and
2. Stability of the tree.

i. Taxa with early fossils: Dating

The following taxa were selected for sequencing because there are either early fossils or possible early fossils.

Penguin (Sphenisciformes) – penguins have a good fossil record, possibly the best for neornithine birds (see Chapters 2 and 3).

Loon (Gaviiformes) – as mentioned above, early loon fossils have been found, and loons have been proposed to be close relatives of penguins (see Chapter 3).

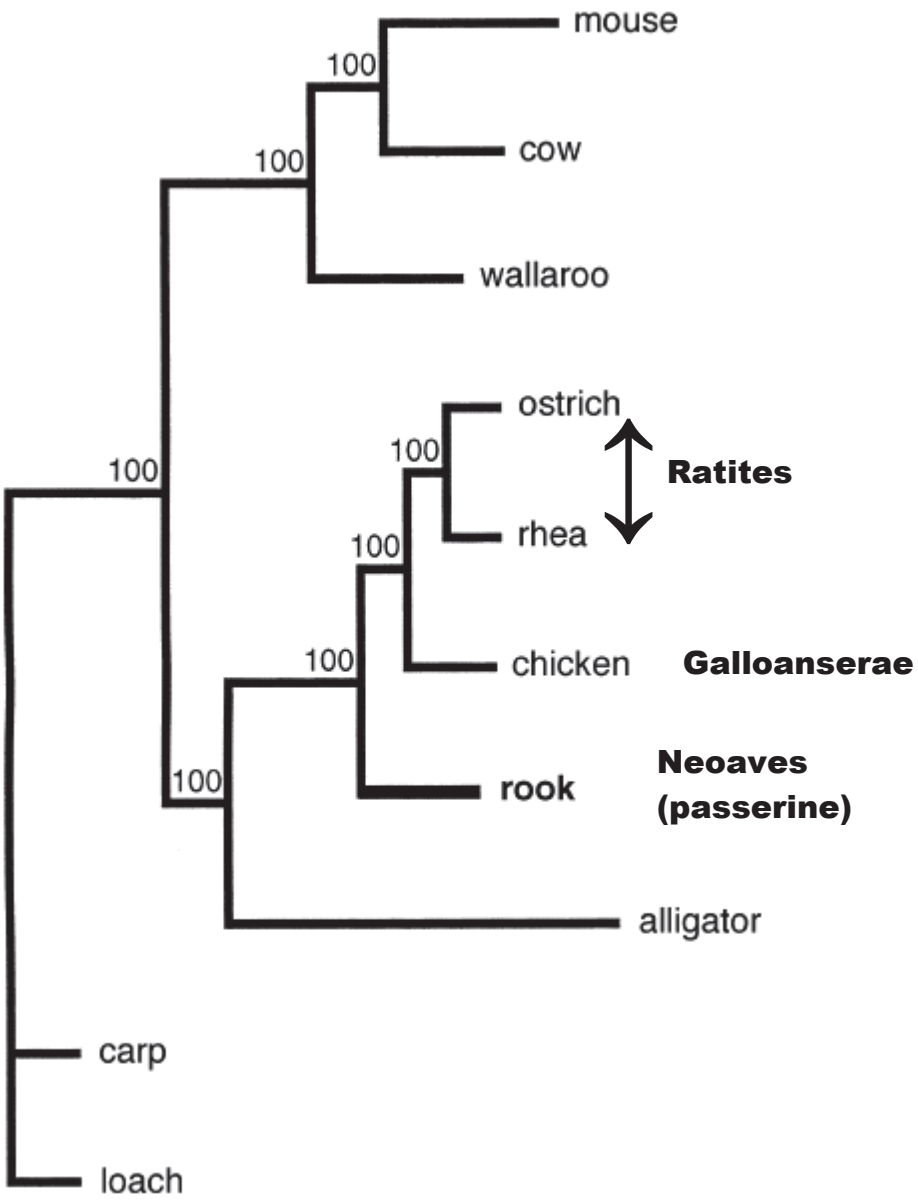


Figure 4. Avian relationships from Figure 1 of Härlid and Arnason (1999). Higher order names have been added for clarity. The tree was generated using maximum likelihood-quartet puzzling (ML-QP) analysis of the combined amino acid sequences of 11 protein-coding mitochondrial genes and was rooted using the carp and the loach. The alligator and mammals were also used as outgroups to the avian taxa. Support values are based on 100 ML replicates and are given above the branches.

Gull (Charadriiformes) – there are early fossil shorebird footprints that may be charadriiform. Indeed, one early hypothesis, involving 'transitional shorebirds', suggested that Charadriiformes were the most deeply diverging group of modern birds (see earlier).

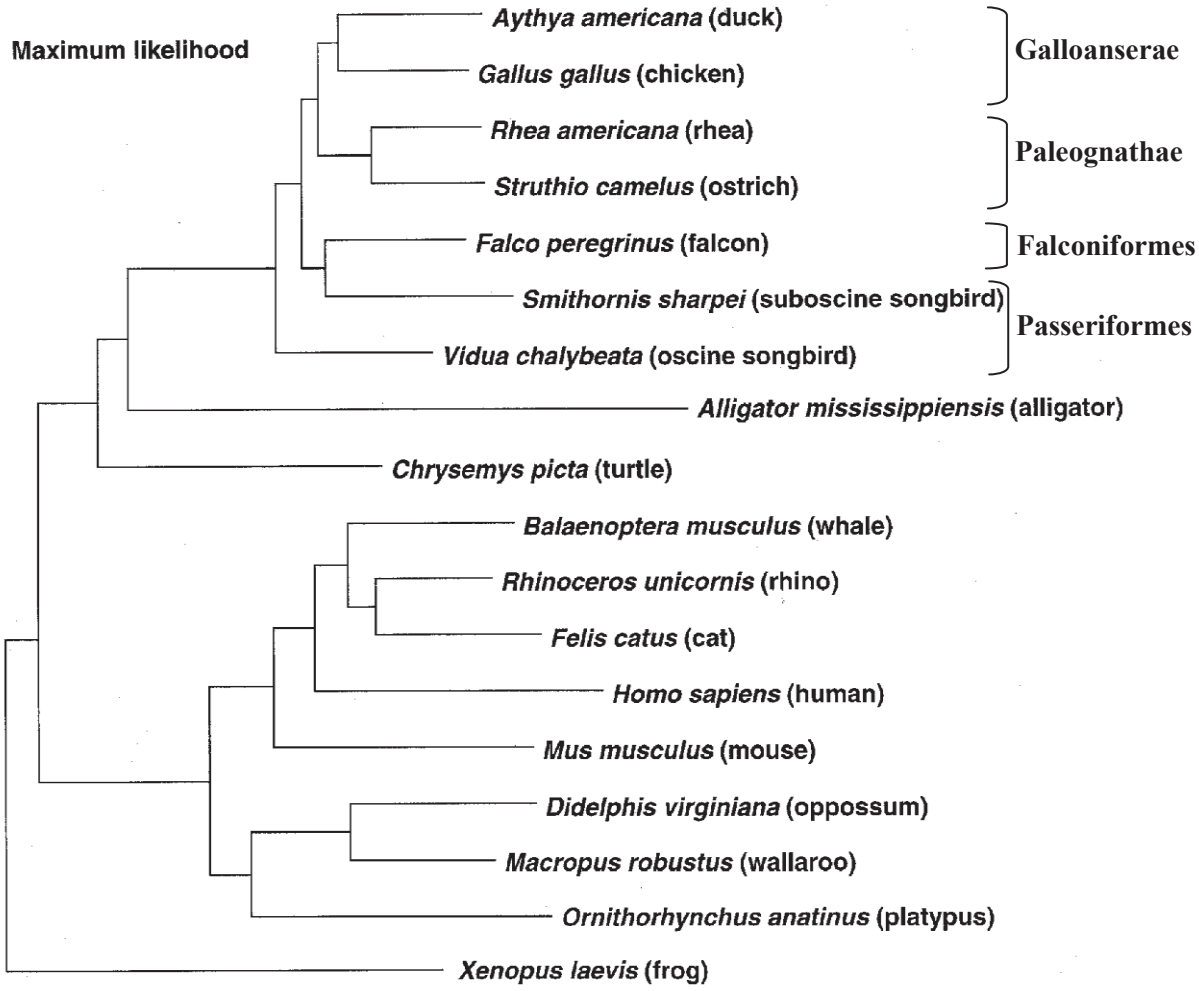


Figure 5. Avian relationships from Figure 2 of Mindell et al. (1999). Higher order avian names have been added for clarity. Optimal maximum likelihood hypothesis based on analysis of the combined nucleotide sequences of 37 mitochondrial genes (13 protein-coding genes, 2 rRNAs and 22 tRNAs), using the general time-reversible model and a discrete approximation to the gamma distribution to help accommodate evolutionary rate heterogeneity across nucleotide sites.

ii. Possible close relatives of taxa with early fossils

The following taxa were selected because they were putative relatives of groups for which there are early fossils.

Albatross (Diomedeidae) and petrel (Procellariidae) (tubenose seabirds; order Procellariiformes). Cooper and Penny (1997) found that, of the birds included in their study, the closest relatives of the penguins were the albatrosses. This hypothesis was somewhat supported by a study (Kinsky 1960) showing that the little blue penguin, at least, has a tubenose when young. An albatross/penguin pairing and early penguin fossils from about 61 MYA give minimal dates for early avian evolution in the Cretaceous (see Chapter 3).

Magpie goose (Anseranatidae; Anseriformes) – which diverged prior to the *Vegavis*/ Anatidae split (see Noriega and Tambussi 1995 and Clarke et al. 2005). The magpie goose was sequenced by colleagues (see Appendix 3).

iii. Root of the avian tree: to test the long branch attraction hypothesis.
Here the criterion for selecting taxa was to help evaluate the early results that suggested Passerines were both basal and paraphyletic. This required taxa that would help break up long branches deep on the Passerine lineages. The following taxa were selected.

Lyrebird (Menuridae; Passeriformes) – a basal oscine.

Flycatcher (Tyrannidae; Passeriformes) – a New World suboscine – sequenced by a colleague (Chapter 3).

Rifleman (Acanthisittidae; Passeriformes) – a New Zealand wren – suggested to be basal to both oscines and suboscines. This was also sequenced by a colleague (Appendix 3).

iv. Galliformes and Anseriformes (testing for long branch attraction)
Along with the magpie goose, additional land- and waterfowl were chosen to test two (mutually exclusive) hypotheses. It was possible that the chicken and duck mt genomes came together from long branch attraction, or that they genuinely formed a natural grouping. The following taxa were selected for sequencing.

Brush-turkey (Megapodiidae, Galliformes) - an early branching galliform (Cracraft 1988).

Goose (Anatidae), order Anseriformes.

v. Other taxa of interest
One additional taxon (a turkey vulture) was selected because of unexpected results involving the placement of the penguin in initial analyses. A strong signal put the penguin and the stork together (Chapters 2 and 3). It has been suggested that the New World vultures (Cathartidae) are closer to storks (Ciconiiformes) than to other birds of prey (Falconiformes). If this were correct, the inclusion of the turkey vulture might break up the penguin/stork grouping and thereby help narrow down the divergence point of the penguins. This taxon was sequenced by a colleague.

1.6 Overview of the thesis

When this study began, only the chicken mt genome was available. Shortly after this, seven more avian mt genomes were published (Figures 4 and 5) and by the time of actual publication of my first paper, 17 had been sequenced in total – 10 of them paleognaths.
In total, I contributed eight new complete avian mt genomes. My two main themes relate to the avian evolutionary tree itself (the relationships of the major orders of birds) and the times of origin of the main orders.

The eight avian taxa for which I sequenced mt genomes were: a penguin, a loon, a gull, a petrel, an albatross, a lyrebird, a brush-turkey and a goose. In parallel with this study, colleagues sequenced the magpie goose, a turkey vulture, a flycatcher and the rifleman.

Perhaps it is not giving away too much to say that the three-way division into Paleognathae, Galloanserae and Neoaves was relatively easy to establish from mt genomes, but the main divisions within the Neoaves were more difficult to resolve. Some groupings of orders were well-supported, but additional support from nuclear DNA would be beneficial. Nevertheless, the times of divergence of the main orders of birds appear to be in the Late Cretaceous, and this aspect of the work of the thesis is strongly supported. The results of this thesis are given in the following Chapters and Appendices and are discussed briefly in Chapter 5 (Summary and Discussion).

Chapter 1. Introduction (this chapter)

Chapter 2. **Slack, K.E**., Janke, A., Penny, D. and Arnason, U. (2003). Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. *Gene* **302**: 43-52.

Chapter 3. **Slack, K.E**., Jones, C.M., Ando, T., Harrison, G.L., Fordyce, R.E., Arnason, U. and Penny, D. (2006). Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Mol. Biol. Evol.* **23**: 1144-1155.

Chapter 4. **Slack, K.E**., Delsuc, F., McLenachan, P.A., Arnason, U. and Penny, D. (2007). Resolving the root of the avian mitogenomic tree by breaking up long branches. *Mol. Phyl. Evol.* **42**: 1-13.

Chapter 5. Summary and discussion

Appendix 1. Statements from contributors

Appendix 2. Ursing, B.M., **Slack, K.E**. and Arnason, U. (2000). Subordinal artiodactyl relationships in the light of phylogenetic analysis of 12 mitochondrial protein-coding genes. *Zool. Scr.* **29**: 83-88.

Appendix 3. Harrison, G.L., McLenachan, P.A., Phillips, M.J., **Slack, K.E.**, Cooper, A. and Penny, D. (2004). Four new avian mitochondrial genomes help get to basic evolutionary questions in the Late Cretaceous. *Mol. Biol. Evol.* **21**: 974–983.

Appendix 4. Tables summarizing features of avian mitochondrial genomes

References

Case, J.A. and Tambussi, C.P. (1999). Maestrichtian record of neornithine birds in Antarctica: comments on a Late Cretaceous radiation of modern birds. *J. Vert. Paleontol.* **19**: 37A.

Chatterjee, S. (1989). The oldest Antarctic bird. *J. Vert. Paleontol.* **9**: 16A.

Clarke, J.A., Tambussi, C.P., Noriega, J.I., Erickson, G.M. and Ketchman, R.A. (2005). Definitive fossil evidence for the extant avian radiation in the Cretaceous. *Nature* **433**: 305-308.

Clements, J.F. (2005). *Birds of the world: a checklist*. *Fifth edition* 2000; Ibis Publishing Company, 3420 Freda's Hill Road, Vista, CA 92084-7466; ISBN 0-934797-16-1, plus updates up until December 2005 (www.ibispub.com).

Cooper, A. and Penny, D. (1997). Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. *Science* **275**: 1109-1113.

Cracraft, J. (1988). The major clades of birds. Pp. 339-361 in *The Phylogeny and Classification of the Tetrapods, Volume 1: Amphibians, Reptiles, Birds* (ed. M. J. Benton). Systematics Association Special Volume **35A**. Clarendon Press, Oxford.

Cracraft, J. (2001). Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. *Proc. R. Soc. Lond. B* **268**: 459-469.

del Hoyo, J., Elliot, A., Sargatal, J. and Christie, D. A. (eds) (1992). *Handbook of the birds of the world. Volume 1*. Barcelona: Lynx Edicions.

Dyke, G.J. and Mayr, G. (1999). Did parrots exist in the Cretaceous period? *Nature* **399**: 317-318.

Ericson, P.G.P. (1997). Systematic relationships of the Paleogene family Presbyornithidae (Aves: Anseriformes). *Zool. J. Linn. Soc.* **121**: 429-483.

Feduccia, A. (1995). Explosive evolution in Tertiary birds and mammals. *Science* **267**: 637-638.

Feduccia, A. (1999). *The origin and evolution of birds. Second edition*. New Haven, CT: Yale University Press.

Groth, J.G. and Barrowclough, G.F. (1999). Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. *Mol. Phyl. Evol.* **12**: 115-123.

Härlid, A. and Arnason, U. (1999). Analyses of mitochondrial DNA nest ratite birds within the Neognathae: supporting a neotenous origin of ratite morphological characters. *Proc. R. Soc. Lond. B* **266**: 305-309.

Hedges, S.B., Parker, P.H., Sibley, C.G. and Kumar, S. (1996). Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**: 226-229.

Hendy, M.D. and Penny, D. (1989). A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38**: 297-309.

Holland, B.R., Penny, D. and Hendy, M.D. (2003). Outgroup misplacement and phylogenetic inaccuracy under a molecular clock – a simulation study. *Syst. Biol.* **52**: 229-238.

Kinsky, F.C. (1960). The yearly cycle of the northern blue penguin (*Eudyptula minor novaehollandiae*) in the Wellington Harbour area. *Rec. Dom. Mus.* **3**: 145-218.

Kvist, L. (2000). Mitochondrial DNA. Pp. 15 - 17 in *Phylogeny and phylogeography of European parids.* Oulu University Library. ISBN 951-42-5536-4/ISBN 951-42-5535-6. ISSN 0355-3191 (http://herkules.oulu.fi/isbn9514255364/isbn9514255364.pdf).

Livezey, B.C. (1997). A phylogenetic analysis of basal Anseriformes, the fossil *Presbyornis*, and the interordinal relationships of waterfowl. *Zool. J. Linn. Soc.* **121**: 361-428.

Lockhart, P.J., Steel, M.A., Hendy, M.D., and Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11**: 605-612.

Lockley, M.G. 1998. The vertebrate track record. *Nature* **396**: 429–432.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., Hasegawa, M., Ast, J.C. and Yuri, T. (1999). Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *Syst. Biol.* **48**: 138-152.

Mindell, D.P. Sorenson, M.D., and Dimcheff, D.E. (1998). Multiple independent origins of mitochondrial gene order in birds. *Proc. Natl Acad. Sci. USA*. **95**: 10693-10697.

Noriega, J.I. and Tambussi, C.P. (1995). A Late Cretaceous Presbyornithidae (Aves: Anseriformes) from Vega Island, Antarctic Peninsula: paleobiogeographic implications. *Ameghiniana* **32**: 57-61.

Olson, S.L. (1989). Aspects of the global avifaunal dynamics during the Cenozoic. Pp. 2023-2029 in *Proceedings of the 19th International Ornithological Congress* (ed. H. Oullet). University of Ottawa Press.

Olson, S.L. (1992). *Neogaeornis wetzeli* Lambrecht, a Cretaceous loon from Chile (Aves: Gaviidae). *J. Vert. Paleontol.* **12**: 122-124.

Olson, S.L. and Parris, D.C. (1987). The Cretaceous birds of New Jersey. *Smithsonian Contrib. Paleobiol.* **63**: 1-22.

Penny, D. (1982). Towards a basis for classification: the incompleteness of distance measures, incompatibility analysis and phenetic classification. *J. Theoret. Biol.* **96**: 129-142.

Penny, D. and Phillips, M.J. (2004). The rise of birds and mammals: are microevolutionary processes sufficient for macroevolution. *Trends Ecol. Evol.* **19**: 516-522.

Rambaut, A. and Bromham, L. (1998). Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15**: 442-448.

Sibley, C.G. and Ahlquist, J.E. (1990). *Phylogeny and classification of birds: a study in molecular evolution*. Yale University Press. ISBN 0-300-04085-7.

Slack, K.E., Janke, A., Penny, D. and Arnason, U. (2003). Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. *Gene* **302**: 43-52.

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

Stidham, T.A. (1998). A lower jaw from a Cretaceous parrot. *Nature* **396**: 29-30.

Van Tuinen, M., Sibley, C. and Hedges, S.B. (2000). The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. *Mol. Biol. Evol*. **17**: 451-457.

Chapter 2.

Slack, K.E., Janke, A., Penny, D. and Arnason, U. (2003). Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. *Gene* **302**: 43-52. (As at August 2011 this paper has been cited 36 times in the Web of Science.)

The two new avian mitochondrial (mt) genomes reported here are the little blue penguin (*Eudyptula minor*) from Wellington harbour (New Zealand), and the greater white-fronted goose (*Anser albifrons*) from Sweden. These two taxa were chosen for the following reasons. Penguins, not surprisingly for marine species, have a relatively good fossil record compared with most birds. Moreover, they have quite distinctive bones that are therefore relatively easy to identify as penguins. Thus it was considered important to include a least one penguin in the mt dataset in order to provide a solid, early calibration point for dating avian divergences.

The goose was chosen because it provided a second representative of Anseriformes (Anatidae). Having both duck and goose mt genomes, together with the magpie goose that was planned to come later, gave a good representation of the Anseriformes. This was important because early studies, such as Härlid and Arnason (1999) and Mindell et al. (1999) (references in Chapter 1) suggested that the bird tree might prove susceptible to 'long branch attraction' problems.

In addition, this paper also reported that I had produced tables of the main features of avian mitochondrial genomes. Part of the reason for these tables was that I found there was considerable inconsistency in the annotations of avian mitochondrial DNA, for example, start/stop codons. These tables have been updated until my last paper in 2007 and are available online.

I did the DNA sequencing and primary analysis, Axel Janke was responsible for important technical advice and laboratory supervision, UA and DP for general oversight and planning – all authors contributed to the final manuscript.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 302 (2003) 43–52
GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES
www.elsevier.com/locate/gene

Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features

Kerryn E. Slack[a,b,*], Axel Janke[a], David Penny[b], Ulfur Arnason[a]

[a]*Division of Evolutionary Molecular Systematics, Department of Cell and Organism Biology, University of Lund, Sölvegatan 29, S-223 62 Lund, Sweden*
[b]*Allan Wilson Center for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand*

Received 22 April 2002; received in revised form 6 September 2002; accepted 23 September 2002
Received by T. Gojobori

Abstract

We report complete mitochondrial (mt) genomes for a penguin (little blue, *Eudyptula minor*) and a goose (greater white-fronted, *Anser albifrons*). A revised annotation of avian and reptile mt genomes has been carried out, which improves consistency of labeling gene start and stop positions. In conjunction with this, a summary of mt gene features is presented and a number of conserved patterns and interesting differences identified. The protein-coding genes from the two new genomes were analysed together with those from 17 other birds plus outgroup (reptile) taxa. The unrooted amino acid tree from 19 avian genomes was locally stable with many high bootstrap values using several maximum likelihood methods. In particular, Anseriformes (goose and duck) grouped strongly with Galliformes (chicken) to form Gallianseres, while the penguin paired firmly with the stork. The position where the outgroup joined the avian tree varied with the combination of outgroup taxa used. The three best supported positions of the root were passerine, but the traditional rooting position between paleognaths and neognaths could not be excluded.

Keywords: Complete mtDNAs; Genome annotation; Avian evolution

1. Introduction

Both the times of origin and relationships between modern avian orders are still uncertain. Nevertheless, there is increasing confidence from molecular studies that these questions are now answerable, given adequate sampling of long DNA sequences. Complete mitochondrial (mt) genomes have been one major source of data for mammalian evolution (for example, see Arnason et al., 2002). However, very few avian orders are currently represented by complete mtDNAs.

Two of the few higher order avian groupings that are reasonably well resolved are the monophyly of the ratites (ostrich, rhea, emu, cassowary, kiwi and extinct moa and elephant bird) and the clustering of ratites plus tinamous (paleognaths) to the exclusion of all other modern birds (neognaths). Both groupings are supported by a number of morphological and molecular studies (see, for example, review in Sheldon and Bledsoe, 1993, and discussion and references in Cracraft, 2001, and Haddrath and Baker, 2001). However, although the paleognaths do appear to

Abbreviations: *12S* and *16S rRNA*, 12S and 16S rRNA subunits; [^{35}S]dATP, radioactively-labeled deoxyadenosine triphosphate; A, adenosine; aa, amino acid(s); Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; *ATPase6* and *8*, ATP synthase subunits 6 and 8; bp, base pair(s); C, cytidine; *COI*, *II* and *III*, cytochrome oxidase subunits I, II and III; CR, control region; Cys, cysteine; *Cytb*, cytochrome *b*; G, guanosine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; H strand, heavy strand; His, histidine; Ile, isoleucine; kb, kilobase(s); K/T, Cretaceous-Tertiary; L strand, light strand; LBP, local bootstrap probability; Leu, leucine; Lys, lysine; Met, methionine; ML, maximum likelihood; mt, mitochondrial; Myr, million years; *NADH1*, *2*, *3*, *4L*, *4*, *5* and *6*, NADH dehydrogenase subunits 1, 2, 3, 4L, 4, 5 and 6; nt, nucleotide(s); PCR, polymerase chain reaction; Phe, phenylalanine; Pro, proline; QP, quartet puzzling; R, purine (A/G); *RAG-1* gene, recombination activating gene (single-copy); RELL method, resampling estimated log-likelihood method; rRNA, ribosomal RNA; S.E., standard error(s); Ser, serine; T, thymidine; Thr, threonine; tRNA, transfer RNA; Trp, tryptophan; Tyr, tyrosine; Val, valine.

* Corresponding author. Tel.: +46-46-222-7862; fax: +46-46-147874.
E-mail address: kerryn.slack@gen.lu.se (K.E. Slack), K.E.Slack@massey.ac.nz (K.E. Slack).

0378-1119/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0378-1119(02)01053-3

form a natural grouping, analyses of complete mt protein-coding sequences cast some doubt on the traditional view that they also constitute the deepest branch of the avian tree (Harlid and Arnason, 1999; Mindell et al., 1999; but see also García-Moreno and Mindell, 2000; van Tuinen et al., 2000).

Another avian relationship that has received support is the early separation of Galliformes and Anseriformes from the remaining neognaths. Galliformes (*sensu lato*) includes pheasants, guineafowl, quail, cracids and megapodes, while Anseriformes consists of geese, ducks, swans and screamers. However, whether these two orders join on a common branch to form Gallianseres (Galli: plural form of Gallus; Anseres: plural of Anser) or constitute separate, adjacent branches has been the subject of some debate (see review by Caspers et al., 1997). Evidence for a gallianserine grouping includes morphological and immunological data, DNA hybridization distances (Sibley and Ahlquist, 1990) and analyses of ribosomal RNA (rRNA), *α-crystallin* and *RAG-1* gene sequences (see Cracraft, 2001). The publication of a duck mt genome (Mindell et al., 1999) provides further support. Nevertheless, the possibility that Galliformes and Anseriformes form independent branches cannot yet be excluded. The mt dataset currently includes only one representative from each order and, knowing that long branches can attract (Hendy and Penny, 1989), it is important to sequence additional members of both groups.

There is a spectrum of views regarding the time of origin of modern bird orders. At one end, a mass extinction of most avian groups at the end of the Cretaceous is assumed, followed by an 'explosive radiation' from one or a small number of surviving bird lineages in the Tertiary (for example, see Feduccia, 1995). However, explosive radiations are unlikely scenarios, requiring extensive morphological change over relatively short periods of evolutionary time and by unknown genetic mechanisms. Furthermore, a recent analysis of the palaeontological record (Kirchner, 2002) found innovation to be generally more continuous, not explosive.

At the other end of the spectrum, Cooper and Penny (1997) suggested that modern bird orders started diverging by the mid-Cretaceous, with at least 22 lineages crossing the Cretaceous-Tertiary (K/T) boundary 65 million years (Myr) ago. Cracraft (2001) recently reviewed the accumulated phylogenetic, palaeontological and biogeographic data, and also concluded that many modern avian orders were present prior to the K/T boundary. Van Tuinen and Hedges (2001) obtained similar results using mt RNA sequences and DNA hybridization and transferrin immunological distances.

The estimation of divergence times using molecular data requires solidly dated fossils that can be identified to present-day groups (calibration points). Few modern avian orders are currently represented by early fossils, however. There is evidence for Cretaceous shorebirds (see discussion and references in Feduccia, 1995), loons (Olson, 1992) and anseriforms (Elzanowski and Brett-Surman, 1995; Noriega and Tambussi, 1995) and Stidham (1998) described a Cretaceous parrot. Penguins have a good palaeontological record (see Fordyce and Jones, 1990), with a range of fossils of different ages. The oldest is of late Early Paleocene age (60.9–63 Myr) and comes from the Waipara Greensand, New Zealand (Jones and Mannering, 1997; C.M. Jones, personal communication). Until now, however, the mitogenomic dataset has included neither penguins nor their closest relatives.

For the present study, we have sequenced the complete mt genomes of a penguin and a goose. The penguin was selected because it will provide an excellent calibration point for the dating of avian divergences (although other seabirds are needed before this can be fully utilized). The goose was chosen because additional gallianserine sequences are required to reduce the risk of artifacts from long branch attraction. During the assembly of the datasets used in this study (bird and bird plus reptile outgroups), it became apparent that there were many discrepancies in bird and reptile mt genome annotation, some of which are the result of early errors being copied and repeated in later work. We have therefore reexamined all published avian and reptilian mt genomes and present a consistent labeling for start and stop positions. These changes have been incorporated into a summary of complete mt genome features in table form that can be used as a basis for further study, including improving the accuracy of mt genome sequencing, annotation and alignment.

2. Materials and methods

2.1. Data collection and sequence alignment

The little blue penguin (*Eudyptula minor*) came from the Nelson conservancy region of New Zealand, while the greater white-fronted goose (*Anser albifrons*) was from the province of Scania in southern Sweden. MtDNA was prepared from approximately 7 g (penguin) and 14.5 g (goose) of liver tissue following the procedure of Arnason et al. (1991). The DNA was restriction enzyme-digested, ligated into the cloning vectors M13mp18/mp19 and/or pUCmp18/mp19 and transformed into *Escherichia coli* strain DH5α. Approximately 540 bp from the goose *NADH5* gene, 575 bp from the penguin *Cytb* gene and a 2.7 kb penguin fragment (*NADH4L* to halfway through *NADH5*) were PCR-amplified prior to cloning. The clones were sequenced using the dideoxy method with [^{35}S]dATP. To ensure that no errors were introduced into the sequence of the amplified regions, a minimum of two (goose) to three (penguin) PCR clones were sequenced.

Seventeen other avian taxa were included in analyses: chicken (*Gallus gallus*; GenBank accession number X52392), ostrich (*Struthio camelus*; Y12025), greater rhea (*Rhea americana*; Y16884), rook (*Corvus frugilegus*; Y18522), redhead duck (*Aythya americana*; AF090337), peregrine falcon (*Falco peregrinus*; AF090338), gray-headed broadbill

(*Smithornis sharpei*; AF090340), village indigobird (*Vidua chalybeata*; AF090341), Oriental white stork (*Ciconia boyciana*; AB026193), giant moa (*Dinornis giganteus*; AY016013), eastern moa (*Emeus crassus*; AY016015), great spotted kiwi (*Apteryx haastii*; AF338708), emu (*Dromaius novaehollandiae*; AF338711), double-wattled cassowary (*Casuarius casuarius*; AF338713), lesser rhea (*Pterocnemia pennata*; AF338709), great tinamou (*Tinamus major*; AF338707) and elegant crested-tinamou (*Eudromia elegans*; AF338710). Six reptile outgroups were also used: American alligator (*Alligator mississippiensis*; Y13113), eastern painted turtle (*Chrysemys picta*; AF069423), green turtle (*Chelonia mydas*; AB012104), blue-tailed mole skink (*Eumeces egregius*; AB016606), common iguana (*Iguana iguana*; AJ278511) and spectacled caiman (*Caiman crocodylus*; AJ404872).

The 12 protein-coding genes encoded by the mt heavy (H) strand were aligned at the amino acid (aa) level using Se-Al version 1.0α1 (http://evolve.zps.ox.ac.uk/software/Se-Al/main.html). The alignments were then edited to remove gaps and adjacent ambiguous sites, using conserved aa columns to define the boundaries of these regions. After editing, the genes were concatenated in PAUP* version 4 (http://paup.csit.fsu.edu/index.html) to give one long sequence (or 'supergene') for each taxon. The light (L) strand-encoded *NADH6* gene was not included as it has different nucleotide (nt) and aa compositions relative to the other protein-coding genes. The two rRNAs and 22 transfer RNAs (tRNAs) were aligned using secondary structure models–the rRNA structures of the Gutell Lab (http://www.rna.icmb.utexas.edu/) and the standard 'cloverleaf' tRNA model (for example, see Kumazawa and Nishida, 1993), but were not analysed using phylogenetic methods.

2.2. Guidelines for consistent annotation of mitochondrial genomes

While aligning the 37 mt genes, we found a large number of discrepancies with regards to start and stop position annotations. Accordingly, we have checked all published bird and reptile mt genomes and standardized their annotations using the following guidelines. First, the start and stop positions of the tRNA genes are identified using both secondary structure and alignment against representatives of the other three main vertebrate groups (mammals, amphibians and fish). Small (1 nt) overlaps between some tRNAs are observed. This first step simultaneously defines where the rRNA genes and the control region begin and end, under the assumption that there are neither intergenic spacers nor overlaps between these regions and the flanking tRNAs. The start positions of protein-coding genes preceded by tRNAs are also limited by the tRNA end points (no overlaps).

In many taxa, some of the protein-coding genes have more than one potential start codon after the end of the flanking tRNA. In these cases, the most likely start codon is determined through alignment against other vertebrates and examination of the start points in those avian/reptile taxa for which there is only one viable start codon. The same procedure is used to identify start codons in protein-coding genes preceded by other protein-coding genes, while the first complete stop codon is accepted in the reciprocal situation (short overlaps are observed in both cases). With protein-coding genes followed by tRNAs, if there is a complete stop codon preceding the tRNA this is accepted. If not, there are two possibilities: an incomplete (T or TA) stop codon immediately adjacent to the start of the tRNA (completed by post-transcriptional polyadenylation; Ojala et al., 1981) or a full stop codon plus a slight overlap with the tRNA. In some genes, both of the latter are possibilities. In the absence of laboratory-based research to determine which alternative is functional, the end point is decided: (i) based on the stop codons of taxa for which there is no alternative, and (ii) on a status quo basis (i.e. following previous work). Having revised the annotations, the comparative patterns observed, including overlaps and intergenic spacers, were recorded in tables (see Section 3.2).

2.3. Phylogenetic datasets and methods

Five protein-coding supergene datasets were analysed using phylogenetic methods: one with the 19 avian taxa only; four with the 19 avian taxa plus different combinations of the outgroups. The bird-only dataset (19bird) was 3534 aa long. The bird plus outgroup datasets were: 19b + croc (3339 aa), birds plus the two crocodilians (alligator and caiman); 19b + liz (3347 aa), birds plus the two lizards (skink and iguana); 19b + turt (3400 aa), birds plus the two turtles (eastern painted and green); and 19b + 6rept (3119 aa), birds plus all six reptile outgroups. The avian-only dataset is longer than those including outgroup taxa because of the higher frequency of gaps and ambiguous sites in the latter. All of the datasets are available at http://awcmee.massey.ac.nz/downloads.htm.

Three types of maximum likelihood (ML) analysis were carried out on the aa datasets:

(i) ProtMLPlus 2.2 (Molphy; ftp://ftp.ism.ac.jp/pub/ISMLIB/MOLPHY/) with the mtREV-22 model of aa evolution. Bootstrap probabilities and estimates of the standard errors (S.E.) of the log-likelihood differences were calculated using the resampling estimated log-likelihood method (RELL; see Adachi and Hasegawa, 1996 and references therein).

(ii) Quartet puzzling (Tree-Puzzle 5.0; http://www.tree-puzzle.de/) with the mtREV-24 model (see discussion and references in Adachi and Hasegawa, 1996) and uniform rates. 1000 quartet puzzling (QP) steps were used to estimate support values for individual branches.

(iii) Local bootstrap probability (LBP)/ProtML 2.3b3

(Molphy) with the mtREV-24 model on both the best QP tree (not the ≥ 50% consensus one) and the best ProtMLPlus tree.

3. Results

3.1. Penguin and goose mitogenomic features

The complete mt genomes of the penguin (GenBank accession number: AF362763) and the goose (AF363031) are 17,611 and 16,737 nt long, respectively. However, both genomes vary in length due to heteroplasmy in the control region (see below). One extra nt is present in *NADH3* in both taxa (penguin: position 9715; goose: position 9724). This extra nt is found in many other birds and some turtles and is thought not to be translated (Mindell et al., 1998). There is some variation in gene order among avian mt genomes (see, for example, Bensch and Harlid, 2000). However, both the penguin and the goose have the standard gene order originally identified in the chicken.

The penguin has one of the longest avian mt genomes sequenced to date, exceeded only by those of the falcon (18,068 nt) and Oriental white stork (17,622 nt). The size of the penguin mt genome is largely due to its control region, which is 2040 nt long and which contains two sets of tandem repeats (positions 16,608–17,367 and 17,457–17,602). Five clones were examined and both sets were heteroplasmic for the number of repeat units. The first set consisted of a 79 nt sequence repeated nine, 12 or 13 times, followed by an incomplete copy (49 nt) of the repeat. The submitted sequence has nine repeats. All but one of these repeats are identical (the fifth repeat has one difference: a T/C transition at position 16,949). The second set of tandem repeats consisted of a varying number (13, 14, 18 or 20) of identical 7 nt repeat units, again followed by an incomplete repeat (6 nt). The sequence presented here has 20 repeats. Interestingly, the control region of another penguin (the Adelie; *Pygoscelis adeliae*) has also been found to contain two sets of repeats – an 81 nt sequence repeated five times and a microsatellite consisting of 30 copies of a 4 nt repeat (Ritchie and Lambert, 2000). The goose control region does not contain repeats, but is heteroplasmic for a string of Cs (positions 15,614–15,624). Four clones were sequenced and the number of Cs varied from 10 to 14. The submitted goose sequence has 11 Cs.

The penguin and goose protein-coding genes conform to those of other birds in terms of their length, start and stop codons (see Table 1) except for the penguin *NADH4* gene, which ends in an incomplete TA stop codon (rather than T as in the other neognaths). In addition, the penguin has one less aa than most of the other birds in both *ATPase8* and *NADH6* and one more in one region of *NADH5*. The penguin also has 8 nt between *tRNA-Met* and *NADH2*, compared to 1 nt in the falcon and no spacer in the other birds. The penguin and goose RNA genes have a number of small differences (base changes, insertions/deletions) compared to those of the other birds, most located in loop regions. Those that occur in stems are very minor: single nt insertions/deletions or changes and some changes from canonical (G:C or A:T) to non-canonical (G:T or A:C) base-pairing.

3.2. Revised annotation and features of avian and reptilian mitochondrial genomes

While aligning the bird and reptile mt sequences and tabulating features of their mt genomes (see below), it was found that there were many annotation inconsistencies regarding gene start and stop positions. For example, among birds alone there are 98 discrepancies involving five of the 13 protein-coding genes, 16 of the 22 tRNAs, both rRNAs and the control region where the start and end points have been recorded differently. This represents almost 6% of the total number of start and stop positions and 62% of the mt genes. Two-thirds of these discrepancies entail only a single nt difference, but 14% involve 2–3 nt and the rest are longer. Some of the latter involve obvious typographical errors where a tRNA has been extended by 1000 nt or where the start/stop positions of a gene have been interchanged in two co-submitted taxa. These discrepancies and revised annotations that provide consistency between avian (and reptile) taxa are given as Supplementary Information at http://awcmee.massey.ac.nz/downloads.htm.

Having made the annotations consistent, a large number of features for both avian and reptile mt genomes were summarized in table form. Tables for the subset of avian taxa analysed in this study are presented here. Table 1 summarizes the lengths, start and stop codons of avian mt protein-coding genes, while Table 2 gives the lengths of the mt control region, intergenic spacers and total genomes. Full tables for all available avian and reptile mt genomes, including additional ones recording the lengths of the RNA genes, are being maintained and updated at the above web site.

The lengths of the protein-coding genes have been given excluding stop codons, i.e. the stop codons have been included in the lengths of the intergenic spacers. There are several reasons for this. Firstly, the protein-coding gene lengths have been expressed in aa (see Table 1) and stop codons neither code for aa nor consist of a multiple of 3 nt in all cases (incomplete stop codons). Secondly, by treating the stop codons as part of the intergenic spacers, overlaps between coding sequences can be clearly identified in Table 2. Thirdly, by calculating the lengths of the features in this way, the numbers entered in the tables for each taxon can easily be double-checked through summing them (i.e. sum of protein-coding gene lengths x three + sum of RNA lengths + sum of control region and intergenic spacer lengths − sum of overlaps = total length of the genome). A further 1 nt must be added for those taxa which have the extra nt in *NADH3* (see Section 3.1).

Table 1
Length (in amino acids), start and stop codons of avian mitochondrial protein-coding genes (consistently annotated[a])

Genes[c]:	Taxa[b]:	Kiwi	Emu	Cassowary	Ostrich	Eastern Moa	Giant Moa	Greater Rhea	Lesser Rhea	Tinamou	Crested-Tinamou	Chicken	Duck	Goose	Penguin	Stork	Falcon	Broadbill	Indigobird	Rook
NADH1	Length	324	323	323	324	323	323	324	324	322	322	324	325	325	325	325	324	325	325	325
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	AGG	TAA	TAA	AGG	TAA	TAA	AGG	AGG	AGA	AGA	TAA	AGG	AGG	AGG	AGG	TAA	TAA	AGA	AGG
NADH2	Length	346	346	346	346	346	346	346	346	346	346	346	346	346	346	346	346	346	346	346
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATA (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAA	TAA	TAA	TAA	TAA
COI	Length	516	516	516	516	516	516	516	516	516	516	515	516	516	516	516	516	516	516	516
	Start codon	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	ATG (Met)	ATG (Met)	GTG (Val)
	Stop codon	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
COII	Length	229	229	229	229	228	228	229	229	228	228	227	228	228	227	227	227	227	227	227
	Start codon	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	ATG (Met)	GTG (Val)	GTG (Val)	ATG (Met)	GTG (Val)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	T–	T–	T–	T–	T–	T–	T–	T–	T–	T–	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
ATPase8	Length	55	55	55	55	55	55	55	55	55	55	54	55	55	54	55	55	55	55	55
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
ATPase6	Length	227	227	227	227	227	227	227	227	227	227	227	227	227	227	227	227	227	227	227
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
COIII	Length	261	261	261	261	261	261	261	261	261	261	261	261	261	261	261	261	261	261	261
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	T–	T–	T–	T–	T–	T–	T–	T–	T–	T–	T–	T–	T–	T–	T–	T–	T–	T–	T–
NADH3	Length	116	116	116	116	116	116	116	116	116	116	116	116	116	116	116	116	116	116	116
	Start codon	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATC (Ile)	ATG (Met)	ATA (Met)	ATC (Ile)	ATT (Ile)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAG	TAA	TAA	TAA	TAA	TAA	TAG	TAA	TAA	TAA	TAG	TAA	TAA
NADH4L	Length	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
NADH4	Length	455	458	458	458	458	458	458	458	458	458	459	459	459	459	459	459	459	459	459
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	AGA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	T–	T–	T–	TA-	T–	T–	T–	T–	T–
NADH5	Length	604	605	605	605	605	605	605	605	603	606	605	607	605	606	603	605	601	605	606
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	GTG (Val)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATA (Met)	ATG (Met)	ATG (Met)	GTG (Val)	GTG (Val)	GTG (Val)	ATG (Met)	ATG (Met)	ATA (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	AGA	AGA	AGA	AGA	AGA	TAA	TAA	TAA	TAA	TAA	TAA	AGA	TAA	AGG	AGA	TAA	AGA	AGA
Cyt*b*	Length	379	379	379	379	380	380	379	379	379	379	380	380	380	380	380	380	380	380	380
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAG	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
NADH6 *(L)*	Length	173	173	173	173	173	173	174	174	173	unknown	173	173	173	172	173	173	173	172	173
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	unknown	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	unknown	TAA	TAG	TAG	TAG	TAG	TAG	AGG	TAG	TAA

[a] See sections 2.2 and 3.2.

[b] See sections 2.1 and 3.1 for scientific names and accession numbers.

[c] Gene length: Does not include stop codons (see section 3.2) or the extra nucleotide often found in NADH3 (see section 3.1). T-- and TA-: Incomplete stop codons. ***(L)***: Encoded by the mitochondrial light strand. **Unknown**: The mt genome of this taxon is incomplete and does not include this region. Light gray shading: This feature (length, start or stop codon) is conserved. Dark gray shading: This feature is conserved in most of the taxa; those that differ are shaded. A color version of this table, with additional patterns and further taxa, is being maintained and updated at http://awcmee.massey.ac.nz/downloads.htm.

Table 2
Length (in nucleotides) of avian mitochondrial control regions, intergenic spacers and complete genomes (consistently annotated[a])

Taxa[b]: / Region[c]:	Kiwi	Emu	Cassowary	Ostrich	Eastern Moa	Giant Moa	Greater Rhea	Lesser Rhea	Tinamou	Crested-Tinamou	Chicken	Duck	Goose	Penguin	Stork	Falcon	Broadbill	Indigobird	Rook
Control Region (CR)	1187	1093	>1138	1031	1501	1508	1171	1202	1102	>383	1227	1066	1174	2040	2053	1510	1453	1295	1339
Additional Non-Coding Region	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	unknown	N/A	N/A	N/A	N/A	N/A	950	301	N/A	N/A
tRNA-Phe/12S rRNA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12S rRNA/tRNA-Val	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
tRNA-Val/16S rRNA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16S rRNA/tRNA-Leu (UUR)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
tRNA-Leu (UUR)/NADH1	9	8	10	9	11	11	11	10	9	9	9	4	6	5	17	15	10	8	9
NADH1/tRNA-Ile	1	3	3	1	4	4	11	11	19	21	3	1	1	1	1	19	24	9	11
tRNA-Ile/tRNA-Gln *(L)*	9	6	5	11	7	8	14	13	7	7	5	8	7	9	8	9	11	5	5
tRNA-Gln *(L)* /tRNA-Met	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	12	1 overlap	1 overlap
tRNA-Met/NADH2	-	-	-	-	-	-	-	-	-	-	-	-	-	8	-	1	-	-	-
NADH2/tRNA-Trp	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	4	2	5
tRNA-Trp/tRNA-Ala *(L)*	1	1	1	-	1	1	1	1	1	1	6	3	5	1	1	10	1	1	1
tRNA-Ala *(L)* /tRNA-Asn *(L)*	1	3	3	4	1	1	2	2	1	-	3	2	1	2	2	10	6	9	9
tRNA-Asn *(L)* /tRNA-Cys *(L)*	2	3	3	1	2	2	3	3	2	2	1	-	3	2	2	2	6	-	2
tRNA-Cys *(L)* /tRNA-Tyr *(L)*	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
tRNA-Tyr *(L)* /COI	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
COI/tRNA-Ser (UCN) *(L)*	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap
tRNA-Ser (UCN) *(L)* /tRNA-Asp	2	2	2	3	2	2	2	2	2	2	2	2	2	4	2	2	2	5	5
tRNA-Asp/COII	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	6	3	10	8
COII/tRNA-Lys	1	1	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4	14	4
tRNA-Lys/ATPase8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ATPase8/ATPase6	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap
ATPase6/COIII	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	12	8	10
COIII/tRNA-Gly	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
tRNA-Gly/NADH3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NADH3/tRNA-Arg	4	4	4	4	4	4	5	5	3	4	4	4	4	7	7	4	7	4	4
tRNA-Arg/NADH4L	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1
NADH4L/NADH4	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap
NADH4/tRNA-His	10	4	4	4	4	4	4	4	5	5	1	1	1	2	1	1	1	1	1
tRNA-His/tRNA-Ser (AGY)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
tRNA-Ser (AGY)/tRNA-Leu (CUN)	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	-	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
tRNA-Leu (CUN)/NADH5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NADH5/Cyt*b*	15	12	12	12	13	13	15	15	2	2	7	2	10	10	13	8	12	11	13
Cyt*b* /tRNA-Thr	4	7	7	5	4	4	8	8	4	4	6	5	5	6	4	5	7	6	6
tRNA-Thr/tRNA-Pro *(L)*	83	82	82	55	26	27	9	9	85	5	-	10	8	9	11	N/A	N/A	12	11
tRNA-Thr/CR	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	-	-	N/A	N/A
CR/tRNA-Pro *(L)* *	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	-	N/A	N/A	N/A	N/A	N/A	-	-	N/A	N/A
tRNA-Pro *(L)* /NADH6 *(L)*	30	23	24	13	14	14	18	17	12	N/A	9	13	13	16	13	18	8	29	12
NADH6 *(L)* /tRNA-Glu *(L)*	3	3	3	3	-	-	3	3	-	unknown	2	-	-	3	3	3	-	1	1
tRNA-Glu *(L)* /CR#	-	-	-	-	-	-	-	-	-	unknown	-	-	-	-	-	-	-	-	-
CR/tRNA-Phe#	-	-	unknown	-	-	-	-	-	-	unknown	-	-	-	-	-	-	-	-	-
Total Length of Genome	16816	16711	>16757	16591	17061	17070	16714	16747	16703	>15302	16775	16616	16737	17611	17622	18068	17344	16895	16931

[a,b] See corresponding footnotes to Table 1.

[c] Intergenic spacer lengths include the stop codons of preceding protein-coding genes (see section 3.2). >: Incomplete region/genome, full length unknown. **N/A**: Not applicable (i.e. this taxon does not have this region/gene arrangement). -: Zero (no spacer). Overlap: These two genes overlap by the specified number of nucleotides (stop codons are not included in the length of the overlap). **CR/tRNA-Pro^{+}**: *tRNA-Pro*/CR in the crested-tinamou. **tRNA-Glu/CR$^{\#}$** and **CR/tRNA-Phe$^{\#}$**: *tRNA-Glu*/Additional Non-Coding Region and Additional Non-Coding Region/*tRNA-Phe* in the falcon and broadbill. (***L***); **unknown**; light and dark gray shading: See Table 1 (footnote c). The apparent conservation of intergenic spacer patterns involving the rRNAs, control region and additional non-coding region is the result of how the boundaries of these regions are determined (see section 2.2). As with Table 1, an extended version of this table is being maintained at http://awcmee.massey.ac.nz/downloads.htm.

A number of conserved patterns and interesting variations have been identified using the tables (see Tables 1 and 2 and full tables at http://awcmee.massey.ac.nz/downloads.htm). Some of the more intriguing of these involve short overlaps between certain genes. These overlaps fall into at least four classes in birds. The first class comprises 'overlaps' between H and L strand-encoded products (1 nt between *tRNA-Gln*/*tRNA-Met* and 6 nt between *COI*/*tRNA-Ser [UCN]*). As different RNA transcripts are involved, these do not comprise genuine overlaps (although it would be interesting to know how different stop points within the same small area are recognized in the complementary transcripts). The second class consists of overlaps involving TAR stop codons: a 1–2 nt overlap between a TAR stop codon in *NADH2* and the start of *tRNA-Trp*, a 1 nt overlap between *ATPase6* (TAA stop codon) and *COIII* and a 1 nt overlap between *NADH5* (TAA stop codon) and *Cytb*. It seems very likely that these are not, in fact, true overlaps but instead represent endonucleolytic cleavage sites producing incomplete stop codons (see Ojala et al., 1981). The third class entails overlaps between the coding sequences of protein-coding genes (7 nt between *ATPase8*/*ATPase6* and 4 nt between *NADH4L*/*NADH4* – excluding stop codons – in all birds). While such overlaps are always associated with different reading frames, more information is required regarding the generation and processing of mt protein-coding transcripts. Finally, the fourth class includes the remaining overlaps and involves an unknown mechanism: 1 nt overlaps between *tRNA-Cys*/*tRNA-Tyr* and *tRNA-Ser* (*AGY*)/*tRNA-Leu* (*CUN*) and a 2 nt overlap between an AGG stop codon in *NADH1* and the start of *tRNA-Ile*.

3.3. Avian phylogenetic relationships

Optimal phylogenetic trees were generated using QP ML, ProtMLPlus and LBP/ProtML for each of the five datasets (avian-only and avian plus four different outgroup combinations). These trees plus the near-optimal ProtMLPlus trees (those within one S.E. of the best tree) are considered here. Only seven unrooted (avian-only) trees were identified as optimal or near-optimal using the different methods/models. Considering that there are approximately 6×10^{18} possible unrooted trees for 19 taxa, this indicates that the avian-only tree was fairly well resolved. Moreover, these seven trees were very similar, all of the differences between them involving local rearrangements (the movement of taxa by one step, or one nearest neighbor interchange, on the tree).

The unrooted QP ML tree for the 19 avian genomes is shown in Fig. 1. Three branches involving our new genomes were well supported in all trees, both unrooted (see Fig. 1) and rooted (see Section 3.4). The first was a penguin/stork grouping (support values: 99 in the unrooted QP tree, 95 in the unrooted LBP/ProtML tree), which is in agreement with Sibley and Ahlquist (1990) for our taxon sample. The second was the pairing of the goose with the duck (support values: 98 and 100, respectively), while the third was the joining of Galliformes (chicken) and Anseriformes (goose and duck) on a common branch to form Gallianseres (support values: 90 and 99). In addition, Gallianseres and the paleognaths (ratites and tinamous) were sister-groups in all of the unrooted trees, albeit with only moderate support in the QP tree (support values: 62 and 100).

The position of the falcon is shown as unresolved in Fig. 1. This taxon was found in two equally well-supported positions in the QP ML analysis: in a sister-group relationship with the penguin/stork (as in Sibley and Ahlquist, 1990) or lying between the penguin/stork and the passerines (broadbill, indigobird and rook). The same two placements were identified in the ProtMLPlus analyses, while the latter grouping was weakly favored in the LBP/ProtML tree (support value: 58). In an earlier mitogenomic study (Mindell et al., 1999), the falcon and broadbill joined on a common branch in some trees. However, using the current dataset with additional avian taxa, there was no support for a (falcon,broadbill) grouping. As a result, the passerines formed a monophyletic group (*sensu lato*) in the unrooted trees (support values: 59 QP and 96 LBP/ProtML).

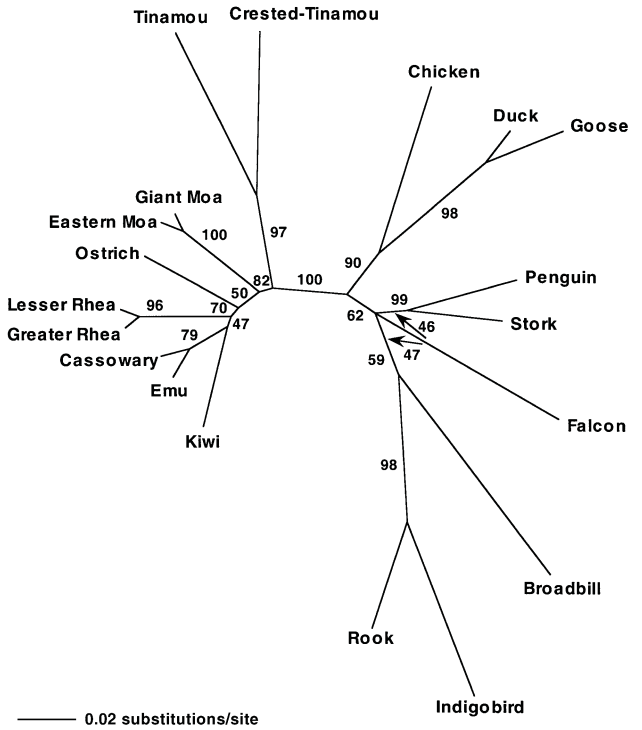


Fig. 1. Unrooted quartet puzzling (QP) maximum likelihood tree for 19 avian taxa. The dataset consisted of the 12 protein-coding genes from the mitochondrial heavy strand and was 3534 amino acids (aa) long. Gaps and adjacent ambiguous sites were excluded. The analysis was carried out using the mtREV-24 model of evolution and uniform rates. All taxa passed a χ^2 test for aa sequence composition at the 5% level. The tree has a log-likelihood value of − 33731.44 and is drawn to scale. QP support values are shown on each branch (% of 1000 puzzling steps). The position of the falcon is shown as unresolved (falcon + passerines = 47%, falcon + penguin/stork = 46%).

The only other variations in the unrooted tree involved paleognath taxa. The paleognaths as a whole grouped together with strong support (100 in both the QP and LBP/ProtML trees). However, the positions of the kiwi, the moas and the ostrich were not completely resolved. The kiwi was found in two places: on the (emu,cassowary) branch (as in Fig. 1) or one step removed to give an ((emu,cassowary),(rheas)) grouping in the LBP/ProtML and some of the ProtMLPlus trees. The moas formed one of two sister-group relationships: with the other ratites in the QP tree and with the tinamous in the LBP/ProtML and ProtMLPlus trees. The ostrich and ((emu,cassowary),kiwi,(rheas)) grouped together in most of the trees, but the ostrich moved one step to form the sister-group of the other paleognaths in one near-optimal ProtMLPlus tree. Further investigation into the paleognath portion of the tree is outside the scope of this study. The issue of paleognath interrelationships will be readdressed once an additional kiwi sequence becomes available (Harrison et al., in preparation).

3.4. Avian and reptilian phylogenetic relationships

Fig. 2 shows the rooted QP ML tree for the 19 avian genomes plus all six reptile taxa. Rooting the tree using the four different outgroup combinations (crocodilians; lizards; turtles; all six reptiles combined) did not affect the following groupings, although the support values were sometimes lower: penguin/stork (support values: 72–82 QP and 87–99 LBP/ProtML), goose/duck (support: 71–99 and 100), Gallianseres (support: 82–96 and 97–98) and paleognaths (support: 78–93 and 100). Comparing the avian-only and rooted QP trees, the main difference is that the falcon stabilized in the rooted trees, grouping with the (penguin, stork). The other changes involved a one-step interchange between the kiwi and the rheas in the 19b + croc QP tree and a lack of resolution regarding the position of the ostrich in the 19b + liz QP tree (45% support for the arrangement shown in Fig. 1; 47% support for an interchange between the ostrich and the moas). The unrooted and rooted LBP/ProtML trees were even more similar, the only difference being that the ostrich and ((moas),(tinamous)) exchanged places in the 19b + liz, 19b + turt and 19b + 6rept trees.

The positions of the falcon, the kiwi, the moas and the ostrich remained unstable in the rooted trees (placements as in the unrooted trees plus two additional one-step interchanges involving the ostrich: one with the kiwi in the optimal 19b + 6rept ProtMLPlus tree and one with the moas in some of the near-optimal 19b + 6rept ProtMLPlus trees and 47% of the time in the 19b + liz QP tree). Adding outgroups to the bird tree also resulted in some variation in the position of Gallianseres. While this taxon continued to be found in a sister-group relationship with the paleognaths in most trees, two alternative positions (both involving a one-step move) were observed among the ProtMLPlus trees. In the first (occurring in one optimal and a number of near-optimal trees), Gallianseres and the ((penguin,stork),falcon) exchanged places. In the second case (one near-optimal tree only), Gallianseres and the ((penguin,stork),falcon) joined on a common branch.

The three best supported positions of the root were passerine: (i) on the suboscine passerine (broadbill) lineage, (ii) with Passeriformes as the sister group of the other birds or (iii) at the base of the oscine passerine (indigobird and rook) branch. The root joined the broadbill branch in all of the 19b + croc and 19b + liz trees (respective support values: 52 and 50 QP; 83 and 78 LBP/ProtML), in the 19b + turt and 19b + 6rept LBP/ProtML trees (respective support values: 39 and 55) and in some of the 19b + turt and 19b + 6rept ProtMLPlus trees. However, a basal split between Passeriformes and (all other birds) was reconstructed in the 19b + turt and 19b + 6rept QP analyses (respective support values: 55 and 39) and in some of the near-optimal 19b + turt and 19b + 6rept ProtMLPlus trees. The third rooting position, at the base of the oscine passerine

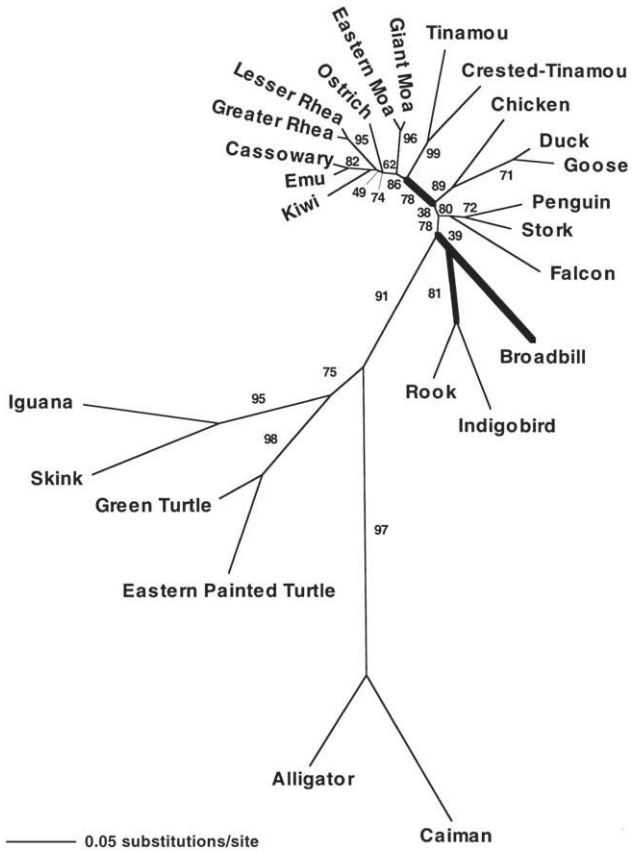


Fig. 2. QP maximum likelihood tree from Fig. 1 rooted using six reptile outgroups. The dataset consisted of the 12 protein-coding genes from the mitochondrial heavy strand and was 3119 amino acids (aa) long. Gaps and adjacent ambiguous sites were excluded. QP analysis was carried out using the mtREV-24 model of evolution and uniform rates. All taxa passed a χ^2 test for aa sequence composition at the 5% level. The tree has a log-likelihood value of −41219.12 and is drawn to scale. QP support values are shown on each branch (% of 1000 puzzling steps). Alternative rooting points (branches) identified in the analyses are shown in bold.

(indigobird and rook) branch, was found in many of the near-optimal 19b + 6rept ProtMLPlus trees. Furthermore, the root fell in the traditional position between paleognaths and neognaths in two of the near-optimal 19b + turt ProtMLPlus trees.

To focus more specifically on the relationships between the ingroup and outgroup taxa, crocodilians were the sister-group of birds in all but one of the 19b + 6rept trees (one near-optimal ProtMLPlus tree had turtles closest to birds). This concurs with the traditional view on bird-reptile relationships (see Janke et al., 2001, and references therein for further discussion on reptile phylogeny and evolution). Conversely, crocodilians were much further from birds in terms of genetic distances than were turtles and lizards (see branch lengths in Fig. 2). The average distance for the 19b + croc dataset as a whole (i.e. the average of all bird–bird, bird–crocodilian and crocodilian–crocodilian pairwise distances) was 0.20160, while the range of bird-crocodilian pairwise distances was 0.44556–0.51872. Turtles were closest to birds on distances (average distance for the 19b + turt dataset: 0.17413; range of bird–turtle pairwise distances: 0.28811–0.36349). Without the inclusion of the other reptiles, however, the two turtles failed a χ^2 test for aa sequence composition at the 5% level. Lizards were also relatively close to birds (average distance for the 19b + liz dataset: 0.17584; range of bird–lizard pairwise distances: 0.32832–0.36932). These results are consistent with those of earlier studies (see, for example, Janke et al., 2001) and strongly suggest that the more slowly evolving outgroup taxa should be used instead of or in combination with the crocodilians because of potential long branch attraction problems.

4. Discussion

4.1. Mitogenomic annotation and features

There has been a marked lack of consistency in recording the start and stop positions of avian and reptile mt genes/regions. All available bird and reptile mt genomes have now been checked, discrepancies eliminated and the revised annotations recorded. In some protein-coding genes, there is more than one possible start/stop codon. Those presented here are now consistent (see Sections 2.2 and 3.2 and Supplementary Information at http://awcmee.massey.ac.nz/downloads.htm) and this is sufficient for the purposes of this study. However, it would be desirable to have confirmation from laboratory research of which alternative is the functional one in order to learn more about the actual processes involved. With uniform genome annotation, a number of conserved patterns and interesting differences can be distinguished (see Tables 1 and 2 and full tables at the above web site). This should help in promoting accuracy of sequencing, alignment and analysis and in laying the framework for eventual automation of mt genome annotation and gene extraction/alignment.

4.2. Phylogenetic relationships

Both the grouping of the two anseriforms (goose and duck) and of the penguin and the stork were well supported in all of the trees. Furthermore, adding the goose to the mt dataset and analyses certainly favored the Gallianseres hypothesis (Galliformes and Anseriformes joined on a common branch with good to strong support in all trees). Additional gallianserine mt genomes are desirable however, particularly from deep-branch taxa. In order to test alternative hypotheses for anseriform relationships, other avian mt genomes are also required – for example, shorebirds (see discussion and references in Feduccia, 1995) and flamingos (Ericson, 1997). We are proceeding with sequencing the mt genomes of appropriate taxa both to investigate these questions further and in order that the penguin sequence can be fully utilized as an avian calibration point.

The avian tree as a whole was fairly well resolved, although several one-step changes involving the positions of the falcon and of certain paleognath taxa were possible (the falcon was not, however, found with the broadbill in any of the trees). Moreover, while Gallianseres was the closest neognath lineage to the paleognaths in the unrooted trees (moderate to strong support), this relationship was occasionally disrupted in the rooted trees and requires further investigation. Although not the primary focus of this paper, it is interesting to note that with the addition of further avian and reptile sequences to the mt dataset, the position of the root of the bird tree is not as clear-cut as in some earlier mitogenomic analyses. This may be due to improved taxon sampling (see Van Tuinen et al., 2000) and a decrease in the effects of long-branch attraction (Hendy and Penny, 1989). This variation of the rooting point does not yet allow us to conclusively define the primary divergence in the avian mt tree. In depth phylogenetic analyses with additional bird mt genomes, once they become available, should aid in clarifying issues in avian evolution.

Acknowledgements

We thank Alan Tennyson and Sandy Bartle (Museum of New Zealand) for supplying the penguin sample and Anette Gullberg (University of Lund) for obtaining the goose tissue and for technical help. This research was supported by a Swedish Institute Guest Scholarship, a Massey University Doctoral Scholarship, a University of Lund (Department of Genetics) Summer Stipendium, a Sven and Lilly Lawski Stipendium, the New Zealand Marsden Fund, the Swedish Science Council and the Nilsson–Ehle Foundation. We would also like to thank two anonymous reviewers for their valuable comments on an earlier draft of the manuscript.

References

Adachi, J., Hasegawa, M., 1996. MOLPHY version 2.3: Programs for molecular phylogenetics based on maximum likelihood. Comput. Sci. Monogr. (Inst. Stat. Math. Tokyo) 28, 1–150.

Arnason, U., Gullberg, A., Widegren, B., 1991. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. J. Mol. Evol. 33, 556–568.

Arnason, U., Adegoke, J.A., Bodin, K., Born, E.W., Esa, Y.B., Gullberg, A., Nilsson, M., Short, R.V., Xu, X., Janke, A., 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc. Natl. Acad. Sci. USA 99, 8151–8156.

Bensch, S., Harlid, A., 2000. Mitochondrial genomic rearrangements in songbirds. Mol. Biol. Evol. 17, 107–113.

Caspers, G.-J., de Weerd, D.U., Wattel, J., de Jong, W.W., 1997. α-Crystallin sequences support a galliform/anseriform clade. Mol. Phylogenet. Evol. 7, 185–188.

Cooper, A., Penny, D., 1997. Mass survival of birds across the Cretaceous–Tertiary boundary: molecular evidence. Science 275, 1109–1113.

Cracraft, J., 2001. Avian evolution, Gondwana biogeography and the Cretaceous–Tertiary mass extinction event. Proc. R. Soc. Lond. B Biol. Sci. 268, 459–469.

Elzanowski, A., Brett-Surman, M.K., 1995. Avian premaxilla and tarsometatarsus from the Uppermost Cretaceous of Montana. Auk 112, 762–767.

Ericson, P.G.P., 1997. Systematic relationships of the paleogene family Presbyornithidae (Aves: Anseriformes). Zool. J. Linn. Soc. 121, 429–483.

Feduccia, A., 1995. Explosive evolution in Tertiary birds and mammals. Science 267, 637–638.

Fordyce, R.E., Jones, C.M., 1990. Penguin history and new fossil material from New Zealand. In: Davis, L.S., Darby, J.T. (Eds.), Penguin Biology, III. Academic Press, San Diego, CA, pp. 419–446.

García-Moreno, J., Mindell, D.P., 2000. Rooting a phylogeny with homologous genes on opposite sex chromosomes (gametologs): a case study using avian CHD. Mol. Biol. Evol. 17, 1826–1832.

Haddrath, O., Baker, A.J., 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. Proc. R. Soc. Lond. B Biol. Sci. 268, 939–945.

Harlid, A., Arnason, U., 1999. Analyses of mitochondrial DNA nest ratite birds within the Neognathae: supporting a neotenous origin of ratite morphological characters. Proc. R. Soc. Lond. B Biol. Sci. 266, 305–309.

Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297–309.

Janke, A., Erpenbeck, D., Nilsson, M., Arnason, U., 2001. The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. Proc. R. Soc. Lond. B Biol. Sci. 268, 623–631.

Jones, C.M., Mannering, A.A., 1997. New Paleocene fossil bird material from the Waipara Greensand, North Canterbury, New Zealand, Miscellaneous Publication, 95A, 88. Geological Society of New Zealand.

Kirchner, J.W., 2002. Evolutionary speed limits inferred from the fossil record. Nature 415, 65–68.

Kumazawa, Y., Nishida, M., 1993. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J. Mol. Evol. 37, 380–398.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., 1998. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. Mol. Biol. Evol. 15, 1568–1571.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., Hasegawa, M., Ast, J.C., Yuri, T., 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst. Biol. 48, 138–152.

Noriega, J.I., Tambussi, C.P., 1995. A Late Cretaceous Presbyornithidae (Aves: Anseriformes) from Vega Island, Antarctic Peninsula: paleobiogeographic implications. Ameghiniana 32, 57–61.

Ojala, D., Montoya, J., Attardi, G., 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature 290, 470–474.

Olson, S.L., 1992. *Neogaeornis wetzeli* Lambrecht, a Cretaceous loon from Chile (Aves: Gaviidae). J. Vertebr. Paleontol. 12, 122–124.

Ritchie, P.A., Lambert, D.M., 2000. A repeat complex in the mitochondrial control region of Adelie penguins from Antarctica. Genome 43, 613–618.

Sheldon, F.H., Bledsoe, A.H., 1993. Avian molecular systematics, 1970s to 1990s. Annu. Rev. Ecol. Syst. 24, 243–278.

Sibley, C.G., Ahlquist, J.E., 1990. Phylogeny and Classification of Birds: A Study in Molecular Evolution, Yale University Press, New Haven, CT.

Stidham, T.A., 1998. A lower jaw from a Cretaceous parrot. Nature 396, 29–30.

Van Tuinen, M., Hedges, S.B., 2001. Calibration of avian molecular clocks. Mol. Biol. Evol. 18, 206–213.

Van Tuinen, M., Sibley, C.G., Hedges, S.B., 2000. The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. Mol. Biol. Evol. 17, 451–457.

Chapter 3.

Slack, K.E., Jones, C.M., Ando, T., Harrison, G.L., Fordyce, R.E., Arnason, U. and Penny, D. (2006). Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Mol. Biol. Evol.* **23**: 1144-1155. (As at August 2011 this paper has been cited 72 times in the Web of Science - including 2 citations with errors in pagination.)

Having already sequenced the complete mitochondrial (mt) genome of a penguin (see previous chapter), for this manuscript I focused on a small albatross (black-browed mollymawk, *Diomedea melanophris*), the Kerguelen petrel (*Pterodroma brevirostris*) and the red-throated loon (*Gavia stellata).* The choice of these three taxa was because they were potentially some of the closest relatives of penguins and thus should aid in tightening the upper bound when using the penguin fossils (see below) as calibration points for dating avian divergences. To clarify, the closer the fossil penguins are to the divergence point between penguin and non-penguin lineages, the more accurate the fossils are as calibration points.

The reason for concentrating on the penguins, and their relatives, was that an older fossil penguin had been found in North Canterbury, but the details (including the dating) had not been published. Because of the importance of this fossil for the dating analysis, Professor Ewan Fordyce and Craig Jones were approached about a joint publication.

Craig Jones (currently of IGNS, Gracefield, Lower Hutt), Tatsuro Ando and Professor Ewan Fordyce (of the Geology Department at the University of Otago) fully analysed and described the penguin fossils (*Waimanu*), as well as the microfossils that were used for dating the fossil penguins. I then used their results as a calibration point for early avian evolution. Some of the early interactions were at IGNS in Gracefield (Lower Hutt), but day-long meetings at Te Papa in Wellington were held twice, with Professor Fordyce coming from Dunedin, Craig Jones from Gracefield, and David Penny and myself from Palmerston North.

I was responsible for the sequencing and analysis of the mitochondrial genomes, including the dating analysis. Again I updated the tables (referred to in the previous chapter) of the properties of avian mt genomes, and placed the updated tables on the Web. Abby Harrison aided with technical and laboratory advice and support. Ulfur Arnason and David Penny were involved at all stages in project design. All authors contributed to the final manuscript.

Early Penguin Fossils, Plus Mitochondrial Genomes, Calibrate Avian Evolution

Kerryn E. Slack,†[1] Craig M. Jones,‡[1] Tatsuro Ando,§[1] G. L.(Abby) Harrison,* R. Ewan Fordyce,§ Ulfur Arnason,† and David Penny**

*Allan Wilson Center for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand; †Division of Evolutionary Molecular Systematics, Department of Cell and Organism Biology, University of Lund, Solvegatan 2, S-223 62 Lund, Sweden; ‡Institute of Geological and Nuclear Sciences, Lower Hutt, New Zealand; and §Department of Geology, University of Otago, Dunedin, New Zealand

Testing models of macroevolution, and especially the sufficiency of microevolutionary processes, requires good collaboration between molecular biologists and paleontologists. We report such a test for events around the Late Cretaceous by describing the earliest penguin fossils, analyzing complete mitochondrial genomes from an albatross, a petrel, and a loon, and describe the gradual decline of pterosaurs at the same time modern birds radiate. The penguin fossils comprise four naturally associated skeletons from the New Zealand Waipara Greensand, a Paleocene (early Tertiary) formation just above a well-known Cretaceous/Tertiary boundary site. The fossils, in a new genus (*Waimanu*), provide a lower estimate of 61–62 Ma for the divergence between penguins and other birds and thus establish a reliable calibration point for avian evolution. Combining fossil calibration points, DNA sequences, maximum likelihood, and Bayesian analysis, the penguin calibrations imply a radiation of modern (crown group) birds in the Late Cretaceous. This includes a conservative estimate that modern sea and shorebird lineages diverged at least by the Late Cretaceous about 74 ± 3 Ma (Campanian). It is clear that modern birds from at least the latest Cretaceous lived at the same time as archaic birds including *Hesperornis*, *Ichthyornis*, and the diverse Enantiornithiformes. Pterosaurs, which also coexisted with early crown birds, show notable changes through the Late Cretaceous. There was a decrease in taxonomic diversity, and small- to medium-sized species disappeared well before the end of the Cretaceous. A simple reading of the fossil record might suggest competitive interactions with birds, but much more needs to be understood about pterosaur life histories. Additional fossils and molecular data are still required to help understand the role of biotic interactions in the evolution of Late Cretaceous birds and thus to test that the mechanisms of microevolution are sufficient to explain macroevolution.

Introduction

The question whether microevolutionary processes (that can be studied in the present) are sufficient to account for all of macroevolution is still debated (Simons 2002) and, in particular, whether mammals and birds would have replaced dinosaurs and mammals even without the extraterrestrial impact that marks the K/T (Cretaceous/Tertiary) boundary (see Penny and Phillips 2004). Patterns of diversity and processes of ecological partitioning in Late Cretaceous birds and flying reptiles are incompletely known. It is clear that early crown birds (Neornithes) overlapped with archaic birds (such as the enantiornithes and the diving birds *Hesperornis* and *Ichthyornis*) as well as with pterosaurs. Furthermore, none of the latter groups—except for the crown birds—survived into the Tertiary. There was a decline to extinction of both pterosaurs and archaic birds such as *Hesperornis* and *Ichthyornis*, but did this precede or follow the rise of modern (crown group) birds (Neornithes)? In order to test models (Penny and Phillips 2004) about extinctions around the end of the Cretaceous we need to understand important evolutionary events *before* the asteroid impact that marks the Cretaceous/Tertiary boundary (Cooper and Penny 1997; Cracraft 2001; Van Tuinen and Hedges 2001; Wilf et al. 2003; Peters 2005). Hypotheses for the origin of modern (crown group) birds vary considerably (Penny and Phillips 2004). Did just one lineage survive the K/T boundary extinctions (Feduccia 2003) and then diversify rapidly, or did many modern lineages diversify from early in the Late Cretaceous, 80–100 Ma (Hedges et al. 1996; Cooper and Penny 1997; Härlid et al. 1997; Cracraft 2001; Paton et al. 2002; Dyke and van Tuinen 2004; Harrison et al. 2004)? To test these hypotheses we require additional well-defined calibration points, and penguins, with their excellent fossil record (Simpson 1975; Fordyce and Jones 1990; Myrcha et al. 2002; Clarke et al. 2003) are important. These large, aquatic, wing-propelled diving birds have solid (nonpneumatic) bones that preserve well, providing a sound fossil record of their lineage and dating their minimum divergence time from related flying birds.

Four associated fossil skeletons representing two new species of Early Paleocene penguins were found near Waipara (North Canterbury, New Zealand) (Fordyce and Jones 1990; Jones and Mannering 1997). They are from the Waipara Greensand, a well-dated unit from a thoroughly described sequence (Vajda et al. 2001; Hollis and Strong 2003) that includes an important K/T site. Indeed it is near one of the three classic K/T sites where the iridium anomaly was first reported (Alvarez et al. 1980). The older of the two new species is the oldest fossil penguin found thus far, and is therefore important for both understanding the origin and evolution of penguins (Baker et al. 2006), as well as providing an improved calibration point for estimating the early divergence times of modern sea and shorebirds. The late Early Paleocene to Late Paleocene age of these new fossils is important because it mitigates against any possible effects from any

[1]K.E.S., C.M.J., and T.A. contributed equally to this work.

E-mail: d.penny@massey.ac.nz.

Mol. Biol. Evol. 23(6):1144–1155. 2006
doi:10.1093/molbev/msj124
Advance Access publication March 13, 2006

(hypothetical) coordinated acceleration of mutation rates of DNA sequences that might affect molecular estimates of divergence times (Conway Morris 1998; Foote et al. 1999; Bromham and Hendy 2000).

Given these excellent early penguin fossils, it is important to establish an evolutionary tree for penguins and their close relatives. We have reported a complete mitochondrial genome from the little blue penguin (*Eudyptula minor*) (Slack et al. 2003) but that study had a restricted number of taxa. Unexpectedly perhaps, though not inconsistent (on the present data set) with some earlier work (Sibley and Ahlquist 1990), storks were the sister taxon of penguins. Previously (Cooper and Penny 1997) we found that procellariforms (such as albatrosses, petrels, and shearwaters) were closer to penguins, but they were not represented at that time by complete mitochondrial genomes. We therefore sequenced the mitochondrial genomes of a small albatross (black-browed mollymawk, *Diomedea melanophris*), Kerguelen petrel (*Pterodroma brevirostris*), and red-throated loon (*Gavia stellata*). Some group other than procellariforms could still be the sister group to penguins (Cracraft 2001), but that does not affect the *minimum* age for the divergence of penguins from related lineages—our estimate is conservative in this respect. By combining the fossil and molecular data we report an excellent calibration point from penguin fossils, and use it to study questions regarding the early evolution of modern birds. To help understand some of the biotic interactions in the Late Cretaceous, we have considered the diversity and lifestyles of birds and of pterosaurs through the Cretaceous. The basic information for birds is from Chiappe and Dyke (2002) and Fountaine et al. (2005), and for pterosaurs from Wellnhofer (1991), supplemented with more-recent publications listed in *Supplementary Material*.

Our approach is in three parts: (1) we describe and analyze some excellent new penguin fossils, (2) we report and analyze three new mitochondrial genomes of penguin relatives, and (3) we analyze some patterns and processes amongst pterosaurs to help elucidate events prior to the K/T boundary.

Description of the Oldest Fossil Penguins, Genus *Waimanu*

The fossil material represents two species of wing-propelled diving bird that are named and described here and illustrated in Figure 1. Each was found as a natural association of bones from a single individual, and judging from fusion of elements (vertebrae in synsacra) and some well-developed articular surfaces, they each represent a mature individual. The larger and older species is almost the size of an Emperor penguin (*Aptenodytes forsteri*); the smaller and geologically slightly younger is estimated at 80 cm tall, about the size of a yellow-eyed penguin (*Megadyptes antipodes*). Most parts of the skeleton are preserved in *Waimanu* (fig. 1), allowing ready comparison with literature on avian cladistics (Cracraft 1988; Mayr and Clarke 2003). Cladistic analyses, given below, place *Waimanu* at the base of the Sphenisciformes (sensu Clarke et al. 2003); and identify characters used in the generic diagnosis below.

Aves Linnaeus 1758

Neornithes Gadow 1893 *sensu* Cracraft 1988

Sphenisciformes Sharpe 1891 (*sensu* Clarke et al., 2003)

Waimanu Jones, Ando and Fordyce, gen. nov.

Type species. *Waimanu manneringi* sp. nov.
Etymology. Maori: *wai* (water), *manu* (bird).
Included species. Type species and *Waimanu tuatahi* sp. nov.
Diagnosis. *Waimanu* is closer to penguins than to any other bird group in the following combination of characters: some thoracic vertebrae are not heterocoelous; synsacrum has 11–12 ankylozed vertebrae; hypotarsal crests and grooves of the tarsometatarsus are not well developed (but medial hypotarsal crest is distinct). *Waimanu* differs from other stem- and crown-Sphenisciformes in that: humerus has an elongated and elevated insertion of supracoracoideus, and scapulotricipital groove does not form a trochlea; ulna has ridged anterior proximal margin; radius lacks anterior angulation and notch; carpometacarpus has stepped anterior margin; scapular blade is evenly wide; thoracic vertebrae are very weakly opisthocoelous and laterally excavated; synsacrum has column-like vertebral bodies; femur has a deep patellar groove; extensor groove of tibiotarsus runs on medial side; and tarsometatarsus is long and waisted, with posterior-directed medial trochlea, distinct medial hypotarsal crest, and very shallow intermetatarsal grooves dorsally.

Waimanu manneringi Jones, Ando and Fordyce sp. nov.

Etymology. Honoring Al Mannering who found and collected the holotype.
Holotype. CM (Canterbury Museum) zfa35: associated part skeleton comprising almost complete right tibiotarsus, proximal half of right fibula, right tarsometatarsus, right pelvis, and synsacrum (with last thoracic vertebra attached to the synsacrum), four caudal vertebrae (fig. 1*A*).
Horizon, Locality, and Age. Basal Waipara Greensand (Wilson 1963; Field and Brown 1989), Waipara River (near 43°04′S, 172°36′E), New Zealand (Fossil Record Number M34/f453, NZ Fossil Record File). Calcareous nannofossils indicate late early Paleocene, 60.5–61.6 Ma (correlations after Cooper 2004); see *Supplementary Material*.
Diagnosis. Larger than *W. tuatahi*, with tarsometatarsus 78 mm long (cf. 65 mm).
Summary description. *Tarsometatarsus*: The tarsometatarsus is longer and more waisted than in more-crownward Sphenisciformes, but is still short and robust compared to volant birds. The prominent intercotylar projection is pointed proximally and dorsally. The proximal vascular foramina are small and, like the distal foramen, open onto the plantar surface. The hypotarsal crests are simplified as in other Sphenisciformes, and not enclosed to form a canal, though the medial crest is more pronounced than in other Sphenisciformes. Trochlea III is the longest, the other trochleae are slightly shorter and of subequal length, and trochlea IV is directed somewhat plantarly. The

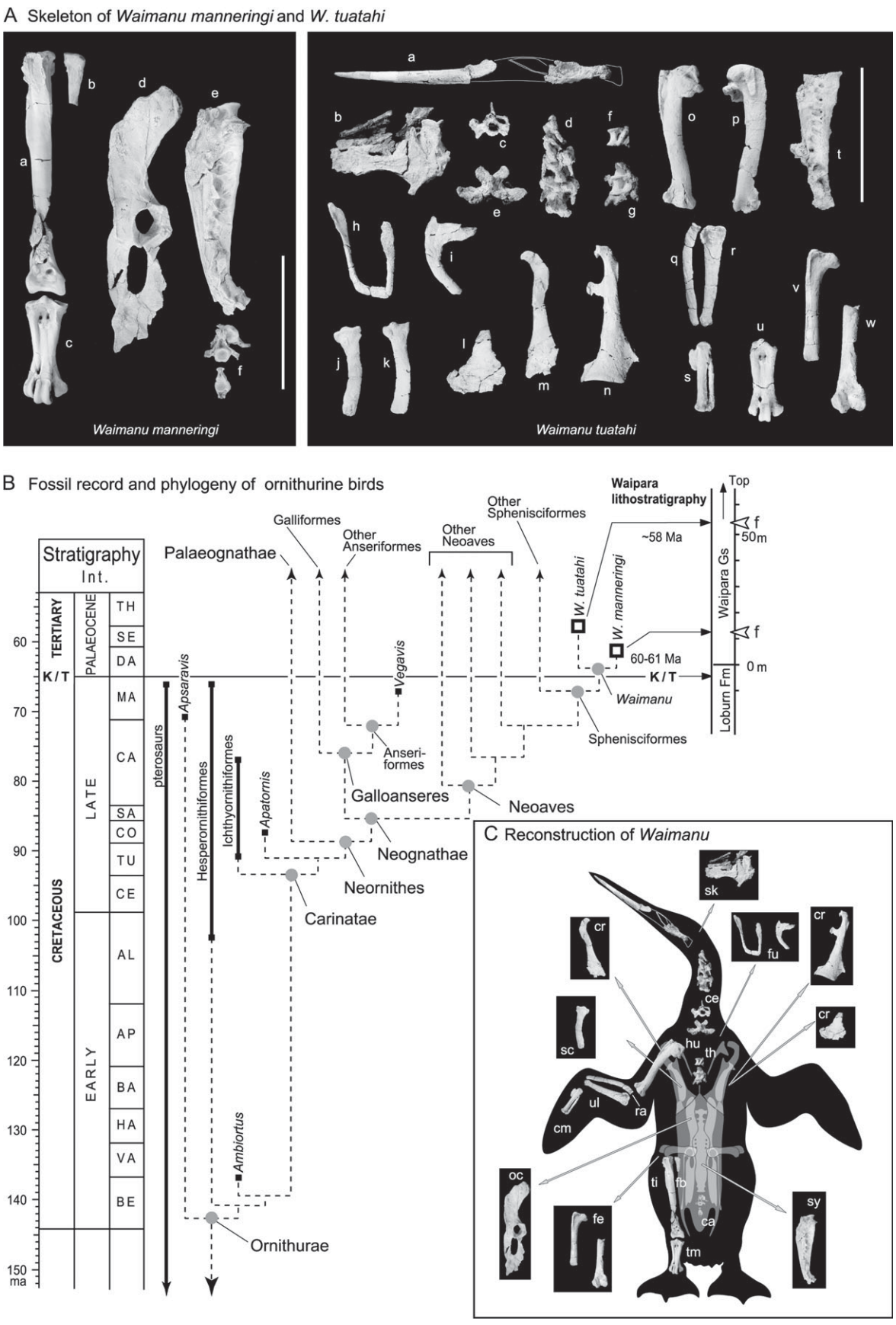
intermetatarsal grooves are separate, as in other Sphenisciformes, but very shallow.
Tibiotarsus: The tibiotarsus is broken in its distal third, revealing dense internal bone, but retains almost its full length; length is intermediate between King penguin (*Aptenodytes patagonicus*) and Emperor penguin (*Aptenodytes forsteri*). Profiles indicate that the cnemial crest protruded proximally. The shaft is more robust than in extant penguins, and the structure resembles that of *Paraptenodytes antarcticus*.
Os coxae: On the relatively complete right os coxae, the ilioischiadic foramen is larger than the acetabulum, and is closed posteriorly. The lateral margin of the preacetabular wing of the ilium has a deep concavity while the dorsal margin is rather straight. The dorsal iliac crest and the spinal crest of the synsacrum are not fused. The dorsolateral crest begins more posteriorly than in extant penguins. The preacetabular process is not well developed; the medial surface of the postacetabular part is rather flat, without a distinct renal fossa. Details are lost for the anterior margin of the ilium, the posterior and ventral margins of the ischium, and the pubis.
Synsacrum and caudal vertebrae: 11 fused vertebrae form this element; it is not fused with the os coxae. The spinal crest is high, robust, and full length. The vertebral bodies do not form a flat ventral surface at the middle of the synsacrum but keep a columnar structure. The synsacrum is intermediate in size between that of King penguin and Emperor penguin. The caudal vertebrae are unsurprising; one has a strong haemal process.

Waimanu tuatahi Ando, Jones and Fordyce sp. nov.

Etymology. Maori: *tuatahi* (first)—the holotype was the first specimen found.
Holotype. OU 12651 [Geology Museum, University of Otago]: associated part skeleton including skull fragments, incomplete mandible, cervical vertebrae, ribs, synsacrum, a furcula, coracoids, anterior half of right scapula, right humerus, distal end of left humerus, an ulna, a radius; proximal end of right femur (fig. 1*A*).
Hypodigm material. Holotype (OU 12651), and other associated part skeletons of: CM zfa 34—skull fragments, cervical and thoracic vertebrae, furcula, scapulae, coracoids, ribs, humeri, a radius, a carpometacarpus, and a tarsometatarsus (fig. 1*A*); CM zfa 33—partial skull and mandible, cervical and thoracic vertebrae, synsacrum, furcula, scapulae, coracoids, humerus, and a femur (fig. 1*A*).
Horizon and locality. All from the middle to upper Waipara Greensand, Waipara River. OU 12651, near 43° 02′ S, 172° 32′ E, fossil record number M34/f138; CM zfa 33, M34/f454; CM zfa 34, M34/f455 (zfa 33 and zfa 34, about 30 m stratigraphically above zfa 35; all zfa specimens are from near 43° 04′ S, 172° 36′ E). Dinoflagellates from M34/f138 indicate early Late Paleocene, 58–60 Ma (Cooper 2004: Fig. 11.5); M34/f454 and M34/f455 have not produced age-diagnostic fossils but are older than Early Eocene, no older than Paleocene, and stratigraphic level in upper Waipara Greensand indicates Late Paleocene; see *Supplementary Material*.
Specific diagnosis. Smaller than *W. manneringi*, with tarsometatarsus 65 mm long (cf. 78 mm long). On the tarsometatarsus, concavities of medial and lateral margins are less distinct, and plantar deflection of trochlea of metatarsal II is relatively weak.
Summary description. *Skull*: Where elements can be compared, there are no significant differences between OU 12651, zfa 34 and zfa 33. There are bilateral nasal gland fossae on the dorsal surface of the frontal. Frontal and parietal are fused completely, and the temporal fossa is large, reaching the top of the skull. The lacrimal has a descending process. The preserved part of the jugal bar is not dorsoventrally curved. On the quadrate, the otic process bears the separated articular surface for the otic and the squamosal. The palatine is present but incomplete.
Mandible: Partial mandibles are preserved. The symphysis is long and completely ossified. The anterior part of each ramus is slender and straight; the rami diverge only slightly posteriorly, but not enough is preserved to judge profiles toward the articulation. The medial mandibular fossa is large.
Furcula: The acromial process is well developed, and makes an acute angle with the rest of the clavicle. The

→

FIG. 1.—The Paleocene penguin *Waimanu*. (A) Skeletal elements of *Waimanu*. Left box, *W. manneringi*. a. right tibiotarsus in anterior view, b. right fibula in anterior view, c. right tarsometatarsus in dorsal view, d. right os coxae in lateral view, e. synsacrum in lateral view, f. caudal vertebrae in anterior view (a–f are CM zfa 35). Right box, *W. tuatahi*. a. mandible in lateral view (anterior part) and in medial view (posterior part), b. cranium in dorsal view, c–e. cervical vertebrae in anterior view (c), and in ventral view (d, e), f–g. thoracic vertebrae in lateral view, h. furcula in anterior view, i. right/left clavicle in lateral view, j–k. scapulae in lateral view, l. right coracoid in ventral view, m–n. left coracoids in ventral view, o. right humerus in ventral view, p. right humerus in dorsal view. q. left radius in dorsal view, r. left ulna in dorsal view, s. left carpometacarpus in dorsal view, t. synsacrum in lateral view, u. right tarsometatarsus in dorsal view, v. right femur in anterior view, w. left femur in posterior view. (a, c, h, j, l–m, p–r, t, and v are OU 12651; f–g, i, k, n, o, s, u, and w are CM zfa 34; b, d, and e are zfa 33). Scale bar = 100 mm. (B) Fossil record and phylogeny of ornithurine birds with the stratigraphy of Waipara region and geological settings for *Waimanu*. Solid line shows geological ranges of taxa with first and last occurrences shown by squares. Dashed line shows postulated phylogeny compiled from literature (Martin and Stewart 1982; Fox 1984; Chiappe 1995, 2003; Elzanowski et al. 2000; Norell and Clarke 2001; Cracraft and Clark 2001; Chiappe and Dyke 2002; Clarke and Norell 2002; Galton and Martin 2003; Clarke 2004). Gray circles indicate possible initial divergence times for clades; known fossils (squares) show constraints on ages. Early divergences within the Carinatae could be older, and we have conservatively placed them later in the Cretaceous to give only one long ghost-lineage between *Ambiortus* and the early Carinatae. The placement of *Waimanu* within Sphenisciformes is evaluated by the cladistic analysis described in the text; see also supplementary figure 4. TH, Thanetian; SE, Selandian; DA, Danian; MA, Maastrichtian; CA, Campanian, SA; Santonian, CO, Coniacian; TU, Turonian; CE, Cenomanian; AL, Albian; AP, Aptian; BA, Barremian; HA, Hauterivian; VA, Valanginian; BE, Berriasian. (C) Reconstruction of *Waimanu* (composite of *W. manneringi* and *W. tuatahi*, based on original art by Chris Gaskin ©Geology Museum, University of Otago). ca, caudal vertebrae; ce, cervical vertebrae; cm, carpometacarpus; cr, coracoid; fb, fibula; fe, femur; fu, furcula; hu, humerus; sk, skull, md, mandible; oc, os coaxae; ra, radius; sc, scapula; sy, synsacrum; ti, tibiotarsus; tm, tarsometartarsus; ul, ulna. In the wing, the dorsal view (left ulna, radius, carpometacarpus) and ventral view (humerus) are combined.


A Skeleton of Waimanu manneringi and W. tuatahi
Waimanu manneringi
Waimanu tuatahi
B Fossil record and phylogeny of ornithurine birds
Stratigraphy
Int.
TERTIARY
CRETACEOUS
PALAEOCENE
LATE
EARLY
TH
SE
DA
MA
CA
SA
CO
TU
CE
AL
AP
BA
HA
VA
BE
K/T
60
70
80
90
100
110
120
130
140
150
ma
Palaeognathae
Galliformes
Other Anseriformes
Other Neoaves
Other Sphenisciformes
Waipara lithostratigraphy
Top
~58 Ma
f
50m
f
0 m
Waipara Gs
Loburn Fm
W. tuatahi
W. manneringi
60-61 Ma
K/T
Waimanu
Sphenisciformes
Vegavis
Anseri-formes
Galloanseres
Neoaves
Neognathae
Neornithes
Carinatae
Apsaravis
Hesperornithiformes
Ichthyornithiformes
Apatornis
Ambiortus
pterosaurs
Ornithurae
C Reconstruction of Waimanu
sk
cr
fu
cr
ce
cr
sc
hu
th
ul
ra
cm
oc
ti
fb
sy
fe
ca
tm

articular facet of the acrocoracoid is projecting. The furcula lacks an interclavicular process, and probably did not abut the carina of the sternum.
Scapula: The acromion and glenoid process are rather symmetrical, not pronounced. The body is evenly wide posteriorly.
Coracoid: The shoulder end is elongate, with a medially bending acrocoracoid process; the scapular cotyla is a round shallow depression. The procoracoid process strongly projects medially, contributing to a thin and distally extended plate-like margin. Profiles indicate that a coracoid foramen was present. The sternal end is flared with a well-developed medial angle, and the base of a lateral process.
Wing elements: Overall, the wing is short relative to body size for birds in general, although relatively longer than in more-crownward penguins. Elements are variably flattened dorsoventrally, more so for the humerus than other bones. Broken sections reveal dense internal bone. The humerus has a sigmoidal shaft, and a well-developed articular surface excavated by a large and deep pneumotricipital fossa which lacks a pneumatic foramen at the base. The insertion of *M. supracoracoideus* is elevated and elongate distally, while a slight impression marks the origin for *M. brachialis* at the distal end of the anterior margin. The dorsal and ventral condyles are relatively prominent, and not flattened. The scapulotricipital and humerotricipital grooves are well developed, but only the latter forms a trochlea with strong ridges on the posterior margin of the distal end. The ulna is straight and broad proximally, while the radius is narrow and curved; these bones are shorter than the humerus (ca. 80% of length). The carpometacarpus is relatively broad and straight, with a stepped anterior margin.
Synsacrum: The synsacrum is smaller than but otherwise similar to that of *W. manneringi*.
Femur: The femur is long and straight, with only a slight posterior bend distally. The trochanteric crest is only weakly projected. The patellar groove is deep with distinct patellar crests. Broken sections reveal dense internal bone.
Tarsometatarsus: A nearly complete right tarsometatarsus (CM zfa 34), damaged proximally, is less waisted than in *W. manneringi* (namely, concavities of medial and lateral margins are less distinct) and is smaller (83% as long). Further, the trochlea of metatarsal II is relatively weakly deflected plantarly.

Remarks

Superficially, *Waimanu* is similar to geologically younger wing-propelled diving birds, such as the extinct Northern Hemisphere auk *Mancalla* and the diving plotopterid pelican *Copepteryx* (see Miller and Howard 1949; Olson 1980). Cladistically, *Waimanu* is closely related to modern penguins and thus belongs in the stem-Sphenisciformes (see below); they are large robust birds (ca. 80–100 cm), with dense heavy bones. Compared with volant birds, and as in other wing-propelled divers, the wing is short relative to body size, with flattened and generally wide bones; however, the distal ulna and radius are not widened. The structure of the humerus suggests limited rotation of the elbow; wing form rules out aerial flight and is consistent with wing-propelled diving. The structure of the humerus head and coracoid are reminiscent of other wing-propelled divers (Miller and Howard 1949; Howard 1966, 1970, 1976; Olson and Hasegawa 1979, 1996; Olson 1980). Structure of the pelvis and legs is similar to that of other penguins, suggesting a marked upright stance, although the tarsometatarsus is longer than in later penguins. The femur is long and straight, as in other flightless wing-propelled divers, unlike the short and often-bent femur of foot-propelled divers such as cormorants, loons, and *Hesperornis* (see Miller and Howard 1949; Howard 1966, 1970, 1976; Olson and Hasegawa 1979, 1996; Olson 1980; Arney and Wise 2003). A long narrow bill occurs in other stem-penguins such as "*Palaeeudyptes*" and *Platydyptes*. The larger species (one specimen) is older at 61 Ma (see *Supplementary Material*); the smaller species (three specimens) is younger, approximately 58 Ma.

Evaluation of the Phylogenetic Position of Waimanu

In such cases, where a disparate new taxon is reported close to the K/T boundary, it should be considered whether the material represents a relict of a much older clade. In this case, *Waimanu* clearly shows synapomorphies for Ornithurae and Carinatae (characters discussed by Cracraft 1988) and, as indicated below, is deeply nested within the crown birds, Neornithes (see supplementary fig. 3).

We evaluated the phylogenetic position of *Waimanu* within Neornithes by cladistic analysis with the published data set of Mayr and Clarke (2003) with additional taxa but no new characters. We include three new taxa (*Waimanu* and the fossil penguins *Platydyptes* and "*Palaeeudyptes*"), giving 46 ingroup and 3 outgroup taxa, and 148 characters. Characters of *Platydyptes* and "*Palaeeudyptes*" were scored from specimens from the OU collections (Fordyce and Jones, 1990, T. Ando, 2006, PhD in progress). The data set was processed with PAUP* 4.0b10 (Swofford 2001) with the same settings as the primary analysis of the original analyses (Mayr and Clarke 2003) including bootstrap analysis.

Results show that *Waimanu* belongs to Sphenisciformes, namely to crown + stem-penguins sensu Clarke et al. (2003). In contrast to the molecular results reported here, the clade of loons and grebes (Gaviiformes + Podicipediformes) appears as the sister taxon to *Waimanu* + other penguins. Of note, *Waimanu* is not close to the Pelecaniformes or Charadriiformes which, according to other studies (Howard 1976, Olson and Hasegawa 1996), include wing-propelled diving clades: the extinct Plotopteridae (Pelecaniformes) and the extinct flightless auks in the Alcidae (Charadriiformes), see *Supplementary Material*.

The position of *Waimanu* is well supported. Unambiguously optimized synapomorphies place it in the Neornithes, the Neognathae, the Neoaves, the clade of Gaviidae + Podicipedidae + Sphenisciformes + Procellariidae, and the Sphenisciformes. Nodes for Neoaves and for the clade of Gaviidae + Podicipediformes + Sphenisciformes + Procellariiformes were not supported by the bootstrap analysis (50%), confirming the results of Mayr and Clarke (2003), but bootstrap analysis (50%) did support other clades including Sphenisciformes (see *Supplementary Material* Figure 4). Nodes and characters are as follows,

with numbers in parentheses from the original data set of Mayr and Clarke (2003) (note that optimization has in some cases allocated state 0 as derived, and state 1 as primitive).
Neornithes: supratendinous bridge on distal end of tibiotarsus is completely ossified (100:1); distal interosseus canal is present in tarsometatarsus (107:0).
Neognathae: frontoparietal suture is closed (32:1); fossa of brachialis muscle of humerus is indistinct (79:1); scapulotricipital groove of humerus is well developed (81:1); ilioischiadic foramen of pelvis is closed posteriorly (94:1).
Neoaves: preacetabular tubercle of pelvis is vestigial (93:1).
Clade of Gaviidae + Podicipediformes + Sphenisciformes + Procellariiformes: nasal gland fossa is present on dorsal surface of frontal (25:1); pneumatic foramen is absent from bottom of pneumotricipital fossa of humerus (77:0); ulna does not distinctly exceed humerus in length (82:0); cnemial crest markedly protrudes proximally (99:1).
Sphenisciformes: some thoracic vertebrae are not heterocoelous (57:0); synsacrum has 11–12 ankylozed vertebrae (91:1); hypotarsal crests and grooves of tarsometatarsus are not well developed (103:0).

Mitochondrial Genomes and Phylogenetic Analysis

DNA extraction, long-range PCR, and subsequent rounds of short-range PCR (including cloning where necessary) and DNA sequencing were by standard methods and are reported in *Supplementary Material*. Accession numbers and sequence lengths for albatross, petrel, and loon are AY158677 (17,026 bp), AY158678 (16,414 bp, control region incomplete), and AY293618 (17,573 bp), respectively. Names and accession numbers of the other taxa analyzed are in *Supplementary Material*. The data set has 25 birds plus an outgroup of six reptiles, and consists of the 12 protein-coding genes from the heavy DNA strand together with 22 combined RNA genes. Alignments are at http://awcmee.massey.ac.nz/downloads.htm. Analysis is by standard programs including ModelTest (Posada and Crandall 1998), PAUP* (Swofford 2001), MrBayes (Huelsenbeck and Ronquist 2001), and Multidivtime (Thorne and Kishino 2002).

Phylogenetic analyses (see *Supplementary Material*) used nucleotide coding for first and second codon positions and for RNA stems and loops, and RY-coding for third codon positions. For both mammals and birds we find that increased taxon sampling (Lin et al. 2002) and RY-coding of third codon positions (Delsuc et al. 2003; Phillips and Penny 2003; Harrison et al. 2004), increases concordance between mitochondrial and nuclear data sets. We ran 1000 unconstrained ML bootstrap replicates with PAUP* on the Helix computing cluster (www.helix.massey.ac.nz), plus a Bayesian analysis using chains of 10^7 replicates. Detailed results are given in *Supplementary Material*. The reptilian outgroup joins between paleognaths and neognaths, consistent with recent work (Harrison et al. 2004). Of the outgroup taxa, crocodilians are closest to birds. Dating estimates using Multidivtime (Thorne and Kishino 2002) under a variety of constraints are also provided in *Supplementary Material* and show the high information content of the data. Due to program limitations of Multidivtime, divergence time estimation could not use RY-coding for the third codon position; therefore third positions were either omitted (as in fig. 2) or coded as nucleotides (as in *Supplementary Material*). The two main calibration points occur in different sections of the avian tree. We used 62 Ma for the divergence of penguins and storks (which are closest to penguins in the present mtDNA analysis) to allow time for divergence from their common ancestor. 66 Ma was used for the magpie goose/*Presbyornis*/*Vegavis* divergence (Kurochkin et al. 2002; Harrison et al. 2004; Clarke et al. 2005). Both calibration points are minimum divergence times on, for example, the penguin lineage. That is, they are estimates of the 'lower bounds' on the divergences, not estimates of the absolute dates. This is what is required to test the models of Penny and Phillips (2004). If the real divergences are older, this would only strengthen our conclusion about modern birds overlapping in time with earlier groups.

Figure 2 shows the avian part of the tree, together with dating estimates. The combined chicken and duck group (Galloanseres) is basal within neognaths. The remaining neognath birds form Neoaves (Cracraft 2001), with the two clades represented here by the passerines and the 'seabirds'/shorebirds/raptors. The division between shorebirds (turnstone and oystercatcher) and the informal group, which for present purposes we call seabirds (albatross, petrel, stork, loon, and penguin) is maintained (Cracraft 2001). The falconiforms (falcon and buzzard) are just basal to these. We do not, as reported earlier with short sequences (Cooper and Penny 1997), find the penguin joining with the procellariforms (petrel and albatross). The most unexpected aspect of the tree is the penguin/stork grouping, but we have omitted the stork, reanalyzed the data and still obtain the ((penguin, loon), (albatross, petrel)) grouping (see later). Thus the stork/penguin grouping does not affect our conclusions, but the relationships in that part of the tree are being addressed by sequencing of other potential relatives for both the stork and the penguin.

The dates of divergence shown in figure 2 are still conservative, but our results are consistent with the fossil record of bones and footprints. There are many fragmentary fossil bird remains in the Late Cretaceous (Hope 2002, Fountaine et al. 2005—and see http://palaeo.gly.bris.ac.uk/data/birds.html), together with a large body of fossil footprint data (Lockley 1998; Hwang et al. 2002; Lockley and Rainforth 2002). Because this material is fragmentary, certainly compared to the preservation of *Waimanu*, it is must be omitted from formal cladistic analyses and so the fossils are difficult to place taxonomically. Nevertheless, the birds existed and would have had ecological requirements and effects! A few new specimens (Kurochkin et al. 2002; Clarke et al. 2005) are important exceptions in that they are well preserved. Because our results come from a combination of excellent fossil material and long DNA sequences, they enhance the value of both these earlier fragmentary fossil bones and footprints.

There is wide debate about the identity and significance of Late Cretaceous Neornithes (modern birds). One recent review gave 44 Late Cretaceous records (Hope 2002, table 2); most are Maastrichtian, but six are Campanian and one (at 84–89 Ma) is Coniacian (giving a lower bound on the divergence of neognaths and paleognaths). Many were

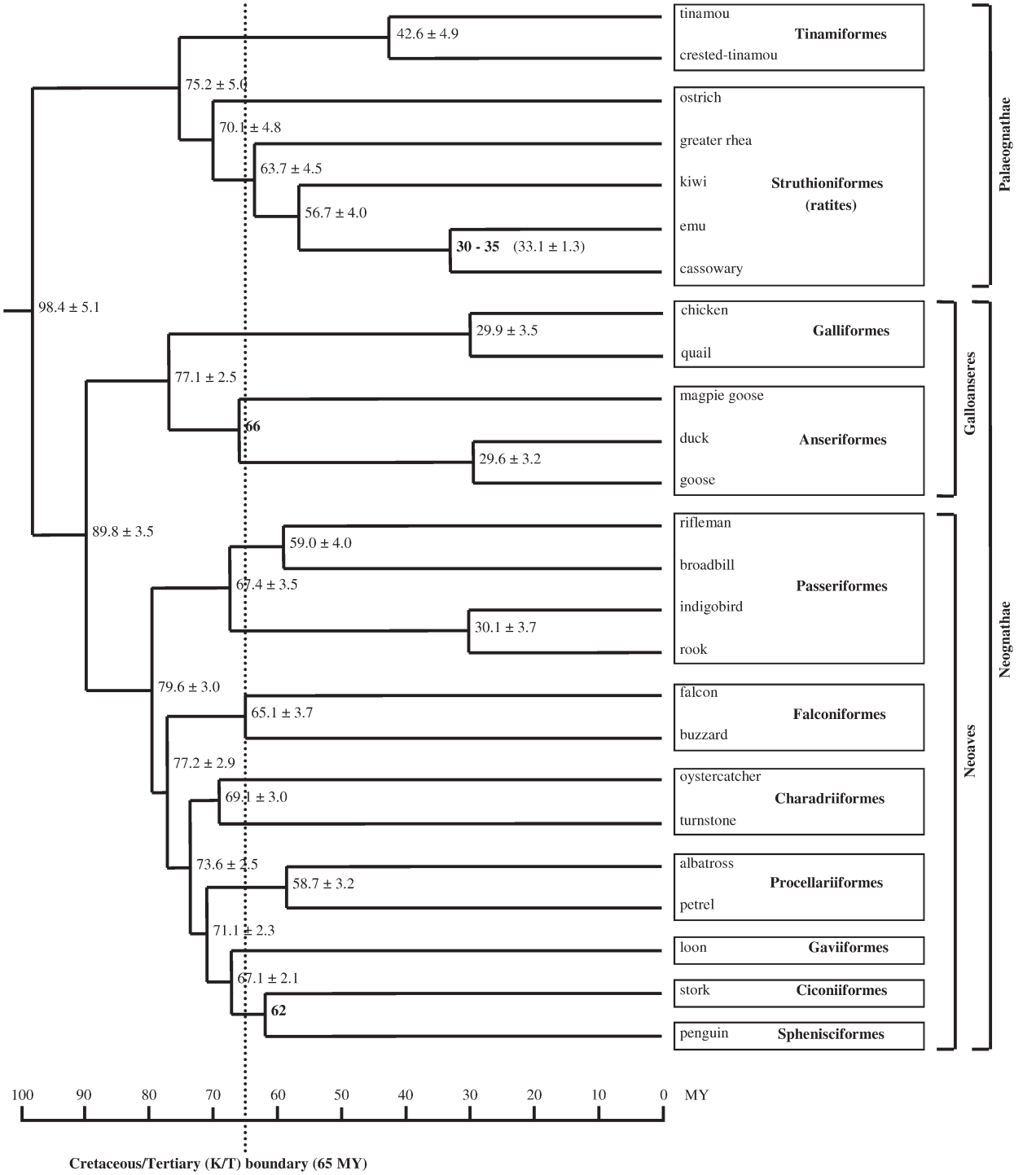


FIG. 2.—Rooted tree for 25 birds showing posterior divergence time estimates and their standard deviations (95% confidence intervals were also calculated but are not shown). The tree is drawn to scale (time in million years [Myr]) and the Cretaceous/Tertiary (K/T) boundary at 65 Ma is marked. Dating estimates were carried out using the tree in Figure S2 (see *Supplementary Material*) and the program packages PAML and Multidivtime. The data set consisted of the 12 protein-coding genes from the mitochondrial heavy strand plus the two ribosomal RNAs and 20 transfer RNAs (tRNA-Phe and tRNA-Glu were excluded due to missing data), coded as nucleotide (nt) data. Due to limitations of Multidivtime, third codon positions (cdp) could not be analyzed as transversion (RY) data and were therefore omitted along with gaps, ambiguous sites around gaps and stop codons, giving a data set 8582 nt long. Two main calibration points were used (indicated in bold): 1. The divergence between magpie goose and (duck, goose) constrained at 66 Ma; 2. The divergence between penguin and stork constrained at 62 Ma. The F84 plus discrete gamma model was used in base_ml (PAML). Details of parameters used in are in *Supplementary Material*.

identified as members of stem lineages of current neornithine groups and, further, Robertson et al. (2004) gives a long list of Neornithes likely to have been present in the Late Cretaceous. Fountaine et al. (2005) recognized 22–23 species of Neornithes published by 2003 (database of Fountaine et al. 2005 at http://palaeo.gly.bris.ac.uk/data/birds.html; derived from Chiappe and Dyke (2002) and Hope's 2002 data). Fountaine et al. (2005) reiterated that the material is largely fragmentary, and indicated that while most specimens are of uncertain taxonomic affinities their ecological habitat may be more certain. Amongst the few Cretaceous neornithines known from reasonable material, *Vegavis* is based on one well-preserved associated skeleton (Clarke et al. 2005), and *Teviornis* is known from associated wing bones (Kurochkin et al. 2002). A third supposed neornithine, the Antarctic loon-like *Polarornis*, is of debatable age (Chiappe and Dyke 2002). In their overview, Fountaine et al. (2005) argued that neornithines really were rare in the Cretaceous, whereas Hope (2002) noted that 20–25 records from the Lance formation are neornithines. For now, our approach is to combine excellent fossil material from the early Paleocene close to the K/T boundary, with long DNA sequences to emphasize that the neornithines radiated in the Late Cretaceous. Our results show the advantage of combining fossil and sequence data.

The Cretaceous radiation of neornithines raises the question of competition, before the K/T boundary, between modern birds and archaic birds and pterosaurs. Competitive interaction could lead to ecological displacement, but the groups under study must have at some point overlapped in space and time and had ecological overlap. Major differences in body size would seem to rule out ecological overlap between any two clades, although disparate size might reflect earlier competitive divergence. If there is a long-term drop in diversity, the time of extinction of the last member in the clade is not as important (in ecological terms) as the preceding pattern of decline (Penny and Phillips 2004). Finally, it must be asked whether low-diversity clades might still be ecologically significant on a regional to global scale.

To consider other Cretaceous birds, the enantiornithines formed the most diverse clade of Mesozoic birds, and are structurally similar to later Neornithes (e.g. Zhou 2004). Enantiornithines apparently did not survive the KT boundary. The enantiornithine stratigraphic record, according to data from Fountaine et al. (2005, http://palaeo.gly.bris.ac.uk/data/birds.html), shows two diversity peaks in the Early and Late Cretaceous (Barremian-Albian and Campanian-Maastrichtian). Although there is an expanding Cretaceous record of footprints (e.g. Kim et al. 2003; Lockley 1998), such traces generally are only identifiable to a higher-level group such as bird or pterosaur, or "ecological" group such as "web-footed bird". The results in figure 2 imply that shorebirds were present in the Campanian and Maastrichtian, and so they are the best candidates for leaving the "web-footed" footprints.

Decline in Pterosaur Diversity

Fossil sites with thousands of footprints (Kim et al. 2003) show that large pterosaurs lived in the same habitat as web-footed birds, although the differences in size, structure, and form of locomotion point to different ecologies for pterosaurs and birds. Wang et al. (2005) suggest that earlier (Aptian) pterosaurs were common. Kim et al. (2003) offer the interpretation that pterosaurs might have fed on small birds—equally plausible to us, however, is that the raptors (Falconiformes; which, as shown in fig. 2 are an early group) could have preyed on young pterosaurs. There are fundamental trends in the long-term record of pterosaurs (fig. 3).

The first aspect of pterosaur diversity is ecological/life history (fig. 3*B*) as revealed by trends in wingspan (see Wellnhofer 1991; Buffetaut and Mazin 2003; Chang et al. 2004; and further details in *Supplementary Material*). From the Late Triassic (Norian, ~225 Ma) until the end of the Jurassic (~145 Ma), the wingspan of pterosaurs was basically less than 2 m (meters), though with an increase toward the end of the Jurassic. From the early Cretaceous onward there was a major increase in maximum wingspan reaching to over 11 m by the end of the Cretaceous. Perhaps critical in relation to bird evolution was the loss of small pterosaurs (<2 m wingspan) from the mid-Cretaceous onward, matching the inferred radiation of Neornithes. (The only exception is a fragmentary record of *Ornithocheirus bunzeli* from 1881 which was <2 m and that needs to be confirmed; see Wellnhofer 1991.) The matter of size is complicated because many pterosaurs had distinct age classes, so that some previously described small 'species' were younger age classes of other species (see Bennett 1996).

The second aspect of pterosaur diversity is taxonomic. Panel A shows the highest diversity of pterosaurs around the end of the Jurassic/early Cretaceous; the later drop in pterosaur diversity matches the inferred radiation of Neornithes. These results are taken from figure 21 in Unwin (2003, shown also in *Supplementary Material*), except that we are not yet able to confirm records of nyctosaurids in the Maastrichtian, and thus terminate that lineage earlier. On quantitative grounds, the azhdarchids are by far the most common form in the Maastrichtian. The conclusion from Panels A and B is that, because the pterosaurs are reducing in diversity over the Late Cretaceous, the impact at the K/T boundary (even if it were the proximate cause) *cannot* be the 'ultimate' cause of pterosaur decline and eventual extinction.

The final small panel (C) is a qualitative reminder that both fossil remains of stem lineages of modern shore birds (Hope 2002) and fossil footprints (e.g. Lockley 1998; Hwang et al. 2002) exist during the later stages of the Cretaceous. The panel illustrates that from cause and effect considerations, modern shorebirds existed before the disappearance of pterosaurs. Similarly, we have reported (Fig. 3 in Penny and Phillips 2004) the decline in small dinosaurs (less than about 2 m long as adults) before the end of the Cretaceous. This decline in both small dinosaurs and pterosaurs is significant because it is different to the current situation in which it is the larger mammals, not the smaller, that are endangered (Cardillo et al. 2005). It is the larger species (not the smaller) that are at most risk of extinction. One inference from this is that pterosaurs were particularly vulnerable to extinction with only large species surviving.

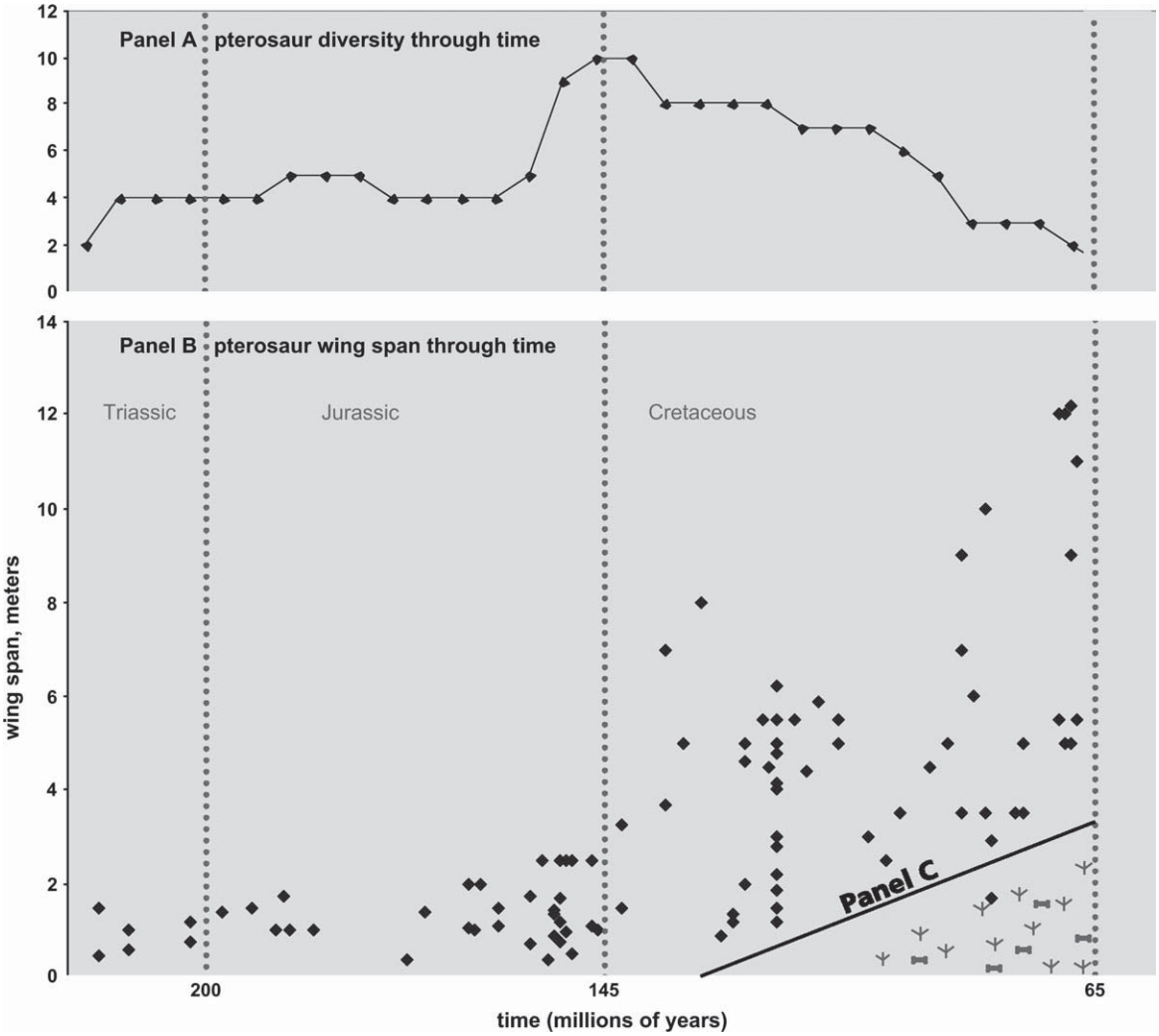


FIG. 3.—Pterosaurs through time. (A) Pterosaur taxonomic diversity from the mid-Triassic until the end of the Cretaceous (calculated at 5 Ma intervals from Fig 21 in Unwin, 2003). (B) Adult wingspans for pterosaurs. Estimates, wherever possible, are for adults and toward the larger size; sources are in *Supplementary Material*. The results indicate a major increase in size throughout the Cretaceous but with an increasing loss of small and medium-sized pterosaurs in the Late Cretaceous—at the same time modern birds are diversifying. (C) Occurrence of fossil footprints and stem-group modern birds in the last quarter of the Cretaceous. The panel is an indication that both fossil footprints of unidentified modern birds, as well as fossils of stem lineages, do occur in the last half of the Late Cretaceous.

Discussion

The four *Waimanu* fossils are some of the oldest well-preserved Neornithine birds. They pre-date by 6–8 Myr the widely cited Early Eocene species from the London Clay (53–55 Ma) (Feduccia 1996, 2003; Chiappe and Dyke 2002; Dyke and Gulas 2002), and are only a little younger than the best-associated partial skeleton of a neornithine, that of *Vegavis* (Clarke et al. 2005) from the Late Cretaceous. *Waimanu* indicates that penguins diverged from other Neornithes and acquired disparate features such as large body size, more upright stance, and wing-propelled diving habits by the early Paleocene—only 3–4 Myr after the K/T boundary event. We suggest that the great disparity between Sphenisciformes and their sister-taxa is consistent with an origin for penguins during the Late Cretaceous neornithine radiation. The time of radiation, which has been debated widely (Cooper and Penny 1997; Bleiweiss 1998; Cracraft 2001; Chiappe and Dyke 2002; Feduccia 2003), is predicted here as starting at 90–100 Ma. Our results do not yet fully resolve ingroup relationships for the seabird clade (see Cracraft 1988; Sibley and Ahlquist 1990; Cooper and Penny 1997; Livezey and Zusi 2001; Mayr and Clarke 2003).

Our results support the recent consensus that modern birds were not restricted to shorebird/seabird niches during the Late Cretaceous (Hope 2002; Dyke and van Tuinen 2004; but see Feduccia 2003). Both morphology (Cracraft 2001) and molecules place the root of the avian tree distant from the marine groups studied here; neither shorebirds nor seabirds are basal among modern birds. Because of the increased chance of finding fossils from an aquatic environment it is easy to under-estimate the relative age of non-aquatic birds (Hope 2002). The fossil record for aquatic birds appears reasonable (Fountaine et al. 2005) but, as judged from figure 2, most terrestrial lineages are not represented in the early fossil record. The best terrestrial bird

record often comes from occasional settings of exceptional preservation (Norell and Clarke 2001; Zhou 2004).

A Late Cretaceous radiation of modern (neornithine) birds prompts questions about the evolution and extinction of earlier stem-birds (such as the enantiornithines, *Hesperornis*, and *Ichthyornis*) and pterosaurs. Were these groups displaced by modern (neornithine) birds, or did they replace the archaic groups after the latter disappeared? It is not clear whether the extinction of archaic birds was abrupt or involved long-term decline over some 30 Myr of the Late Cretaceous (Unwin 1988; Chiappe and Dyke 2002; see also Ward et al. 2005), when diverse crown-lineage birds were already present. Fossil footprints indicate that pterosaurs certainly coexisted with web-footed birds around 80–96 Ma (Hwang et al. 2002; Kim et al. 2003) (Campanian-Cenomanian), implying that some specialized birds occupied fresh-water settings in the Late Cretaceous. Some archaic forms clearly survived quite late—the flightless marine bird *Hesperornis* until early Maastrichtian (Hills et al. 1999), and *Ichthyornis*-like birds until the latest Cretaceous (Dyke et al. 2002; but debated by Clarke 2004). Following Penny and Phillips (2004) our focus is on when a group started to decline, not when the last member of a clade went extinct (which is intrinsically uncertain because of sampling issues from the Signor-Lipps effect [Wagner 2000]). Because archaic and modern birds overlapped for many million years, we focus on the ecological implications of this. When, for example, did modern birds start displacing/replacing earlier groups such as ichthyornithids; when did the earlier groups start losing niche-space to modern birds (Penny and Phillips 2004)?

Models for potential interactions between modern birds, archaic birds, and pterosaurs require life history information. For example, early pterosaurs show both slow bone growth and year classes (Bennett 1996), thus they may have been K-selected. Early birds, such as enantiornithines, show comparable patterns, with bone having distinct lines of arrested growth (Chinsamy et al. 1998). In contrast, *Hesperornis* and *Ichthyornis*, like neornithines, had rapid and sustained bone growth (Chinsamy et al. 1998), suggesting that later bird clades were physiologically advanced over earlier lineages. Given the recent focus (Holdaway and Jacomb 2000; Johnson 2002; Cardillo et al. 2005) on the susceptibility to extinction of large, K-selected animals (with slow growth and reproductive rates), life history patterns are important. Traditionally, there has been an apparent conflict between paleontological and molecular estimates of divergences for both mammals and birds. In contrast, we see important interactions between molecular and paleontological information; collaborative work is important. We need to integrate studies on paleontology (including footprints), DNA sequences (nuclear and mitochondrial), ecology, and physiology to develop testable models of past extinctions and radiations. Macroevolutionary models can then lead to testable models (Penny and Phillips 2004). One of the simplest predictions is an increase in the size of fossil pterosaur footprints during the Late Cretaceous, at the same time fossil footprints occur of shore birds. Similarly, we might expect the food sources of pterosaurs (as indicated by aspects of the fossils themselves, see Wellnhofer 1991) to decline in the Late Cretaceous. More predictions and tests are required.

Supplementary Material

Supplementary Material are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/). For the penguin fossils it includes information on the statigraphy, and character states for the cladistic analysis. For the mitochondrial genomes it includes details of GenBank numbers, selection of taxa, and results of Bayesian runs. For pterosaurs it includes a table of all fossil records used in calculating Figure 3, together with references.

Acknowledgments

We thank Dick Gill (Department of Conservation Waikanae office) for the petrel sample, Alan Tennyson and Sandy Bartle of Te Papa (Museum of New Zealand) for the albatross, Einar Arnason and Thorvaldur Björnsson [Reykjavik, Iceland] for the loon, Trish McLenachan, Alicia Gore, and Gillian Gibb for some laboratory work, Matt Phillips and Fred Delsuc for assistance with analysis, David Unwin and Peter Wellnhofer for advice on pterosaur evolution, Gerald Mayr for providing a cladistics matrix for birds, Graeme Wilson for dinoflagellate dating, Norton Hiller for allowing study of the Canterbury Museum material, Te Papa for the use of facilities, and the Massey interloan service for assistance beyond the call of duty. The NZ Marsden Fund, the Otago Research Committee and the Swedish Science Council provided financial support. K.E.S. and G.L.H. carried out the DNA work; C.M.J., T.A., and R.E.F. the fossil part of the study (including naming and describing of *Waimanu* and both its species), T.A. the cladistic analysis of fossils, K.E.S. the phylogenetic and dating analysis, D.P. the pterosaur analysis, D.P. and U.A. designed the project, and D.P. and K.E.S. developed the first draft. We thank the landowners (Forrester and Goorde families) for access to localities. Finally, we thank Brad Field for finding the first specimen (OU 12651), and we particularly acknowledge the sterling efforts of Al Mannering in finding and preparing the Canterbury Museum specimens.

Funding to pay the Open Access publication charges for this article was provided by the Allan Wilson Center for Molecular Ecology and Evolution.

Literature Cited

Alvarez, L. W., W. Alvarez, F. Asaro, and H. V. Michel. 1980. Extraterrestrial cause for the Cretaceous-Tertiary extinction—experimental results and theoretical interpretation. Science **208**:1095–1108.

Arney, J. E., and S. W. Wise Jr. 2003. Paleocene-Eocene nannofossil biostratigraphy of ODP Leg 183, Kerguelen Plateau. Proc. Ocean Drilling Program, Sci. Results **183**:1–59.

Baker, A. J., S. L. Pereira, O. P. Haddrath, and K.-A. Edge. 2006. Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. Proc. R. Soc. Lond. B **273**:11–17.

Bennett, S. C. 1996. Year-classes of pterosaurs from Solnhofen limestone of Germany: taxonomic and systematic implications. J. Vert. Paleontol. **16**:432–444.

Bleiweiss, R. 1998. Fossil gap analysis supports early Tertiary origin of trophically diverse avian orders. Geology **26**:323–326.
Bromham, L. D., and M. D. Hendy. 2000. Can fast early rates reconcile molecular dates to the Cambrian explosion? Proc. R. Soc. Lond. B **267**:1041–1047.
Buffetaut, E., and J.-M. Mazin. 2003. Evolution and Palaeobiology of Pterosaurs. Geol. Soc. Lond. Spec. Publ. 217.
Cardillo, M., G. M. Mace, K. E. Jones, J. Bielby, O. R. P. Bininda-Emonds, W. Sechrest, C. D. L. Orme, and A. Purvis. 2005. Multiple causes of high extinction risk in large mammal species. Science **309**:1239–1241.
Chang, M-M., ed. 2004. The Jehol Biota: the emergence of feathered dinosaurs, beaked birds and flowering plants. Shanghai Scientific and Technical Publishers, Shanghai.
Chiappe, L. M. 1995. The first 85-million years of avian evolution. Nature **378**:349–355.
Chiappe, L. M. 2003. Basal bird phylogeny. Pp. 448–472 *in* L. M. Chiappe and L. M. Witmer, eds. Mesozoic birds: above the heads of Dinosaurs. University of California Press, Berkeley.
Chiappe, L. M., and G. J. Dyke. 2002. The Mesozoic radiation of birds. Annu. Rev. Ecol. Syst. **33**:91–124.
Chinsamy, A., L. D. Martin, and P. Dodson. 1998. Bone microstructure of the diving *Hesperornis* and volant *Ichthyornis* from the Niobrara chalk of western Kansas. Cretac. Res. **19**:225–235.
Clarke, J. A. 2004. Morphology, phylogenetic taxonomy, and systematics of *Ichthyornis* and *Apatornis* (Avialae: Ornithurae). Bull. Am. Mus. Nat. Hist. **286**:1–179.
Clarke, J. A., and M. A. Norell. 2002. The morphology and phylogenetic position of *Apsaravis ukhaana* from the Late Cretaceous of Mongolia. Am. Mus. Novit. **3387**:1–46.
Clarke, J. A., E. B. Olivero, and P. Puerta. 2003. Description of the earliest fossil penguin from South America and the first Paleogene vertebrate locality of Tierra del Fuega, Argentina. Am. Mus. Novit. **3423**:1–18.
Clarke, J. A., C. P. Tambussi, J. I. Noriega, G. M. Erickson, and R. A. Ketchman. 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. Nature **433**:305–308.
Conway Morris, S. 1998. Early metazoan evolution: reconciling paleontology and molecular biology. Am. Zool. **38**: 867–877.
Cooper, A., and D. Penny. 1997. Mass survival of birds across the Cretaceous/Tertiary boundary. Science **275**:1109–1113.
Cooper, R. A. 2004. The New Zealand geological timescale. Inst. Geol. Nucl. Sci. Monogr. **22**:1–284.
Cracraft, J. 1988. Pp. 339–361 *in* M. J. Benton, ed. The Phylogeny and Classification of the Tetrapods. Clarendon Press, Oxford.
Cracraft, J. 2001. Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. Proc. R. Soc. Lond. B **268**:459–469.
Cracraft, J., and J. A. Clarke. 2001. The basal clades of modern birds. Pp. 143–156 *in* J. Gauthier and L. F. Gall, eds. New perspectives on the origin and early evolution of birds: proceedings of the international symposium in honor of John H. Ostrom. Peabody Museum of Natural History, Yale University.
Delsuc, F., M. J. Phillips, and D. Penny. 2003. Technical comment on "Hexapod Origins: monophyletic or paraphyletic?" Science **301**:1482d.
Dyke, G. J., R. W. Dortangs, J. W. M. Jagt, E. W. A. Mulder, A. S. Schulp, and L. M. Chiappe. 2002. Europe's last Mesozoic bird. Naturwissenschaften **89**:408–411.
Dyke, G. J., and B. E. Gulas 2002. The fossil galliform bird *Paraortygoides* from the Lower Eocene of the United Kingdom. Am. Mus. Novit. **3360**:1–14.
Dyke, G. J., and M. van Tuinen. 2004. The evolutionary radiation of modern birds (Neornithes) reconciling molecules, morphology and the fossil record. Zool. J. Linn. Soc. **141**:153–177.
Elzanowski, A., G. S. Paul, and T. A. Stidham. 2000. An avian quadrate from the late Cretaceous Lance formation of Wyoming. J. Vert. Paleontol. **20**:712–719.
Feduccia, A. 1996. The Origin and Evolution of Birds. Yale University Press, New Haven.
Feduccia, A. 2003. 'Big bang' for Tertiary birds? Trends Ecol. Evol. **18**:172–176.
Field, B. D., and G. H. Browne. 1989. Cretaceous and Cenozoic sedimentary basins and geological evolution of the Canterbury Region, South Island, New Zealand. N. Z. Geol. Survey Basin Studies **2**:1–94.
Foote, M., J. P. C. Hunter, M. Janis, and J. J. Sepkowski. 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. Science **283**:1310–1314.
Fordyce, R. E., and C. M. Jones 1990. Penguin history and new fossil material from New Zealand. Pp. 419–446 *in* L. S. Davis and J. T. Darby, eds. Penguin Biology. Academic Press, San Diego.
Fountaine, T. M. R., M. J. Benton, G. J. Dyke, and R. L. Nudds. 2005. The quality of the fossil record of Mesozoic birds. Proc. R. Soc. Lond. B **272**:289–294.
Fox, R. C. 1984. *Ichthyornis* (Aves) from the early Turonian (Late Cretaceous) of Alberta. Can. J. Earth Sci. **21**: 258–260.
Galton, P. M., and L. D Martin. 2003. *Enaliornis*, an Early Cretaceous hesperornithiform bird from England, with comments on other hesperornithiformes. Pp. 317–338 *in* L. M. Chiappe and L. M. Witmer, eds. Mesozoic Birds: Above the heads of Dinosaurs. University of California Press, Berkeley.
Härlid, A., A. Janke, and U. Arnason. 1997. The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. Mol. Biol. Evol. **14**:754–761.
Harrison, G. L., P. A. McLenachan, M. J. Phillips, K. E. Slack, A. Cooper, and D. Penny. 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the Late Cretaceous. Mol. Biol. Evol. **21**:974–983.
Hedges, S. B., P. H. Parker, C. G. Sibley, and S. Kumar. 1996. Continental breakup and the ordinal diversification of birds and animals. Nature **381**:226–229.
Hills, L. V., E. L. Nicholls, L. M. Nunez-Betelu, and D. J. McIntyre. 1999. *Hesperornis* (Aves) from Ellesmere Island and palynological correlation of known Canadian localities. Can. J. Earth Sci. **36**:1583–1588.
Holdaway, R. N., and C. Jacomb. 2000. Rapid extinction of the moas (Aves: Dinornithiformes) Model, test, and implications. Science **287**:2250–2254.
Hollis, C. J., and C. P. Strong. 2003. Biostratigraphic review of the Cretaceous/Tertiary boundary transition, mid-Waipara River section, North Canterbury, New Zealand. N. Z. J. Geol. Geophys. **46**:243–253.
Hope, S. 2002. The Mesozoic radiation of Neornithes. Pp. 339–388 *in* L. M. Chiappe and L. M. Witmer, eds. Mesozoic Birds: Above the heads of dinosaurs. University California Press, Berkeley.
Howard, H. 1966. A possible ancestor of the Lucas auk (family Mancallidae) from the Tertiary of Orange County, California. Los Angeles County Mus. Contrib. Sci. **101**:1–8.
Howard, H. 1970. A review of the extinct avian genus, *Mancalla*. Los Angeles County Mus. Contrib. Sci. **203**:1–12.
Howard, H. 1976. A new species of flightless auk from the Miocene of California (Alcidae: Mancallinae). Smithsonian Contrib. Paleobiol. **27**:141–146.

Huelsenbeck, J., and F. Ronquist. 2001. MrBayes: Bayesian inference on phylogenetic trees. Bioinformatics **17**:754–755.

Hwang, K.-G., H. Huh, M. G. Lockley, D. M. Unwin, and J. L. Wright. 2002. New pterosaur tracks (Pteraichnidae) from the Late Cretaceous Uhangri Formation, southwestern Korea. Geol. Mag. **139**:421–435.

Johnson, C. N. 2002. Determinants of loss of mammal species during the late Quaternary 'megafauna' extinctions: life history and ecology, but not body size. Proc. R. Soc. Lond. B **269**: 2221–2227.

Jones, C. M., and A. Mannering. 1997. New Paleocene fossil bird material from the Waipara Greensand, North Canterbury, New Zealand. Geol. Soc. N. Z. Misc. Publ. **95A**:88.

Kim, C. B., M. Huh, C. S. Cheong, M. G. Lockley, and H. W. Chang. 2003. Age of the pterosaur and web-footed bird tracks associated with dinosaur footprints from South Korea. Island Arc **12**:125–131.

Kurochkin, E. N., G. J. Dyke, and A. A. Karhu. 2002. A new presbyornithid bird (Aves, Anseriformes) from the late Cretaceous of Southern Mongolia. Am. Mus. Novit. **3386**:1–11.

Lin, Y.-H., P. A. McLenachan, A. R. Gore, M. J. Phillips, R. Ota, M. D. Hendy, and D. Penny. 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Mol. Biol. Evol. **19**:2060–2070.

Livezey, B. C., and R. L. Zusi. 2001. Higher-order phylogenetics of modern Aves based on comparative anatomy. Netherlands J. Zool. **51**:179–205.

Lockley, M. G. 1998. The vertebrate track record. Nature **396**:429–432.

Lockley, M. G., and E. C. Rainforth. 2002. The track record of Mesozoic birds and pterosaurs: an ichnological and paleoecological perspective. Pp. 405–418 *in* L. M. Chiappe and L. M. Witmer, eds. Mesozoic Birds: Above the heads of dinosaurs. University of California Press, Berkeley.

Martin, L. D., and J. D. Stewart. 1982. An ichthyornithiform bird from the Campanian of Canada. Can. J. Earth Sci. **19**:324–327.

Mayr, G., and J. Clarke. 2003. The deep divergences of neornithine birds: a phylogenetic analysis of morphological characters. Cladistics **19**:527–553.

Miller, L., and H. Howard. 1949. The flightless Pliocene bird *Mancalla*. Pp. 201–228 *in* R. W. Wilson, ed. Some Tertiary mammals and birds from North America. Carnegie Institution, Washington.

Myrcha, A., P. Jadwiszczak, C. P. Tambussi, J. I. Noriega, A. Gazdzicki, A. Tatur, and R. A. del Valle. 2002. Taxonomic revision of Eocene Antarctic penguins based on tarsometatarsal morphology. Polish Polar Res. **23**:5–46.

Norell, M. A., and J. A. Clarke. 2001. Fossil that fills a critical gap in avian evolution. Nature **409**:181–184.

Olson, S. L. 1980. A new genus of penguin-like pelecaniform bird from the Oligocene of Washington (Pelecaniformes: Plotopteridae). Contrib. Sci. Nat. Hist. Mus. Los Angeles County **330**:51–57.

Olson, S. L., and Y. Hasegawa. 1996. A new genus and two new species of gigantic Plotopteridae from Japan (Aves: Pelecaniformes). J. Vert. Paleontol. **16**:742–751.

Olson, S. L., and Y. Hasegawa. 1979. Fossil counterparts of giant penguins from the North Pacific. Science **206**:688–689.

Paton, T., O. Haddrath, and A. J. Baker. 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proc. R. Soc. Lond. B **269**:839–846.

Penny, D., and M. J. Phillips. 2004. The rise of birds and mammals: are microevolutionary processes sufficient for macroevolution. Trends Ecol. Evol. **19**:516–522.

Peters, S. E. 2005. Geologic constraints on the macroevolutionary history of marine animals. Proc. Natl Acad. Sci. USA **102**:12326–12331.

Phillips, M. J., and D. Penny. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phyl. Evol. **28**:171–185.

Posada, D., and K. A. Crandall. 1998. MODELTEST: Testing the model of DNA substitution. Bioinformatics **14**:817–818.

Robertson, D. S., M. C. McKenna, O. B. Toon, S. Hope, and J. A. Lillegraven. 2004. Survival in the first hours of the Cenozoic. Geol. Soc. Am. Bull. **116**:760–768.

Sibley, C. G., and J. E. Ahlquist. 1990. Phylogeny and Classification of Birds: A Study in Molecular Evolution. Yale University Press, New Haven.

Simons, A. M. 2002. The continuity of microevolution and macroevolution. J. Evol. Biol. **15**:688–701.

Simpson, G. G. 1975. Fossil Penguins. Pp. 19–41 *in* B. Stonehouse, ed. The Biology of Penguins. Macmillan, London.

Slack, K. E., A. Janke, D. Penny, and U. Arnason. 2003. Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene **302**:43–52.

Swofford, D. L. 2001. PAUP* (phylogenetic analysis using parsimony (*and other methods) ver 4.0b8. Sinauer Associates, Sunderland, Mass.

Thorne, J. L., and H. Kishino. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. **51**:689–702.

Unwin, D. M. 2003. On the phylogeny and evolutionary history of pterosaurs. Pp. 139–190 *in* E. Buffetaut and J.-M. Mazin, eds. Evolution and Palaeobiology of Pterosaurs. Geol. Soc. Lond. Spec. Publ. 217.

Unwin, D. M. 1988. Extinction and survival in birds. Pp. 295–318 *in* G. P. Larwood, ed. Extinction and survival in the fossil record. Clarendon Press, Oxford.

Vajda, V., J. I. Raine, and C. J. Hollis. 2001. Indication of global deforestation at the Cretaceous-Tertiary boundary by New Zealand fern spike. Science **294**:1700–1702.

Van Tuinen, M., and S. B. Hedges. 2001. Calibration of avian molecular clocks. Mol. Biol. Evol. **18**:206–213.

Wagner, P. J. 2000. The quality of the fossil record and the accuracy of phylogenetic inferences about sampling and diversity. Syst. Biol. **49**:65–86.

Wang, X., A. W. A. Kellner, Z. Zho, and D. S. Campos. 2005. Pterosaur diversity and faunal turnover in Cretaceous terrestrial ecosystems in China. Nature **437**:875–879.

Ward, P. D., J. Botha, R. Buick, M. O. De Kock, D. H. Erwin, G. H. Garrison, J. L. Kirschvink, and R. Smith. 2005. Abrupt and gradual extinction among Late Permian land vertebrates in the Karoo Basin, South Africa. Science **307**:709–714.

Wellnhofer, P. 1991. The Illustrated Encyclopedia of Pterosaurs. Salamander Books, London.

Wilf, F., K. R. Johnston, and B. T. Huber. 2003. Correlated terrestrial and marine evidence for global climate changes before mass extinction at the Cretaceous-Paleogene boundary. Proc. Natl Acad. Sci. USA **100**:599–604.

Wilson, D. D. 1963. The geology of Waipara subdivision. N. Z. Geol. Surv. Bull. **64**:1–122.

Zhou, Z. H. 2004. The origin and early evolution of birds: discoveries, disputes, and perspectives from fossil evidence. Naturwissenschaften **91**:455–471.

William Martin, Associate Editor

Accepted March 1, 2006

Supplementary Information

Early penguin fossils, plus mitochondrial genomes, give a firm calibration point for avian evolution

Kerryn E Slack[♦1,2], Craig M Jones[♦3], Tatsuro Ando[♦4], G. L.(Abby) Harrison[1], R. Ewan Fordyce[4], Ulfur Arnason[2] and David Penny[1]

[1] Allan Wilson Center for Molecular Ecology and Evolution
Institute of Molecular BioSciences
Massey University
Palmerston North
New Zealand

[2] Division of Evolutionary Molecular Systematics
Department of Cell and Organism Biology
University of Lund
Solvegatan 29
S-223 62 Lund
Sweden

[3] Institute of Geological and Nuclear Sciences
Lower Hutt
New Zealand

[4] Department of Geology,
University of Otago
Dunedin
New Zealand

♦ These authors contributed equally to this work.

□escription of the oldest fossil penguins, genus *Waimanu*

Age determination

The Waipara Greensand yields microfossils (foraminifera, radiolaria and calcareous nannofossils) [1,2,3] (Jenkins 1971□Strong 19□4□Hollis and Strong 2003) indicating a wholly Paleocene age, clearly younger than the K/T boundary[4] (□a□da et al. 2001). The Greensand lies within the local (New Zealand) Teurian stage. The top of the Waipara Greensand marks the Teurian - Waipawan local stage boundary, which is correlated internationally with the Paleocene - Eocene boundary[5] (Cooper 2004).

The stratigraphically lowest stem-penguin, CM □fa 35 (fossil record number M34/f453, holotype of *Waimanu manneringi*), provides a well-preserved calcareous nannofossil assemblage. The sample includes two key age-diagnostic ta□a, *Chiasmolithus bidens* and *Hornibrookina teuriensis*. Based on the latest correlations of New Zealand sequences to the international timescale[5] (Cooper 2004□Fig 11.1, 11.4, 11.10), these nannofossils indicate that the age is high in the lower Teurian local stage. Of note, the top of the range of *Hornibrookina teuriensis* is lower Selandian, no younger than about 60.5Ma. The overlap of *Chiasmolithus bidens* and *Hornibrookina teuriensis* indicates an age of about 60.5-61.6Ma.

Matri□from the holotype of *W. tuatahi*, OU 12651 (fossil record number M34/f13□), produced Paleocene dinoflagellates including *Palaeocystodinium golzowense* and *Deflandrea foveolata* (G.J. Wilson, personal communication)□the former indicates local Teurian Stage, Paleocene, while the latter indicates the *Deflandrea foveolata* □one, roughly Selandian or early Late Paleocene[5] , 5□-60 Ma (Cooper 2004□Fig 11.5). Such an age is consistent with known origin from the middle to upper Waipara Greensand. The other specimens of *W. tuatahi*, CM □fa 34 (M34/f454) and CM □fa 33 (M34/f455), produced no age-diagnostic fossils□based on superposition and known age of overlying strata, they are from the upper Waipara Greensand and no younger than Late Paleocene. There is no reason to propose that they are significantly different in age from the holotype of *W. tuatahi* (OU 12651, M34/f13□).

□ itochondrial genomes and phylogenetic analysis

Mitochondrial Genomes: Material and Methods

DNA was e□tracted from muscle or liver using High Pure™ PCR Template Purification Preparation Kit (Roche). In order to avoid amplifying nuclear copies, mitochondrial DNA was amplified first as two main fragments from □-12.3Kb, using the E□pand™ Long template PCR kit (Roche). This could leave small gaps that were covered by fragments □3.5Kb that had e□tensive overlap with the longest fragments. The long-range PCR fragments were used as template for short-range PCR reamplifications (1□2kb). We used the Fasta search in the GCG program (Wisconsin Package, version 10.0) to search our primer database for appropriate targets for primer walking. Where possible, primers from Kocher[6], Cooper[7] or Sorensen□ were used. Otherwise, new primers were designed using Oligo□4.03 (National BioSciences, Inc.). Sequencing

reactions were done according to standard protocols and run on a 377 ABI Applied BioSystems DNA sequencer. Some rechecking was done on an ABI 3730. Sequences were checked and assembled using Sequencher™ 4.1 (Gene Codes Corp).

Mitochondrial Genomes: □election of ta□a

In addition to the three mt genomes reported here, 15 other neognath ta□a are included in the analyses□ chicken (*Gallus gallus*□GenBank accession number □52392), quail (*Coturni□□a□onica*, AP003195), redhead duck (*Aythya americana*□AF090337), greater white-fronted goose (*Anser albifrons*, AF363031), magpie goose (*Anseranas semi□almata*□A□309455), rook (*Corvus frugilegus*□□1□522), village indigobird (□*idua chalybeata*□AF090341), gray-headed broadbill (□*mithornis shar□ei*□AF090340), rifleman (N Z wren, *Acanthisitta chloris*□A□325307), bu□□ard (□*uteo buteo*, AF3□0305), peregrine falcon (□*alco* □*eregrinus*□AF09033□), ruddy turnstone (*Arenaria inter□res*, A□074□□5), blackish oystercatcher (*Haemato□us ater*, A□074□□6), Oriental white stork (*Ciconia boyciana*□AB026193), and little blue penguin (□*udy□tula minor*, AF362763). In addition, seven paleognaths (five ratites and two tinamou) were included. They are, Great spotted kiwi (*A□tery□haastii*, AF33□70□)□emu (*Dromaius novaehollandiae*, AF33□711)□double-wattled cassowary (*Casuarius casuarius*, AF33□713)□ostrich (□*truthio camelus*, □12025)□greater rhea (□*hea americana*□AF33□713)□great tinamou (□*inamus ma□or*, AF33□707)□and elegant crested-tinamou (□*udromia elegans*, AF33□710). This gives a total of 25 avian mt genomes.

Within the Neoaves, an owl and a parrot were omitted. In the current dataset both are isolated ta□a in the Neoavian portion of the tree[9] and their positions are still a little unstable – though none come within the seabird/shore bird group that is the focus of this study. The owl and parrot can group together weakly, or the owl comes close to the passerines, and the parrot towards the falconiforms group. We are in the process of sequencing a barn owl (□*yto*), an African parrot (lovebird, *Aga□ornis*) and a forest falcon (*Micrastur*) (www.awcmee.massey.ac.n□/mt□genomes). Based on previous e□perience of improved ta□on sampling[9,10], we e□pect that these new sequences will help stabili□e the position of both the owls and parrots. Once again, parrots and owls are not the focus here, rather it is the time of origin of modern seabird and shore bird lineages and whether they predate the decline of hesperornids, ichthyornids and pterosaurs.

Si□reptiles were used as outgroups□American alligator (*Alligator mississi□□iensis*□□13113), eastern painted turtle (*Chrysemys □icta*□AF069423), green turtle (*Chelonia mydas*□AB012104), blue-tailed mole skink (□*umeces egregius*□AB016606), common iguana (□*guana iguana*□AJ27□511) and spectacled caiman (*Caiman crocodylus*□AJ404□72).

Mitochondrial Genomes: Phylogenetic analysis

In previous work[11,12,13] we found that R□-coding of nucleotide data both increases the proportion of changes on internal branches of the tree (treeness) and decreases the differences in nucleotide composition

(Relative Compositional Variability, RCV). Consequently this is our preferred method of analysis of animal mitochondrial data, and the tree reported here (Figure S2) has the third codon positions recoded as R & Y. The full data set is available from www.awcmee.massey.ac.nz/downloads.

Dating estimates were made on the tree shown in Figure Suppl_3 using the program packages PAML and Multidivtime. However, because of limitations of the dating program, third codon positions (3rd cdp) could not be analyzed as transversions (RY-coded). Two variants of the 25 bird + 6 reptile protein-coding + RNA dataset were therefore analyzed: one without 3rd cdp (8582 nucleotides [nt]) and one with 3rd cdp coded as nt data (11709 nt). The T3: (Thornian Time Traveler version http://abacus.gene.ucl.ac.uk/; see also Yang & Yoder[14]) of the Multidivtime program package[15] was used to estimate divergence times. Multidivtime uses Markov chain Monte Carlo (MCMC) procedures for Bayesian estimation of evolutionary divergence times and rates. The program baseml in the PAML version 3.14beta package[16] (http://abacus.gene.ucl.ac.uk/software/paml.html) was used to estimate base frequencies, the ratio of within to between group substitution rates (kappa) and the relative substitution rate of each category (r values) under the F84 plus discrete gamma model (ncatG = 8). These values were used in the Multidivtime program est_branches to estimate branch-lengths for the specified tree (ingroup only) and to produce a variance-covariance matrix for the branch-length estimates. The multidivtime program was then used to estimate divergence times (plus standard deviations and 95% confidence intervals) for each node in the tree (without assuming a molecular clock).

We estimated both the prior (no data) and posterior distribution of divergence times, analyzed the dataset both with and without third codon positions and constrained the three calibration points to different extents. We also ran the same analysis five times (Fig Suppl_2) with different random seed numbers to check that the results were similar, therefore indicating convergence of the MCMC. Table S1 summarizes our estimates of avian divergence times.

Three calibration points were used (indicated in bold in table S1): 1. The divergence between magpie goose and (duck, goose): lower bound set at 66 MY in all analyses, upper bound set at 66 MY in FIXED analyses and unconstrained in VARIABLE analyses; 2. The divergence between penguin and stork: lower bound set at 62 MY in all analyses, upper bound set at 62 MY in FIXED analyses and unconstrained in VARIABLE analyses; 3. The divergence between emu and cassowary: when used the lower bound was set at 30 million years (MY) and upper bound at 35 MY. Note that the in the 'variable' analyses the calibration point is used in the mathematical sense as a bound, not as a point estimate. The numbers in the first (lefthand) column of the table are those shown in Figure 3. The parameters used in the analyses were as for Figure 3, except that rt_rate was 0.010582 and rtrate_sd was 0.005291 when 3rd cdp were included. 95% confidence intervals were calculated but are not shown. SD: standard deviation.

A priori estimates for multidivtime parameters were as follows: 84 MY for the mean of the prior distribution for the time separating the ingroup root and the present (rt_tm) and 42 MY for the standard deviation of rttm (rttmsd); 0.001687 for the mean of the prior distribution for the rate at the ingroup root node (rtrate) and 0.000844 for the standard deviation of rtrate (rtratesd); 0.01 for both the mean of the prior distribution for Brownian motion constant 'nu' (brown_mean) and the standard deviation of brown_mean (brown_sd); 500 MY for the length of time between the root and the present (big_time). The Markov chain was sampled 10,000 times (num_samps), the number of cycles between samples (samp_freq) was 100 and the number of cycles before the first sample of the Markov chain (burnin) was 100,000. The results of other dating analyses (approximation of prior distributions of divergence times; with 3rd cdp, coded as nt data; without an upper bound on the penguin/stork and magpie goose/(duck + goose) calibration points) are summarized in Table S1.

Table S1. Summary of avian divergence time estimates.

	Posterior approximation								Prior approximation							
	Without 3rd cdp				With 3rd cdp				Without 3rd cdp				With 3rd cdp			
	FIXED		VARIABLE		FIXED		VARIABLE		FIXED		VARIABLE		FIXED		VARIABLE	
Divergences	Age	SD	Age	SD	Age	SD	Age	SD	Age	SD	Age	SD	Age	SD	Age	SD
emu/cassowary	**33.1**	**1.3**	**33.2**	**1.3**	**33.0**	**1.4**	**33.3**	**1.3**	**32.4**	**1.4**	**32.4**	**1.4**	**32.4**	**1.4**	**32.5**	**1.4**
kiwi/emu+cassowary	56.7	4.0	59.3	4.8	56.7	3.2	59.1	3.8	56.2	21.9	63.2	27.9	56.5	22.2	62.4	26.8
greater rhea/kiwi+emu+cassowary	63.7	4.5	67.2	5.7	65.7	3.9	69.4	4.8	80.3	29.2	93.9	37.2	80.7	29.6	93.4	36.2
ostrich/greater rhea+kiwi+emu+cassowary	70.1	4.8	74.5	6.2	71.9	4.2	76.3	5.4	103.8	33.2	124.7	42.1	104.7	33.7	124.0	41.2
tinamou/crested-tinamou	42.6	4.9	45.3	5.7	44.1	4.2	46.7	4.8	64.3	42.5	76.4	51.1	64.7	42.4	77.4	51.4
Struthioniformes (ratites)/Tinamiformes (tinamous)	75.2	5.0	80.3	6.6	80.2	4.6	85.6	6.0	128.0	35.3	154.9	43.9	128.9	36.1	155.0	44.1
chicken/quail	29.9	3.5	32.9	4.5	29.9	2.7	33.7	3.8	51.5	34.0	68.7	45.6	52.0	34.0	70.0	45.9
duck/goose	29.6	3.2	32.7	4.4	25.2	2.2	28.9	3.4	33.1	19.1	53.6	36.2	33.3	19.0	53.9	36.2
magpie goose/duck+goose	**66.0**	**0.0**	**72.5**	**5.0**	**66.0**	**0.0**	**74.2**	**5.1**	**66.0**	**0.0**	**107.2**	**32.9**	**66.0**	**0.0**	**107.1**	**33.1**
Galliformes/Anseriformes	77.1	2.5	84.1	5.7	75.5	2.0	83.6	5.4	102.4	28.1	138.8	38.8	102.8	28.5	139.3	38.9
penguin/stork	**62.0**	**0.0**	**66.7**	**4.1**	**62.0**	**0.0**	**65.9**	**3.6**	**62.0**	**0.0**	**77.6**	**15.6**	**62.0**	**0.0**	**77.6**	**15.6**
loon/penguin+stork	67.1	2.1	72.5	4.7	65.1	1.3	69.7	4.0	74.4	12.7	93.0	21.8	75.0	12.9	93.1	21.9
albatross/petrel	58.7	3.2	63.4	5.0	56.1	2.2	60.0	4.0	43.9	27.4	54.0	35.0	44.1	27.6	54.1	35.2
Procellariiformes/loon+penguin+stork	71.1	2.3	76.8	5.0	67.2	1.5	72.0	4.1	87.3	18.3	108.7	26.6	88.1	18.4	108.7	27.1
oystercatcher/turnstone	69.1	3.0	74.8	5.3	61.9	2.5	66.5	4.4	49.8	31.9	62.6	40.0	50.4	32.0	61.9	40.0
Charadriiformes/Procellariiformes+loon+penguir	73.6	2.5	79.6	5.2	71.3	1.8	76.8	4.5	100.1	22.7	124.1	30.6	100.8	22.5	124.5	31.5
falcon/buzzard	65.1	3.7	70.5	5.6	66.5	2.7	71.6	4.8	56.5	36.1	69.6	44.6	56.8	36.0	68.4	45.3
Falconiformes/Charadriiformes+Procellariiforme +penguin+stork	77.2	2.9	83.6	5.6	74.7	2.1	80.6	4.8	112.8	26.2	139.8	34.4	113.6	26.1	140.1	35.2
indigobird/rook	30.1	3.7	32.7	4.5	34.2	3.1	37.0	3.9	42.3	32.0	51.5	39.9	42.2	32.2	52.0	40.5
rifleman/broadbill	59.0	4.0	63.9	5.7	62.3	3.2	67.4	5.0	42.1	31.9	51.4	40.0	42.5	31.9	51.7	40.0
indigobird+rook (oscines)/broadbill (suboscine)+rifleman	67.4	3.5	73.0	5.6	71.1	2.7	76.9	5.1	83.8	36.0	103.5	45.8	84.4	36.3	104.4	45.9
Passeriformes/other Neoaves	79.6	3.0	86.2	5.7	80.1	2.4	86.8	5.2	125.4	29.6	155.3	37.7	126.1	29.5	155.8	38.5
Gallianseres/Neoaves	89.8	3.5	97.4	6.6	91.0	2.9	99.1	6.2	138.0	32.6	171.0	40.9	138.9	32.7	171.6	41.7
Palaeognathae/Neognathae	98.4	5.1	106.1	7.8	102.5	4.7	110.6	7.3	151.7	36.1	185.4	44.3	152.9	36.5	185.9	44.7

Table S2. Character matrix of 148 morphological character for □ □*i*□□□□, □□□□□□□□□□*p*□□□□, and □□□□□□□□*p*□□□.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Waimanu	□	□	□	□	0	□	□	□	□	□	□	0	0	□	□	□	□	□	□	□	□	□	□	□	1
"Palaeeudyptes"	1	0	0	1	□	□	□	□	□	□	□	0	□	□	□	□	□	□	□	□	□	□	□	□	□
Platydyptes	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Waimanu	□	□	□	□	□	□	1	□	1	□	□	1	0	□	0	1	1	0	□	□	□	□	□	□	□
"Palaeeudyptes"	□	□	□	□	□	□	□	□	1	0	0	1	0	□	0	□	1	0	0	0	0	0	□	□	□
Platydyptes	□	□	□	□	□	□	□	□	□	□	□	□	□	□	0	□	1	0	0	0	□	□	□	□	□
	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
Waimanu	□	□	□	0	□	□	0	1	1	□	□	0	0	0	0	0	0	0	□	□	□	□	□	□	0
"Palaeeudyptes"	□	□	□	□	□	□	0	□	1	0	0	0	0	0	0	0	0	0	0	□	1	0	1	□	0
Platydyptes	□	□	0	□	□	□	0	□	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1	□	0
	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Waimanu	0	0	0	1	0	1	0	1	0	0	0	□	□	□	0	1	0	1	1	0	0	0	0	1	1
"Palaeeudyptes"	0	0	0	1	0	1	0	□	□	0	0	□	□	□	0	1	0	1	1	0	0	0	0	1	1
Platydyptes	0	0	0	1	0	1	0	□	0	0	0	□	□	□	0	□	0	1	1	0	0	0	0	1	1
	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125
Waimanu	0	0	0	0	0	0	0	1	0	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
"Palaeeudyptes"	0	0	0	0	0	0	1	1	0	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
Platydyptes	0	0	0	0	0	0	1	1	0	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148		
Waimanu	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□		
"Palaeeudyptes"	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□		
Platydyptes	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□		

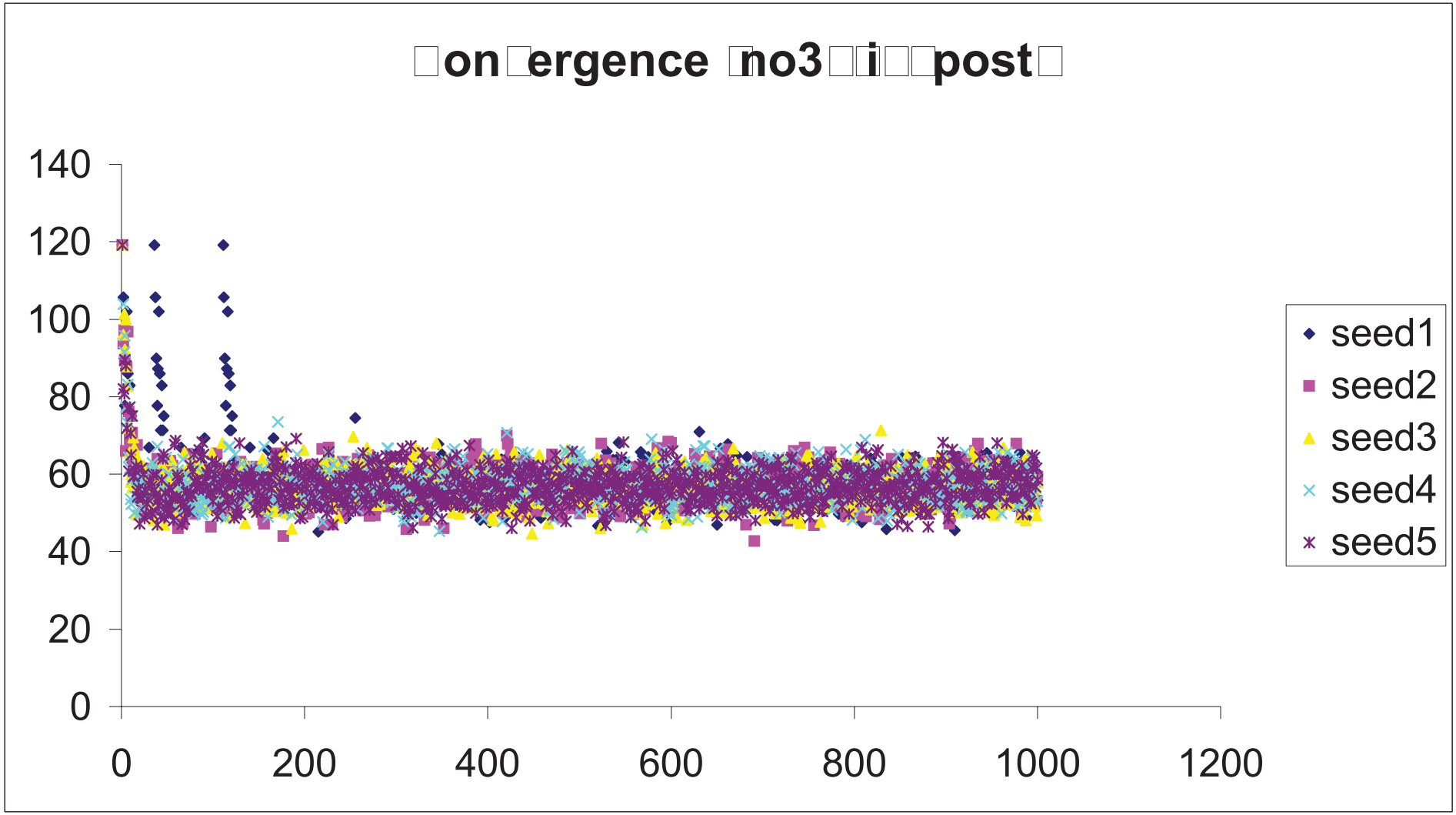


Figure S1. □utput from five MCMC runs to show that the runs both had converged and to similar values. Every 1000th step of each chain is plotted

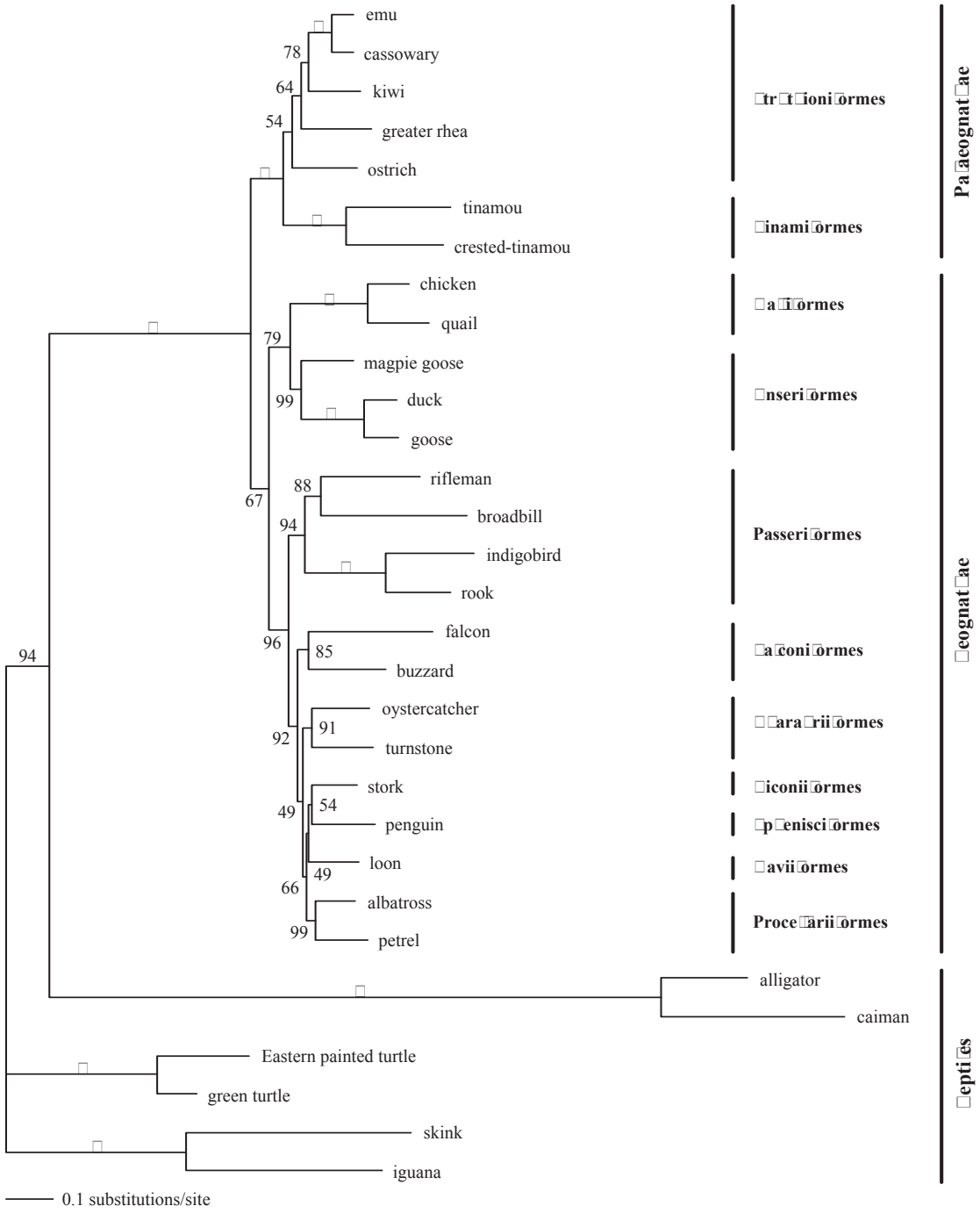


Figure S2. Maximum likelihood tree for 25 birds and six reptiles. The dataset is the 12 protein-coding genes from the mitochondrial heavy strand plus 22 RNAs (two ribosomal RNAs and 20 transfer RNAs - tRNA-Phe and tRNA-Glu were excluded due to missing data). Gaps, ambiguous sites around gaps, and stop codons were removed, giving a dataset of 11709 nt. First and second codon positions (cdp) and RNAs were analyzed as nucleotide (nt) data, third cdp were analyzed as transversion (RY) data. The analysis used the likelihood settings from the best-fit (GTR+I+G) model selected by the Akaike Information Criterion (AIC) in Modeltest Version 3.06. The tree has a -Ln likelihood value of 107275.44 and is to scale. Numbers represent bootstrap values (1000 independent bootstrap replicates); 100% bootstrap support is indicated by □

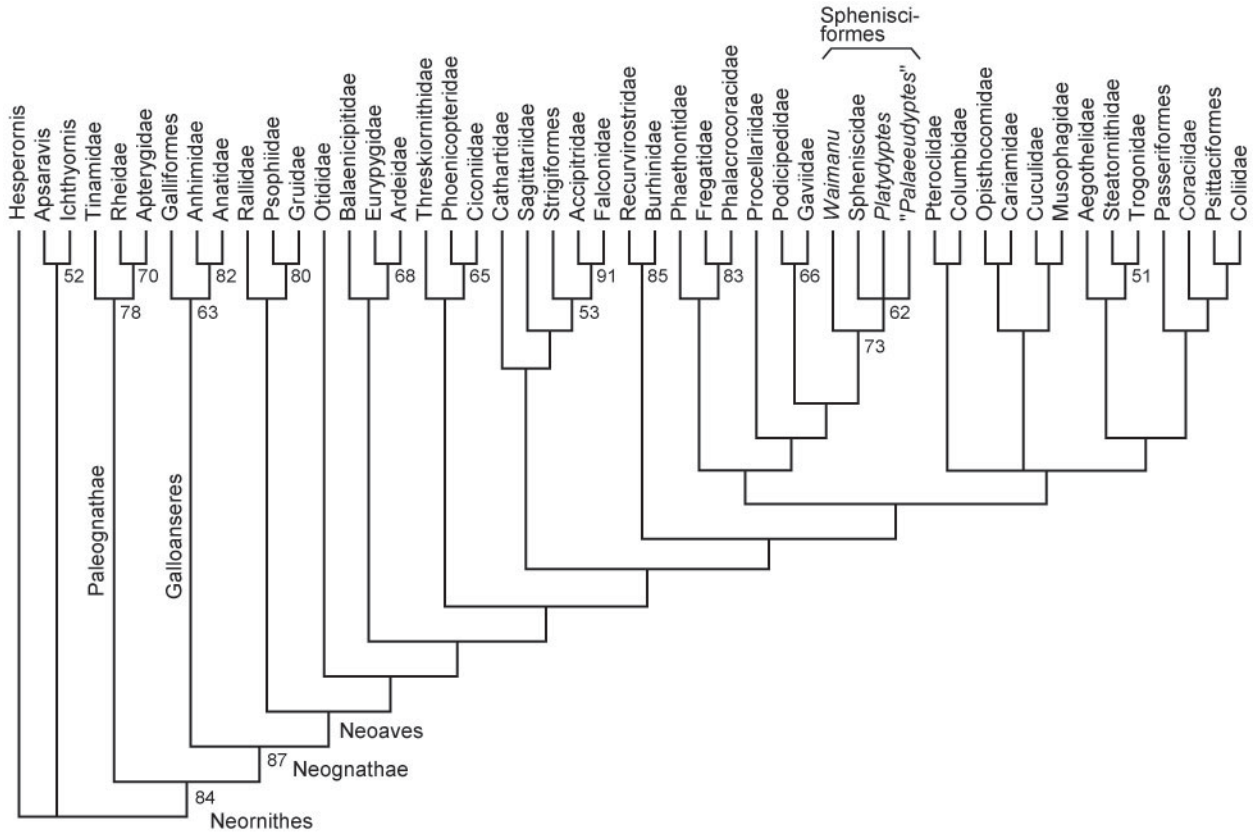


Figure S3. Phylogenetic position of □□*i*□□□□within Sphenisciformes in the strict consensus cladogram of 6 most parsimonious trees resulting from the analysis of morphological data (Length = 822, CI = 0.3228, RI = 0.4871, RC = 0.1573). Bootstrap support values more than 50% are indicated below and to the right of corresponding nodes. Waimanu and the penguins are not close to the Charadriiformes (represented here by Burhinidae and Recurvirostridae), nor to the Pelecaniformes (such as Fregatidae, Phalacrocoracidae and the Phaethontidae)

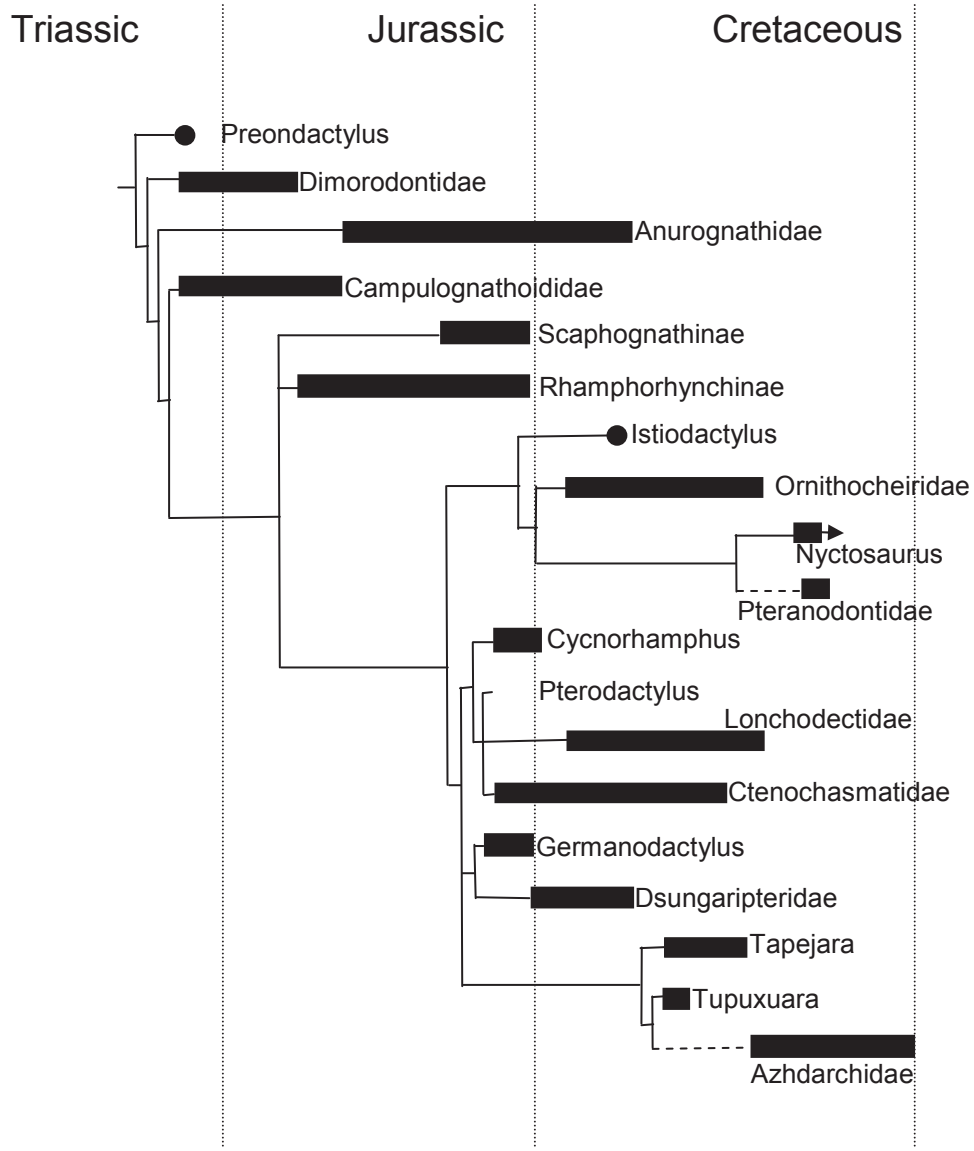
Table S3, Data and references for pterosaur (figure 3) - Data

			Age □a		span□ m			
□udimorphodon ranzii	□taly	p60	220	10	1	p60	□amphorhynch	*eudimorpodont*
□udimorphodon cf. ranzii	Austria	□□□24	210	20	0.75	□□□7	□amphorhynch	*eudimorpodont*
Peteinosaurus zambellii	Tethys	p60	220	10	0.6	p62	□amphorhynch	*dimorpodont*
Preondactylus (Padian)	□taly	p64	225	5	1.5	p64	□amphorhynch	*ramphorhynch*
Preondactylus buffarini		p64	225	5	0.45	p65	□amphorhynch	*ramphorhynch*
Austriadactylus cristatus	Austria	□□□27	210	20	1.2	□□□□ □□02	□amphorhynch	*?eudimorpodont?*
□imorphodon macrony□	□ngland □ias	p70	205	25	1.4	p70	□amphorhynch	*dimorpodont*
□amphinion □enkinsi	Arizona	p74	200	30	1.5	p78		*undetermined*
□orygnathus banthensis	Germany	p74	196	34	1	p74	□amphorhynch	*ramphorhynch*
Campylognathoides liasicus	Germany	p74	194	36	1	p76	□amphorhynch	*ramphorhynch*
Campylognathoides zitteli	Germany	p74	195	35	1.75	p76	□amphorhynch	*ramphorhynch*
Parapsicephalus purdoni	□ngland	p74	190	40	1	p78	□amphorhynch	*ramphorhynch*
agnurognathid	□ongolia	□96□297	175	55	0.4	□96□298	□amphorhynch	*anurognathidae*
□uizacal pterosaur	□e□ico	□96□298	172	58	1.4	□96-298	□amphorhynch	*campylognathoid?*
□amphocephalus bucklandii	□ngland	p81	165	65	1.05	p80	□amphorhynch	*ramphorhynch*
□amphocephalus depressirostris	□ngland	p81	164	66	1	□96	□amphorhynch	*ramphorhynch*
□amphocephalus	□ngland	□96□293	163	67	2	□96□295	□amphorhynch	*ramphorhynch*
□amphorhynchus	□ngland	□96□295	160	70	1.1	□96□296	□amphorhynch	*ramphorhynch*
Angustinaripterus longicephalus	China	p86	165	65	2	□□□□ □	□amphorhynch	*ramphorhynch*
□erbstosaurus pigmaeus	Argentina	p81	160	70	1.5	□96□299	Pterodacyloidea	*dsungaripteridae*
□amphorhynchus longicaudus	Solnhofen	p86	152	78	0.4	p84	□amphorhynch	*ramphorhynch*
□amphorhynchus longiceps	□ □urope	p86	155	75	1.75	p84	□amphorhynch	*ramphorhynch*
Pterodactylus longicollum	□ □urope	p86	151	79	1.45	p88	Pterodacyloidea	*pterodactylidae*
Pterodactylus grandis	□ □urope	p86	149	81	2.5	p88	Pterodacyloidea	*pterodactylidae*
Pterodacyloidea indet	Switzerland	□ □□99	151	79	3.5	□ □□99	Pterodacyloidea	*indeterm*
Scaphognathus crassirostris	Solnhofen	p90	151	79	0.9	p93	□amphorhynch	*scaphogn*
Anurognathus ammoni	Solnhofen	p90	148	82	0.5	p93	□amphorhynch	*anurognathidae*
Germanodactylus cristatus	Solnhofen	p95	149	81	0.98	p95	Pterodacyloidea	*germanodactylidae*
Germanodactylus ramphastinus	Solnhofen	p95	145	85	1.08	p95	Pterodacyloidea	*germanodactylidae*

Gallodactylus suevicus	Solnhofen	p95	151	79	1.35	p97	Pterodacyloidea	*gallodactylidae*			
Ctenochasma roemeri	Fr/Germ	p98	150	80	1.2	p98	Pterodacyloidea	*ctenochasmatidae*			
Gnathosaurus subulatus	Solnhofen	p98	150	80	1.7	p100	Pterodacyloidea	*ctenochasmatidae*			
□uanhepteris quingyangensis	China	p98	150	80	2.5	p105	Pterodacyloidea	*ctenochasmatidae*			
Gen et sp indet	north France	□□01	148	82	2.5	□□01	Pterodacyloidea	*pterodactylidae*			
□ermodactylus montanus	Colorado	p102	144	86	1	p105	Pterodacyloidea	*undetermined*			
Comodactylus ostromi	Colorado	p102	145	85	2.5	p105	□amphorhynch	undetermined			
□arpactognathus gentryii	□ yoming	□□□46	153	77	2.5	□□□51	□amphorhynch	*scaphogn*			
□atrachognathus volans	□azakhstan	p102	150	80	0.75	□□□423	□amphorhynch	*anurognathidae*			
Sordes pilosus	□azakhstan	□□□51	155	75	0.7	□□□51	□amphorhynch	*scaphogn*			
□rnithocheirus sp	France	p117	133	97	3.7	p117	Pterodacyloidea	*ornithocheiridae*			
Criorhynchus simus	□ngland	p110	130	100	5	p111	Pterodacyloidea	*criorhychidae*		& other spp	
Ornithodesmus latidens	Isle of Wight	p114	120	110	5	p115	Pterodacyloidea	*ornithodesmidae*			
Dsungaripterus weii	China	p118	140	90	3.25	p119	Pterodacyloidea	*dsungaripteridae*			
Phobetor parvus	China	p118	140	90	1.5	p119	Pterodacyloidea	*dsungaripteridae*		Dsungaripterus parvus	
Toxaster caste	France	Buff04	133	97	7	Buff04	Pterodacyloidea	*ornithocheiridae*			
Isle of Wight specimen	England	MFGG96	127	103	8	MFGG96	Pterodacyloidea	*undetermined*			
Noripterus complicidens	China	p118	120	110	2	p120	Pterodacyloidea	*dsungaripteridae*			
Arthurdactylus conan-doylei	Brazil	FM94	120	110	4.6	FM94	Pterodacyloidea	*ornithocheiridae*			
Jeholopterus ningchengensis	Jehol	J_104	124	106	0.9	J_100	Ramphorhynch	*anurognathidae*			
Eosipterus yangi	Jehol	J_22	122	108	1.2	J_100	Pterodacyloidea	*pterodacylidae*			
Haopterus gracilis	Jehol	J_22	122	108	1.35	J_100	Pterodacyloidea			subadult?	
Sinopterus dongi	Jehol	J_106	115	115	1.2	J_104	Pterodacyloidea	*tapejaridae*	young	indiv	Well04
Chaoyangopterus zhangi	Jehol	J_106	115	115	1.85	J_104	Pterodacyloidea	*nyctosauridae*			
Liaoningopterus gui	Jehol	J_106	115	115	5	J_104	Pterodacyloidea	*anhangueridae*			
Anhanguera blittersdorffi	Brazil	p128	115	115	4	p129	Pterodacyloidea	*anhangueridae*			
Anhanguera santanae	Brazil	p128	115	115	4.15	p125	Pterodacyloidea	*anhangueridae*			
Anhanguera piscator	Brazil	KT2000	105	125	5	Kell2004a	Pterodacyloidea	*anhangueridae*		very young indiv	
Cearadactylus atrox	Brazil	p128	115	115	5.5	p129	Pterodacyloidea	*ctenochasmatidae*		Unwin02	
Cearadactylus' ligabuei	Brazil	KT2000	105	125	5.5	KT2000	Pterodacyloidea	*ctenochasmatidae*			
Tropeognathus mesembrinis	Brazil	p128	115	115	6.2	p129	Pterodacyloidea	*criorhychidae*			
Tropeognathus robustus	Brazil	p128	117	113	5.5	p129	Pterodacyloidea	*criorhychidae*		in proportion	
Santanadactylus araripensis	Brazil	p128	115	115	3	p124	Pterodacyloidea	*ornithocheiridae*			
Tapejara wellnhoferi	Brazil	p128	115	115	1.5	Kell04	Pterodacyloidea	*tapejaridae*			

Tupuxuara longicristatus	Brazil	p130	115	115	2.8	Kell04	Pterodacyloidea	*tapejaridae*	toothless
Tupuxuara leonardii	Brazil	Well04	116	114	4.5	Kell04	Pterodacyloidea	*tapejaridae*	
Araripesaurus castilboi	Brazil	p124	115	115	2.2	p124	Pterodacyloidea		
Araripedactylus debmi	Brazil	p124	115	115	4.8	p124	Pterodacyloidea	indeterm	
ornithocheirid'	Mongolia	U&B426	112	118	5.5	U&B426	Pterodacyloidea	*ornithocheiridae*	
Bennett 1989	Argentina	p134	115	115	4	p134	Pterodacyloidea		
Coloborhynchus spielbergi	Brazil	eld03	108	122	5.9	eld03	Pterodacyloidea	*anhangueridae*	
Thalassodromeus sethi	Brazil	KC02	110	120	4.4	KC02	Pterodacyloidea	*tapejaridae*	
Pterodaustro guinazui	Argentina	Chiappe04	100	130	3	Chiappe04	Pterodacyloidea	*pterodaustridae*	filter feeder
Ornithostoma orientalis	England	p110	97	133	2.5	p111	Pterodacyloidea	*pteranodontidae*	
cf. Anhanguera	Russia	U&B425	95	135	3.5	U&B425	Pterodacyloidea	*ornithocheiridae*	
cf. Anhanguera	Russia	U&B425	95	135	3.5	U&B425	Pterodacyloidea	*ornithocheiridae*	
Ornithocheirus sp	aire	p121	90	140	4.5	p121			
Pteranodon sternbergi	Kansas	p134	85	145	9	p139	Pterodacyloidea	*pteranodontidae*	toothless
Pteranodon ingens	Kansas	p134	85	145	7	p139	Pterodacyloidea	*pteranodontidae*	toothless
Pteranodon sp	Japan	p145	83	147	6	UBLML	Pterodacyloidea	*azhdarchid*	
Haenamichnus footprint	Korea	KHCLC03	81	149	10	KHCLC03	Pterodacyloidea	*footprints!*	
Ornithocheirus bunzeli	Austria	p116	80	150	1.7	117	Pterodacyloidea	*ornithocheiridae*	may be Azdharch
Azdharcho lancicollis	Uzbekistan	U&B426	87	143	5	U&B426	Pterodacyloidea	*azhdarchid*	
Azdharchid	Russia	U&B427	85	145	3.5	U&B427	Pterodacyloidea	*azhdarchid*	
Wairoa specimen	N	WM88	76	154	3.5	WM88	Pterodacyloidea	indeterm	
Ornithocheiridae indet	Russia	A 04	75	155	5	A 04	Pterodacyloidea	*ornithocheiridae*	
Nyctosaurus gracilis	Kansas	p138	80	150	2.9	p139	Pterodacyloidea	*nyctosauridae*	
Nyctosaurus lamegoi	Brazil	p145	75	155	3.5	p122	Pterodacyloidea	*nyctosauridae*	
uetzalcoatlus northropi	Texas	p142	66	164	11	p143	Pterodacyloidea	*azhdarchid*	
uetzalcoatlus sp	Texas	p142	66	164	5.5	p142	Pterodacyloidea	*azhdarchid*	p142
Phosphatodraco mauritanicus	Morocco	B_M80	68	162	5	B_M87	Pterodacyloidea	*azhdarchid*	
Arambourgiania philadelphiae	Jordan	MSFK96	69	161	12	FM96	Pterodacyloidea	*azhdarchid*	formerly Titanopteryx B_M86
hejiangopterus linhaiensis	China	U_J97	81	149	3.5	B_M87	Pterodacyloidea	*azhdarchid*	
cf. uetzalcoatlus	France	BLLB97	67	163	9	BLLB97	Pterodacyloidea	*azhdarchid*	
Azhdarchidae indet.	Spain	Company99	69	161	5.5	Company99	Pterodacyloidea	*azhdarchid*	
Hatzegopteryx thambema	Romania	Buffet02	68	162	12	Buffet02	Pterodacyloidea	*azhdarchid*	
Montplaisir pterosaur	France	Buffet96	67	163	5	Buffet96	Pterodacyloidea	*azhdarchid*	
largest ever'	Spain	Company01	67	163	12.2	Company01	Pterodacyloidea	*azhdarchid*	

□ata □r □gure 3, □anel □, pter□aur □□er□ty t□r□ug□t□me□□e □ata □n □gure 3□ □a□ta□en □re□tly r□m □gure 21 □□□n□□n □2003□, □ut t□at □gure □□re□ra□n an□ □□□□n □ere □□r □□n□en□en□e□



□e □ll□□ng re□eren□e □□ere u□e □□r t□e □ata □n □gure 3 □t□er re□eren□e □□ere □n□ulte□, □ut n□t u□e□ □□r t□e pr□mary □ata□

□er□an□, □□□an □□□□ar □□□□2004 □n t□e □□urren□e □□a g□ant □ly□ng rept□le □□ter□aur□a□n t□e term□nal □ate □reta□e□u□□□t□e l□□er □□lga reg□n□ale□nt □□3□, 66□671□

□ennet, S□□1 □6 □ear□la□e □□pter□aur □r□m t□e S□ln□□en l□me t□ne □□□ermany□ta□n□m□□an□ □y□temat□□mpl□at□n□, □□□ert□al□16, 432□444□

□erggren □ □, □ent □□, S□□□er □□□□□, □u□y □ □□ 1□□5□□ re□□e□□en□□□□ge□□r□n□l□gy an□ □r□n□trat□grap□y□n □e□□r□n□l□gy, □me □ale□, an□□l□al Strat□grap□□□□rrelat□n □e□ □erggren □ □, □ent □ □, □u□y □ □□, □ar□en□1 □□12□212 S□□□ S□□ety □r Se□imentary □e□l□gy□□□la□□ma□

□u□etaut, □□2004□□ g□ant pter□aur □r□m t□e □□□er □reta□e□u□□□t□e ea□tern □ar□□a□n□□ull S□□□ge□1□ □ran□e 175, 573□577□

□u□etaut, □□, □□□□lar□e an□□□□e □□eu□□1□□6□□ term□nal □reta□e□u pter□aur □r□m t□e □□r□ere□ □□ut□ern □ran□e an□t□e pr□lem □□□□n□aur e t□n t□□n□□ull S□□□□e□1□□ran□e 167, 753□75□□

□u□etaut, □□, □□□r□g□re□u an□□□□□□□□2002□□ ne□ g□ant pter□aur □t□a r□□u□t □□ull r□m t□e late□t □reta□e□u□□□□man□a□□atur□□□en□□a□ten□□□, 1□0□1□4□

□u□etaut, □□, □□□r□g□re□u an□□□□□□□□2003□□ant a□□ar□□□□pter□aur□r□m t□e term□nal □reta□e□u□□□ □ran□yl□an□a □□e□tern □□man□a□*In:* □u□etaut, □□an□□□□□ □□ a□n□2003 □□□lut□□n an□□alae□□□□1□gy □□□ter□aur□□□e□1□S□□□□□□n□□n Spe□□□u□1□217, □1□104□

□u□etaut, □□, □□□□epage an□□□□epage□1□□□□□ ne□ pter□a□tyl□□□r□m t□e □mmer□g□an □□t□e □ap □e la □□□e □□□rman□y, □ran□e□□□e□1□□ ag□135, 71□722□

□u□etaut, □□an□□□□□ □□ a□n□2003 □□□lut□□n an□□alae□□□□1□gy □□□ter□aur□□□e□1□S□□□□□□n□□Spe□□ □u□1□217□

□u□etaut, □□, □□□aurent, □□□e □□eu□an□□ □□□1□□tte□1□□7□□ term□nal □reta□e□u□g□ant pter□aur □r□m t□e □ren□□□yrenee□□□□e□1□□ ag□134,553□556□

□arpenter, □□, □□□□n□□n, □□□□l□□ar□, □□□ le□am□□□□□ le□2003 □ ne□ □□ap□gnat□□ne pter□aur □r□m t□e □pper □ura□□□□ □rr□□□n □□rmat□□n □□□ y□m□ng, □S□□*In:* □u□etaut, □□an□□□□□ □□ a□n□2003 □□□lut□□n an□□alae□□□□1□gy □□□ter□aur□□□e□1□S□□□□□□n□□n Spe□□□u□1□217, 45□54□

□□ang, □ □m, □e□2004□□e□e□1□□□ta□t□e emergen□e □□□eat□ere□□□n□aur, □ea□e□□r□□an□□1□□er□ng plant□S□ang□a S□ent□□□an□□e□n□al □u□l□□er, S□ang□a□□

□□appe, □□□ □, □□□□r□rn□u, □□□rellet□□nner, an□□□□□ar□la□2004□□rgent□na un□at□e□ pter□aur □□□□l□□ature 432, 571□572□

□□mpany, □, □□□□u□□□□mena□a an□□□□□Su□er□□la□1□□□□□ l□ng□ne□e□pter□aur □ter□a□tyl□□ea, □□□ar□□□ae□r□m t□e □pper □reta□e□u□□□□alen□a, Spa□n□□e□1□g□e en □ □n□u□ 7□, 31□333□

□□mpany, □, □□□ □□n□□n, □□□□u□□□□mena□a an□□□□ere□a Su□er□□la□2001□□ g□ant a□□ar□□□□pter□aur □r□m t□e late□t □reta□e□u□□□□alen□a, Spa□n – t□e large□t □ly□ng □reature e□er□□□ert□al□21, 41□□ 42□□

□alla □e□□a, □□□ □2003 □e□ m□rp□1□g□al □□□er□at□n□□n □ra□□□pter□aur□*In:* □u□etaut, □□an□ □□□□ □□ a□n□e□□□□□□lut□□n an□□alae□□□□1□gy □□□ter□aur□□□e□1□S□□□□□□n□□n Spe□□□u□1□217, 23□ 44□

□alla □e□□a, □□□ □, □□□r□u□n□, an□□□□ □□□ellner□2001□□e □r□t pter□aur □r□m t□e □en□man□an □□ate □reta□e□u□□ager□t□tten □□□e□an□n□□reta□□□e□22, 21□225□

□alla □e□□a, □□□ □, □□□□ 1□, □□, □□□□p□, □ an□□□□e□tner□2002□□ re□te□r□amp□□r□yn□□□□pter□aur □r□m t□e □ate □ra□□□□□□u□tra□□□ert□al□22, 1□6□1□□□

□rey, □□an□□□□ □□ art□ll□1□□6□□ reappra□al □□□ram□urg□an□a □□ter□aur□a, □ter□a□tyl□□ea□□□ne □□t□e □□rl□□□large□t □ly□ng an□mal□□□□□□□e□1□□alaent□1□□□□1□□, 221□247

□rey, □□an□□□□ □□ art□ll□1□□4□□ ne□ pter□aur □r□m t□e □rat□□□rmat□n □□□□er □reta□e□u□, □pt□an□□□ □ra□ll□□□□□□□e□1□□alaent□1□□□□1□□4, 37□412□

□rey, □□, □□□ □□ art□ll, an□□ □□□□u□y□2003□□ ne□ □re□te□□rn□t□□□□e□r□□r□m t□e □□□er □reta□e□u□□□ n□t□ea□tern □ra□ll an□t□e unu□ual □eat□□an unu□ual pter□aur□*In:* □u□etaut, □□an□□□□□ □□ a□n□ 2003 □□□lut□□n an□□alae□□□□1□gy □□□ter□aur□□□e□1□S□□□□□□n□□n Spe□□□u□1□217, 55□63□

rey, , art ll, an uy 2003 ne pe e tape ar pter aur t t ue ea re t *In:* u etaut, an a n 2003 lut n an alae l gy ter aur e l S n n Spe u l 217, 65 72
a par n, , ernan e an e la uente 2004 ne pter aur r m t e ura u a alae nt l gy 47, l 27
ang, , u, , ley, , n n, an r g t, 2002 e pter aur tra tera n ae r m t e ate reta e u angr rmat n, ut e tern rea e l ag 13, 421 435
en n, , S u n, S ate y an a an 2001 m nut e pter aur ter aur a u m rp nt ae r m t e reenlan ra ull u mp l 156, 151 170
u e, S 2004 e r pt n t e ull a *n* ter aur a r m t e late t ura ea tern ran e, t a ta n m re n ur pean t n an ter a tyl ea ert ale nt 24, 542 554
ellner, 2004 e n rmat n n t e ape ar ae ter aur a, ter atyl ea an u n t e relat n p t la e meg n ana 41, 521 534
ellner, 2004a e an le tru ture t pter a tyl pter aur r m t e Santana rmat n er reta e u , ra l ull m u at t 2 5, 25 35
ellner, an amp 2002 e un t n t e ran al re t an a a un ue pter aur r m t e arly reta e u ra l S en e 2 7, 3 3 2
ellner, an ang t n l 6 ran al rema n uet al atlu ter aur a, ar ae r m ate reta e u e ment g en at nal ar , e a ert al 16, 222 231
ellner, an y 2003 ter aur teran nt a, ter a tyl ea apul ra r m t e arly reta e u ene uela *In:* u etaut, an a n 2003 lut n an alae l gy ter aur e l S n n Spe u l 217, 73 77
ellner, an m a 2000 e r pt n a ne pe e n anguer ae ter a tyl ea t mment n pter aur auna r m t e Santana rmat n pt an l an , n rt ea ter ra l at nal S en e u eum, y
m, , u, S e ng, ley an ang 2003 ge pter aur an e te r tra a ate t n aur tpr nt r m S ut rea e lan r 12, 125 131
n ll, an u eur 2001 large ura pter a tyl n n rt ern ran e an a re e t e ren pter aur re r ull S e l ran e 172, 447 454
ley, an a n rt e tra re r e r an pter aur an n l g al an pale e l g al per pe t e n,
ley, , S ang, at u a a, lem ng an S m l 2 e tra re r e r an pter aur e en e an mpl at n l ran S n n 336, 113 134
a , , at e an Sun 2004 ne ungar pter pter aur r m t e l er reta e u t e ut ern unggar a n, n rt e t na ret e 25, 625 634
art ll, , rey, reen an reen l 6 ant pter aur r m t e er reta e u t e le g t, e l alae nt 11, 672 6 3
art ll, , rey, Sa a a, an ury l ery t e l type t e g ant pter aur *n* 1 5, an t e tatu *n* an 207, 57 76 eue a r u ur e l g e un alae nt l gy
eyer, an unt l e r t pter aur r m t e ate ura S t erlan e en e r t e large t ura ly ng mammal ry t 2, 111 116
ere a Super la, , ar et, S u e, ar ene an mag a 2003 ne a ar pter aur r m t e ate reta e u p p ate r *In:* u etaut, an a n 2003 lut n an alae l gy ter aur e l S n n Spe u l 217, pp 7 0
Sull an, l 7 rea e ment rept l an er ty a r t e reta e u ert ary un ary ntr ut n n S en e, 3 1, 1 26 at t u eum ngele unty
n n, l 7 ter aur e t n t n nature an au e em S ge l ran e, S 150, 105 111
n n, l 6 e l re r m le ura pter aur n ral e e nt nental ura u rt ern r na ull 60, 2 l 304
n n, 2003 *n* an t e early t ry pter aur u S at a 22, 3 46

□n□n, □□□ □2003□□n t□e p□yl□geny an□e□□lut□nary □□t□ry □□pter□aur□*In*, □u□etaut, □□an□ □□□□ □ □a□n□2003 □□□lut□n an□ □alae□□□□l□gy □□□ter□aur□□□e□l□S□□□□□n□□n Spe□□u□1□217, pp 13□□ 1□0□

□n□n, □□□ □an□ □□□□ □□a□ur□na 2000□□ter□aur□r□m □u□□a, □ □□□le □□a an□□ □ng□l□a□n □ □□□ ent□n, □ □□□ S□□□□□n, □□□ □r□n an□ □□□□□ur□□□□n □e□□□□e □ge □□□n□aur□n □u□□a an□ □□ng□l□a□ □am□r□ge □n□er□ty □re□□□pp 420□433□

□n□n, □□□ □, □□□□ □a□ur□na, □ □□ □□□□□ley, □ □□ ana□e an□□□□u□l□7□□ter□aur□r□m □□a□ ale□nt□ S□□□□□rea Spe□□u□1□□□2□43□65□

□n□n, □□□ □an□□□□u□l□7□□n □□□□□*n*□□□□□□□an□t□e relat□n□□p□□□pter□a□tyl□□pter□aur□□□□t□ □□1□12, 1□□210□

□el□me□er, □□□2003□□e□r□pt□n □□□□l□□□r□yn□u□□p□el□erg□p□□□□□□ter□a□tyl□□□ea□r□m t□e □l□an□□□□er□reta□e□u□□□□□ra□l□S□r□pta□e□l□g□a 125, 35□13□

□ang, □□□□an□□□□□□□u□2004□□ter□aur□□p□□107, □n□□e□e□1□□□ta□t□e emergen□e□□□eat□ere□ □n□aur□, □ea□e□□r□□an□□l□□er□ng plant□□□□ang, □□m, □e□□S□ang□a□S□ent□□□□an□□e□n□al □u□1□□□er□, S□ang□a□□

□ang, □□□□an□□□□□□□u□2003□□ ne□ pter□aur□□ter□□a□tyl□□□ea, □ape□ar□□ae□r□m t□e early □reta□e□u□□□u□□tang□□rmat□n□□□e□tern□□a□n□ng, □□na an□□t□□mpl□at□n□□□r□□□□trat□grap□y□ □□n□S□□□□ull□4□, 16□23□

□ elln□□er, □□1□□1□□□e□llu□trate□□n□y□1□pe□a□□□ter□aur□□Salaman□er□□□□□, □□n□□n□

□ elln□□er, □□2003□□ □ate□r□a□□□□pter□aur□r□m t□e n□rt□ern□al□are□u□alp□□□yr□1, □u□tr□a□□*In:* □u□etaut, □□an□□□□□ □□ a□n□2003□□□lut□n an□□alae□□□□l□gy □□□ter□aur□□□e□l□S□□□ □n□□n Spe□□u□1□217, 5□22□

□ □□en, □□an□□□□□□ □lnar□1□□□□□□r□t pter□aur□r□m □e□ □ealan□□□l□er□nga 12, 53□5□□

[1] □en□n□□□ □1□71□□e□ □ealan□ □en□□□□plan□t□n□ □ram□n□era□□ □ □e□Sur□ □ull 42□1□27□□

[2] Str□ng □□ □1□□4□□reta□e□u□□ert□ary □□un□ary, □ □□□□a□para □□er Se□t□□n, □□rt□ □anter□ury, □e□ □ealan□□□ □ □□e□l □e□p□y□27□231□234□

[3] □□ll□□□, Str□ng □□ □2003□□□□trat□grap□□re□e□ □□t□e □reta□e□u□/□ert□ary □□un□ary tran□□t□□n, m□□□□a□para □□er □e□t□□n, □□rt□ □anter□ury, □e□ □ealan□□□ □ □□e□l □e□p□y□46□ 243□253□

[4] □a□□a □, □a□ne □□, □□ll□□□□□2001□□n□□□at□□n □□gl□□al □e□□re□tat□□n at t□e □reta□e□u□□ert□ary □□un□ry □y □e□ □ealan□ □ern □p□□e□ S□□en□e 2□4□1700□1702□

[5] □□□per □□□ et al□2004□□□e □e□ □ealan□ge□l□g□□al t□me□□ale□ □n□t□tute □□□e□l□g□□al an□ □u□lear S□□en□e □m□n□grap□22□1□2□4□

[6] □□□□er □□, □□□ma□□ □, □eyer □, □□□ar□□S□, □□□□□S, □□lla□lan□a □□, □ □1□□n□□ □1□□□□□□ynam□□□ □□m□t□□□□n□r□al □□□□e□□lut□□n □n an□mal□□ampl□□□□at□□n an□ □e□uen□□ng □□t□ □□n□er□e□pr□mer□□r□ □atl □□a□S□□□S□ □6□61□6□6200□

[7] □□□per □, □alue□a□□□□□, □n□er□□n S, □am□aut □, □u□t□n □, □ar□□ □2001□□□mplete m□t□□□□n□r□al gen□me □e□uen□e□□□t□ □e□t□n□t m□a□□lar□y rat□te e□□lut□□n □ature 40□704–707□

□S□ren□en □ □, □□t □□, □□m□□e□□□□, □ur□□, □ □n□ell □□□1□□□□□□r□mer□□□r a □□□□□a□e□appr□a□□t□ m□t□□□□n□r□al gen□me □e□uen□□ng □n □□r□□an□ □t□er □erte□rate□□□ □1□□□yl□g□□□1□ 12□105–114□

□□arr□□□n, □ □, □□ □ □□ena□□an, □ □□□□ll□p□, □ □ Sla□□, □ □□□per an□ □ □enny 2004 □□ur ne□ a□□an m□t□□□□n□r□al gen□me□ □elp get t□ □a□□e□□lut□□nary □ue□t□□n□ □n t□e □ate □reta□e□u□□□ □1 □□□l □□□1 21□□74□□3□

[10] □□n □□□, □ □□ena□□an □□, □□re □□, □□□ll□p□□ □, □ta □, □en□y □ □, □enny □ □2002□□□ur ne□ m□t□□□□n□r□al gen□me□an□ t□e □n□rea□e□ □ta□□l□ty □□e□□lut□□nary tree□ □□mammal□□r□m □mpr□□e□ ta□□n □ampl□ng□□ □1 □□□l □□□1 1□□2060□2070□

[11] □□□ll□p□□ □, □enny □ □2003□□□e r□□t □□t□e mammal□an tree □n□erre□ □r□m □□ □□e□m□t□□□□n□r□al gen□me□□ □ □1 □□yl □□□1 2□□171 1□5□

[12] □el□u□□, □□□ll□p□□ □, □enny □ □2003□□□□mment □n □□e□ap□□ □r□g□n□□m□n□p□ylet□□□r parap□ylet□□□□ S□□en□e 301□14□2□□

[13] □□□ll□p□□ □, □el□u□□, □enny □ □2004□□en□me□□□ale p□yl□geny□□ampl□ng an□ □y□temat□□err□r□are □□t□mp□rtant□□ □1 □□□l □□□1 21□1455□145□□

[14] □ang □, □□□er □□ □2003□□□□mpar□□□n □□1□□el□□□□□an□ □aye□□an met□□□□□□r e□t□mat□ng □□□ergen□e t□me□□u□□ng mult□ple gene l□□□an□ □al□□rat□□n p□□nt□, □□t□appl□□at□□n t□a ra□□at□□n □□□□ute□l□□□ng m□u□e lemur □pe□□e□□Sy□t □□□1 52□705□716□

[15] □□□rne □□, □ □□□n□ □ □2002□□□□ergen□e t□me an□e□□lut□□nary rate e□t□mat□□n □□t□mult□l□□u□□ata□ Sy□t □□□1 51□6□□702 □ttp□//□tatgen□n□□u□e□u/t□□rne/mult□□□□t□me□□tml□□

[16] □ang, □ □□□ □□a pr□gram pa□□age □□r p□yl□genet□□analy□□□□y ma□□mum l□□el□□□ □ □ □ □□ □□□S 13□ 555□556 □1□ □7□□ttp□//a□□a□□u□□gene□□u□1a□□u□/□□□t□are/paml□□tml□□

Chapter 4.

Slack, K.E□, el u, □, □ ena an, □□□, rna n, □ an enny, □□2007□□e□l ng t e r□ t t e a an m t gen m tree y rea ng up l ng ran e□□ □□□□□□□□□□□□**42**□l□13□ □□at ugu t 2011 t□□paper a een te□35 t me n t e □ e□□S en e □n lu ng 3 tat n□□ t err r n pag nat n, et□□

□t e t tle □t e paper ugge t, t e □□□□e □□ta a □ tr□□manu r pt a me at mpr□□ng t e ta l ty □t e a an m t□□□n r al mt tree y re u ng t e e e t □□ l ngun an e□ l neage□□□e ta a □r □□□□ □mplete mt gen me□□ere e uen e are t e u tral an ru□□□ tur ey □□□□□□□□□□□□□□□□□, t e uper lyre r□ □□ *n* □□□ *n* □□□□□□□□ *n* □□□□, te u u□ ly at er □ *n* □□□□□□□□□□□□□□□□□□□□□, t e ut ern la□□□a e gull □□□□□□□□□□□ *n* □□ *n* □□□ a□ t e tur ey ulture □□□□□□□□□□□□□□□□

n r er t□□l ly te t t e all□□rme /□n er□□rme gr up ng, ne urt er early gall□□rm ta□n a re u re□□□lrea y a a la le ere t e magp e g□□e □eep n t e □n er□□rme□ l neage, ee ppen□□□3□, t m re re ent an er□□rm□□u□an g□□e, an t□□m re re ent gall□□rm□□□□□□en an□ ua l□□□□u a megap e t e ru□□□tur ey, □□□□□□a eep gall□□ m, a ele te□□

e au e □t e r g er rate □m le ular e lut n, t e pa ern ne a e een e pe ally pr lemat t pla e n t e a an tree □ee r e ample, gure 4 an 5 □□□apter 1□□□e lyre r□ t ug t t e ne □□e eepe t ran ng □□□ne pa ern ne□□□e ly at er □a eeply ran ng u□□□□□ne pa ern ne, an t u rea up t e l ng ran t t e ngle u□□□□ne, t e r a□□ll □□□□□□ *n* □□□□□□□□□□□□□□e r leman, a e ealan ren, □□n t e eepe t pa ern ne ran , an a een e uen e pre u ly ee ppen□□3□□erall, a ng t e lyre r an ly at er e uen e nt t e mt ata et meant all □□e t e eepe t ran ng pa ern ne l neage ere repre ente□

e gull a mp rtant because it allowed a direct test of Feduccia's hypothesis (1995; 1999) that the oldest modern birds were 'transitional shorebirds' p□□□ly □ara r□□□rme□□n a□t n, y a ng t e tur ey ulture t t e mt ata et a yp t e□□e relat n p et een t an t e t r□ ul e te te, p entally rea ng up a p□□□□le penguin/t r gr up ng een n earl er analy e□□ee □apter 2 an 3□□

□a re p n□□le r e uen ng t e ru□□tur ey, t e lyre r, an t e gull□r re er□□ el u, n□ t e n er ty □□ ntpell er n ut ern ran e, e uen e t e ly at er an t e tur ey ulture□r□□ □ ena an a al ay m t elp ul n t e la rat ry, an al ay a□□te□ t n nal ng an a em ng t e □mplete mt gen me□□□a pr mar ly re p n□□le r t e analy an a t ely part pate n t e r t ng t e manu r pt Ulfur Arnason and David Penny were involved at all stages in project design. All authors contributed to the final manuscript.

Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 42 (2007) 1–13
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Resolving the root of the avian mitogenomic tree by breaking up long branches

Kerryn E. Slack [a,b], Frédéric Delsuc [a,1], Patricia A. Mclenachan [a], Ulfur Arnason [b], David Penny [a,*]

[a] *Allan Wilson Center for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand*
[b] *Department of Genetics, Cell and Organism Biology, University of Lund, Lund, Sweden*

Received 1 March 2006; revised 28 May 2006; accepted 1 June 2006
Available online 7 June 2006

Abstract

Incomplete taxon sampling has been a major problem in resolving the early divergences in birds. Five new mitochondrial genomes are reported here (brush-turkey, lyrebird, suboscine flycatcher, turkey vulture, and a gull) and three break up long branches that tended to attract the distant reptilian outgroup. These long branches were to galliforms, and to oscine and suboscine passeriformes. Breaking these long branches leaves the root, as inferred by maximum likelihood and Bayesian phylogenetic analyses, between paleognaths and neognaths. This means that morphological, nuclear, and mitochondrial data are now in agreement on the position of the root of the avian tree and we can, move on to other questions. An overview is then given of the deepest divisions in the mitogenomic tree inferred from complete mitochondrial genomes. The strict monophyly of both the galloanseres and the passerines is strongly supported, leaving the deep six-way split within Neoaves as the next major question for which resolution is still lacking. Incomplete taxon sampling was also a problem for Neoaves, and although some resolution is now available there are still problems because current phylogenetic methods still fail to account for real features of DNA sequence evolution.

Keywords: Avian evolution; Long branches; Galloanseres; Passeriformes; Neoaves; Cracrafti

1. Introduction

After a period of relative quiescence following the publication of Sibley and Ahlquist's DNA/DNA hybridization tapestry (1990) the ever decreasing cost of DNA sequences has led to a rebirth of deep-level avian systematics. As a part of this effort to acquire new data aiming at resolving the phylogeny of avian orders, complete mitochondrial genomes had been sequenced for an increasing number of taxa (for example, Mindell et al., 1999; Paton et al., 2002; Slack et al., 2003; Harrison et al., 2004). The amount of sequence data accumulated so far permits the evaluation of a range of prior hypotheses primarily proposed on morphological, paleontological, and biogeographical grounds (for example, Cracraft, 2001; Cracraft et al., 2004). In the case of placental mammals, as expected on theoretical grounds, increasing the number of taxa led to an excellent agreement between nuclear and mitochondrial datasets (Lin et al., 2002; Reyes et al., 2004). Such agreement between different datasets is essential for corroboration (Penny et al., 1982) because model misspecification (Buckley, 2002) can be hard to detect. Systematic biases can lead to 100% bootstrap support for conflicting trees (Phillips et al., 2004), and even maximum likelihood methods can become inconsistent when the model is not specified accurately (Chang, 1996).

An early problem was that mitogenomic-based studies of avian phylogenetic relationships appeared to place the

* Corresponding author. Fax: +64 6 350 5626.
E-mail address: D.Penny@massey.ac.nz (D. Penny).
[1] Present address: Laboratoire de Paléontologie, Phylogénie et Paléobiologie, Institut des Sciences de l'Evolution UMR5554-CNRS, Université Montpellier II, Montpellier, France.

1055-7903/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2006.06.002

reptilian outgroup within Passeriformes (Härlid and Arnason, 1999; Mindell et al., 1999). Such a rooting was at odds with data from morphology (Sibley and Ahlquist, 1990; Cracraft, 2001; Livezey and Zusi, 2001; Mayr and Clarke, 2003), nuclear DNA (Garcia-Moreno and Mindell, 2000), and genomic DNA strings (Edwards et al., 2002). These all supported a major division between paleognaths (tinamous and ratites) and neognaths (all other birds including the species-rich Passeriformes). The passerine rooting of the mitochondrial tree has been suggested to be the result of a possible long-branch attraction effect between the distant reptilian outgroup and the fast-evolving passerines (Braun and Kimball, 2002; Slack et al., 2003). Indeed, in the unrooted avian tree the passerines were grouped together on the tree, but could become paraphyletic, or even diphyletic when the outgroup was added (Slack et al., 2003). Such disruption of the ingroup by a distant outgroup is known from simulation studies (Holland et al., 2003) and here the ingroup tree by itself was more likely to be correct. A similar ingroup disruption was also found with mammals (Lin et al., 2002). In general, distant outgroups are hard to place correctly into the ingroup tree (Holland et al., 2003). Taxon sampling from Passeriformes has been sparse with only one suboscine (broadbill) and two oscines (rook and indigobird) included at that time (Braun and Kimball, 2002). Base composition bias might also have played a role since phylogenetic reconstructions under a non-homogeneous model dealing with base composition heterogeneity (Paton et al., 2002) and RY-coding analyses (Braun and Kimball, 2002) supported the classical rooting of the avian tree between paleognaths and neognaths. However, the position of the root of the avian tree is still difficult to resolve clearly with complete mitochondrial data and must be tested by the inclusion of more basal neognaths (Galloanserae) as well as Passeriformes that appear especially fast evolving (Harrison et al., 2004). Ideally, it is preferable for a result to be found by all good methods of phylogenetic analysis, not just by one specialized analysis. In difficult cases, it is desirable for both theoretical and practical reasons to first build an unrooted tree, and then to test the placement of the root secondarily.

The phylogeny of paleognaths, with a particular emphasis on both living and extinct ratites, has been well studied from the mitogenomic viewpoint (Härlid and Arnason, 1999; Härlid et al., 1999; Cooper et al., 2001; Haddrath and Baker, 2001). Their origins and biogeographic evolution in relation to the tectonic fragmentation of Gondwana are now relatively well understood (Cooper et al., 2001; Haddrath and Baker, 2001), although some relationships within ratites and between deep ratites and tinamous are still uncertain. Turning to neognaths, they are usually divided into Galloanserae (chicken, ducks, and allies) and Neoaves (all other neognath birds). This subdivision is now well supported from molecular data by both mitochondrial (Paton et al., 2002; Harrison et al., 2004) and nuclear sequences (Van Tuinen et al., 2000; Garcia-Moreno et al., 2003; Chubb, 2004). We have recently added the magpie goose (*Anseranas semipalmata*) a basal member of the water fowl lineage (Anseriformes) to the complete mitochondrial dataset, establishing an important calibration point for avian evolution (Harrison et al., 2004). In the present study, we include a basal member (Ericson et al., 2002b) of the chicken-related lineage (Galliformes) by sequencing the complete mitochondrial genome of the Australian brush-turkey (*Alectura lathami*, Megapodiidae). This is expected to break the relatively long ancestral lineage leading to chicken and quail, and to test the relationships within Galloanserae further.

Phylogenetic relationships among the six or more major groups of Neoaves that contain the vast majority of extant bird species remain elusive and are usually represented as a multifurcation (Cracraft, 2001; Cracraft et al., 2004). Within this unresolved phylogeny, Passeriformes are by far the most speciose group. The current taxonomy of oscines, based mainly on the results of DNA/DNA hybridizations (Sibley and Ahlquist, 1990), has recently been challenged by analyses of both nuclear and mitochondrial genes (Irestedt et al., 2001; Ericson et al., 2002a,b; Ericson and Johansson, 2003; Barker et al., 2004). Based on these new results a new taxonomy of major passerine groups has been proposed (Ericson et al., 2002b). Three major lineages have been distinguished: Acanthisittia (New Zealand wrens), suboscines (Eurylaimides or Old World suboscines and Tyrannides or New World suboscines), and oscines (Menurae and Euoscines). Oscines and suboscines are grouped together into Eupasseres to the exclusion of New Zealand wrens, the latter representing the most basal lineage of Passeriformes. The early emergence of New Zealand wrens within Passeriformes is consistent with a Gondwanan origin for the whole passerine group (Cracraft, 2001; Barker et al., 2002; Edwards and Boles, 2002; Ericson et al., 2002a; Fuchs et al., 2006).

The passerines have been represented in the mitogenomic database by only three taxa: the gray-headed broadbill (*Smithornis sharpei*; Eurylemidae) belonging to the Old World suboscines, and two Euoscines: the village indigobird (*Vidua chalybeata*; Estrildidae) and the rook (*Corvus frugilegus*; Corvidae). Under the new classification (Ericson et al., 2002b), rook and indigo bird are both in the Euoscines, leaving the Menurae unrepresented. A representative of New Zealand wrens, the rifleman (*Acanthisitta chloris*; Acanthisittidae) has been recently added to the complete mitochondrial dataset, but its position was still locally unstable within passerines (Harrison et al., 2004). By incorporating the new complete mitochondrial genomes of the superb lyrebird (*Menura novaehollandiae*; Menuridae) representing the second major group of oscines, and the fuscous flycatcher (*Cnemotriccus fuscatus*; Tyrannidae) a member of the New World suboscines, we now have a much more representative taxon sampling of Passeriformes. We expect from theory (Hendy and Penny, 1989) and from simulations (Holland et al., 2003) that these key taxa will stabilize the relationships within Passeriformes by subdividing the two long branches leading to oscines and suboscines, respectively. This is especially important in order to tackle issues such as the position of the rifleman within

passerines, and especially the central question of the position of the root of the avian tree.

Another interesting aspect of the current mitochondrial tree is the continuing difficulty (Haring et al., 2001; Slack et al., 2003; Harrison et al., 2004) in recovering the expected monophyly of birds of prey (Falconiformes)—currently represented by the buzzard (*Buteo buteo*; Accipitridae) and the falcon (*Falco peregrinus*; Falconidae). Falconiformes are part of a large group of Neoaves including seabirds, shorebirds, doves, cranes, rails, flamingos, penguins, loons, and grebes (Cracraft, 2001), a major grouping we informally call the Conglomerati (or Cracrafti). The buzzard and falcon share the same alternative mitochondrial gene order, but are fairly divergent from each other and seem to represent an early split in the raptor lineage (Haring et al., 2001). Thus we again have sparse taxon sampling. Previously, only partitioned-likelihood phylogenetic analyses (with RY-coding third codon positions of the mitochondrial proteins plus nucleotides from RNAs) appearing to support their monophyletic origin (Harrison et al., 2004).

As a first step we have sequenced the complete mitochondrial genome of the turkey vulture (*Cathartes aura*; Cathartinae). This provides an opportunity to test the Sibley and Ahlquist hypothesis (1990) that New World vultures (Cathartinae) are closer to storks (Ciconiinae) than to other birds of prey. It has long being realized that New World and Old World vultures may not share a most recent common ancestor, but may instead represent an example of morphological convergence resulting from adaptation to a scavenging way of life (see Sibley and Ahlquist, 1990 for an historical review). However, the question of whether New World vultures are more closely related to storks, as suggested by behavioral resemblances (Rea, 1983) and DNA/DNA hybridizations (Sibley and Ahlquist, 1990), or to birds of prey (including Old World vultures) is still ambiguous (Seibold and Helbig, 1995). Resolving this question certainly requires more molecular data (Helbig and Seibold, 1996).

The order Charadriiformes represents a very large and diversified group of shorebirds which also belongs to the Conglomerati/Cracrafti (Cracraft, 2001). Shorebirds are usually divided in three major clades: Charadrii (oystercatchers, thick-knees, sheathbills, plovers, and allies), Scolopaci (turnstones, sandpipers, and jacanas), and Lari (gulls, coursers, pratincoles, terns, skimmers, and skuas) (see Van Tuinen et al., 2004). DNA/DNA hybridization suggested an early emergence of Scolopaci with a sister-group relationship between Charadrii and Lari (Sibley and Ahlquist, 1990). The first two shorebird mitochondrial genomes to be sequenced were the blackish oystercatcher (*Haematopus ater*; Haematopodidae) and the ruddy turnstone (*Arenaria interpres*; Scolopacidae) (Paton et al., 2002). Since then, data from nuclear genes have explored the relationships among shorebirds families (Ericson et al., 2003; Paton et al., 2003; Thomas et al., 2004). These have challenged the DNA/DNA hybridization results by finding a closer relationship between Scolopaci and Lari with an early divergence of Charadrii. The sequencing of the southern black-backed gull (*Larus dominicanus*; Laridae) mitochondrial genome offers the opportunity to test the new nuclear-based hypothesis by adding a representative of the previously unsampled third major group of Charadriiformes (Lari, gulls).

Here we report mitochondrial genomes for five birds chosen to help clarify the deepest divisions, and test specific phylogenetic hypotheses, in the avian tree. The sequences have been determined for the brush-turkey (*Alectura lathami*, Megapodiidae), two passerines with a suboscine flycatcher (*Cnemotriccus fuscatus*; Tyrannidae) and the superb lyrebird (*Menura novaehollandiae*; Menuridae), the turkey vulture (*Cathartes aura*; Cathartinae) and a gull (*Larus dominicanus*; Laridae). The phylogenetic analyses using maximum likelihood and Bayesian methods support a growing consensus from nuclear, mitochondrial, and morphological data for the position of the root of the avian tree and for its first main divisions. The turkey vulture is not positioned with the stork and the gull joins with the turnstone.

2. Materials and methods

2.1. Tissue samples

The southern black-backed gull (*Larus dominicanus* [Lichtenstein, 1823]) came from Waikanae, New Zealand, and was supplied by the New Zealand Department of Conservation (Kapiti Area branch). The Australian brush-turkey (*Alectura lathami* [Gray, 1831]) and superb lyrebird (*Menura novaehollandiae* [Latham, 1802]) samples came from Australia. Darryl Jones (Australian School of Environmental Studies, Griffith University, Brisbane) and Ian Owens (Queensland University) provided the brush-turkey, and Cathy Nock (Centre for Animal Conservation Genetics, Southern Cross University [Lismore campus], New South Wales) the lyrebird. The turkey vulture (*Cathartes aura* [Linnaeus, 1758]) came from Texas, USA, and the fuscous flycatcher (*Cnemotriccus fuscatus* [Wied-Neuwied, 1831]) from Peru. Both samples were supplied by the Louisiana State University Museum of Natural Science Collection of Genetic Resources (sample numbers: turkey vulture LSUMNS B-17242, flycatcher LSUMNS B-7284).

2.2. DNA extraction and sequencing

Genomic DNA was extracted from 25 to 50 mg of sample tissue using the High Pure™ PCR Template Preparation Kit (Protocol Vb; Boehringer–Mannheim). To minimize the possibility of obtaining nuclear copies of mitochondrial (mt) genes (numts), the five mitochondrial genomes were amplified in two (turkey vulture, flycatcher) or three (gull, brush-turkey, lyrebird) overlapping fragments using long range PCR. The long range products (see below) were recovered from agarose gels using an appropriate kit (e.g., Concert™ Rapid Gel Extraction Sys-

tem kit [Gibco-BRL/Life Technologies], AccuPrep™ Gel Purification Kit [Bioneer]). They were then used as templates in a second round of PCR amplification (overlapping fragments of 0.3–3.1 kb in length). PCRs were set up using the long range and reamplification conditions described in Slack et al. (2003). Annealing temperatures for the reamplifications ranged from 48 to 60 °C and extension times from 1 to 3 min. The conserved primers used were from Kocher et al. (1989), Sorenson et al. (1999), Cooper et al. (2001), and Cooper et al. (unpublished), with some being modified from the original, or were designed from alignments of bird mitochondrial genes. Where necessary (e.g., control region), specific primers were designed for individual birds using the Oligo® 4.03 program (National Biosciences, Inc.).

The positions and sizes of the long range PCR products and the annealing temperatures, extension times, and primers used to generate them are given below. Primer nomenclature: Av (avian) position gene direction (F: forward; R: reverse). Positions and genes indicate where the 3′ end of the primer binds in the chicken mitochondrial genome. To avoid confusion, 'gene' and 'direction' are inverted for rRNA primers.

12S or 16S rRNA to tRNA-Leu(CUN):
Turkey vulture and flycatcher (11.3 kb): 60 °C, 8 min
Gull, brush-turkey, and lyrebird (9.3 kb): 53 °C, 9.5 min, and 2 × 54 °C, 10 min, respectively

Turkey vulture and flycatcher: Av1753F12S (25 nt): 5′-AAACTGGGATTAGATACCCCACTAT-3′
Gull, brush-turkey, and lyrebird: Av3725F16S-LR (32 nt): 5′-AATAGGGTTTACGACCTCGATGTTGG ATCAGG-3′
All five birds: Av13026tLeuR2-LR (37 nt): 5′-CTTGG AKTTGCACCAAGRTDVTTGGTTCCTAAGACCA-3′

COIII to tRNA-Pro:
Gull and brush-turkey (5.5 kb): 54 °C, 5.5 and 7 min, respectively; lyrebird (7.6 kb): 52 °C, 8 min
Av10647COIIIF (23 nt): 5′-TTTGAAGCAGCAGCCT-GATAYTG-3′
Av16137tProR (23 nt): 5′-ARAATRCCAGCTTTGG GAGTTGG-3′

Cytb to tRNA-Met:
Gull (6.2 kb): 53 °C, 8 min
Av15656CytbF (20 nt): 5′-AACCTGTTAGGRGAYCC AGA-3′
Av5201tMetR (20 nt): 5′-CCATCATTTTCGGGGTA TGG-3′

tRNA-Leu(CUN) to 16S rRNA:
Brush-turkey (7.4 kb) and lyrebird (8.5 kb): 55 and 54 °C, respectively, 10 min
Both: Av13063tLeuF-LR (38 nt): 5′-TGGTCTTAGGA RCCATCTATCTTGGTGCAAMTCCAAGT-3′
Brush-turkey: Av3782R16S (22 nt): 5′-CGGTCTGAA CTCAGATCACGTA-3′
Lyrebird: Av3797R16S (22 nt): 5′-CGACCTGGATTT CTCCGGTCTG-3′

COIII to 16S rRNA:
Turkey vulture (9.9 kb) and flycatcher (10.3 kb): Av10647COIIIF and Av3797R16S (60 °C, 8 min)

The reamplification products were purified by treatment with 2 U of shrimp alkaline phosphatase (SAP) and 10 U of Exonuclease I (incubated at 37 °C for 30 min, then 80 °C for 15 min) or were recovered from agarose gels as above. Two regions of the lyrebird mitochondrial genome were cloned: a 0.9 kb fragment spanning the end of tRNA-Phe plus the start of 12S rRNA and a 1.4 kb fragment covering part of the control region, tRNA-Pro and part of NADH6. The PCR products were ligated into the pGem®-T Easy vector (Promega) and transformed into MAX Efficiency® DH5α™ competent cells (Invitrogen) following manufacturer's instructions. Plasmid DNA was isolated from transformants containing inserts using the GenElute™ Plasmid Miniprep Kit (Sigma). Two clones were sequenced for the first region (plus direct sequence from a PCR product generated using a specific control region primer) and three clones for the second. Sequencing reactions were run on ABI 377 or ABI 3730 sequencers and were set up according to manufacturer's instructions. Sequences were manually checked/corrected (including the removal of any primer sequence) and assembled using Sequencer™ 4.1 (Gene Codes Corp.). Overlaps between sequences were sufficient to ensure synonymy (usually $\geqslant$ 100 nt between individual sequences; a total of from 1 to 4 kb between the different long range products). Sequence identity was confirmed through Fasta searches of the EMBL database (http://www.ebi.ac.uk/fasta33/nucleotide.html).

2.3. Dataset assembly

In addition to the five new birds from this paper, 25 other complete avian mitochondrial genomes were included in the analyses (18 neognaths and 7 paleognaths). The 18 neognath taxa are: chicken (*Gallus gallus*; GenBank Accession number AP003317), Japanese quail (*Coturnix japonica*; AP003195), magpie goose (*Anseranas semipalmata*; AY309455), redhead duck (*Aythya americana*; AF090337), greater white-fronted goose (*Anser albifrons*; AF363031), rifleman (NZ wren, *Acanthisitta chloris*; AY325307), grey-headed broadbill (*Smithornis sharpei*; AF090340), village indigobird (*Vidua chalybeata*; AF090341), rook (*Corvus frugilegus*; Y18522), peregrine falcon (*Falco peregrinus*; AF090338), Eurasian buzzard (*Buteo buteo*; AF380305), blackish oystercatcher (*Haematopus ater*; AY074886), ruddy turnstone (*Arenaria interpres*; AY074885), Oriental stork (*Ciconia boyciana*; AB026193), red-throated loon (*Gavia stellata*; AY293618), little blue penguin (*Eudyptula minor*; AF362763), black-browed albatross (*Diomedea*

melanophris; AY158677) and Kerguelen petrel (*Pterodroma brevirostris*; AY158678). The 7 paleognath taxa are: emu (*Dromaius novaehollandiae*; AF338711), southern cassowary (*Casuarius casuarius*; AF338713), great spotted kiwi (*Apteryx haastii*; AF338708), greater rhea (*Rhea americana*; Y16884), ostrich (*Struthio camelus*; Y12025), great tinamou (*Tinamus major*; AF338707), and elegant crested tinamou (*Eudromia elegans*; AF338710). The most up-to-date version of each GenBank file was used and the original chicken mitochondrial genome (X52392) used in previous analyses (e.g., Harrison et al., 2004; Slack et al., 2006) has been replaced by a more recent sequence.

Six reptiles were used as outgroups: American alligator (*Alligator mississippiensis*; Y13113), spectacled caiman (*Caiman crocodylus*; AJ404872), eastern painted turtle (*Chrysemys picta*; AF069423), green turtle (*Chelonia mydas*; AB012104), blue-tailed mole skink (*Eumeces egregius*; AB016606), and common iguana (*Iguana iguana*; AJ278511). This is the same set of outgroups used previously (for example, Harrison et al., 2004; Slack et al., 2006).

Two neoavian taxa (an owl and a parrot; Harrison et al., 2004) and three paleognaths (extinct NZ moas; Cooper et al., 2001; Haddrath and Baker, 2001) were omitted from the full analyses until paired taxa are available. In the meantime, the owl and parrot do not affect the position of the root of the avian tree that is the focus of this study. We are in the process of sequencing a barn owl (*Tyto*), an African parrot (lovebird, *Agapornis*), and a forest falcon (*Micrastur*) (www.awcmee.massey.ac.nz/mt_genomes). Based on previous experience of improved taxon sampling (Lin et al., 2002; Delsuc et al., 2003; Harrison et al., 2004) the addition of these three mitochondrial genomes will help stabilize the position of both owls and parrots within Neoaves for future analyses. Similarly, there is some instability within paleognaths when moas are included (Slack et al., 2003), and the issue of fine-tuning paleognath interrelationships will be readdressed once additional kiwi sequences become available (G.C. Gibb, in preparation).

2.4. Phylogenetic analysis

Sequences were aligned in SeAl v1 (Rambaut, 1996), at the amino acid level for protein-coding genes, and based on secondary structure for RNA genes. Each dataset has 12 protein-coding genes, two rRNAs and 21 tRNAs (lacking tRNA-Phe because it is still not available for the rifleman). Gaps, ambiguous sites adjacent to gaps, the NADH6 (light-strand encoded), and stop codons (often incomplete in the DNA sequence), were excluded from the alignment.

We made two alignments; with and without the six outgroup taxa. The 'birds-only' dataset was used first to study relationships within the ingroup (birds) in order to test whether there were any changes to the tree when the outgroup was added. In practice, the tree from the birds-only dataset was compared with the tree using only the birds from the full alignment (aligned including the outgroup). This allowed the separation of any effects of adding the outgroup, from any changes from the alignment. The full dataset had 11,737 unambiguously aligned nucleotide sites, the birds-only dataset is 15% longer (13,440 nucleotides).

In previous work, RY-coding of the most variable partitions of the nucleotide data (especially the 3rd codon position) has been shown to be advantageous (Delsuc et al., 2003; Phillips and Penny, 2003; Phillips et al., 2006). The recoding increases the proportion of changes on internal branches of the tree (treeness) and alleviates the differences in nucleotide composition (Relative Compositional Variability, RCV). It also increases concordance between mitochondrial and nuclear datasets. RY-coding does increase the ML scores, but because RY-coding is not strictly nested within nucleotide coding (M.A. Steel, pers. comm.) it is not valid to compare their ML scores directly. However, because of the better fit of the data to the model (higher treeness, and less variability in nucleotide composition (lower RCV)), this is our preferred method of analysis of vertebrate mitochondrial data. Thus the trees reported here were inferred with the third codon positions recoded as R and Y. Analysis is by standard programs including ModelTest (Posada and Crandall, 1998), PAUP* (Swofford, 1998), and MrBayes (Ronquist and Huelsenbeck, 2003). We ran 500 unconstrained ML bootstrap replicate with PAUP* on the Helix computing cluster (www.helix.massey.ac.nz), plus a partitioned Bayesian analysis using four MCMCMC chains of 10^7 generations. The full data sets and command blocks for both PAUP* and MrBayes are available in nexus format at www.awcmee.massey.ac.nz/downloads. These files illustrate the procedures used both in finding optimal estimates for gamma and invariant sites for the different data partitions, as well running subsets of the data as RY coded.

3. Results

The five new avian complete mitochondrial genomes are deposited in GenBank under the following accession numbers: Australian brush-turkey (*Alectura lathami*, AY346091); superb lyrebird (*Menura novaehollandiae*, AY542313); fuscous flycatcher (*Cnemotriccus fuscatus*, AY596278); southern black-backed gull (*Larus dominicanus*, AY293619); and turkey vulture (*Cathartes aura*, AY463690). The gull, brush-turkey, and turkey vulture have the standard avian gene order first identified in the chicken (Desjardins and Morais, 1990). This is expected in that the other galliforms in the dataset (chicken and quail) and the two other shorebirds (oystercatcher and turnstone) are already known to have the standard gene order. However, the turkey vulture also appears to have the standard avian gene order, and thereby differs from the two falconiform mitochondrial genomes (falcon and buzzard) already available. By itself, this is certainly of interest, but not sufficient yet to exclude it from the falconiforms, even though for characters with an extremely high number of states parsimony is an ML estimator (Steel and Penny, 2004).

The lyrebird and flycatcher genomes have the alternative avian gene order first recognized by Mindell et al. (1998). The flycatcher result is as expected; Mindell et al. (1998) examined nine major suboscine lineages (including a tyrannid flycatcher) and found that all had the alternative gene order. On the other hand, Mindell et al. (1998) found that a set of 106 oscine species had the standard avian gene order. However, that set was mostly Passeridae (90 taxa) and Fringillidae (11 taxa) and the remaining five families (Corvidae, Sturnidae, Cisticolidae, Certhiidae, and Estrildidae) were each represented by a single taxon. Although Mindell et al. (1998) concluded "that the alternative mitochondrial gene orders distinguish the two primary groups of songbirds (order Passeriformes), oscines, and suboscines" Bensch and Härlid (2000) subsequently identified the alternative gene order in an oscine group (*Phylloscopus* warblers, Sylviidae). The lyrebird (Menuridae) now provides the second example of the alternative gene order in oscine passerines. The duplicated control region associated with the alternative gene order has been reduced to a short non-coding region in both the lyrebird and the *Phylloscopus* warblers. However, if this non-coding region is still similar to the control region then the sequence fragment (most of NADH6, tRNA-Glu, part of the control region) that Mindell et al. (1998) used to examine most of their oscine species (102 of 106) might not detect the alternative gene order when it is present. This could happen, for example, where the rearrangement is recent, or if a form of concerted gene evolution homogenized the duplicates (G.C. Gibb et al. in preparation).

3.1. Unrooted tree

We know from both simulations (Holland et al., 2003) and empirical studies on mammals (Lin et al., 2002) and birds (Slack et al., 2003) that an outgroup can disrupt a previously established ingroup tree (see Section 1). Therefore, we first report an unrooted ML tree for the 30 bird sequences, together with the results of 500 full ML bootstrap runs. The data is the combined protein-coding and RNA genes for both datasets—aligned with or without the six outgroup taxa, giving a combined length of 11,737 and 13,440 nucleotides, respectively. Again, we find that re-coding the 3rd position (cdp) as R/Y characters markedly reduces the relative compositional variability (RCV), and increases the signal on the internal branches compared with the external branches (the treeness value is increased). As before, the models were optimized separately for each of the four data partitions using ModelTest (Posada and Crandall, 1998). The four partitions were: codons 1 and 2; codon 3; RNA stems; and RNA loops. The unrooted maximum likelihood tree for the birds-only dataset is shown in Fig. 1.

The position of the megapode (brush-turkey) is, as expected, a deep lineage in Galliformes, and the three galliforms (chicken, quail, and brush-turkey) then group with the three anseriforms (duck, goose, and magpie goose [*Anseranas*]), forming Galloanserae. The tyrant flycatcher, as expected, joins with the other suboscine (broadbill) and their union is quite deep. Similarly, the lyrebird is deeper in the oscines than the indigo bird/rook divergence. It is worth noting that with the new oscine and suboscine included the NZ wren (rifleman) now comes basal to the passerines. Overall, the results strongly support the revised classification of oscines (Ericson et al., 2002b) into Menurae (which includes lyrebird) and Euoscines (which includes rook and indigo bird). These three new taxa (brush-turkey, tyrant flycatcher, and lyrebird) were selected from prior knowledge in anticipation that they would break up long branches that could be attracting the outgroup (a long-branch attraction problem, Hendy and Penny, 1989) when rooting the avian tree. This will be re-examined later when examining the position of the root.

The gull joins strongly with the turnstone, rather than with the other shorebird (oystercatcher). More formally, the Lari (gulls) are closer to the Scolopaci (turnstones) than to the Charadrii (oyster catcher). The result is in agreement with the results from nuclear data of Ericson et al. (2003) and Paton et al. (2003). It is not in agreement with Sibley and Ahlquist (1990), but it is only a single interchange (nearest neighbor interchange) on the tree. We say that a tree is 'locally stable' if taxa shift no more than one branch on the tree (Cooper and Penny, 1997). Again, the agreement of nuclear and mitochondrial data is pleasing, but expected.

The fifth new species is the turkey vulture—a New World vulture. The two issues here are whether the falconiforms (in a broad sense) are monophyletic, and if not, whether the turkey vulture goes with the stork. On the present dataset, turkey vulture does not come with the two other falconiforms, but nor is it particularly close to the stork. Thus there is some support for the independent origin of core falconiforms and New World vultures, but we need a wider selection of taxa, such as flamingos, grebes, herons, rails, etc. before firmly identifying the closest relative of New World vultures. However, it is also desirable to have additional core falconiforms, because the falcon still has a tendency to go deeper in the tree, disrupting the falcon/buzzard group. Similarly, Fain and Houde (2004) do not recover the falcon/buzzard clade, emphasizing the need for additional taxa. On the present data set the falconiforms (in the broad sense, including New World vultures) are not monophyletic, but we cannot distinguish yet between their being polyphyletic (diphyletic in this case) or paraphyletic. This latter case is potentially interesting from the ecological/life histories viewpoint in that many of the sea and shorebirds may have diversified from carnivores (raptors) into a related niche in a marine environment. At present, the position of both the peregrine falcon and the stork are still locally unstable and improved taxon sampling is thus required.

3.2. Rooted tree

The incorporation of the brush-turkey, the tyrant flycatcher and the lyrebird breaks up long unbranched edges situated deep in the avian tree. Therefore, the next

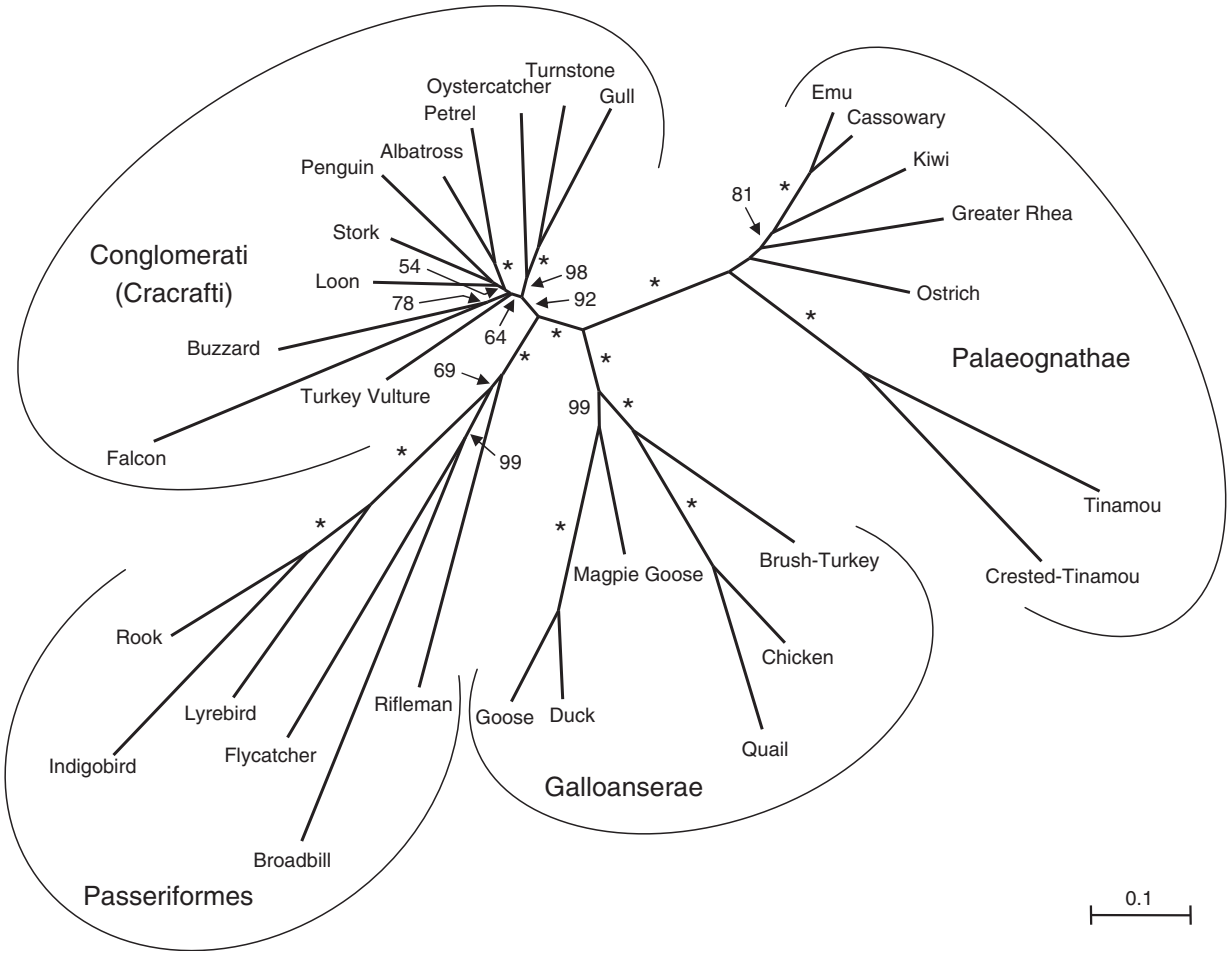


Fig. 1. Unrooted tree on complete mitochondrial genomes from 30 birds. The 3rd codon position was RY coded with other sites (1st and 2nd positions for protein genes, stems and loops for RNA genes) remained as nucleotides. Bootstrap values are from 500 runs.

step is to identify what (if any) effect this has for locating the position of the root in the avian tree. Fig. 2 is the rooted tree, using the six taxa outgroup constituted of two crocodilians, two turtles, and two lizards. This reptilian outgroup joins with strong support between paleognaths and neognaths (Fig. 2), and is thus consistent with most recent work on both nuclear genes (Van Tuinen et al., 2000; Chubb, 2004; Garcia-Moreno et al., 2003) and mitochondrial genomes (Paton et al., 2002; Harrison et al., 2004). Of the outgroup taxa, crocodilians are again closest to birds.

An important observation is that the addition of the outgroup does not make major disruptions to the ingroup tree as was previously the case with fewer taxa (Slack et al., 2003). As mentioned earlier, simulations (Holland et al., 2003) have shown that the addition of the outgroup can lead to disruption of the previously established relationships within the ingroup as it has been observed in placental mammal mitogenomic trees for example (Lin et al., 2002). It is therefore pleasing that the basic ingroup tree (from the unrooted tree of Fig. 1) is not altered when the outgroup is included, giving us more confidence in the present rooting. As a result, the paleognath/neognath division is supported as well as the strict monophyly of passerines. As before, there is increased support for the chicken/duck grouping (Galloanseriforms), again in agreement with morphological (Cracraft, 2001), nuclear (Van Tuinen et al., 2000; Chubb, 2004; Garcia-Moreno et al., 2003), and mitochondrial (Paton et al., 2002; Harrison et al., 2004) datasets. Within the Neoaves, there is strong support for Passeriformes and Conglomerati/Cracrafti being on opposite sides of the Neoaves tree, though significant groups such as rails and pigeons are not yet represented, and there are differences in predictions between Cracraft (2001) and Cracraft et al. (2004). There is one interesting difference between the birds-only results (Fig. 2), and the tree derived from the alignment that includes the six outgroup taxa (Fig. 1). This is that the first divergence is between shorebirds and raptors/seabirds on the birds-only dataset, but the falconiforms (buzzard and falcon in particular) are deeper on the tree from the alignment containing the 6-taxon outgroup. This difference, especially the deeper divergence of shorebirds is worth following up, given the comment of Feduccia (2006) that shorebirds appear to be a late Cretaceous lineage, and from the results of Slack et al. (2006) that discuss the agreement between estimates of times of divergence from molecular data and the existence of early fossil footprints of birds. However, some relationships within the Conglomerati remain unresolved with low bootstrap values

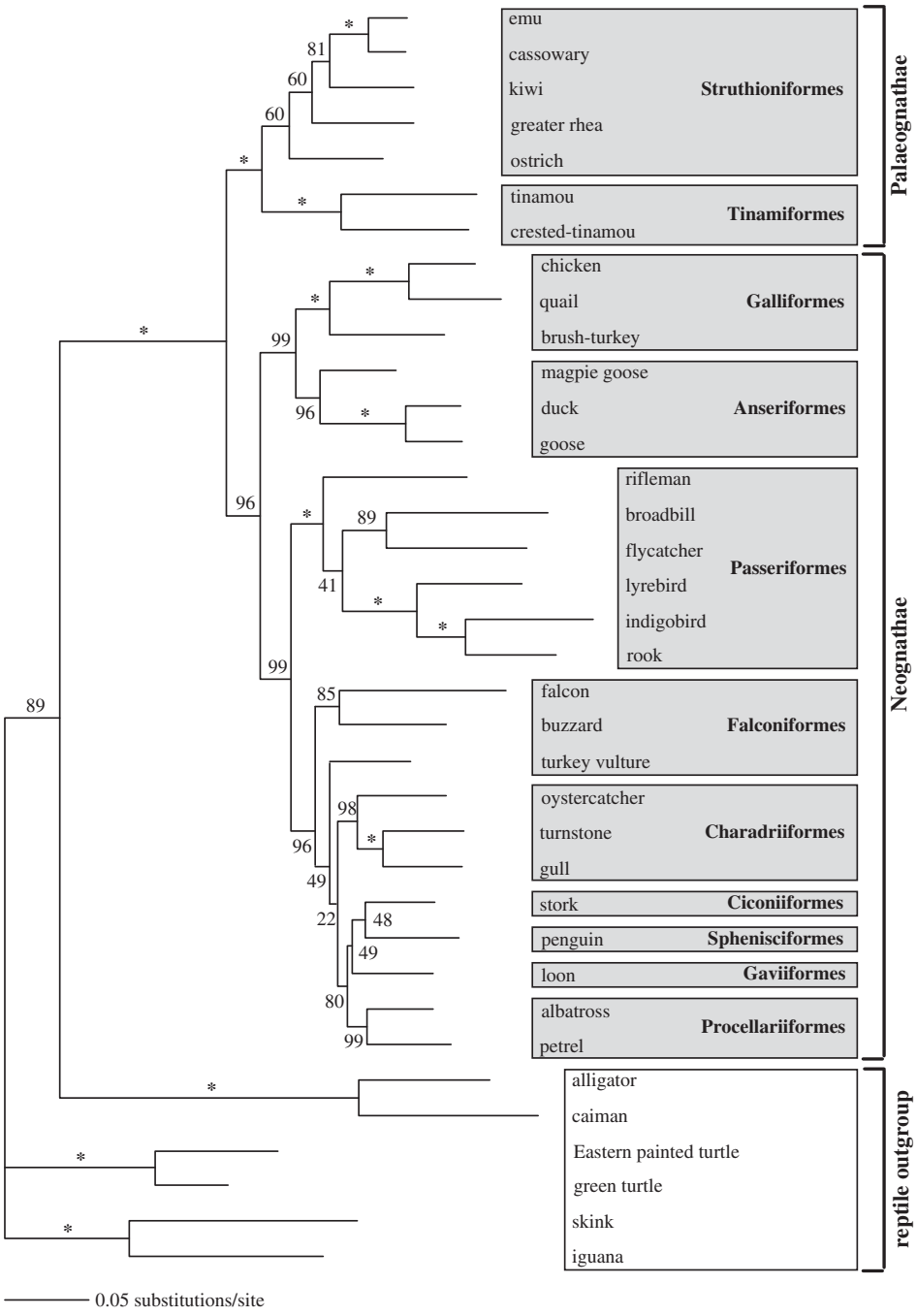


Fig. 2. Rooted tree using six outgroup taxa. The same coding as for Fig. 1. Bootstrap values are shown; * is 100%.

except for the monophyly of Charadriiformes and for the putative grouping of Procellariformes with Gaviiformes, Sphenisciformes, and Ciconiiformes (Figs. 2 and 3). The lack of conflicting signal for this latter grouping (see Fig. 3) is interesting in that it argues against there being any strong systematic bias (Phillips et al., 2004).

3.3. Increased stability from breaking up long branches (edges)

A primary interest here is testing whether the improved taxon sampling (breaking up long branches within the avian tree) leads to improved stability. Qualitatively it appears so since the addition of brush turkey, tyrant flycatcher, and lyrebird, has significantly increased the strength of the rooting between paleognaths and neognaths. To make this conclusion quantitative we deleted each of the 30 avian taxa in turn, and ran 100 bootstrap samples on each of the reduced 30 data sets, using ML. Thus each dataset had 29 birds and the six outgroup taxa. This is essentially a jackknife approach (Lanyon, 1985), deleting taxa sequentially and measuring the effect (Penny and Hendy, 1985). For the 30 bootstrap runs, the outgroup was constrained so that the six outgroup taxa always grouped

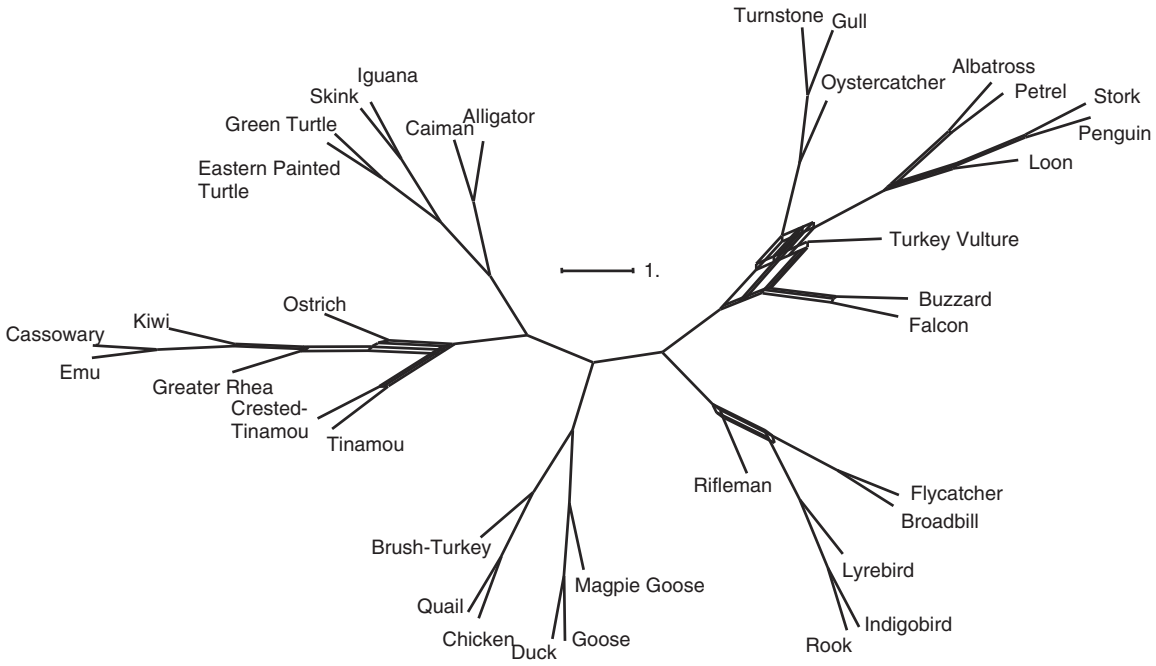


Fig. 3. Phylogenetic super-network obtained by applying the Z-closure method to the 30 jackknife ML trees where each individual bird taxon has been removed in turn from the complete dataset. This figure represents a graphical summary of multiple phylogenies reconstructed from the different sets of taxa. It shows on which parts of the phylogeny all partial trees are in agreement (dichotomous parts) and where there exist contradicting signals (networking parts).

together, as well as the two turtles, the two crocodilians and the two lizards. There were no constraints on the avian tree. For each of the 30 datasets, the percentage of trees from the 100 bootstrap samples were summed under the categories of paleo/neognath rooting, rooting within paleognaths, galloanseriform rooting, passerine rooting, or within the Neoaves generally (Table 1). For both the galloanseriform and passerine cases the results include the rooting either within, for example, galloanseriforms or between galloanseriforms and all other birds.

With phylogenetic trees (including our own studies) it is traditional to give the results first and subsequently think up explanations! Here, we reverse the process and give our predictions first (Table 1, column 2) and then run the bootstrap samples before comparing the predictions and the results. Because our theme has been that breaking up long branches improves the stability of the tree, we focus our predictions on how we expect the additional sequences to affect the bootstrap values for different rootings. The second column in Table 1 indicates our predictions on how the bootstrap values might change for each jackknife sample, relative to the bootstrap values with the full 30 bird dataset. The predictions are whether we expect that removal of a particular avian taxon will increase the bootstrap values for a competing placement of the root. For example, we expect that removing the brush-turkey will lead to the root appearing significantly more often on the galloanseriform lineage, indicated as '>Ga'. Concomitably, there would have to be a reduction in cases where the root comes into the expected position between paleognaths and neognaths. The magnitude of the effect is hard to predict, but based on prior experience (for example, Slack et al., 2003) we would expect bootstrap values to decrease by 10–30%, and possibly even more. Where it seemed that a smaller effect could occur we have added question marks, for example '>Ga??'. Note that to estimate the increased stability from adding a taxon, we are measuring the decrease in bootstrap values from removing the taxon. More than one value can occur in a column, for example in column 3, the first value is the paleognaths occurring as a monophyletic group; the second for the neognaths. Similarly, within the galloanseriforms, the root could come on the branch basal to the group, or within the chicken or the duck group. One aberrant bootstrap sample could affect several internal branches on the tree. For example, a '1 + 1 + 1' will usually be one bootstrap sample that changes three internal branches. Having given our predictions, the actual results are given in Table 1.

Surprisingly, our predictions were relatively poor; there was virtually no loss of stability by removing any individual taxon. That is, the results were considerably more robust than we expected. Overall this implies that there is some cumulative effect from adding a range of taxa that break up long branches, leading to the increased stability. Nevertheless there are interesting individual effects especially evidenced when the rows are summed to give the 'total' effect of removing each taxon. Removal of the New Zealand wren (rifleman) had the largest effect, by which we infer that adding this taxon into the dataset contributed the most to stabilizing the tree. As expected, removing a deeper branching taxon had more effect than removing one member of a pair of more closely related taxa. For example, removing the brush-turkey had more effect than deleting either the

Table 1
Predicted and actual effects on the position of the root from taxon removal

Species omitted	Predicted	Paleo/neognaths[a]	Paleognaths Pg	Galloanseres Ga	Passerines Pa	Neoaves	Total
None, all 30 birds	n.a.						
Emu	—	100/99	—	1	—	—	1
Cassowary	—	99/97	—	2 + 1	1 + 1 + 1	—	6
Kiwi	—	100/97	—	1 + 1	1	1	4
Rhea	>Pg??	100/95	—	1 + 1	3 + 4 × 1	4	13
Ostrich	>Pg??	100/93	—	3 + 1	3 + 1	4	12
Tinamou	>Pg	100/91	—	1	2 + 2 + 2	8	15
gs_tinamou	>Pg	100/95	—	3 + 1	1 + 1 + 1	2	9
Magpie goose	>Ga	100/94	—	3	1 + 1	3	8
Goose	—	100/98	—	1 + 1	1 + 1 + 1	1	6
Duck	—	99/96	1	2	—	2	5
Brush-turkey	>Ga	100/92	—	7 + 3	—	1	11
Chicken	—	100/99	—	1 + 1	—	—	2
Quail	—	100/97	—	2	1 + 1	1	5
Rifleman	>Pa?	100/91	—	2	5 + 5 + 3	7	22
Broadbill	>Pa	100/94	—	3 + 1	2	3	9
Flycatcher	>Pa	98/96	2	—	2 + 1 + 1	4	10
Lyrebird	>Pa	99/93	1	3 + 1	3 + 2 + 2	4	16
Rook	>Pa??	100/94	—	4 + 3 × 1	1 + 1	2	11
Indigobird	>Pa??	100/93	—	3 + 1	1 + 2 × 1[b]	4	11
Falcon	—	99/97	1	3 + 2	—	—	6
Buzzard	>Oth??	100/96	—	—	2 + 1 + 1	4	8
Turkey vulture	—	99/94	—	2 + 1	1	4	8
Oystercatcher	—	100/95	—	2 + 1	1	3	7
Turnstone	—	100/97	—	2 + 3 × 1	1	1	7
Gull	—	99/98	1	1 + 1 + 1	—	—	4
Stork	—	100/98	—	1	1	1	3
Penguin	—	100/91	—	4 + 1	3 + 1 + 1	5	15
Albatross	—	100/98	—	1	—	1	2
Petrel	—	100/95	—	2	—	3	5
Loon	—	100/95	—	3	1 + 1	2	7

[a] Expected position of the root from nuclear and morphological data.
[b] Crested tinamou joins rhea, inside ratites.

chicken or quail. Overall, removal of a passerine had a relatively large effect, a result that might be explained by their higher evolutionary rates. However, we were surprised that the removal of the penguin has also a significant effect, despite its relatively short branch and its internal position in the Neoavian tree. We had assumed that the problem was that the stork was relatively unstable on the tree, but we must consider penguin as a potentially difficult taxon.

In an additional analysis, we computed a Z-closure super-network (Huson et al., 2004), using SplitsTree 4.1 (Huson and Bryant, 2006), from the 30 jackknife ML trees where each avian species was omitted in turn from the rooted tree (Fig. 3). This method offers a graphical summary of the topologies of the 30 jackknifed trees and allows identification of the parts of the phylogeny which were in agreement for all the 30 trees with 29 avian taxa. Areas of conflict are represented as rectangles in the super-network, reflecting any instability caused by removing individual taxa. The results in Fig. 3 are striking in that there is no incongruence detected with respect to the position of the root of the avian part of the tree. This again shows the lack of effect of removing any single taxon and illustrates that a robust rooting of the avian tree can be obtained by using an expanded number of mitochondrial genomes. In contrast, uncertainty is detected in three places within the avian tree corresponding to areas of instability identified previously (see Fig. 2). The first is between the tinamou and the deeper ratites (ostrich and rhea) where the tree is not completely stable and therefore removing taxa has an effect. Similarly, there is a tendency for the rifleman to come onto the basal suboscine branch as represented by a cycle in Fig. 3. However, the major effect was inside the Conglomerati/Cracrafti with a series of boxes involving the raptors (including turkey vulture) and the sea birds and shore birds. Either the core raptors (falcon and buzzard) or the shore birds can come out basal in this Conglomerati group. This relatively unstable part of the avian tree is being studied further by sequencing additional mitochondrial genomes, including osprey, a forest falcon, and additional potential relatives of stork and penguin. However, although the variability within the Conglomerati (Cracrafti) is very interesting, it is not relevant to the question at hand of demonstrating the stability of the root of the avian tree.

4. Discussion

The long-term goal of this project is to use nuclear and mitochondrial sequences, together with fossil data, to test

modes of macroevolution in the Late Cretaceous (Penny and Phillips, 2004). The primary aim of this study is to determine whether the breaking up of some long branches on the avian tree leads to agreement between nuclear and mitochondrial data on the position of the root. In this respect, the position of the root now appears in agreement between morphological, nuclear, and mitochondrial data, and thus it is time to move on to other questions. Data from complete mitochondrial genomes obviously takes longer to obtain for each taxon than sequences from a single nuclear gene. This has lead to some false starts concerning the position of the avian root due in part to incomplete taxon sampling (Härlid and Arnason, 1999; Mindell et al., 1999), though earlier analyses by Braun and Kimball (2002) and Slack et al. (2003) indicated that the paleognath/neognath division could not be rejected. There appears to have been an 'urban myth' that mitochondrial genomes could not recover the same avian root as morphological and nuclear data. In fact, the main problem was incomplete taxon sampling, a problem that is better solved by more data collection than by polemics. Indeed, it is hard to imagine theoretical reasons that would lead to different roots from different datasets. It is unlikely that the trees would be fundamentally different and therefore additional sequences and better methods of analysis are a more likely solution to the problem. We should all aim at improving data sets to test whether they lead to consensus.

With the major early divisions having been resolved, perhaps the next step on the avian tree is to provide resolution within Neoaves and especially within the Conglomerati/Cracrafti. For this specific question, breaking up long branches does appear to be an effective strategy, but obviously depends on appropriate taxa being available. When there was only a single suboscine (broadbill) and a single Falconiform (falcon) in the dataset, there was a strong tendency for them to come together (see discussion in Slack et al., 2003). However, with additional passerine being incorporated, we now have strong support for the respective monophyly of both Passeriformes and Conglomerati. The falcon and the stork have both been difficult to place in the mitogenomic tree in that both still appear to be uncertain about their final position. In early datasets, the falcon could even occur at the base of the passerines (Slack et al., 2003). The addition of the buzzard has reduced the wandering of the falcon, but the falcon/buzzard grouping is relatively weak, and the falcon can go deeper within the neoaves. We are currently sequencing a forest falcon (*Micrastur*) to determine its effect because in Sibley and Ahlquist (1990) this was the deepest divergence among relatives of the falcon. The present data set has no species from within the proposed Metaves group of Fain and Houde (2004), and thus we cannot yet comment on that hypothesis. Mitochondrial genomes from members of that group are being completed and then we will be able to test the predicted distinction of Neoaves into Metaves and Coronaves.

The problem of the early divisions of Neoaves is going to be difficult. Cracraft (2001) proposed six unresolved groups, and this increased to nine in Cracraft et al. (2004). It has been suggested that the early divergence of neoavian birds was an 'explosive radiation' (as just one example, see Poe and Chubb, 2004). However, for a radiation to be 'explosive' it requires both a rapid series of lineage divergences, combined with simultaneous morphological and/or ecological adaptations. It would not be sufficient just for divergences of lineages to be close together—that could occur very easily by a rapidly dispersing taxon even though the subsequent ecological and morphological divergences occurred many millions of years later. Such a delayed adaptation would scarcely be an 'explosive' radiation. The first aspect (rate of diversification of lineages) is best studied from molecular data (as is done here). The second, the timing of adaptations to new environments/niches is perhaps better studied from the fossil record. As yet, in the present data we do not see any evidence whatsoever for an 'explosive radiation' of neoavian birds, even though the early divergences may have been relatively rapid in Neoaves. We are more cautious, and want to see data on the speed of morphological and ecological changes before coming to such dramatic conclusion about an explosive radiation. We prefer at present to consider the early diversification of Neoaves as an adaptive radiation, indicating that it is a relatively fast radiation but strictly according to known microevolutionary principles.

Indeed, as mentioned earlier, one of our main goals is to determine the extent to which the processes of microevolution are sufficient to explain macroevolution. This is the theme behind our models of avian and mammal diversification in the Late Cretaceous (Penny and Phillips, 2004), and whether (by comparing the human and chimpanzee genomes) there is any aspect of the human genome that is not a normal microevolutionary (genetic) process (Penny, 2004). We think that the 'explosive radiation' should be to restrict to possible cases where normal microevolutionary process are clearly insufficient to account for macroevolution. To test such hypotheses we need a tree sufficiently stable in order that good timing estimates are possible, preferably on combined nuclear and mitochondrial data in order to test predictions with alternative data sets (Penny et al., 1991). All progress in this direction is welcomed.

Acknowledgments

We especially thank Dick Gill, Darryl Jones, Ian Owens, Cathy Nock, Fred Sheldon, and Donna Dittman for the samples. Tim White and Warwick Allen assisted with computer programs for running analyses on the Helix supercomputer and Gillian Gibb for considerable assistance in manuscript preparation. F.D. acknowledges the support of a Lavoisier Postdoctoral Grant from the French Ministry of Foreign Affairs. Both the New Zealand Marsden Fund and the Centre for Research Excellence funds supported the work.

References

Barker, F.K., Barrowclough, G.F., Groth, J.G., 2002. A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implica-

tions of an analysis of nuclear DNA sequence data. Proc. R. Soc. Lond. B 269, 295–308.
Barker, F.K., Cibois, A., Schikler, P., Feinstein, J., Cracraft, J., 2004. Phylogeny and diversification of the largest avian radiation. Proc. Natl. Acad. Sci. USA 101, 11040–11045.
Bensch, S., Härlid, A., 2000. Mitochondrial genomic rearrangements in songbirds. Mol. Biol. Evol. 17, 107–113.
Braun, E.L., Kimball, R.T., 2002. Examining basal avian divergences with mitochondrial sequences: model complexity, taxon sampling, and sequence length. Syst. Biol. 51, 614–625.
Buckley, T.R., 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51, 509–523.
Chang, J.T., 1996. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. Math. Biosci. 134, 189–215.
Chubb, A.L., 2004. New nuclear evidence for the oldest divergence among neognath birds: the phylogenetic utility of ZENK (i). Mol. Phylogenet. Evol. 30, 140–151.
Cooper, A., Penny, D., 1997. Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. Science 275, 1109–1113.
Cooper, A., Lalueza-Fox, C., Anderson, S., Rambaut, A., Austin, J., Ward, R., 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature 409, 704–707.
Cracraft, J., 2001. Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. Proc. R. Soc. Lond. B 268, 459–469.
Cracraft, J., Barker, F.K., Braun, M. et al., 2004. Phylogenetic relationships among modern birds (Neornithes). In: J. Cracraft, M.J. Donoghue (Eds.), Assembling the Tree of Life. Oxford University Press, pp. 468–489.
Delsuc, F., Phillips, M.J., Penny, D., 2003. Comment on "Hexapod origins: monophyletic or paraphyletic?". Science 301, 1482d.
Desjardins, P., Morais, R., 1990. Sequence and gene organization of the chicken mitochondrial genome—a novel gene order in higher vertebrates. J. Mol. Biol. 212, 599–634.
Edwards, S.V., Boles, W.E., 2002. Out of Gondwana: the origin of passerine birds. Trends Ecol. Evol. 17, 340–347.
Edwards, S.V., Fertil, B., Giron, A., Deschavanne, P.J., 2002. A genomic schism in birds revealed by phylogenetic analysis of DNA strings. Syst. Biol. 51, 599–613.
Ericson, P.G.P., Johansson, U.S., 2003. Phylogeny of Passerida (Aves: Passeriformes) based on nuclear and mitochondrial sequence data. Mol. Phylogenet. Evol. 29, 126–138.
Ericson, P.G.P., Christidis, P.L., Cooper, A., Irestedt, M., Jackson, J., Johansson, U.S., Norman, J.A., 2002a. A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. Proc. R. Soc. Lond. B 269, 235–241.
Ericson, P.G.P., Christidis, P.L., Irestedt, M., Norman, J.A., 2002b. Systematic affinities of the lyrebirds (Passeriformes: Menura), with a novel classification of the major groups of passerine birds. Mol. Phylogenet. Evol. 25, 53–62.
Ericson, P.G.P., Envall, I., Irestedt, M., Norman, J.A., 2003. Inter-familial relationships of the shorebirds (Aves: Charadriiformes) based on nuclear DNA sequence data. BMC Evol. Biol. 3, 16.
Fain, M.G., Houde, P., 2004. Parallel radiations in the primary clades of birds. Evolution 58, 2558–2573.
Feduccia, A., 2006. Mesozoic aviary takes form. Proc. Natl. Acad. Sci. USA 103, 5–6.
Fuchs, J., Fjeldsa, J., Bowie, R.C.K., Voelker, G., Pasquet, E., 2006. The African warbler genus *Hyliota* as a lost lineage in the oscine songbird tree. Mol. Phylogenet. Evol. 39, 186–197.
Garcia-Moreno, J., Mindell, D.P., 2000. Rooting a phylogeny with homologous genes on opposite sex chromosomes (gametologs): a case study using avian CHD. Mol. Biol. Evol. 17, 1826–1832.
Garcia-Moreno, J., Sorenson, M.D., Mindell, D.P., 2003. Congruent avian phylogenies inferred from mitochondrial and nuclear DNA sequences. J. Mol. Evol. 56, 1–11.
Haddrath, O., Baker, A.J., 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. Proc. R. Soc. Lond. B 268, 939–945.
Haring, E., Kruckenhauser, L., Gamauf, A., Riesing, M.J., Pinsker, W., 2001. The complete sequence of the mitochondrial genome of *Buteo buteo* (Aves, Accipitridae) indicates an early split in the phylogeny of raptors. Mol. Biol. Evol. 18, 1892–1904.
Härlid, A., Arnason, U., 1999. Analyses of mitochondrial DNA nest ratite birds within the "Neognathae" supporting a neotenous origin of ratite morphological characters. Mol. Phylogenet. Evol. 12, 115–123.
Härlid, A., Janke, A., Arnason, U., 1999. The complete mitochondrial sequence of *Rhea americana* and early avian divergences. J. Mol. Evol. 46, 669–679.
Harrison, G.L., McLenachan, P.A., Phillips, M.J., Slack, K.E., Cooper, A., Penny, D., 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the Late Cretaceous. Mol. Biol. Evol. 21, 974–983.
Helbig, A.J., Seibold, I., 1996. Are storks and New World vultures paraphyletic? Mol. Phylogenet. Evol. 6, 315–319.
Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297–309.
Holland, B.R., Penny, D., Hendy, M.D., 2003. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock: a simulation study. Syst. Biol. 52, 229–238.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.
Huson, D.H., Dezulian, T., Kloepper, T., Steel, M.A., 2004. Phylogenetic super-networks from partial trees. IEEE/ACM Trans. Comput. Biol. Bioinform. 1, 151–158.
Irestedt, M., Johansson, U.S., Parsons, T.J., Ericson, P.G.P., 2001. Phylogeny of major lineages of suboscines (Passeriformes) analysed by nuclear DNA sequence data. J. Avian Biol. 32, 15–25.
Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86, 6196–6200.
Lanyon, S.C., 1985. Detecting internal inconsistencies in distance data. Syst. Zool. 34, 397–403.
Lin, Y.-L., McLenachan, P.A., Gore, A.R., Phillips, M.J., Ota, R., Hendy, M.D., Penny, D., 2002. Four new mitochondrial genomes, and the stability of evolutionary trees of mammals. Mol. Biol. Evol. 19, 2060–2070.
Livezey, B.C., Zusi, R.L., 2001. Higher-order phylogenetics of modern Aves based on comparative anatomy. Netherlands J. Zool. 51, 179–205.
Mayr, G., Clarke, J., 2003. The deep divergences of neornithine birds: a phylogenetic analysis of morphological characters. Cladistics 19, 527–553.
Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., 1998. Multiple independent origins of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. USA 95, 10693–10697.
Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., Hasegawa, M., Ast, J.C., Yuri, T., 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst. Biol. 48, 138–152.
Paton, T., Haddrath, O., Baker, A.J., 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proc. R. Soc. Lond. B 269, 839–846.
Paton, T., Baker, A.J., Groth, J.G., Barrowclough, G.F., 2003. RAG-1 sequences resolve phylogenetic relationships within Charadriiform birds. Mol. Phylogenet. Evol. 29, 268–278.
Penny, D., 2004. Our relative genetics. Nature 427, 208–209.
Penny, D., Hendy, M.D., 1985. The use of tree comparison metrics. Syst. Zool. 34, 75–82.
Penny, D., Phillips, M.J., 2004. The rise of birds and mammals: are microevolutionary processes sufficient for macroevolution. Trends Ecol. Evol. 19, 516–522.
Penny, D., Foulds, L.R., Hendy, M.D., 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. Nature 297, 197–200.
Penny, D., Hendy, M.D., Steel, M.A., 1991. Testing the theory of descent. In: Miyamoto, M.M., Cracraft, J. (Eds.), "Phylogenetic Analysis of DNA Sequences". Oxford University Press, pp. 155–183.

Phillips, M.J., Penny, D., 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogenet. Evol. 28, 171–185.

Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny: sampling and systematic errors are both important. Mol. Biol. Evol. 21, 1455–1458.

Phillips, M.J., McLenachan, P.A., Gibb, G.C., Penny, D., 2006. Combined nuclear and mitochondrial protein-coding DNA sequences resolve the interrelations of the major Australasian marsupial radiations. Syst. Biol. 55, 122–137.

Poe, S., Chubb, A.L., 2004. Birds in a bush: five genes indicate explosive evolution of avian orders. Evolution 58, 404–415.

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

Rambaut, A., 1996. Se-Al: Sequence Alignment Editor. Available at <http://evolve.zoo.ox.ac.uk//>.

Rea, A.M., 1983. Cathartid affinities: a brief overview. In: Wilbur, S.R., Jackson, J.A. (Eds.), Vulture Biology and Management. University of California Press, Berkeley, California, pp. 26–54.

Reyes, A., Gissi, C., Catzeflis, F., Nevo, E., Pesole, G., Saccone, C., 2004. Congruent mammalian trees from mitochondrial and nuclear genes using Bayesian methods. Mol. Biol. Evol. 21, 397–403.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Seibold, I., Helbig, A.J., 1995. Evolutionary history of New and Old World vultures inferred from nucleotide sequences of the mitochondrial cytochrome *b* gene. Phil. Trans. R. Soc. Lond. B 350, 163–178.

Sibley, C.G., Ahlquist, J.E., 1990. Phylogeny and classification of birds. Yale University Press, New Haven, CT.

Slack, K.E., Janke, A., Penny, D., Arnason, U., 2003. Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene 302, 43–52.

Slack, K.E., Jones, C.M., Tatsuro, A., Harrison, G.L., Fordyce, R.E., Arnason, U., Penny, D., 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. Mol. Biol. Evol. 23, 1144–1155.

Sorenson, M.D., Ast, J.C., Dimcheff, D.E., Yuri, T., Mindell, D.P., 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Mol. Phylogenet. Evol. 12, 105–114.

Steel, M.A., Penny, D., 2004. Two further links between MP and ML under the Poisson model. Appl. Math Lett. 17, 785–790.

Swofford, D.L., 1998. PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods). Sinauer Associates Sunderland, Massachusetts.

Thomas, G.H., Wills, M.A., Székely, T., 2004. Phylogeny of shorebirds, gulls, and alcids (Aves: Charadri) from the cytochrome c gene: parsimony, Bayesian inference, minimum evolution and quartet puzzling. Mol. Phylogenet. Evol. 30, 516–526.

Van Tuinen, M., Sibley, C., Hedges, S.B., 2000. The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. Mol. Biol. Evol. 17, 451–457.

Van Tuinen, M., Waterhouse, D., Dyke, G.J., 2004. Avian molecular systematics on the rebound: a fresh look at modern shorebird phylogenetic relationships. J. Avian Biol. 35, 191–194.

Chapter 5. Summary and discussion

In a powerful sense the results of my published papers, including the appendices, stand alone as descriptions of my work. Nevertheless, and because my last paper was published in 2007, it is useful to have a short summary that covers my three main focuses of:

the position of the root of the avian tree,
solving some deep relationships within the avian tree, and
the timing of the origin/divergences of a number of the main groups of birds.

For each aspect I will first give a quick overview of my results with respect to avian evolution and, second, comment on what has been done more recently – are my main conclusions still supported? I will largely combine the first two topics because they are best considered together. Finally, I will briefly comment on possible new directions.

The first topic is relatively easy. When I began this research, some uncertainty had arisen over the position of the root of the avian tree. The classical hypothesis certainly placed the root between Paleognathae and Neognathae (see references and discussion in Sibley and Ahlquist 1990). The first alternative hypothesis was the 'Eoaves' suggestion of Sibley and Ahlquist (1990, see Figure 2 of my Introduction), which had Paleognathae and Galloanserae on one side of the root and Neoaves on the other side. Subsequently to Sibley and Ahlquist's work, some early trees from mitochondrial (mt) sequence data (see Figures 4 and 5 of the Introduction) placed the root within the Passeriformes.

In order to test the passerine rooting, I sequenced a lyrebird (a deep oscine), and was involved in the selection of a flycatcher (an early suboscine) and a rifleman (a New Zealand wren, and thought to be the deepest passerine branch) in order to break up long branches among Passeriformes. In the event, all my results supported the classic hypothesis of a Paleognathae/Neognathae split. The most likely explanation for the passerine rooting reconstructed by initial mt data analyses is simply insufficient taxon sampling (see Hendy and Penny 1989). This is a perfectly natural phenomenon when a new class of data is being analysed, and the number of taxa for which the new data is available is initially limited. In the passerine case, the situation was probably exacerbated by an apparent increase in mutation rate in that group – thus increasing the chance for 'long branch attraction' (Hendy and Penny 1989). Another potential confounding effect is a possible linkage between the faster rate of evolution in passerines and their higher speciation rate (Lanfear et al. 2010). Whatever the original reasons, the improved taxon sampling certainly strengthened the traditional placement of the root of the avian tree. Further investigation was deemed to be overkill - especially in light of the many other interesting hypotheses that remained to be tested.

I was not directly involved in sequencing any of the paleognaths and will therefore not discuss this group, except to mention the following. Both Harshman et al. (2009), using nuclear DNA sequences, and Phillips et al. (2010), using complete mt genomes, have shown that, whilst the tinamous do all group together they do <u>not</u> form the sister group to ratites -

but instead fall within ratites. Because the rate of molecular evolution in tinamous appears to be somewhat accelerated, previous molecular studies placing tinamous as the sister group to the ratites may be the result of long branch attraction. The clustering of tinamous within the ratites requires the reevaluation of two main areas of paleognath evolution – firstly, whether the ancestor of ratites was or was not flightless, and secondly, in understanding biogeographic distribution of the paleognaths across the Southern Hemisphere – important questions for the future.

Given the support for the traditional placement of the root, and turning now to some of the other deep divisions within birds, it was initially unclear whether the landfowl and the waterfowl orders united to form Galloanserae, or whether they formed separate branches. I sequenced a goose for the Anseriformes lineage, and a brush turkey (a megapode) as a relatively deep representative of the Galliformes lineage. In addition, I helped arrange for a magpie goose (*Anseranas*) to be sequenced in order to both further test the Galliformes/ Anseriformes grouping and to help provide a good calibration point with the *Vegavis* fossil from Vega Island, Antarctica (see Chapter 1). All my results supported the union of the land- and waterfowl lineages to form Galloanserae. This gives a main three-way split within birds; firstly the Paleognathae/Neognathae split and secondly with Neognathae being split into Galloanserae/Neoaves. I am pleased to have assisted in this conclusion. Perhaps the only deep division left to test within Galloanserae is the 'screamers' (Anhimidae; Anseriformes) and it does seem that the complete mt genome of at least one representative of this group should be sequenced as soon as possible.

Within the Neoaves, my main interest was in sequencing mt genomes from birds with good fossil dates, and their close relatives. The more closely the taxa in the dataset bracket the fossils (the tighter the bounds), the more accurate the fossils are as calibration points for avian evolution. It is clear that the penguin lineage existed over 60 MYA, and that places a good lower bound on its age. But what is the upper bound? The lower bound is important with respect to the theory, discussed again later, that modern birds were diversifying in the Late Cretaceous, at the same time smaller pterosaurs, in particular, were declining. An upper bound is more difficult to establish, but determining the closest relatives of penguins helps significantly. I therefore sequenced several complete mt genomes including the little blue penguin (*Eudyptula minor*), a small albatross (black-browed mollymawk, *Diomedea melanophris*), a Kerguelen petrel (*Pterodroma brevirostris*) and the red-throated loon (*Gavia stellata*). In the case of the loons, there is also the open question of the nature and age of their early fossils (see Chapter 1), and although outside the scope of my study, it is important to further clarify this.

In addition, I also sequenced the complete mt genome of the southern black-backed gull (*Larus dominicanus*), but this was more to test the prediction, in this case, from fossil avian footprints, that some 'shorebird-like' birds were present in the Late Cretaceous (Lockley 1998, see also Chapter 1). In this respect, my results for the shorebird divergence times

certainly mean that it is possible, even likely, that the avian fossil footprints were indeed generated by early shorebirds (Charadriiformes). This is certainly the simplest hypothesis - the existence of any hypothetical non-charadriiform group that left neither modern descendents nor any fossils other than footprints does not need to be predicted. There is no evidence from molecular studies for the suggestion (Feduccia 1995; 1999) that all modern birds (or at least Neoaves) are derived from 'transitional shorebirds'.

My conclusion, particularly using the *Vegavis* and penguin (*Waimanu*) fossils as calibration points, is that the main divisions/Orders of birds arose in the Late Cretaceous, concurrent with the decline of the smaller pterosaurs (Slack et al. 2006, Chapter 3). A reason for focusing on this area was that it reflects on the mechanisms leading to macroevolution. Are there any special requirements for macroevolution, or are the normal microevolutionary processes sufficient (see Penny and Phillips 2004)? There are still some authors who apparently see 'explosive radiations', by unknown mechanisms, as necessary for macroevolution (see Gibb et al. 2007 for a critique). Similarly, other authors seem to search for a physical factor to be a 'driver' of evolution (including extraterrestrial impacts and/or climate change), a somewhat 'deterministic' view of evolution. My findings support the equivalence of micro- and macroevolution. However, further testing is certainly another interesting area for future research.

As far as dates and timings of early Neoavian divergences are concerned, recent work has tended to support my interpretations of a Late Cretaceous divergence of quite a few Neoavian groups. Brown et al. (2007) showed that it was important to use good fossil calibrations to support conclusions, and followed this up (Brown et al. 2008) with additional sequence data. The most recent major publication is Pacheco et al. (2011), who add 17 additional avian mt genomes. Their findings are fully concordant with mine in that the major lineages of birds appear to arise during the Late Cretaceous. Even with incomplete sampling (for example, rails are not included), their results suggest that least 30 lineages of Neoaves seem to have survived from the Cretaceous to the present. The results of Pacheco et al. (2011) are supported by several other studies. For example, Jiang et al. (2010) sequenced two additional anseriforms, and used Bayesian methods to conclude that at least one anseriform divergence occurred more than 90 MYA. Kan et al. (2010) carried out a similar study on Galliformes, and reported that the megapodes, at least, began diverging in the Late Cretaceous.

It is important to try new forms of sequence data, especially nuclear data, to evaluate the hypotheses generated here. Unfortunately, the first major set of nuclear data used to study deep avian lineages was from the 7th intron of the β-fibrinogen gene (Fain and Houde 2004), They reported a new supposed division within the Neoaves – a suggested 'Coronaves/ Metaves' split. However, Morgan-Richards et al. (2008) challenged this main conclusion because the 7th intron of the β-fibrinogen gene nearly doubled in length during Fain and Houde's alignment procedures, leaving a dataset with no constant sites. I always used a conservative alignment procedure (and others in the Massey laboratory followed my

procedures) and deleted sites around a gap until two or more consecutive constant sites were obtained. If my more conservative alignment procedures were used on the Fain and Houde (2004) data, there would be no sites left for analysis! Despite this, a major study using many intron sequences (but also with some exon data) was undertaken by Hackett et al. (2008). While much longer nuclear sequences are highly desirable, perhaps a more productive way to go in the future is to test deeper avian divergences using exon data, and only to use introns for more recent divergences. Finally, Suh et al. (2011) have very recently used retroposon data from nuclear sequences to start resolving splits within the Neoaves – one very interesting relationship that they inferred was a parrot/passerine grouping.

Future technological developments are expected to markedly increase the amount of DNA sequence data available for analysis. Next generation (Next-Gen) sequencing techniques are becoming widespread, and third generation DNA sequencing should also become available soon. There are already some interesting studies published (e.g. Kunster et al. 2010) involving sequencing large volumes of transcriptome data for a range of avian taxa. Next-gen sequence data for the great tit (*Parus major*, Passeriformes, Santure et al. 2011) has also been generated. Providing that the same tissues (or early embryos, as in the case of kiwi [Subramanian et al. 2010]) are sequenced, directly comparable sequences for many genes will be available. These advances and the refinement of techniques (based on the number of trees) for testing prior hypotheses (see Pratt et al. 2009) will allow further independent testing of the primary conclusions developed here – which is certainly good science. It is also to be hoped that having much longer datasets will enable conflicting hypotheses with regards to avian relationships to be resolved (see for example, Kennedy et al. 2005).

Far more interesting in the long run will be the chance to move the next step beyond phylogeny, and learn much more about biogeographic distributions and the macroecological changes that have occurred over time. There have already been some studies along these lines - for example, Kennedy et al. (2000) considered how phylogeny might relate to life history parameters, while Gibb et al. (2012) suggest that a group of ‘water carnivore’ birds (including Pelecaniformes and Ciconiiformes) appear to have occupied a somewhat similar niche since the Late Cretaceous and discuss long-term niche stability. Thus while it has been very interesting solving some aspects of avian phylogeny, I expect that the new questions of the future will be equally as interesting and that these will eventually be answered as more data becomes available.

References

Brown, J.W., Payne, R.B., and Mindell, D.P. (2007). Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on Ericson et al. *Biol. Lett*. **3**: 257-259.

Brown, J.W., Rest, J.S., Garcia-Moreno, J., Sorenson, M.D., Mindell, D.P. (2008). Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biol*. **6**: 6.

Fain, M.G., and Houde, P. (2004). Parallel radiations in the primary clades of birds. *Evolution* **58**: 2558-2573.

Feduccia, A. (1995). Explosive evolution in Tertiary birds and mammals. *Science* **267**: 637-638.

Feduccia, A. (1999). *The origin and evolution of birds. Second edition*. New Haven, CT: Yale University Press.

Gibb, G.C, Kardailsky, O, Kimball, R., Braun, E. and Penny, D. (2007). Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. *Mol. Biol. Evol*. **24**: 269-280.

Gibb, G.C, Kennedy, M, and Penny, D. (2012). Beyond phylogeny: Pelecaniform and Ciconiiform birds, and long-term niche stability. *Syst. Biol*. (in revision).

Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, E.L. et al. (2008). A phylogenomic study of birds reveals their evolutionary history. *Science* **320**:1763-1768.

Harshman, J., Braun, E.L., Braun, M.J., Huddleston, C.J., Bowie, R.C.K., et al. 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. *Proc. Natl Acad. Sci. USA* **105**:13462-13467.

Hendy, M.D. and Penny, D. (1989). A framework for the quantitative study of evolutionary trees. *Syst. Zool*. **38**: 297-309.

Jiang, F.W., Miao, Y.W., Liang, W., Ye, H.Y., Liu, H.L., Liu, B. (2010). The complete mitochondrial genomes of the whistling duck (*Dendrocygna javanica*) and black swan (*Cygnus atratus*): dating evolutionary divergence in Galloanserae. *Mol. Biol. Rep*. **37**: 3001-3015.

Kan, X.Z., Li, X.F., Lei, Z.P., Chen, L., Gao, H., et al. (2010). Estimation of divergence times for major lineages of galliform birds: evidence from complete mitochondrial genome sequences. *Afri. J. Biotech*. **9**: 3073-3078.

Kennedy, M., Gray, R.D., Spencer, H.G. (2000). The phylogenetic relationships of the shags and cormorants: can sequence data resolve a disagreement between behavior and morphology? *Mol. Phylog. Evol*. **17**: 345-359.

Kennedy, M., Holland, B.R., Gray, R.D., Spencer, H.G. (2005). Untangling long branches: identifying conflicting phylogenetic signals using spectral analysis, neighbor-net, and consensus networks. *Syst. Biol*. **54**: 620-633.

Kunstner, A., Wolf, J.B.W., Backstrom, N., Whitney, O., Balakrishnan, C.N., et al. (2010). Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Mol. Ecol*. **19**: S266–S276.

Lanfear, R., Ho, S.Y.W., Love, D., Bromham, L. (2010). Mutation rate is linked to diversification in birds. *Proc. Natl Acad. Sci. USA*. **107**: 20423-20428.

Lockley, M.G. (1998). The vertebrate track record. *Nature* **396:** 429–432.

Morgan-Richards, M., Trewick, S.A., Bartosch-Härlid, A., Kardailsky, O., Phillips, M.J., McLenachan, P.A., and Penny, D. (2008). Bird evolution: testing the Metaves clade with six new mitochondrial genomes. *BMC Evol. Biol*. **8**: 20.

Pacheco, M.A., Battistuzzi, F.U., Lentino, M., Aguilar, R.F., Kumar, S., and Escalante, A.A. (2011). Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. *Mol. Biol. Evol*. **28**: 1927–1942.

Penny D. and Phillips, M.J. (2004). The rise of birds and mammals: are microevolutionary processes sufficient for macroevolution. *Trends Ecol. Evol.* **19**: 516-522.

Phillips, M.J., Gibb, G.C., Crimp, E.A. and Penny, D. (2010). Tinamous and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites. *Syst. Biol*. **59**: 90–107.

Pratt, R.C., Gibb, G.C., Morgan-Richards, M., Phillips, M.J., Hendy, M.D. and Penny, D. (2009). Towards resolving deep Neoaves phylogeny: data, signal enhancement and priors. *Mol. Biol. Evol*. **26**: 313-326.

Santure, A.W, Gratten, J., Mossman, J.A., Sheldon, B.C., Slate, J. (2011) Characterisation of the transcriptome of a wild great tit *Parus major* population by next generation sequencing. *BMC Genomics*. **12**: 283.

Sibley, C.G. and Ahlquist, J.E. (1990). *Phylogeny and classification of birds: a study in molecular evolution*. New Haven, CT: Yale University Press. ISBN 0-300-04085-7.

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

Subramanian, S., Huynen, L., Millar, C.D., Lambert, D.M. (2010). Next generation sequencing and analysis of a conserved transcriptome of New Zealand's kiwi. *BMC Evol. Biol*. **10**: 387

Suh, A., Paus, M., Kiefmann, M., Churakov, G., Franke, F.A., Brosius, J., Kriegs, J.O. and Schmitz J. (2011). Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat. Comm*. **2:** 44.. DOI: 10.1038/ncomms1448.

Appendix 1. Statements from collaborators

Professor Margaret Tennant, Chair
Doctoral Research Committee,
Massey University,
Palmerston North.

Dear Margaret,

Kerryn Slack PhD thesis

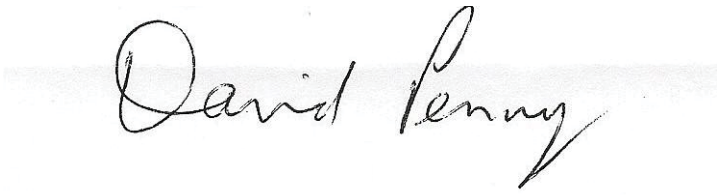
This letter is to confirm the statements that Kerryn has made in this thesis about her contributions to the published papers. She has incorporated them with each publication (chapter or appendix). My statement is also made on behalf of the New Zealand and Australian based molecular systematists, namely Abby Harrison, Trish McLenachan, Matt Phillips, and Alan Cooper.

I have read all the statements that Kerryn makes about her contributions and certify that they are accurate. She was always been fully involved in the selection of taxa for sequencing, searched for taxa that split up 'long branches' that might lead to artefacts in the analysis, did the extensive DNA sequencing that she describes, carried out the analyses, and was always fully involved in writing up the manuscripts. We have been very pleased with her performance, and it is a tribute to her skills and analysis that the papers have been so widely cited – she has obviously met a need for good accurate information that other scientists require.

I have also asked Professor Bengtsson of Lund University (Sweden) and Professor Fordyce of Otago University for equivalent statements on behalf of the Swedish collaborators (Prof Bengtsson) and the New Zealand paleontologists (Prof Fordyce) respectively.

I would be happy to provide any additional information that you might require.

Yours sincerely,



David Penny, PhD (Yale), FRSNZ, CNZM
Distinguished Professor of Theoretical Biology

LUND
UNIVERSITY

Department of Biology
Genetics

Professor Margaret Tennant, Chair,
Doctoral Research Committee,
Massey University,
Palmerston North,
New Zealand

Re: Kerryn Slack's PhD thesis

Dear Professor Tennant,

I was Kerryn's co-supervisor while she was enrolled for a PhD at Lund University. Professor Ulfur Arnason was her primary supervisor, but he has retired and now unfortunately does not have any association with the university (just in case you need it, his e-mail address is ulfur.arnason@gmail.com). Dr Axel Janke was also involved in Kerryn's supervision, but he has now moved to Germany (Goethe University, Senckenberg, Frankfurt).

I have read the statements of the contributions of Professor Arnason and Drs Axel Janke and Bjorn Ursing to Kerryn's manuscripts, and can confirm that Kerryn did all the avian DNA sequencing and analysis herself under Prof Arnason's (and Axel Janke's) guidance. In addition to the avian work, and as Kerryn says in her acknowledgements, Bjorn Ursing completed and largely wrote up the alpaca mitochondrial genome that appears in Appendix 2.

Perhaps most importantly, I can confirm, that for personal reasons Kerryn had to return to New Zealand and that she never completed and submitted her PhD thesis to Lund University. However, she did extremely well in her studies here and was close to submission when she departed. I am therefore very pleased that she is now finishing her thesis and I wish her all the best for the future.

Lund, 5 September 2011

Bengt Olle Bengtsson
Professor of Genetics
Department of Biology
Lund University

(A signed copy has been air mailed directly to the chair of the Doctoral Research Committee)

Professor Margaret Tennant, Chair,
Doctoral Research Committee,
Massey University,
Palmerston North,
New Zealand

Dear Professor Tennant,

Kerryn Slack PhD thesis

This letter is to confirm our relative contributions to our joint paper on penguin evolution. Perhaps the most direct approach is for me to first quote the two paragraphs from Kerryn's description of our paper.

"The reason for concentrating on the penguins, and their relatives, was that we were aware that an older fossil penguin had been found in North Canterbury, but the details (including the dating) had not been published. Because of the importance of this fossil for the dating analysis, we approached Professor Ewan Fordyce and Craig Jones about a joint publication.

Craig Jones (currently of IGNS, Gracefield, Lower Hutt), Tatsuro Ando and Professor Ewan Fordyce (of the Geology Department at the University of Otago) fully analysed and described the penguin fossils (*Waimanu*), as well as the microfossils that were used for dating the fossil penguins. We then used their results as a calibration point for early avian evolution. Some of the early interactions were at IGNS in Gracefield (Lower Hutt), but we twice held day-long meetings at Te Papa in Wellington where Professor Fordyce came from Dunedin, Craig Jones from Gracefield, and David Penny and myself from Palmerston North."

Yes, this is certainly an accurate description of our contribution. As paleontologists, the three of us did the full analysis of the macro and microfossils (and their drawings), but we certainly did not get involved in the DNA analysis! It is good to see collaboration between disciplines.

Yours sincerely

Prof Ewan Fordyce
Head, Department of Geology, University of Otago

Appendix 2.

Ursing, B. M., **Slack, K. E**. and Arnason, U. (2000). Subordinal artiodactyl relationships in the light of phylogenetic analysis of 12 mitochondrial protein-coding genes. *Zool. Scr.* **29**: 83-88. (As at August 2011 this paper has been cited 17 times in the Web of Science.)

In this paper the alpaca mt genome was reported. While at the University of Lund in Sweden I learned the basic techniques for sequencing whole mitochondrial genomes, using alpaca DNA that I had extracted. Once I had mastered those techniques, I changed to my primary interest of avian evolution. Bjorn Ursing then completed the alpaca mt genome, and did the main writing of the manuscript (assisted by myself). Bjorn, deservedly, is the senior author on this manuscript. Ulfur Arnason was involved in project design, supervision, and writing.

Subordinal artiodactyl relationships in the light of phylogenetic analysis of 12 mitochondrial protein-coding genes

BJÖRN M. URSING[1], KERRYN E. SLACK & ULFUR ARNASON

Accepted 15 October 1999

Ursing, B. M., Slack, K. E. & Arnason, U. (2000) Subordinal artiodactyl relationships in the light of phylogenetic analysis of 12 mitochondrial protein-coding genes. — *Zoologica Scripta*, *29*, 83–88.

Extant artiodactyls (even-toed hoofed mammals) are traditionally divided into three main lineages: Suiformes (pigs, peccaries and hippopotamuses), Tylopoda (camels and llamas) and Ruminantia (bovids, deer, tragulids and giraffes). Recent molecular studies have not supported a close relationship between pigs and hippopotamuses, however, instead grouping hippopotamuses with Cetacea (whales, dolphins and porpoises). In this study we have sequenced the complete mitochondrial genome of a tylopod — the alpaca (Lama pacos), the only artiodactyl suborder not previously represented by a complete mitochondrial sequence. This sequence was included in phylogenetic analyses together with the complete mitochondrial protein-coding sequences of other artiodactyls plus two cetaceans. Despite the length of the data set, the relationship between Suina (Suiformes sine Hippopotamidae), Tylopoda and Ruminantia/Hippopotamidae/Cetacea could not be fully resolved, however, a basal position of the alpaca (Tylopoda) relative to the other artiodactyls/cetaceans was unsupported.

Björn M. Ursing, Kerryn E. Slack & Ulfur Arnason, Department of Genetics, Lund University, Sölvegatan 29, S-223 62 Lund, Sweden.

Introduction

There are approximately 80 genera and 10 families of extant artiodactyls (even-toed hoofed mammals). In classical systematics, Artiodactyla is usually divided into three suborders: Suiformes (pigs, peccaries and hippopotamuses), Tylopoda (camels and llamas) and Ruminantia (bovids, deer, tragulids and giraffes) (Colbert & Morales 1991). Artiodactyl relationships at various taxonomic levels have attracted considerable attention (e.g. see Beintema *et al*. 1977; Miyamoto *et al*. 1993; Douzery & Catzeflis 1995; Randi *et al*. 1996; Kleineidam *et al*. 1999) and the relationship between artiodactyls and cetaceans (whales, dolphins and porpoises) has also been the subject of a number of molecular studies (Graur & Higgins 1994; Irwin & Arnason 1994; Arnason & Gullberg 1996; Gatesy *et al*. 1996; Gatesy 1997; Montgelard *et al*. 1997; Shimamura *et al*. 1997; Ursing & Arnason 1998a).

Molecular studies have, almost without exception, identified a close relationship between artiodactyls and cetaceans and some have placed extant cetaceans within Artiodactyla as the sister group of Hippopotamidae (Irwin & Arnason 1994; Arnason & Gullberg 1996; Gatesy 1997; Montgelard *et al*. 1997; Ursing & Arnason 1998b; Gatesy *et al*. 1999). These studies have also disrupted the traditional Suiformes grouping by removing Hippopotamidae from the suiform lineage. The sister group relationship between Hippopotamidae and Cetacea received very strong support in a recent study based on the analysis of the complete set of 12 heavy (H) strand encoded mitochondrial (mt) protein-coding genes (Ursing & Arnason 1998b). In the present study, the order Cetacea, the suborder Ruminantia and the family Hippopotamidae are regarded as a monophyletic group, which is referred to as Cetruminantia in conjunction with the phylogenetic nomenclature of de Queiroz & Gautier (1994). This group includes all descendants of the most recent common ancestor of Ruminantia and Hippopotamidae/Cetacea. The term Suina is used for Suiformes *sine* Hippopotamidae.

The artiodactyls included in Ursing & Arnason's (1998b) study were the pig, two ruminants (the cow and the sheep) and the hippopotamus. Thus one of the three artiodactyl suborders — Tylopoda — was not represented. The inclusion of a complete tylopod mt sequence in the artiodactyl/cetacean analysis is of considerable interest, as molecular and morphological studies

[1]Present address: Department of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. E-mail: ursing@chem.rug nl

to date have to some extent provided conflicting theories on the relationship between Suiformes, Tylopoda and Ruminantia.

Webb & Taylor (1980) grouped Tylopoda and Ruminantia together into Neoselenodonta, to the exclusion of Suiformes, on the basis of morphological data. This relationship was supported by the molecular studies of Miyamoto *et al.* (1993) and Kleineidam *et al.* (1999). Gatesy *et al.* (1999), on the other hand, in a combined study of mt and nuclear data favoured the positioning of Camelidae (Tylopoda) as the sister group of the remaining cetartiodactyls (cetaceans and artiodactyls). Still other molecular studies (Arnason & Gullberg 1996; Montgelard *et al.* 1997; Shimamura *et al.* 1997) have left the basal artiodactyl divergences unresolved. In order to examine these relationships further without the stochastic effects associated with limited sequence data (Cao *et al.* 1994), we have sequenced the complete mt genome of a tylopod — the alpaca (*Lama pacos*). Thus this study is based on the largest mt data set used to date to examine basal artiodactyl relationships.

Materials and methods

An enriched mtDNA fraction was extracted from a juvenile alpaca kidney following the procedure of Arnason *et al.* (1991). The tissue was a gift from Ernst Jung of Rolsberga, Sweden. The mtDNA was digested separately with *Spe* I and *Bln* I and the resulting fragments were ligated into phage M13mp18 and/or M13mp19 and transformed into *E. coli* strain DH5α. Regions not successfully cloned in this manner were PCR-amplified and then cloned. Sequencing was carried out using two approaches: (i) 'Manual' sequencing of single-stranded DNA using the dideoxy method with ^{35}S-dATP (Sanger 1981) and (ii) 'Automatic' sequencing using the Thermo-Sequenase fluorescent-labelled primer cycle-sequencing kit (Amersham) with 7-deaza-dGTP and DNA sequencer model 4000 L (LICOR Inc.). Both universal and numerous specific oligonucleotide primers were employed. The sequences of the regions covered by cloned PCR products represent the consensus of a minimum of three clones.

The complete mtDNA sequence of the alpaca has been deposited at the EMBL database under accession number Y19184. Users of this sequence are kindly requested to refer to this publication and not to the accession number alone.

The data set used in this study consisted of sequences from 18 eutherian mammals, seven of which represented Artiodactyla/Cetacea. These taxa were: alpaca, *Lama pacos* (this study); pig, *Sus scrofa* (Ursing & Arnason 1998a); hippopotamus, *Hippopotamus amphibius* (Ursing & Arnason 1998b); cow, *Bos taurus* (Anderson *et al.* 1982); sheep, *Ovis aries* (Hiendleder *et al.* 1998); fin whale, *Balaenoptera physalus* (Arnason *et al.* 1991); blue whale, *B. musculus* (Arnason & Gullberg 1993); harbour seal, *Phoca vitulina* (Arnason & Johnsson 1992); grey seal, *Halichoerus grypus* (Arnason *et al.* 1993); cat, *Felis catus* (Lopez *et al.* 1996); horse, *Equus caballus* (Xu & Arnason 1994); donkey, *E. asinus* (Xu *et al.* 1996a); Indian rhinoceros, *Rhinoceros unicornis* (Xu *et al.* 1996b); white rhinoceros, *Ceratotherium simum* (Xu & Arnason 1997); mole, *Talpa europea* (Mouchaty *et al.*, in press); armadillo, *Dasypus novemcinctus* (Arnason *et al.* 1997); human ('Lund'), *Homo sapiens* (Arnason *et al.* 1996); mouse, *Mus musculus* (Bibb *et al.* 1981).

The alignment used consisted of the concatenated sequences of the 12 protein-coding genes encoded by the mt H strand. The light (L) strand encoded NADH6 gene was not included as the nucleotide (nt) composition of this gene differs from that of the other mt protein-coding genes, thereby violating the assumptions of some of the phylogenetic algorithms used. Overlapping regions of ATPase subunits 6 and 8 and of NADH subunits 4 and 4 L were also excluded. After these exclusions and the removal of gaps and ambiguous sites adjacent to gaps, the length of the alignment was 10 554 nt or 3518 amino acids (aa).

Analyses were carried out on both the nt and the aa data, using three different approaches to phylogenetic reconstruction: maximum likelihood (ML; Felsenstein 1981), neighbour joining (NJ; Saitou & Nei 1987) and maximum parsimony (MP; Fitch 1971). The program packages used were PUZZLE version 4.0 (Strimmer & von Haeseler 1996), PHYLIP version 3.52c (Felsenstein 1991) and MOLPHY version 2.3 (Adachi & Hasegawa 1996a). The nt analyses included all non-synonymous 1st codon position changes, all 2nd codon position changes and 3rd codon position transversions.

The NJ analyses were carried out using the HKY (Hasegawa-Kishino-Yano) model (Hasegawa *et al.* 1985) for nt sequence evolution and the Dayhoff matrix (Dayhoff 1978) for aa sequence evolution. In both the NJ and MP analyses, support values were calculated from 1000 nt and 100 aa bootstrap replicates. The ML/QP (quartet puzzling) analyses were based on 1000 puzzling steps, using the TN (Tamura-Nei) model (Tamura & Nei 1993) for nt sequence evolution and the mtREV-24 model (Adachi & Hasegawa 1996b) for aa sequence evolution.

Results

The organization of the alpaca mt genome, which is 16 652 nt long, is consistent with that of other complete eutherian mt genomes. All of the alpaca protein-coding genes have a methionine (ATG) start codon except for NADH3, which starts with isoleucine (ATT), and NADH4L, which starts with valine (GTG). As in a number of other mammals, the stop codons of the COIII, NADH3 and NADH4 genes are incomplete (TA or T).

The control region of the alpaca contains a 30 nt tandem repeat motif. This motif occurred six times in the sequenced clone, with five of the repeats being identical while the sixth differed from the others by three transitions. Control region repeats are also found in the sheep (Hiendleder *et al.* 1998) and the pig (Ursing & Arnason 1998a), but are not present in the cow (Anderson *et al.* 1982), the hippopotamus (Ursing

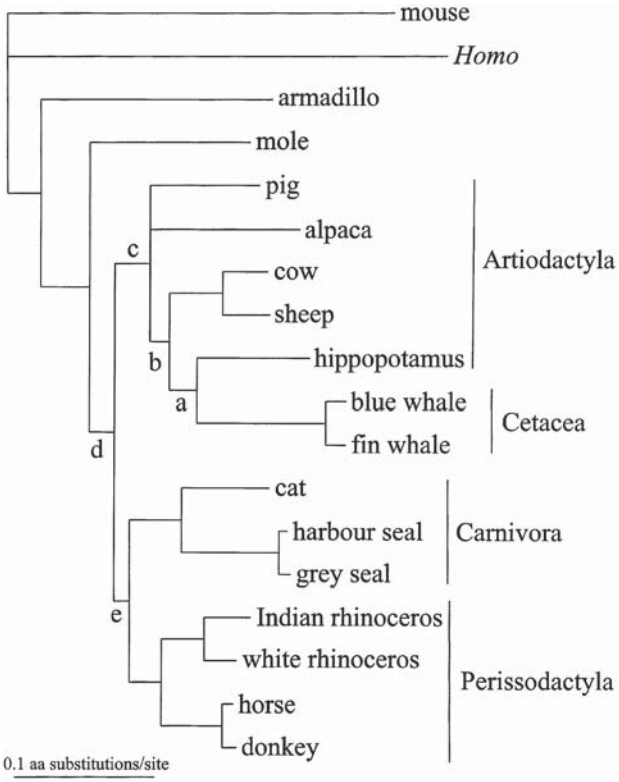


Fig. 1 Maximum likelihood tree from amino acid (aa) data from the 12 H strand-encoded mitochondrial protein-coding genes. The alignment used was 3518 aa long and did not include the overlapping regions of ATPase subunits 6 and 8 and NADH subunits 4 and 4 L nor gaps and ambiguous sites adjacent to gaps. The tree was reconstructed using PUZZLE version 4.0 (Strimmer & von Haeseler 1996) and the mtREV-24 model for aa sequence evolution (Adachi & Hasegawa 1996b). Support values for the branches labelled a-e are given in Table 1. The lengths of the branches are proportional to the number of amino acid substitutions per site. The relationship between Suina, Tylopoda and Cetruminantia has been shown as unresolved as the same relationship was not reconstructed in all analyses.

Table 1 Support values for the labelled branches (A-E) in Fig. 1.

Method	Data set	A	B	C	D	E
ML/QP	aa	91	74	85	87	81
	nt	96	46	75	77	62
MP	aa	97	85	89	96	90
	nt	100	65	73	95	90
NJ	aa	100	99	98	98	96
	nt	100	92	89	100	100

The ML/QP support values were established using 1000 puzzling steps. The MP and NJ support values were calculated from 1000 nt bootstrap replicates and 100 aa bootstrap replicates.

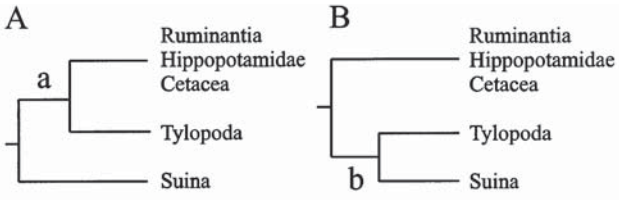


Fig. 2 Two of the three possible rooted topologies for the relationship between Suina, Tylopoda and Cetruminantia. Support values for the branches labelled a and b are given in Table 2.

Table 2 Support values for the branches labelled a and b in Fig. 2.

Method	Data set	a	b
QP/ML	aa	37	51
	nt	10	73
MP	aa	25	66
	nt	73	8
NJ	aa	76	6
	nt	98	0

The support values were established as for Table 1.

& Arnason 1998b) or the fin and blue whales (Arnason *et al*. 1991; Arnason & Gullberg 1993). The repeats of the alpaca, the sheep and the pig are highly dissimilar and therefore have limited phylogenetic value for subordinal artiodactyl comparisons. The alpaca L strand origin of replication forms a hairpin structure with an 11-bp stem and a 13 nt loop, similar to the structures formed by the other artiodactyls.

The same phylogenetic tree was reconstructed in all analyses (see Fig. 1) with the exception of the Suina/Tylopoda/Cetruminantia relationship (which has therefore been shown as unresolved in Fig. 1). Table 1 shows the support values from the different methods of analysis/data sets for the branches labelled a-e in Fig. 1. Of the three possible rooted topologies for the relationship between Suina, Tylopoda and Cetruminantia, the topology with Tylopoda as the sister group of Suina/Cetruminantia was virtually unsupported (see Fig. 2 and Table 2).

Six selected cetartiodactyl topologies were tested against each other using both ML (Kishino-Hasegawa test; Kishino & Hasegawa 1989) and MP (Templeton test; Templeton 1983). The tested topologies are shown in Fig. 3. The first four (topologies A-D) are identical except for the position of the alpaca. Topologies A-C represent the three possible rooted relationships between Suina, Tylopoda and Cetruminantia, while topology D places Tylopoda and Ruminantia (the cow and the sheep) in a sister group relationship. The last two topologies (E and F) represent two of the traditional views

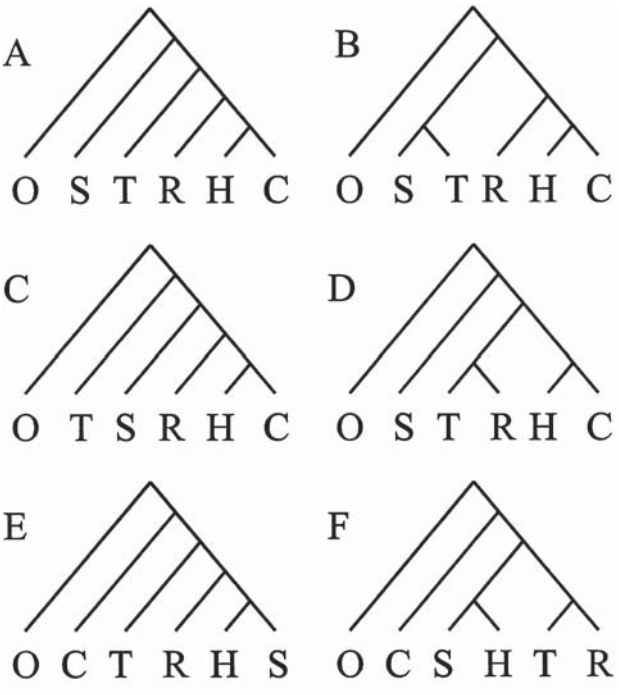


Fig. 3 Six selected rooted topologies for cetartiodactyl relationships. The relative level of support for each of these topologies is shown in Table 3. O: outgroup; S: Suina (pig); T: Tylopoda (alpaca); R: Ruminantia (cow, sheep); H: Hippopotamidae (hippopotamus); C: Cetacea (fin whale, blue whale).

on intraordinal artiodactyl relationships, with Artiodactyla forming a monophyletic group.

The results of the tests are given in Table 3. Topology A (Suina (Tylopoda, Cetruminantia)) was optimal in three out of the four tests while topology B ((Suina, Tylopoda) Cetruminantia) was optimal in the fourth. Topology C (Tylopoda (Suina, Cetruminantia)) could be significantly rejected (> 2 S. E. or S. D. in three of the four tests and > 1 S. E. in the fourth) and topology D, which grouped Tylopoda and Ruminantia together, was rejected in the Templeton-aa test and in both ML tests. The remaining topologies (E and F) could be conclusively rejected in all four tests.

Consistent with an earlier study based on complete mt protein-coding genes (Ursing & Arnason 1998b), a sister group relationship between Hippopotamidae and Cetacea was strongly supported, with support values ranging from 91 to 100 (see Table 1 and Fig. 1, branch a). Thus the inclusion of the alpaca did not affect this particular relationship. The Cetruminantia (Ruminantia/Hippopotamidae/Cetacea grouping) was also reasonably to strongly supported in most of the analyses (see Table 1 and Fig. 1, branch b).

The remaining cetferungulate (Carnivora, Perissodactyla, Artiodactyla and Cetacea) relationships in the tree shown in Fig. 1, and the support values for these relationships (Table 1), were consistent with those found in other studies of complete mt genomes. The sister group relationship between Perissodactyla and Carnivora (Xu *et al*. 1996b) received a generally high level of support (see Table 1 and Fig. 1, branch e).

Discussion

This study unequivocally supports the findings of Irwin & Arnason (1994) and other subsequent molecular studies that Cetacea and Hippopotamidae form a sister group relationship to the exclusion of the other artiodactyls and that Cetacea is nested deeply within Artiodactyla. Thus the inclusion of a

Table 3 Phylogenetic tests evaluating the relative level of support for the six topologies (A-F) shown in Fig. 3.

	Maximum Likelihood (Kishino-Hasegawa) test						**Maximum Parsimony (Templeton) test**			
	amino acid			nucleotide			amino acid		nucleotide	
	ΔlnL	S.E.	pBoot	ΔlnL	S.E.	pBoot	Δsteps	S.D	Δsteps.	S.D
A	<–39674.1>		0.499	< – 34813.9 >		0.707	8	8.5	<12457>	
B	– 2.7	18.5	0.439	– 12.6	18.0	0.226	<5701>		20	15.0
C	– 20.9	16.6	0.027	– 32.6	15.5	0.002	16	7.7	36	14.5
D	– 31.5	18.2	0.035	– 22.1	15.3	0.065	14	11.0	4	12.0
E	– 215.6	42.1	0.000	– 221	35.7	0.000	72	14.7	175	23.9
F	– 201.9	41.9	0.000	– 194	36.3	0.000	64	14.9	149	23.7

The ML (Kishino-Hasegawa) tests (Kishino & Hasegawa 1989) were carried out using the mtREV-24 model (Adachi & Hasegawa 1996b) for amino acid sequence evolution and the TN (Tamura-Nei) model (Tamura & Nei 1993) for nucleotide sequence evolution. Values in angled brackets represent the log likelihood (lnL) values of the best tree. ΔlnL shows the difference between the lnL value of the best tree and that of each of the other trees, followed by the standard error (S.E.) (Kishino & Hasegawa 1989). The bootstrap probability (pBoot) (Kishino *et al*. 1990) for each topology is also shown. The lnL, S.E. and pBoot values were calculated using NucML and ProtML in the MOLPHY program package, version 2.3 (Adachi & Hasegawa 1996a). The Templeton tests (Templeton 1983) were performed using ProtPars and DNAPars in the phylip program package, version 3.52c (Felsenstein 1991). Values in angled brackets represent the number of steps required for the best tree, while Δsteps indicates how many more steps are required for each nonoptimal tree and S. D. gives the standard deviation.

member of the third artiodactyl suborder (the alpaca; Tylopoda) in the mt analyses does not affect these relationships. These findings indicate that both Artiodactyla and Suiformes are paraphyletic and challenge the earlier hypothesis that Artiodactyla and Cetacea arose independently from the mesonychids (Van Valen 1967).

While the monophyly of both Cetruminantia and Cetartiodactyla (Table 1, branches b and c) are reasonably well supported in this study, the basal divergences among Cetartiodactyla, i.e. the branching order of Suina, Tylopoda and Cetruminantia, could not be conclusively resolved. Of the three possible rooted topologies for the relationship between the three groups, one (a sister group relationship between Tylopoda and Suina/Cetruminantia) was significantly worse than the other two. In addition to other phylogenetic conclusions this finding further underlines the molecular distinction between Suina and Hippopotamidae, two groups which are both included in Suiformes under the traditional classification scheme.

One possible explanation to the lack of complete resolution of the relationship between Suina, Tylopoda and Cetruminantia may be a relatively rapid artiodactyl divergence compared to the subsequent time of separate evolution (≈ 60 million years). This may also aid in explaining why previous analyses based on shorter sequences (Irwin & Arnason 1994; Arnason & Gullberg 1996) failed to resolve the basal artiodactyl divergences even though these studies identified the sister group relationship between Hippopotamidae and Cetacea.

The findings of this study do not agree with those of two other recent studies: Kumar & Hedges (1998), who reconstructed a sister group relationship between pigs and cetaceans, and Gatesy *et al*. (1999), who identified Camelidae as the sister group of the other cetartiodactyls. Neither of these topologies was favoured in the present study, which was based on a data set consisting of considerably longer sequences than those analysed by Kumar & Hedges (1998) and Gatesy *et al*. (1999). While the pig/cetacean relationship identified by Kumar & Hedges (1998) has not been supported by other recent analyses, it remains to be seen whether additional nuclear data from a broader range of taxa will tip the scales in favour of a sister group relationship between Tylopoda and the other extant cetartiodactyls, as reconstructed by Gatesy *et al*. (1999), or whether a less basal position for Tylopoda will be supported, as in this study.

Acknowledgements

We would like to express our gratitude to Ernst Jung of Rolsberga, Sweden for the alpaca tissue and to Anette Gullberg and Hjörvar Pétursson for their help with lab. work. This study was supported by the Swedish National Sciences Research Council, the Jörgen Lindström Foundation, the Swedish Institute (scholarship to K. E. Slack), the New Zealand Marsden Fund (K. E. Slack), the Nilsson-Ehle Foundation, and by contract ERBCHR XCT 930254 from the European Commission.

References

Adachi, J. & Hasegawa, M. (1996a). MOLPHY, Version 2.3: Programs for molecular phylogenetics based on maximum likelihood. *Computer Science Monograph*, *28*, 1–150.

Adachi, J. & Hasegawa, M. (1996b). Model of amino acid substitution in proteins encoded by mitochondrial DNA. *Molecular Biology and Evolution*, *42*, 459–468.

Anderson, S., de Brujin, M. H. L., Coulson, A. R., Eperon, I. C., Sanger, F. & Young, G. (1982). Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. *Molecular Biology and Evolution*, *156*, 683–717.

Arnason, U. & Gullberg, A. (1993). Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature. *Molecular Biology and Evolution*, *37*, 312–322.

Arnason, U. & Gullberg, A. (1996). Cytochrome b nucleotide sequences and the identification of five primary lineages of extant cetaceans. *Molecular Biology and Evolution*, *13*, 407–417.

Arnason, U., Gullberg, A., Janke, A. & Xu, X. (1996). Pattern and timing of evolutionary divergences among hominoids based on analyses of complete mtDNAs. *Molecular Biology and Evolution*, *43*, 650–661.

Arnason, U., Gullberg, A. & Janke, A. (1997). Phylogenetic analyses of mitochondrial DNA suggest a sister group relationship between Xenarthra (Edentata) and ferungulates. *Molecular Biology and Evolution*, *14*, 762–768.

Arnason, U., Gullberg, A., Johnsson, E. & Ledje, C. (1993). The nucleotide sequence of the mitochondrial DNA molecule of the grey seal, *Halichoerus grypus*, and a comparison with mitochondrial sequences of other true seals. *Molecular Biology and Evolution*, *37*, 323–330.

Arnason, U., Gullberg, A. & Widegren, B. (1991). The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *Molecular Biology and Evolution*, *33*, 556–568.

Arnason, U. & Johnsson, E. (1992). The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina*. *Molecular Biology and Evolution*, *34*, 493–505.

Beintema, J. J., Gaastra, W., Lenstra, J. A., Welling, G. W. & Fitch, W. M. (1977). The molecular evolution of pancreatic ribonuclease. *Molecular Biology and Evolution*, *10*, 49–71.S.

Bibb, M. J., Van Etten, R. A., Wright, C. T., Walberg, M. W. & Clayton, D. A. (1981). Sequence and organization of mouse mitochondrial DNA. *Cell*, *26*, 167–180.

Cao, Y., Adachi, J., Janke, A., Paabo, S. & Hasegawa, M. (1994). Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: Instability of a tree based on a single gene. *Molecular Biology and Evolution*, *39*, 519–527.

Colbert, E. H. & Morales, M. (1991). *Evolution of the Vertebrates*. New York: Wiley-Liss Inc.

Dayhoff, M. O. (1978). Atlas of protein sequence and structure. Washington, DC: National Biomedical Research Foundation.

Douzery, E. & Catzeflis, F. M. (1995). Molecular evolution of the mitochondrial 12S rRNA in Ungulata (Mammalia). *Molecular Biology and Evolution*, *41*, 622–636.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Molecular Biology and Evolution*, *17*, 368–376.

Felsenstein, J. (1991). PHYLIP: Phylogeny inference package, Version 3.52c. Seattle: University of Washington.

Fitch, W. M. (1971). Towards defining the course of evolution, minimum change from a specific tree topology. *Systematic Zoology*, *20*, 406–415.

Gatesy, J., Hayashi, C., Cronin, M. A. & Arctander, P. (1996). Evidence from milk casein genes that cetaceans are close relatives of hippopotamid artiodactyls. *Molecular Biology and Evolution*, *13*, 954–963.

Gatesy, J. (1997). More DNA support for a Cetacea-Hippopotamidae clade: The blood-clotting protein gene gamma-fibrinogen. *Molecular Biology and Evolution*, *14*, 537–543.

Gatesy, J., Milinkovitch, M., Waddell, V. & Stanhope, M. (1999). Stability of cladistic relationships between Cetacea and higher-level artiodactyl taxa. *Systematic Biology*, *48*, 6–20.

Graur, D. & Higgins, D. G. (1994). Molecular evidence for the inclusion of cetaceans within the order Artiodactyla. *Molecular Biology and Evolution*, *11*, 357–364.

Hasegawa, M., Kishino, H. & Yano, T. A. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Molecular Biology and Evolution*, *22*, 60–174.

Hiendleder, S., Lewalski, H., Wassmuth, R. & Janke, A. (1998). Complete mitochondrial DNA sequence of a domestic sheep (*Ovis aries*) and comparative data from another major haplotype. *Molecular Biology and Evolution*, *47*, 441–448.

Irwin, D. M. & Arnason, U. (1994). Cytochrome *b* gene of marine mammals: phylogeny and evolution. *Journal of Mammalian Evolution*, *2*, 37–55.

Kishino, H. & Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution*, *29*, 170–179.

Kishino, H., Miyata, T. & Hasegawa, M. (1990). Maximum Likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution*, *31*, 153–160.

Kleineidam, R. G., Pesole, G., Breukelman, H. J., Beintema, J. J. & Kastelein, R. A. (1999). Inclusion of cetaceans within the order Artiodactyla based on phylogenetic analysis of pancreatic ribonuclease genes. *Molecular Biology and Evolution*, *48*, 360–368.

Kumar, S. & Hedges, S. B. (1998). A molecular timescale for vertebrate evolution. *Nature*, *392*, 917–920.

Lopez, J. V., Cevario, S. O. & Brien, S. J. (1996). Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. *Genomics*, *33*, 229–246.

Miyamoto, M. M., Kraus, F., Laipis, P. J., Tanhauser, S. M. & Webb, S. D. (1993). Mitochondrial DNA phylogenies within Artiodactyla. In F. S. Szalay, M. J. Novacek & M. C. McKenna (Eds) *Mammal Phylogeny* (pp. 268–281). New York: Springer-Verlag New York Inc.

Montgelard, C., Catzeflis, F. M. & Douzery, E. (1997). Phylogenetic relationships of artiodactyls and cetaceans as deduced from the comparison of cytochrome *b* and 12S rRNA mitochondrial sequences. *Molecular Biology and Evolution*, *14*, 550–559.

Mouchaty, S. K., Gullberg, A., Janke, A. & Arnason, U. (2000). The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. *Molecular Biology and Evolution*, *17*, 60–67.

de Queiroz, K. & Gautier, J. (1994). Towards a phylogenetic system of biological nomenclature. *Trends in Ecology and Evolution*, *9*, 27–30.

Randi, E., Lucchini, V. & Dijong, C. H. (1996). Evolutionary genetics of the Suiformes as reconstructed using mtDNA sequencing. *Journal of Mammalian Evolution*, *3*, 163–194.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*, 406–425.

Sanger, F. (1981). Determination of nucleotide sequences in DNA. *Science*, *214*, 1205–1210.

Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M. & Munechika, I. (1997). Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature*, *388*, 666–670.

Strimmer, K. & von Haeseler, A. (1996). Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution*, *13*, 964–969.

Tamura, K. & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*, 512–526.

Templeton, A. R. (1983). Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the humans and the apes. *Evolution*, *37*, 221–244.

Ursing, B. M. & Arnason, U. (1998a). The complete mitochondrial DNA sequence of the pig (*Sus scrofa*). *Molecular Biology and Evolution*, *47*, 302–306.

Ursing, B. M. & Arnason, U. (1998b). Analyses of mitochondrial genomes strongly support a hippopotamus-whale clade. *Royal Society, Proceedings: Biological Sciences*, *265*, 2251–2255.

Van Valen, L. (1967). Monophyly or diphyly in the origin of whales. *Evolution*, *22*, 37–41.

Webb, S. D. & Taylor, B. E. (1980). The phylogeny of hornless ruminants and a description of the cranium of *Archaeomeryx*. *Bulletin of American Museum of Natural History*, *167*, 117–158.

Xu, X. & Arnason, U. (1994). The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. *Gene*, *148*, 357–362.

Xu, X. & Arnason, U. (1997). The complete mitochondrial DNA sequence of the white rhinoceros, *Ceratotherium simum*, and comparison with the mtDNA sequence of the Indian rhinoceros, *Rhinoceros unicornis*. *Molecular Phylogenetics and Evolution*, 7, 189–194.

Xu, X., Gullberg, A. & Arnason, U. (1996a). The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. *Molecular Biology and Evolution*, *43*, 438–446.

Xu, X., Janke, A. & Arnason, U. (1996b). The complete mitochondrial DNA sequence of the greater Indian rhinoceros, *Rhinoceros unicornis*, and the phylogenetic relationship among Carnivora, Perissodactyla, and Artiodactyla (+ Cetacea). *Molecular Biology and Evolution*, *13*, 1167–1173.

Appendix 3.

Harrison, G.L., McLenachan, P.A., Phillips, M.J., **Slack, K.E**., Cooper, A. and Penny, D. (2004). Four new avian mitochondrial genomes help get to basic evolutionary questions in the Late Cretaceous. *Mol. Biol. Evol.* **21**: 974–983. (As at August 2011 this paper has been cited 63 times in the Web of Science.)

This paper represents another important step in breaking up long branches in order to get more stability in the avian phylogenetic tree.

Based on the literature and the results of analyses in Chapter 3 (Slack et al. 2003), I proposed several hypotheses that required testing:

- That long branch attraction was a potential problem in birds, and therefore additional taxon sampling was required (the rifleman [a New Zealand wren] to aid in breaking up the long branch to the oscine/suboscine passerines, and the magpie goose as a very deeply diverging anseriform),
- That owls (Strigiformes) and/or parrots (Psittaciformes) - unrepresented in the mt genome dataset at the time - might constitute the deepest avian divergence/s among the Neoaves.
- An owl was also needed in the avian mt dataset to test the relationship between nocturnal and diurnal avian predators.

Abby (G.L.) Harrison and Trish (P.A.) McLenachan carried out the primary sequencing, while Matt Phillips had a major role in the analysis. My primary contributions were four fold:

- discussion over which taxa to sequence,
- sequence alignment and editing,
- a major updating of the tables of avian mitochondrial features, and
- the identification of interesting features of tRNA-Phe (see Figure 1 of the manuscript, the data came from my tables in Appendix 4).

Alan Cooper was especially involved with developing the long range PCR techniques that were used in this manuscript, as well as in project design. David Penny was involved in all aspects of project design. All authors approved the final manuscript.

Four New Avian Mitochondrial Genomes Help Get to Basic Evolutionary Questions in the Late Cretaceous

G. L. (Abby) Harrison,[1] P. A. McLenachan,* M. J. Phillips,* Kerryn E. Slack,*† Alan Cooper,‡ and David Penny**

*Allan Wilson Center for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand; †Department of Genetics, Cell and Organism Biology, University of Lund, Lund, Sweden; and ‡Department of Zoology, Oxford University, Oxford, United Kingdom

Good phylogenetic trees are required to test hypotheses about evolutionary processes. We report four new avian mitochondrial genomes, which together with an improved method of phylogenetic analysis for vertebrate mt genomes give results for three questions in avian evolution. The new mt genomes are: magpie goose (*Anseranas semipalmata*), an owl (morepork, *Ninox novaeseelandiae*); a basal passerine (rifleman, or New Zealand wren, *Acanthisitta chloris*); and a parrot (kakapo or owl-parrot, *Strigops habroptilus*). The magpie goose provides an important new calibration point for avian evolution because the well-studied *Presbyornis* fossils are on the lineage to ducks and geese, after the separation of the magpie goose. We find, as with other animal mitochondrial genomes, that RY-coding is helpful in adjusting for biases between pyrimidines and between purines. When RY-coding is used at third positions of the codon, the root occurs between paleognath and neognath birds (as expected from morphological and nuclear data). In addition, passerines form a relatively old group in Neoaves, and many modern avian lineages diverged during the Cretaceous. Although many aspects of the avian tree are stable, additional taxon sampling is required.

Good evolutionary trees are required to test hypotheses. For example, we wish to know how many lineages of birds survived from the Cretaceous to the present (Cooper and Penny 1997) in order to test models of apparent "mass extinctions" and "explosive radiations" (Feduccia 1995, 2003). A well-resolved avian tree is also required for testing biogeographic (Cracraft 2001; Ericson et al. 2002) and/or ecological hypotheses (Cooper and Penny 1997; see later).

It is almost an offense against birds that the deep mammalian evolutionary tree is virtually resolved (Waddell, Kishino, and Ota 2001; Lin et al. 2002; Springer et al. 2003) whilst there are still major uncertainties about many aspects of the avian evolutionary tree (see for example Cracraft 2001). A major uncertainty is the position of the root of the avian tree; mitochondrial (mt) data sets tend to place the root within the passerine birds (Mindell et al. 1999; Härlid, Janke, and Arnason 1999, although see Braun and Kimball 2002). In contrast, morphological and nuclear sequences tend to place the root between paleognath birds (ratites and tinamous) and all other birds (neognaths). There is also uncertainty over the time of origin of passerines (perching birds and/or song birds); Feduccia (1995, 2003) places them as a recent order of modern birds, other authors place their origin before the diversification of shore birds (Barker, Barrowclough, and Groth 2002; Ericson et al. 2002).

Part of our confidence that the higher-level mammalian tree is now quite accurate is that highly similar trees are being found using independent data sets—nuclear (for example, Springer et al. 2003) and mitochondrial (Lin et al. 2002). Agreement can be treated quantitatively; in the mammal example, a deep four-way split in the eutherian tree was defined with nuclear data sets. The probability of randomly selecting a tree with this same four-way split from mitochondrial data, given the number of taxa, was less than $\approx 10^{-14}$. The four-way split was found with mitochondrial data, confirming the high similarity of trees from the two data sets. We expect that a combination of mitochondrial and nuclear data should eventually give similar confidence in avian trees.

There is good progress toward resolving the avian tree using both nuclear and mitochondrial sequences (Sibley and Ahlquist 1990; Van Tuinen, Sibley, and Hedges 2000; Cracraft 2001; Cooper et al. 2001; Haddrath and Baker 2001; Ericson et al. 2002; Paton, Haddrath, and Baker 2002; Barker, Barrowclough, and Groth 2002; Garcia-Mareno, Sorenson, and Mindell 2003). In an unrooted avian tree, as expected from morphological data, ratites and tinamous unite to form paleognaths, and all remaining birds are neognaths (and separate into Gallianseres [chicken, geese, and relatives] and Neoaves (Cracraft 2001). The succession of divergences within Neoaves, which contains the vast majority of living birds, remains unclear. Cracraft (2001) has a six-way split between the following groups:

Passerines
Parrots
Cuckoos
Woodpeckers, rollers, bee-eaters, kingfishers, jacanas, and mousebirds (four orders)
Owls, nightjars, swifts, and turacos
Seabirds, shore birds, doves, cranes, raptors, rails, penguins, storks, loons, and grebes (a very diverse group, ~10 orders)

Despite this lack of resolution, we use the Cracraft (2001) tree as an informal prior for evaluating results. Of the six Neoaves groups, only two (passerines and the seabird/shorebird alliance) are currently represented in the complete mitochondrial set, showing the need for increased taxon sampling. The species sequenced here, together with

[1] Present address: Fiji School of Medicine, Private Mail Bag, Suva, Fiji.

Key words: avian evolution, mitochondrial genomes, Anseranas (Anseriformes), morepork (owl, Strigiformes), kakapo (parrot Psittaciformes), rifleman (N Z Wren, Passeriformes), RY-coding.

E-mail: d.penny@massey.ac.nz.

Mol. Biol. Evol. 21(6):974–983. 2004
DOI:10.1093/molbev/msh065
Advance Access publication January 22, 2004

some reasoning for the choices, is given below. All are Australasian taxa, which helps avoid duplication of effort (a list of taxa being sequenced by our group is available at http://awcmee.massey.ac.nz/mt_genomes.htm).

We have sequenced mt genomes of two of the four unrepresented groups namely an owl and a parrot. Parrots are a distinct and old group for which a Late Cretaceous fossil has been reported (Stidham 1998; Hope 2002); a kakapo (owl-parrot or night parrot, *Strigops habroptilus*: fam. Psittacidae) was selected for this study. Owls are another distinct avian group, and a New Zealand owl (morepork or ruru, *Ninox novaeseelandiae*: fam. Strigidae) was chosen.

The rifleman (a New Zealand wren, *Acanthisitta chloris*: fam Acanthisittidae) is a basal passerine. New Zealand wrens do not really fit the oscine/suboscine classification. Cracraft (2001) shows an unresolved three-way split between oscines (which form the large majority of passerine birds), suboscines, and New Zealand wrens. Ericson et al. (2002) reports nuclear sequences for the rifleman, analysis of which places it basal to all other passerines—oscine and suboscine. A rifleman mt genome should also help resolve the position of passerines within the avian tree, including the position of the root. In the earliest mitochondrial data sets (with only a small number of genomes) the root of the avian tree tended to fall within passerines (Mindell et al. 1997; Härlid, Janke, and Arnason 1999), rather than in the expected position between neognaths and paleognaths. Recently, and with more taxa in the data sets, it has not been possible to reject the classical (neognath/paleognath) rooting (Paton, Haddrath, and Baker 2002; Braun and Kimball 2002; Slack et al. 2003). In one case, with transversion likelihood, the results rejected the passerine rooting (Braun and Kimball 2002). In our work with eutherian mammals (Lin et al. 2002) we found that increased taxon sampling led to agreement between trees from nuclear and mitochondrial data.

A magpie goose (*Anseranas semipalmata*) was chosen because two major morphological studies (Ericson 1997; Livezey 1997) conclude that *Presbyornis* fossils are on the lineage to geese and ducks—after the divergence of the magpie goose lineage. Goose and duck mitochondrial genomes are available (see Slack et al. 2003), and the addition of a magpie goose mt genome therefore establishes an important calibration point for avian evolution. Some molecular results are available for the magpie goose (see Sibley and Ahlquist 1990; Sraml et al. 1996; Mindell et al. 1997) and support its placement outside geese and ducks, but still within Anseriformes. With respect to dates, Ericson (1997) places the *Anseranas*/*Presbyornis* divergence at least 60 MYA (Paleocene) but some older *Presbyornis* fossils are reported from about 66–67 MYA, in the very Late Cretaceous of Antarctica (Noriega and Tambussi 1995; Case and Tambussi 1999). These fossils are not yet fully published, and in the interim we use both the 60 MYA date as a lower bound on the time of divergence, and compare results using this to those with the older (66 MYA) calibration point.

As mentioned above, the rooting point of the avian tree is controversial. We take the view that, although the data are correct, inadequacies in analytical methods can lead to different results from nuclear and mitochondrial data. Rather than "blame the data," the onus is on theoreticians to improve techniques of analysis to reflect the unusual nucleotide composition of some vertebrate mitochondrial genomes. This includes differences between pyrimidines (C&T) and between purines (A&G) (see Schmitz et al. 2002; Phillips and Penny 2003). We also require criteria to evaluate which techniques are more powerful in capturing information in the data. One such measure is the *treeness* statistic, the sum of internal internode (branch) lengths divided by the sum of all internodes on the tree (see Lanyon 1988; Phillips and Penny 2003). Treeness increases when apparently saturated sites are omitted—such as third codon positions or, especially (for mitochondrial data), by reducing the nucleotides (A,G,C,T) to pyrimidines and purines (RY-coding). RY-coding reduces the effect of differences in nucleotide composition between species resulting from C-T differences (pyrimidine bias), or the lesser differences between A and G (purine bias) (Phillips and Penny 2003). The reduced bias is measured by the *relative compositional variability* (RCV, the average variability for character states between taxa). For nucleotides, RCV is defined as:

$$\mathrm{RCV} = \Big(\sum |\mathrm{A}_i, -\mathrm{A}^*| + |\mathrm{T}_i, -\mathrm{T}^*,| + |\mathrm{C}_i - \mathrm{C}^*| + |\mathrm{G}_i - \mathrm{G}^*|\Big)/n.t$$

(see Phillips and Penny 2003)

A_i, T_i, C_i, and G_i are the total numbers of each nucleotide for the ith taxon; A*, T*, C*, and G* are the averages for the n taxa, and t the number of sites. RCV allows direct comparison of the extent of composition bias for data sets and data treatments.

In summary, for data partitions or codings compared on the same tree, higher treeness and/or lower RCV values indicate a stronger phylogenetic signal and/or a lower composition bias that can mislead phylogenetic inference. Phylogeny estimates from data treatments (such as partitioning and/or coding) that have the highest treeness/RCV values are expected to be the least susceptible to composition bias (Phillips and Penny 2003). We find that treeness and RCV values are preferable to using chi-squared values of deviations in amino acid (aa) composition, because the chi-squared test loses sensitivity when coding sequences are expressed as amino acids. For the same original amount of data, the number of degrees of freedom is increased markedly, whereas the number of sites is reduced by two-thirds, making the analysis much less powerful. RY-coding is effective for mitochondrial sequences in that it results in more agreement between data sets. For example, monotremes (platypus and echidna) were placed just outside the therians (marsupial plus placental mammals; Phillips and Penny 2003), and the Hexapoda clade of insects plus Collembollans was recovered (Delsuc, Phillips, and Penny 2003; see also Nardi et al. 2003).

Materials and Methods

The owl was from Nga Manu Bird Sanctuary, Waikanae, New Zealand; Trevor Worthy provided a

rifleman sample from Northwest Nelson, N.Z.; David Lambert, Massey University, donated kakapo tissue; and Peter Whitehead and Julian Gorman, Northern Territory University, Darwin, Australia, provided magpie goose tissue. DNA was extracted from muscle, liver, or blood using standard kits. Mitochondrial DNA was amplified in fragments longer than 5 kb (to minimize the risk of amplifying nuclear copies) using the Expand Long template polymerase chain reaction (PCR) kit (Roche).

For the owl, parrot, and rifleman, long PCR DNA fragments were sequenced directly or used as template for a second round of short-range PCR of 1~2 kb. Primers were designed to match conserved regions of avian mtDNA genomes, allowing 0–3 degenerate sites to maximize their usefulness for other species. We used the Fasta search in the GCG program (Wisconsin Package, version 10.0) to search our primer database for appropriate targets for primer walking. Where possible, primers from Sorenson et al. (1999) and Cooper et al. (2001) were used. Any new primers required were designed using Oligo 4.03 (National Biosciences, Inc.). Sequencing reactions followed standard protocols for Applied BioSystems 377 and 3730 Sequencers. Sequences were assembled and checked using Sequencing Analysis and MT Navigator programs (ABI) and Sequencher 4.1 (Gene Codes Corp.).

For magpie goose, long-range PCR products were pooled and fragmented pneumatically with a nebulizer for 40 s at 40 psi into pieces about 2 kb in length, then cloned and sequenced using the TOPO Shotgun Subcloning Kit Version D (Invitrogen). This involved ligation into pCR 4Blunt-TOPO and transformation into TOPO10 *E. coli*. Plasmid DNA was extracted and purified using the GenElute Plasmid Miniprep Kit (Sigma), and insert size was determined by restriction digest. Plasmids containing inserts >800 bp were sequenced with the universal forward and reverse primers. The sequences were edited and assembled in Sequencher; any gaps were filled with short-range reamplifications from the appropriate long fragments.

In addition to the four new mt genomes, we used 20 other avian taxa: chicken (*Gallus gallus*; GenBank accession number X52392), quail (*Coturnix japonica*; AP003195), redhead duck (*Aythya americana*; AF090337), greater white-fronted goose (*Anser albifrons*; AF363031), rook (*Corvus frugilegus*; Y18522), gray-headed broadbill (*Smithornis sharpei*; AF090340), village indigobird (*Vidua chalybeata*; AF090341), peregrine falcon (*Falco peregrinus*; AF090338), common buzzard (*Buteo buteo*; AF380305), Oriental white stork (*Ciconia boyciana*; AB026193), ruddy turnstone (*Arenaria interpres*; AY074885), blackish oystercatcher (*Haematopus ater*; AY074886), little blue penguin (*Eudyptula minor*; AF362763), great spotted kiwi (*Apteryx haastii*; AF338708), emu (*Dromaius novaehollandiae*; AF338711), double-wattled cassowary (*Casuarius casuarius*; AF338713), ostrich (*Struthio camelus*; Y12025), greater rhea (*Rhea americana*; Y16884), great tinamou (*Tinamus major*; AF338707), and elegant crested-tinamou (*Eudromia elegans*; AF338710).

Six reptiles were used as outgroups: American alligator (*Alligator mississippiensis*; Y13113), eastern painted turtle (*Chrysemys picta*; AF069423), green turtle (*Chelonia mydas*; AB012104), blue-tailed mole skink (*Eumeces egregius*; AB016606), common iguana (*Iguana iguana*; AJ278511), and spectacled caiman (*Caiman crocodylus*; AJ404872). Data sets were prepared both with and without outgroups because in preliminary studies we found that the avian tree could change when the outgroup was added. A similar phenomenon has been reported with eutherian mammals (Lin et al. 2002).

Sequences were aligned manually in Se-Al version 1.0 a1 (http://evolve.zps.ox.ac.uk/Se-Al/Se-Al.html). rRNA sequences were aligned on the basis of secondary structure (www.rna.icmb.utexas.edu/RNA/) to maximize homologous positions. Data are available at http//imbs.massey.ac.nz/downloads.htm. Standard programs were used for all analysis, including PAUP* 4.0b8 (Swofford 1998), MOLPHY (Adachi and Hasegawa 1996), and MrBayes 3.0b4 (Huelsenbeck and Ronquist 2001). Maximum parsimony, minimum evolution (with ML distances), maximum likelihood (ML), and Bayesian methods were employed on both the avian and avian plus outgroup data sets. Optimal parameters for the maximum likelihood models were determined using Modeltest (Posada and Crandall 1998). The hierarchical and AIC tests were in agreement for Modeltest.

Results

Description of mt Genomes

The GenBank numbers and lengths of the four new sequences are as follows:

Acanthisitta chloris, rifleman—AY325307 (missing tRNA-Phe and some control region)
Anseranas semipalmata, magpie goose—AY309455, 16,869 bp (complete)
Ninox novaeseelandiae, owl—AY309457, 17,122 bp (missing part of tRNA-Phe and control region)
Strigops habroptilus, parrot—AY309456, 16,311 bp (missing part of control region)

Each mitochondrial genome was sequenced from $tRNA^{Phe}$ or 12S RNA through to $tRNA^{Thr}$/$tRNA^{Pro}$/ ND6/ $tRNA^{Glu}$ and into the control region. These genomes have the same gene order as the chicken and not the alternative avian gene order ($tRNA^{Thr}$ /control region/ $tRNA^{Pro}$ / ND6/$tRNA^{Glu}$/noncoding; Mindell, Sorenson, and Dimcheff 1998). The sequences for ND6 and t-Glu in the kakapo appear to encode functional genes; this as well as the fact that $tRNA^{Pro}$ follows $tRNA^{Thr}$ indicates that the kakapo does not have the same rearrangement as found in the parrot genus *Amazona* (Eberhard, Wright, and Bermingham 2001). Unfortunately, for political reasons, we are unable to clone DNA fragments from native birds in New Zealand. Hence, parts of $tRNA^{Phe}$ and the control region are missing from all three native birds, as they have proved difficult to sequence without cloning, on account of the presence of repeats and heteroplasmy. Features such as start and stop positions for each gene (as in Slack et al. 2003) are reported in http://awcmee.massey.ac.nz/downloads.htm. The TψC loop of $tRNA^{Phe}$ has three variants that are potentially informative within birds. Paleognaths (ratites and tinamou), galliformes, anseriformes, and the owl have the same pattern (see fig. 1A). Other Neoaves (except penguin) have an

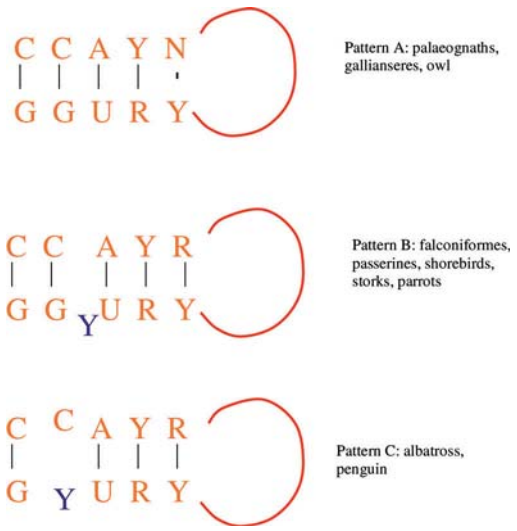


FIG. 1.—TψC stem patterns among birds for tRNA-Phe. Pattern A appears to be ancestral for modern birds, being shared by the four most basal groups (tinamous, ratites, anseriformes, and galliformes), as well as the owl. The inferred ancestral bases are in red. An additional unpaired base (indicated in blue) occurs in all other birds examined, for which pattern B appears primitive. A third pattern (C, in the penguin and petrel) can easily be derived from pattern B by the loss of a guanine.

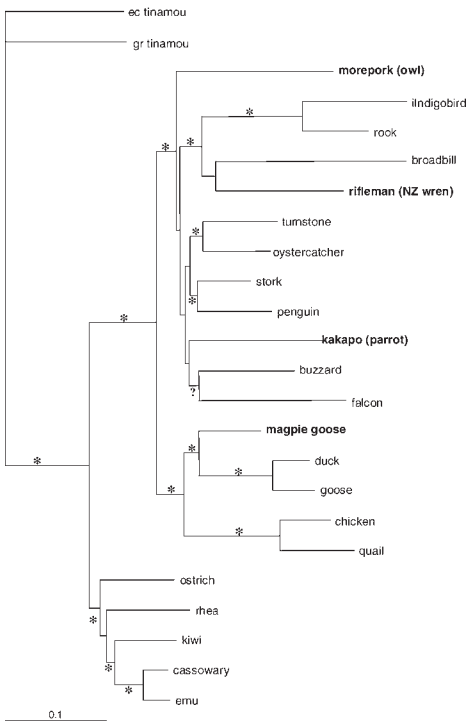


FIG. 2.—Unrooted MrBayes tree for 24 avian mt genomes. The data represent 13,439 base pairs of the combined protein and RNA data sets using $12_n3_rSL_n$ coding. Each of the resulting five partitions is optimized for its GTR + I + Γ_4 model. An asterisk indicates the groupings that have both high bootstrap support by several analyses on this data set, and also on prior information (nuclear and morphological). The falconiform (buzzard/falcon) group is marked by a question mark because it is only weakly supported on the present analyses, even though it is well-supported on other data. The four new taxa from this study are indicated in bold.

inserted pyrimidine, usually "C," which is unpaired. This is illustrated as pattern B in figure 1. The penguin/petrel pattern is similar to 1B but has lost a "G" and has one less set of paired bases (pattern C).

Two data sets were used for phylogenetic analysis: a 24 taxon bird-only data set; and a 30-taxon data set with the 24 birds plus the six outgroups (two turtles, two crocodilians, and two lizards) as used in Slack et al. (2003). To check the relative size of signal in different data partitions, we calculated RCV values on the data, and treeness values on the tree in figure 2 (see later). The results are given in table 1. The first, second, and third codon positions of the 12 proteins encoded on the heavy strand are indicated by 1, 2, and 3, and the values for coding as nucleotides (or as RY-coding) are shown by a subscript "n" (or "r"), respectively. RNA coding genes (ribosomal and transfer) were partitioned into stems (S) and loops (L). Protein-coding genes were also translated to amino acids. The number of nucleotides was 10,338 protein-coding sites (3,446 amino acids) and 3,101 RNA-coding sites, giving a combined total of 13,439 nucleotides (excluding gaps).

The most important conclusion from table 1 is that, compared to amino acid coding, or omitting the third codon position, RY-coding the third position improves the signal-to-noise ratio, thus retaining more phylogenetic information. This result is seen in several comparisons, for example, the values for third position coded as nucleotides (3_n), and as RY-coding (3_r). The 3_n value (0.70) is the lowest treeness/RCV value in the table; the 3_r value (2.76) is one of the highest. Similarly, it is informative to compare 12_n with 123_n and with 12_n3_r (that is, adding the third codon position to the first two, first as nucleotides, then as RY-coded). By itself, adding the third position as nucleotides reduces the treeness/RCV value from 1.74 to 1.15—consistent with the experience of many authors that the third codon position is "saturated." Adding the third position as RY-coded enhances the value from 1.74 to 2.32. Thus coding the third position as RY both increases the signal in the internal edges (branches) of the tree and reduces the variability of nucleotide composition between taxa. This can only happen if there is a large amount of information as purines and pyrimidines that is masked by within-purine and within-pyrimidine biases (Phillips and Penny 2003). Most of our results are therefore given under the $12_n3_rSL_n$ coding scheme, although others have been used (data not shown), such as reducing the first position to RY-coded (1_r), loops to RY-coding (L_r), etc. For analysis, each of the five partitions (codons one, two, and three, stems, and loops) has its own optimized model (including for gamma distribution and proportion of invariant sites).

Unrooted Trees

We find cases with both real (Lin et al. 2002; Slack et al. 2003) and simulated data (Holland, Penny, and Hendy 2003), where the unrooted tree changes when the outgroup is added. We therefore examine the unrooted tree first, and only include the outgroup taxa later. Figure 2 shows the unrooted (avian-only) tree for the combined protein and

Table 1
Treeness and RCV Values for Partitions of the 24 Taxon Avian-Only Data Set

Partition	treeness	RCV	treeness/RCV
$12_n3_rSL_n$	0.1376	0.0495	2.7798
12_n	0.0888	0.0511	1.7378
3_n	0.0652	0.0931	0.7003
3_r	0.1603	0.0580	2.7638
SL_n	0.1393	0.0530	2.6283
123_n	0.0829	0.0724	1.1450
12_n3_r	0.1384	0.0596	2.3221
PTN_{aa}	0.1584	0.0787	2.0127

NOTE.—Partitions of the data coding the third codon position as RY (3_r) rather than nucleotide (3_n) improves the signal-to-noise ratio (treeness/RCV). In addition, it retains all alignable sites. Values for other partitions are shown; PTN_{aa} are the protein-coding genes as amino acids. For the main analysis, other positions were retained as nucleotides (first and second codon positions and stems (S) and loops (L), that is, $12_n3_rSL_n$).

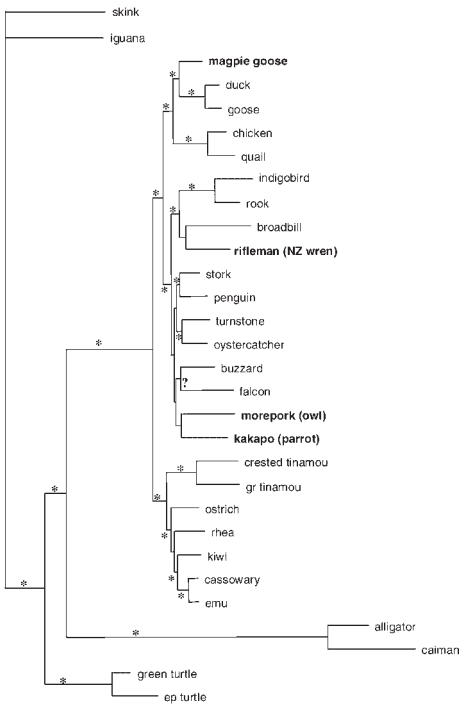


FIG. 3.—Avian MrBayes tree rooted by six outgroup taxa. The data is the combined protein and RNA genes using $12_n3_rSL_n$ coding, and with each partition optimized for its GTR + I + Γ_4 model. Posterior probabilities are >0.98 for all internal edges, except for the following groupings of the owl/parrot/falconiformes (0.82) and rhea/emu/kiwi/cassowary (0.97). These values do not include effects from model misspecification. Asterisks, question mark, and bold font are as in figure 2.

RNA data ($12_n3_rSL_n$), using MrBayes. Because we do not have any new paleognath taxa, we do not discuss them in detail. We simply note that they have the standard subdivision into tinamou and ratites, and that some details of the ratite subtree are not robust. Looking at the new taxa, the magpie goose (as expected on the basis of prior information) is always deep on the duck/goose lineage. These three anseriforms join with the two galliforms (chicken and quail), to form Gallianseres. This grouping has increasingly been supported in recent years by both molecular and morphological data (for example, Cracraft 2001; Livezey and Zusi 2001; Slack et al. 2003; Sorenson et al. 2003). Thus the unrooted avian tree has the predicted strong three-way subdivision into paleognaths (ratites and tinamou) and the two neognath subdivisions (Gallianseres and Neoaves).

With respect to the four passerines (rifleman, broadbill, indigobird, and rook), the first important point is that they are united on the unrooted tree. Slack et al. (2003) found that the passerines (then without the rifleman) grouped together on the unrooted mt tree; it was only on addition of the reptilian outgroup that the passerine grouping became (at best) paraphyletic. A second point is that, given the expected rooting point between paleognaths and neognaths, passerines appear to form an early division of Neoaves. As discussed in Boles (1997), passerines have traditionally been considered relatively recent within extant birds (see also Livezey and Zusi 2001; Feduccia 2003). Because there are still major Neoavian groups missing from this data set (cuckoos, woodpeckers, mouse birds, etc.; Cracraft 2001) it is possible these could form earlier divisions within Neoaves. Nevertheless, the deep placement of passerines is worth noting.

Turning to the new passerine, the rifleman (as expected) is deep within the passerines—either ancestral to all passerines (as in Ericson et al. 2002) or basal on the suboscine (broadbill) lineage (Sibley and Ahlquist 1990). In the present study, the highest bootstrap (PAUP*) and posterior support values (MrBayes) favor the broadbill/rifleman association. However, this latter grouping should be treated cautiously. Both the rifleman and broadbill are long branches in the tree and with, at most, a short edge between them. This pattern fits the classic long-branch attraction case, where misleading results are found even without differences in rates of evolution (Hendy and Penny 1989); this affects all tree selection criteria where the model is to some extent mis-specified. To check this possibility, we are now sequencing another suboscine (a New World tyrant flycatcher) to break up the long broadbill branch. However, in our terminology (Cooper and Penny 1997), the tree is locally stable; the alternatives are rearrangements around a single internal branch of the tree.

In the current data set, neither owl nor parrot has any strong associations. Both fall within the Neoaves (which is supported by 100% bootstrap, or posterior probabilities of 1.0, in all our analyses). In general, the owl is either deeper in the tree than the parrot (as in figure 2), or they have a weak association (as in fig. 3, see later). The respective positions of owl and parrot are only preliminary until taxon sampling is increased. In the meantime, however, it is noteworthy that both come within Neoaves as ancient and distinctive lineages.

This leaves the six representatives of the large Neoavian group that includes seabirds, shorebirds, and raptors. Four taxa that come together on nearly all trees are penguin, stork, and the shore birds (oystercatcher and turnstone). The buzzard is usually adjacent to this group of four (although it might be expected to be one step closer to the stork/penguin; Sibley and Ahlquist 1990). However, the position of the falcon is quite variable; only with

the RNA data set does it come strongly with the buzzard (as in fig. 2), although both have the same duplication/rearrangement of gene order (Mindell, Sorenson, and Dimcheff 1998; Haring et al. 2001). In earlier work, with smaller numbers of mt genomes, the falcon even came with passerines (see Discussion in Slack et al. 2003). Given our emphasis on increased taxon sampling as solving many problems (Hendy and Penny 1989; Lin et al. 2002), the frequent separation of falcon and buzzard, especially on protein-coding genes, is unexpected. At the moment we can only fall back on the "still better taxon sampling" argument and perhaps a kite, osprey, or especially a forest-falcon would strengthen the position of falcon and buzzard on the tree. Whether this then moves the raptors closer to penguin and stork could then be evaluated.

Finally, we use an asterisk to indicate those aspects of the tree that are well supported by these and other data. With the ever-increasing number of analysis methods available, it appears arbitrary selecting one set of, say, bootstrap or posterior probability values. Our approach is to identify the groupings that have strong support values from several methods and which are supported by other data (nuclear and/or morphological). The weakest association we have marked with an asterisk is the stork/penguin pairing, which has only 95% bootstrap support under ML, but was in our prior tree (Cracraft 2001). In addition, although we would be surprised if the buzzard and falcon continued to be separate when more falconiformes are available, it is hard to recover this grouping from our data. Consequently, falconiformes is indicated by a question mark. For other weaker groupings, we simply show results from the MrBayes tree. However, we would not be surprised, with more data, if these groupings changed on the tree—though usually not by more than a single interchange on the tree.

Rooted Trees

As in Slack et al. (2003), we used six reptilian sequences to root the avian tree, two turtles, two crocodilians, and two lizards. Figure 3 shows a MrBayes analysis on the combined protein and RNA data, with third codon positions reduced to RY-coding (that is, $12_n3_rSL_n$). Again, the model was optimized for each of the five partitions. Unlike previous analyses of avian mt genomes, this straightforward analysis gave the root between paleognaths and neognaths (as expected from morphological and nuclear data). The bootstrap value for this position of the root is 96% with ML, but less with minimum evolution (78% with ML distances). Reducing other partitions to RY-coding also placed the root in the same position (and increased the treeness/RCV ratio, indicating a higher signal-to-noise ratio). Because the position of the root in figure 3 is found from mitochondria data with analyses giving the strongest signal-to-noise ratio, and because nuclear and morphological data give the same rooting, we consider this the accepted rooting for birds. This is basically the argument of congruence between independent data sets (Penny, Foulds, and Hendy 1982). In addition, finding crocodilians closest to birds (Archosauria) has been difficult to obtain with mitochondrial data (see Cao et al. 2000) but is recovered easily with the present RY-coding.

Apart from the position of the owl, the ingroup is unchanged from figure 2. Among the paleognaths, the same relationship holds between ratites and tinamous, although again deeper divergences within ratites are not highly stable. The passerines still come together and rifleman can still occur as the deepest division within Passerines. The penguin/stork/shorebird group is unchanged, but the owl has moved across to the parrot, and they come within the expected association with the seabird/shorebird/raptor group. However, the placement of owl and parrot is not strong, and a Shimodaira/Hasegawa test (1999) shows that at least 10 trees involving the deeper Neoavian lineages cannot be rejected (even at $P = 0.50$). The 10 trees all have either owls or passerines as the deepest division within Neoaves. Irrespective of the placement of the root, the passerines appear to be a very old Neoavian group, and this point needs more emphasis (see Boles 1997). As in figure 2, we indicate with an asterisk the groupings that are both expected on prior information and are stable over a variety of analyses; we would be surprised if these changed with additional data.

Analyses of the present data without RY-coding of the third position can still place the root within passerines. In such trees the oscine songbirds were separated from suboscines, and they are usually the first avian branch (although the rest of the tree was virtually unchanged). In such trees the passerines are at best paraphyletic, at worst polyphyletic. One possibility is that in earlier analyses of the avian mitochondrial data, the outgroup came into the traditional position, and that the long edge to the oscines (rook and indigobird) was secondarily attracted to the long edge of the outgroup. In any case, adding the outgroup can lead to a rearrangement of the unrooted avian tree. We have reported such effects in mammals (Lin et al. 2002) and in simulations (Holland, Penny, and Hendy 2003). Although it is possible in simulations for the addition of the outgroup to correct an error in the ingroup, it is much more common for the outgroup to disrupt a correct ingroup (Holland, Penny, and Hendy 2003). This is additional grounds for accepting the root in figure 3 as highly likely to be correct. Finding a taxon (such as lyrebird) that breaks up the long branch to the oscines is a priority, and our prediction is that, even without RY-coding, the root will then come between paleognaths and neognaths.

Some preliminary results on dating are given here, basically comparing results with two new calibration points. The first is a new penguin date taken at 62 MYA (Jones and Mannering 1997; Slack et al. in preparation). Good fossils (Jones and Mannering 1997) of at least two species of early penguins dated at between 61 and 63 MYA have been found in North Canterbury, New Zealand. This calibration point may be conservative because the closest bird to penguin in the present data set is a stork. The second calibration point is the Presbyornis/magpie goose divergence estimated at either 60 MYA (Ericson 1997) or (with the discovery of new fossils on Vega Island, Antarctica) 66–67 MYA (Case and Tambussi 1999). This latter site was discovered relatively recently and has the remains of at least five different species that fall within modern birds (J. Case, personal communication). Our aim here is to compare the divergence times estimated from the

two potential Presbyornid dates with those found using the penguin date.

Divergence times estimated by the Sanderson (1997) method that allows rate variation, are given in table 2. We have deliberately omitted confidence intervals to focus on the issue of the two Presbyornis/magpie goose divergences; the older divergence (66 MYA) is in agreement with the penguin/stork date. Although this is encouraging, it is preferable that these Vega Island fossils (Case and Tambussi 1999) be fully described, because they will give additional calibration points, including burhinid (stone curlew/thick knees) shorebirds. In general, our results are about 10% younger than those of Van Tuinen and Hedges (2001). They used an external calibration point approach with the avian/mammalian divergence at 310 MYA as their primary point. A combination of internal and external calibration points may be preferable, because interpolating between points can give an unbiased estimate (M.A. Steel and M.D. Hendy, personal communication).

Our preliminary analyses support at least 13 lineages of modern birds surviving from the Cretaceous to the present. These include two lineages each of ratites, anseriformes, and passerines; plus at least one lineage each of tinamou, galliformes, owls, parrots, shorebirds, falconiformes, and stork/penguin. In Cooper and Penny (1997) we report 22 lineages, none of which are contradicted on this present data set with fewer taxa but longer sequences. Although several methods of estimating divergence have been tried in the present work, our preference at present is to resolve the avian tree further before returning to date estimates.

Table 2
Dating Estimates for Early Avian Divergences Based on Two Calibration Points

	Penguin/Stork @ 62 MYA	Magpie Goose/Duck @ 60 MYA
Within birds		
Paleognaths/neognaths	101	92
Ratites/tinamou	84	77
Ostrich/other ratites	75	69
Gallianseres /Neoaves	90	82
Galliforms/ anseriforms	76	70
Magpie goose/duck+goose	66	**60** (fixed)
Owl/other neoavian birds	80	73
Passerines/other neoavians	78	71
Oscines/suboscines	70	64
Falconiforms+parrot/rest	74	68
Falconiforms/parrot	72	66
Shorebirds/penguin,stork	74	68
Penguin/stork	**62** (fixed)	57
Outside birds		
Birds/crocodilians	183	167
Archosaurs/turtles	199	182
Turtles (green/painted)	79	72
Iguana/skink	146	134

NOTE.—Calibration points (boldface) are a penguin/stork divergence of 62 MYA (left column) and a Presbyornis/magpie goose divergence of 60 MYA (the conservative estimate for Presbyornis, right column). The less conservative Presbyornis/magpie goose divergence (66 MYA) gives the same estimates as the penguin calibration point. Standard errors are omitted to permit focus on the comparisons.

Discussion

An important reason to establish a good phylogeny of modern birds (the crown group) and then estimate divergences times is that fundamental evolutionary models can be tested (see fig. 4). Our underlying interest here is whether the diversifying lineages of modern birds were competing with (and possibly outcompeting) pterosaurs and earlier avian groups during the Late Cretaceous. In other words, can we use dated trees to infer evolutionary processes? If all lineages of modern avian orders only diverged and diversified in the Tertiary (after the extinction of the earlier groups) then modern birds cannot have affected these earlier groups, either directly or indirectly. This example, basically the Feduccia model (1995; 2003), is Model A in figure 4—all modern birds have a common ancestor in the Tertiary. On this model, all ecological, morphological, and taxonomic differentiation of birds (ratites and raptors, swifts and seabirds, penguins and parrots, owls and oystercatchers) occurred early within the Tertiary, and by unknown genetic mechanisms.

There is a range of alternative models. One (4B) is that many lineages of modern birds diverged in the Cretaceous, but diversification into the range of forms and niches we see today only occurred in the Tertiary. Here we distinguish *divergence* of lineages, and *diversification* into a range of ecological, taxonomic, and morphological forms. This includes both short fuse and long fuse models (Cooper and Fortey 1998; Springer et al. 2003). Under this model, there may have been little competition during the Late Cretaceous between modern birds and earlier birds; each could still be in a separate niche. In contrast, the third model (4C) proposes that phylogenetic divergence and ecological/morphological diversification both occurred in the Late Cretaceous. This does not mean that all orders of birds diverged and diversified in the Late Cretaceous, but that most of them did. In 4C, major ecological transitions occurred during the Cretaceous, and we expect that modern birds were competing in the same niche as some earlier birds (such as enantiornithines, *Hesperornis*, and *Ichthyornis*) and pterosaurs. There is a range of intermediates between 4A, 4B, and 4C.

It is premature to decide which model is closest to being correct, and more detailed treatments are needed (Phillips and Penny 1998; Penny and Phillips, in preparation). Although results clearly contradict model A—all divergences in the Tertiary—the present evidence is insufficient to decide between B and C. Both models have early divergences, but current results do not tell directly about diversification. The eventual goal is to understand interactions in the Late Cretaceous between modern birds and the earlier groups mentioned above. It is helpful to separate the process into three steps: the phylogeny, comparison of divergence times based on molecular and fossil data, and ecological transformations (if any) in the Late Cretaceous. For the first, we do not require a complete phylogeny of all modern birds; just a robust phylogeny of major avian groups, especially those for which the oldest fossils are available. Our priority is to improve taxon

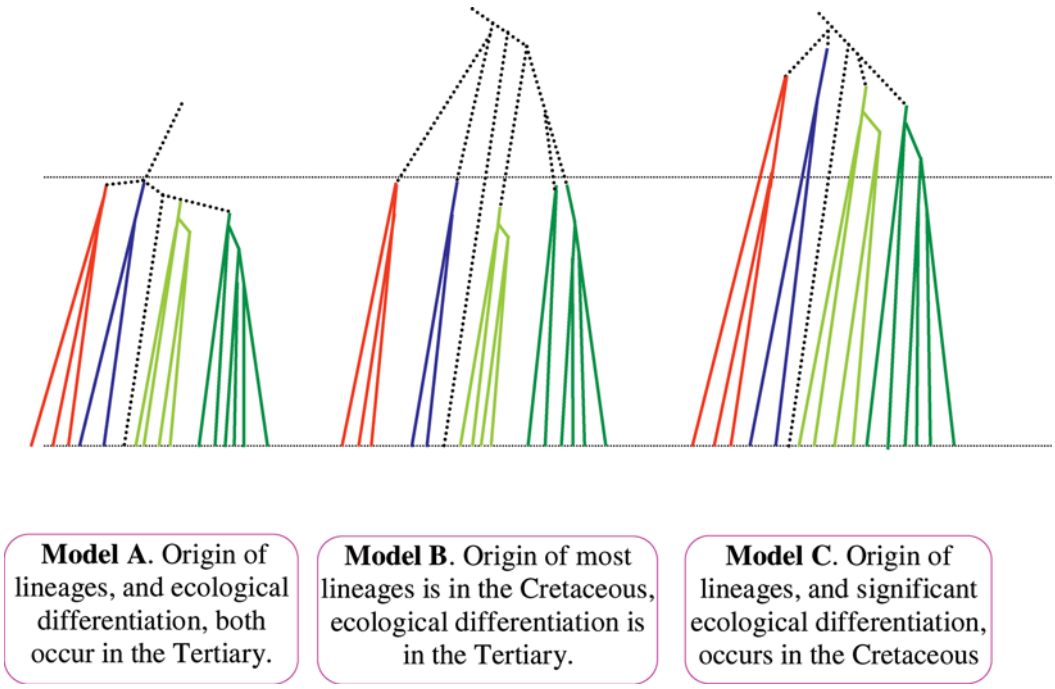


FIG. 4.—Three general models that need to be evaluated for both avian and mammalian evolution. In model A, modern orders of birds both originate and diversify ecologically in the Tertiary. In model B, many lineages diverge in the Cretaceous, but ecological diversification is in the Tertiary. In model C, both the origin of lineages and significant ecological diversification occurs in the Cretaceous. The models differ in their implications about mechanisms of evolution leading to extinctions, and they illustrate how trees can be used to study evolutionary mechanisms. (In each model, dashed lines represent a group still within the ancestral niche.)

sampling until a more stable tree is obtained for both nuclear and mitochondrial data. This will allow stronger estimates of the divergence of major avian lineages. Comparing divergence times with molecular and fossil data (Bromham and Penny 2003; Smith and Peterson 2002) still requires careful work, but there is a large body of evidence for fossil of modern birds toward the end Cretaceous (Hope 2002). This includes newer fossil discoveries in Antarctica such as reported in Case and Tambussi (1999). The final aspect is evaluating the fossil record for the ecological role of modern birds in the Late Cretaceous (Chiappe and Dyke 2002), including evidence from fossil footprints (Lockley and Rainforth 2002). Only after this analysis can our models be evaluated thoroughly.

So far the discussion has been on general issues of avian evolution, not specifically on the other claim in Feduccia (2003), the relatively recent origin of passerines. There is no fossil information available to substantiate that claim. Recent molecular work places early oscine evolution in Australia (Barker, Barrowclough, and Groth 2002; Edwards and Boles 2002), and this is consistent with the earliest known passerine fossil being Australian (Boles 1995). Unfortunately, no land vertebrate fossil beds are known from Australia between the early Eocene (54 MYA) until the Early Cretaceous (105 MYA). Gurnis, Müller, and Moresi (1998) report that plate tectonic processes raised the Australian continent during the mid-late Cretaceous by up to 250 m, leaving few areas for net deposition and fossilization. The absence of fossil beds means that it is unlikely there will be fossil evidence, for or against, the older origin of passerines, and so the molecular data stand alone.

There appears to be sufficient information in the mitochondrial data to recover a good avian phylogeny, especially with RY-coding. Although our results support the avian root between paleognaths and neognaths, it can appear arbitrary if some analyses are favored over others, even if the rooting is supported by prior information. For this reason the treeness/RCV ratio is helpful in evaluating which method of analysis gets the most phylogenetic signal. There are many signals in DNA sequence data (Penny et al. 1993). There is no guarantee that the largest signals are always the correct phylogeny, and in the present data there is some signal from a particular form of nucleotide bias (such as within pyrimidines) which has to be reduced.

There is always a tendency to "blame the data" if a predicted result is not obtained. On the contrary, we suggest the data are neutral; it is the methods of analysis that are inadequate. It is important to develop improved methods that more accurately reflect the underlying mutational mechanisms; an "optimal" model can still give a wrong tree. Thus we require methods that determine which aspects of the mutational mechanism and/or selection are accounted for, and which are not. With both mammals and birds it appears that improved taxon sampling and RY-coding are key factors in obtaining highly congruent trees between different data types (nuclear and mitochondrial). Of course, there will be cases where the appropriate taxa no longer exist and improved taxon sampling will not be

possible. Overall, the results are extremely encouraging that the avian tree is being resolved, and will then allow improved estimates of the survival of bird lineages through the Cretaceous/Tertiary boundary (Cooper and Penny 1997).

Acknowledgments

We thank Trevor Worthy for the rifleman sample, Nga Manu Bird Sanctuary for morepork, David Lambert for kakapo, and Peter Whitehead and Julian Gorman, Darwin, Australia, for the magpie goose sample. We also thank Craig Jones and Judd Case for sharing information on new avian fossils (penguin and *Presbyornis*, respectively). The New Zealand Marsden Fund supported the work.

Literature Cited

Adachi, J., and M. Hasegawa. 1996. MOLPHY. Computer program published by the authors. Tokyo. http://bioweb.pasteur.fr/seqanal/interfaces/MolPhy.html.

Barker, F. K., G. F. Barrowclough, and J. G. Groth. 2002. A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data. Proc. R. Soc. Lond. Ser. B **269**:295–308.

Boles, W. E. 1995. The world's oldest songbird. Nature **374**: 21–22.

———. 1997. Fossil songbirds (Passeriformes) from the Early Eocene of Australia. Emu **97**:43–50.

Braun, E. L., and R. T. Kimball. 2002. Examining basal avian divergences with mitochondrial sequences: model complexity, taxon sampling, and sequence length. Syst. Biol. **51**:614–625.

Bromham, L. D., and D. Penny. 2003. The modern molecular clock. Nat. Rev. Genet. **4**:216–224.

Cao, Y., M. D. Sorenson, Y. Kumazawa, D. P. Mindell, and M. Hasegawa. 2000. Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes. Gene **259**:139–148.

Case, J. A., and C. P. Tambussi. 1999. Maestrichtian record of neornithine birds in Antarctica: comments on a Late Cretaceous radiation of modern birds. J. Vert. Paleontol. **19**:37A.

Chiappe, L. M., and G. J. Dyke. 2002. The Mesozoic radiation of birds. Annu. Rev. Ecol. Syst. **33**:91–124.

Cooper, A., and R. Fortey. 1998. Evolutionary explosions and the phylogenetic fuse. Trends Ecol. Evol. **13**:151–156.

Cooper, A., C. Lalueza-Fox, S. Anderson, A. Rambaut, J. Austin, and R. Ward. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature **409**:704–707.

Cooper, A., and D. Penny. 1997. Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. Science **275**:1109–1113.

Cracraft, J. 2001. Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. Proc. R. Soc. Lond. Ser. B **268**:459–469.

Delsuc, F., M. J. Phillips, and D. Penny. 2003. Technical comment on "Hexapod origins: monophyletic or paraphyletic?" Science **301**:1482d.

Eberhard, J. R., T. F. Wright, and E. Bermingham. 2001. Duplication and concerted evolution of the mitochondrial control region in the parrot genus Amazona. Mol. Biol. Evol. **18**:1330–1342.

Edwards, S. V., and W. E. Boles. 2002. Out of Gondwana: the origin of passerine birds. Trends Ecol. Evol. **17**:347–349.

Ericson, P. G. P. 1997. Systematic relationships of the paleogene family Presbyornithidae (Aves: Anseriformes). Zool. J. Linn. Soc. **121**:429–483.

Ericson, P. G. P., L. Christidis, A. Cooper, M. Irestedt, J. Jackson, U. S. Johansson, and J. A. Norman. 2002. A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. Proc. R. Soc. Lond. Ser. B **269**:235–241.

Feduccia, A. 1995. Explosive evolution in Tertiary birds and mammals. Science **267**:637–638.

———. 2003. "Big bang" for tertiary birds? Trends Ecol. Evol. **8**:172–176.

Garcia-Moreno, J., M. D. Sorenson, and D. P. Mindell. 2003. Congruent avian phylogenies inferred from mitochondrial and nuclear DNA sequences. J. Mol. Evol. **57**:27–37.

Gurnis, M., R. D. Müller, and L. Moresi. 1998. Cretaceous vertical motion of Australia and the Australian-Antarctic discordance. Science **279**:1499–1504.

Haddrath, O., and A. J. Baker. 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. Proc. R. Soc. Lond. Ser. B. **268**:939–945.

Haring, E., L. Kruckenhauser, A. Gamauf, M. J. Riesing, and W. Pinsker. 2001. The complete sequence of the mitochondrial genome of *Buteo buteo* (Aves, Accipitridae) indicates an early split in the phylogeny of raptors. Mol. Biol. Evol. **18**:1892–1904.

Härlid, A, A. Janke, and U. Arnason. 1999. The complete mitochondrial sequence of Rhea americana and early avian divergences. J. Mol. Evol. **46**:669–679.

Hendy, M. D., and D. Penny. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. **38**:297–309.

Holland, B. R., D. Penny, and M. D. Hendy. 2003. Accuracy of tree reconstruction under equal rates. Syst. Biol. **52**:229–238.

Hope, S. 2002. The Mesozoic radiation of Neornithes. Pp. 339–388 *in* L. M. Chiappe and L. M. Witmer, eds., Mesozoic birds: above the heads of dinosaurs. University of California Press, Berkeley, Calif.

Huelsenbeck, J., and F. Ronquist. 2001. MrBayes: Bayesian inference on phylogenetic trees. Bioinformatics **17**:754–755.

Jones, C. M., and A. Mannering 1997. A new Paleocene fossil bird material from the Waipara Greensand, North Canterbury, New Zealand. Geol. Soc. N. Z. Misc. Publ. 95A, 88.

Lanyon, S. 1988. The stochastic mode of molecular evolution: what consequences for systematic investigators? Auk **105**: 565–573.

Lin, Y-L., P. A. McLenachan, A. R. Gore, M. J. Phillips, R. Ota, M. D. Hendy, and D. Penny. 2002. Four new mitochondrial genomes, and the stability of evolutionary trees of mammals. Mol. Biol. Evol. **19**:2060–2070.

Livezey, B. C. 1997. A phylogenetic analysis of basal Anseriformes, the fossil *Presbyornis*, and the interordinal relationships of waterfowl. Zool. J. Linn. Soc. **12**:361–428.

Livezey, B. C., and R. L. Zusi. 2001. Higher-order phylogenetics of modern Aves based on comparative anatomy. Neth. J. Zool. **51**:179–205.

Lockley, M. G., and E. C. Rainforth. 2002. The track record of Mesozoic birds and pterosaurs. Pp 405–418 *in* L. M. Chiappe and L. M. Witmer, eds., Mesozoic birds: above the heads of dinosaurs. University of California Press, Berkeley, Calif.

Mindell, D. P., M. D. Sorenson, and D. E. Dimcheff. 1998. Multiple independent origins of mitochondrial gene order in birds. Proc. Natl Acad. Sci. USA **95**:10693–10697.

Mindell, D. P., M. D. Sorenson, D. E. Dimcheff, M. Hasegawa, J. C. Ast, and T. Yuri. 1999. Interordinal relationships of

birds and other reptiles based on whole mitochondrial genomes. Syst. Biol. **48**:138–152.

Mindell, D. P., M. D. Sorenson, C. J. Huddleston, H. C. Miranda, A. Knight, S. J. Sawchuk, and T. Yuri. 1997. Phylogenetic relationships among and within select avian orders based on mitochondrial DNA. Pp. 00–00 *in* D. P. Mindell, ed., Avian Molecular Evolution and Systematics. Academic Press, San Diego, Calif.

Nardi, F., G. Spinsanti, J. L. Boore, A. Carapelli, R. Dallai, and F. Frati. 2003. Hexapod origins: monophyletic or paraphyletic? Science **299**:1887–1889.

Noriega, J. I., and C. P. Tambussi. 1995. A Late Cretaceous Presbyornithidae (Aves Anseriformes) from Vega Island, Antarctic Peninsula; biogeographic interpretations. Ameghiniana **32**:57–61.

Paton, T., O. Haddrath, and A. J. Baker. 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proc. R. Soc. Lond. Ser. B. **269**:839–846.

Penny, D., L. R. Foulds, and M. D. Hendy. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. Nature **297**: 197–200.

Penny, D., E. E. Watson, R. E. Hickson, and P. J. Lockhart. 1993. Some recent progress with methods for evolutionary trees. N. Z. J. Bot. **31**:275–288.

Phillips, M. J., and D. Penny. 1998. Mass survival of tetrapods across the Cretaceous-Tertiary boundary. Pp 63–65 *in* R. A. Cooper and C. M. Jones, eds., Geology and Genes. Geol. Soc. N. Z. Publ. 97(ISBN 0–908678–69-X), Wellington, New Zealand.

———. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogenet. Evol. **28**:171–185.

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

Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. **14**:1218–1231.

Schmitz, J., M. Ohme, B. Suryobroto, and H. Zischler. 2002. The Colugo (*Cynocephalus variegatus*, Dermoptera): the primates' gliding sister? Mol. Biol. Evol. **19**:2308–2312.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic inference. Mol. Biol. Evol. **16**:1114–1116.

Sibley, C. G., and J. E. Ahlquist. 1990. Phylogeny and classification of birds. Yale University Press, New Haven, Conn.

Slack, K. E., A. Janke, D. Penny, and U. Arnason. 2003. Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene **302**:43–52.

Smith, A. B., and K. J. Peterson. 2002. Dating the time of origin of major clades. Annu. Rev. Earth Planet. Sci. **30**: 65–88.

Sorenson, M. D., J. C. Ast, D. E. Dimcheff, T. Yuri, and D. P. Mindell. 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Mol. Phylogenet. Evol. **12**:105–114.

Sorenson, M. D., C. Oneal, J. Garcia-Moreno, and D. P. Mindell. 2003. More taxa, more characters: the hoatzin problem is still unresolved. Mol. Biol. Evol. **20**:1484–1498.

Springer, M. S., W. J. Murphy, E. Eizirik, and S. J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc. Natl Acad. Sci. USA **100**:1056–1061.

Sraml, M., L. Christidis, S. Easteal, P. Horn, and C. Collet. 1996. Molecular relationships between Australian waterfowl (Anseriformes). Aust. J. Zool. **44**:47–58.

Stidham, T. A. 1998. A lower jaw from a Cretaceous parrot. Nature **396**:29–30.

Swofford, D. L. 1998. PAUP*, Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Mass.

Van Tuinen, M., and S. B. Hedges. 2001. Calibration of avian molecular clocks. Mol. Biol. Evol. **18**:206–213.

Van Tuinen, M., C. Sibley, and S. B. Hedges. 2000. The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. Mol. Biol. Evol. **17**:451–457.

Waddell, P. J., H. Kishino, and R. Ota. 2001. A phylogenetic foundation for comparative mammalian genomics. Genome Informatics **12**:141–154.

William Martin, Associate Editor

Accepted November 26, 2003

Appendix 4. Tables summarizing avian mitochondrial (mt) genome features

These tables were first published in Slack et al. (2003; see Chapter 2) and have here been updated to include all of the birds analysed in Chapter 4 (Slack et al. 2007) plus 11 additional taxa: a second greater rhea (Greater Rhea 2; *Rhea americana*; AF090339), lesser rhea (*Pterocnemia pennata*; AF338709), a second ostrich (Ostrich 2; *Struthio camelus*; AF338715), two eastern moas (*Emeus crassus*; Eastern Moa 1: AY016015; Eastern Moa 2: AF338712), little bush moa (*Anomalopteryx didiformis*; AF338714), giant moa (*Dinornis giganteus*; AY016013), the original chicken mt sequence (Chicken 1; *Gallus gallus*; X52392), morepork (New Zealand owl; *Ninox novaeseelandiae*; AY309457), kakapo (New Zealand ground parrot; *Strigops habroptilus*; AY309456) and white stork (*Ciconia ciconia*; AB026818). The information given in these tables incorporates a number of changes made to the start and stop points of these genes in order to provide consistency between avian mt genomes - see 'Revised Bird Annotations' online at: http://www.allanwilsoncentre.ac.nz/massey/research/centres-research/allan-wilson-centre-molecular-ecology-and-evolution/teachers--students/download/supplementary_information/en/supplementary_information_home.cfm.

There are three sets of tables:

Table 1. Length (in amino acids), start and stop codons of avian mitochondrial protein-coding genes

The length of each gene includes the start but not the stop codon (i.e. these genes are treated as ending after the last amino acid-encoding codon). The length of NADH3 does not include the extra nucleotide found in many of the taxa. T-- and TA- = incomplete stop codons. R and Y (in the range columns) indicate purine (A/G) and pyrimidine (C/T) respectively.

Table 2. Length (in nucleotides) of avian mitochondrial control regions, intergenic spacers and complete genomes

The lengths of the intergenic spacers include the stop codons of the protein-coding genes (i.e. protein-coding genes are treated as ending after the last amino acid-encoding codon). **N/A** = Not applicable (i.e. this taxon does not have this gene/region arrangement). + = gene/region order is reversed in the crested-tinamou (i.e. tRNA-Pro ***(L)***/Control Region). # = 'tRNA-Glu ***(L)***/Additional Non-Coding Region' and 'Additional Non-Coding Region/tRNA-Phe' in the crested-tinamou, broadbill, flycatcher, lyrebird, falcon and buzzard. - = 0 (except in the range columns, in which '0' is used while '-' is used to indicate the range). x overlap = an overlap of x nucleotides in length. Light gray shading - these features are conserved as a result of the way the start and stop positions of the rRNA genes, the control region and the additional non-coding region are determined.

Both the single-letter and the three-letter code are given for each tRNA (e.g. F, Phe).

(L) = encoded by the L strand. * = incomplete gene/region/genome, full length unknown. Gray shading - this feature (length, start or stop codon) is conserved among paleognaths or among neognaths. Purple shading - this feature is conserved in most paleognath or neognath taxa (≥85% of species or major groups); variations are shaded and given in purple in the range columns. Three range columns are given: one for paleognaths, one for neognaths and one for all birds. The ranges do not include values from incomplete genes/regions.

□able □ Length (in amino acids) and start and stop codons of avian mt protein-coding genes

Genes (13)		Emu	Cassowary	Kiwi	Greater Rhea 1	Greater Rhea 2	Lesser Rhea	Ostrich 1
NADH1	**Length**	323	323	324	324	324	324	324
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	AGG	AGG	AGG	AGG	AGG
NADH2	**Length**	346	346	346	346	346	346	346
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAG	TAG	TAG	TAG	TAG	TAG	TAG
COI	**Length**	516	516	516	516	516	516	516
	Start codon	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)
	Stop codon	AGG	AGG	AGG	AGG	AGG	AGG	AGG
COII	**Length**	229	229	229	229	229	229	229
	Start codon	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)
	Stop codon	T--	T--	T--	T--	T--	T--	T--
ATPase8	**Length**	55	55	55	55	55	55	55
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAA
ATPase6	**Length**	227	227	227	227	227	227	227
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAA
COIII	**Length**	261	261	261	261	261	261	261
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	T--	T--	T--	T--	T--	T--	T--
NADH3	**Length**	116	116	116	116	116	116	116
	Start codon	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)
	Stop codon	TAA	TAA	TAA	TAG	TAG	TAA	TAA
NADH4L	**Length**	98	98	98	98	98	98	98
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAA
NADH4	**Length**	458	458	455	458	458	458	458
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	AGA	TAA	TAA	TAA	TAA
NADH5	**Length**	605	605	604	605	605	605	605
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	GTG (Val)
	Stop codon	AGA	AGA	TAA	TAA	TAA	TAA	AGA
Cyt*b*	**Length**	379	379	379	379	379	379	379
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAG	TAA	TAA	TAA	TAA	TAA
NADH6 *(L)*	**Length**	173	173	173	174	174	174	173
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAG	TAG	TAG	TAG	TAG	TAG	TAG

□able □continued

Ostrich 2	Eastern Moa 1	Eastern Moa 2	Little Bush Moa	Giant Moa	Great Tinamou	Crested-Tinamou	Range (paleognaths)
324	323	323	323	323	322	322	322 - 324
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
AGG	TAA	TAA	TAA	TAA	AGA	AGA	AGR, TAA
346	346	346	346	346	346	346	346
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATA (Met)	ATG (Met); ATA (Met)
TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG
516	516	516	516	516	516	516	516
GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)
AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
229	228	228	228	228	228	228	228 - 229
GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)
T--	T--	T--	T--	T--	T--	T--	T--
55	55	55	55	55	55	55	55
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
227	227	227	227	227	227	227	227
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
261	261	261	261	261	261	261	261
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
T--	T--	T--	T--	T--	T--	T--	T--
116	116	116	116	116	116	116	116
ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)	ATA (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA; TAG
98	98	98	98	98	97	98	97; 98
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
458	458	458	458	458	458	458	455; 458
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA; AGA
605	605	605	605	605	603	606	603 - 606
GTG (Val)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATA (Met)	ATG (Met)	ATG (Met); GTG (Val), ATA (Met)
AGA	AGA	AGA	AGA	AGA	TAA	TAA	TAA, AGA
379	380	380	380	380	379	379	379; 380
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA; TAG
173	173	173	173	173	173	173	173; 174
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG

□able □continued

Genes (13)		Chicken 1	Chicken 2	Quail	Brush-turkey	Magpie Goose	Duck	Goose
NADH1	**Length**	324	324	324	325	324	325	325
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	AGG	AGG	AGG	AGG
NADH2	**Length**	346	346	346	346	346	346	346
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAG	TAG	TAG	TAG	TAA	TAG	TAG
COI	**Length**	515	516	516	516	516	516	516
	Start codon	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)
	Stop codon	AGG	AGG	AGG	AGG	AGG	AGG	AGG
COII	**Length**	227	227	227	227	228	228	228
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	GTG (Val)	ATG (Met)	GTG (Val)	GTG (Val)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAA
ATPase8	**Length**	54	54	55	55	55	55	55
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAA
ATPase6	**Length**	227	227	227	227	227	227	227
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAA
COIII	**Length**	261	261	261	261	261	261	261
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	T--	T--	T--	T--	T--	T--	T--
NADH3	**Length**	116	116	116	116	116	116	116
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATA (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAG
NADH4L	**Length**	98	98	98	98	98	98	98
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	GTG (Val)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA	TAA
NADH4	**Length**	459	459	459	459	459	459	459
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	T--	T--	T--	T--	T--	T--	T--
NADH5	**Length**	605	605	606	604	605	607	605
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	GTG (Val)	GTG (Val)
	Stop codon	TAA	TAA	TAA	TAA	AGA	TAA	AGA
Cyt*b*	**Length**	380	380	380	380	380	380	380
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TA-	TAA	TAA	TAA
NADH6 *(L)*	**Length**	173	173	173	173	173	173	173
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	AGG	TAG	TAG	TAG	TAG

□able □continued

Morepork	Kakapo	Rifleman	Broadbill	Flycatcher	Lyrebird	Indigobird	Rook
324	326	325	325	325	325	325	325
ATG (Met)	ATG (Met)	ATA (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
AGG	AGG	AGG	TAA	AGA	AGG	AGA	AGG
348	346	346	346	346	346	346	346
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAG	TAA	TAG	TAA	TAG	TAA	TAA	TAA
516	515	516	516	516	516	516	516
GTG (Val)	GTG (Val)	GTG (Val)	ATG (Met)	ATG (Met)	GTG (Val)	ATG (Met)	GTG (Val)
AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
226	227	227	227	227	227	227	227
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
55	55	55	55	55	55	55	55
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
227	227	227	227	227	227	227	227
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
261	261	261	261	261	261	261	261
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
T--	T--	T--	T--	T--	T--	T--	T--
116	116	116	116	116	116	116	116
ATA (Met)	ATA (Met)	ATC (Ile)	ATC (Ile)	ATT (Ile)	ATC (Ile)	ATT (Ile)	ATG (Met)
TAA	TA-	TAA	TAG	TAA	TAA	TAA	TAA
98	98	98	98	98	98	98	98
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
459	464	455	459	459	458	459	459
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
T--	T--	AGA	T--	T--	TAG	T--	T--
606	604	605	601	605	604	605	606
ATG (Met)	GTG (Val)	ATG (Met)	ATA (Met)	GTG (Val)	ATG (Met)	ATG (Met)	ATG (Met)
AGA	TAG	TAA	TAA	AGA	TAA	AGA	AGA
380	379	380	380	380	380	380	380
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
173	171	173	173	173	172	172	173
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAG	TAG	TAG	AGG	AGG	AGG	TAG	TAA

□able □continued

Genes (13)		Falcon	Buzzard	Turkey Vulture	Oyster-catcher	Turnstone	Gull
NADH1	**Length**	324	325	325	325	325	325
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	AGG	AGG	AGG	AGG	AGG
NADH2	**Length**	346	346	346	346	346	346
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAG	TAG	TAG	TAG	TAG
COI	**Length**	516	516	516	516	516	516
	Start codon	GTG (Val)	GTG (Val)	ATG (Met)	ATG (Met)	GTG (Val)	GTG (Val)
	Stop codon	AGG	AGG	AGG	AGG	AGG	AGG
COII	**Length**	227	227	227	227	227	227
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA
ATPase8	**Length**	55	55	56	55	55	55
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA
ATPase6	**Length**	227	227	227	227	227	227
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA
COIII	**Length**	261	261	261	261	261	261
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	T--	T--	T--	T--	T--	T--
NADH3	**Length**	116	116	116	116	116	116
	Start codon	ATA (Met)	ATC (Ile)	ATC (Ile)	ATT (Ile)	ATA (Met)	ATT (Ile)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA
NADH4L	**Length**	98	98	98	98	98	98
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	TAA	TAA	TAA	TAA
NADH4	**Length**	459	459	459	459	459	459
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	T--	T--	T--	T--	T--	T--
NADH5	**Length**	605	605	604	604	604	604
	Start codon	ATG (Met)	GTG (Val)	ATG (Met)	GTG (Val)	GTG (Val)	GTG (Val)
	Stop codon	AGA	TAA	AGA	TAA	AGA	AGA
Cyt*b*	**Length**	380	380	381	380	380	380
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAA	TAA	T--	TAA	TAA	TAA
NADH6 *(L)*	**Length**	173	172	173	173	173	173
	Start codon	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
	Stop codon	TAG	TAG	TAG	TAA	TAG	TAG

□able □continued

White Stork	Oriental Stork	Loon	Penguin	Albatross	Petrel	Range (neognaths)	Range (all)
325	325	325	325	325	325	324 - 326	322 - 326
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met); ATA (Met)	ATG (Met); ATA (Met)
AGG	AGG	AGG	AGG	AGG	AGG	AGR, TAA	AGR, TAA
346	346	346	346	346	346	346; 348	346; 348
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met); ATA (Met)
TAA	TAA	TAG	TAG	TAG	TAG	TAR	TAR
516	516	516	516	516	516	515; 516	515; 516
GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val)	GTG (Val), ATG (Met)	GTG (Val); ATG (Met)
AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
227	227	227	227	227	227	226; 227; 228	226 - 229
GTG (Val)	GTG (Val)	ATG (Met)	ATG (Met)	GTG (Val)	GTG (Val)	ATG (Met), GTG (Val)	ATG (Met), GTG (Val)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	T--, TAA
55	55	55	54	55	55	54; 55; 56	54; 55; 56
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
227	227	227	227	227	227	227	227
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
261	261	261	261	261	261	261	261
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
T--	T--	T--	T--	T--	T--	T--	T--
116	116	116	116	116	116	116	116
ATC (Ile)	ATG (Met)	ATC (Ile)	ATC (Ile)	ATC (Ile)	ATC (Ile)	ATR (Met), ATY (Ile)	ATR (Met), ATY (Ile)
TAA	TAA	TAA	TAA	TAA	TAA	TAA; TAG, TA-	TAA; TAG, TA-
98	98	98	98	98	98	98	97; 98
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met); GTG (Val)	ATG (Met); GTG (Val)
TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAA
460	459	459	459	459	459	455, 458; 459; 460, 464	455 - 464
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
T--	T--	T--	TA-	T--	T--	T--; AGA, TAG, TA-	TAR, AGA, T--, TA-
603	603	604	606	604	604	601 - 607	601 - 607
ATG (Met)	ATG (Met)	ATG (Met)	GTG (Val)	ATG (Met)	ATG (Met)	ATR (Met), GTG (Val)	ATR (Met), GTG (Val)
AGG	AGG	TAA	TAA	AGA	AGA	TAR, AGR	TAR, AGR
380	380	380	380	380	380	379; 380; 381	379 - 381
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAA	TAA	TAA	TAA	TAA	TAA	TAA; TA-, T--	TAA; TAG, TA-, T--
173	173	173	172	175	173	171 - 175	171 - 175
ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)	ATG (Met)
TAG	TAG	TAG	TAG	TAG	TAG	TAR, AGG	TAR, AGG

□able □ Length (in nucleotides) of avian mt control regions, intergenic spacers and complete genomes

	Emu	Cassowary	Kiwi	Greater Rhea 1	Greater Rhea 2	Lesser Rhea	Ostrich 1
Control Region (CR)	1093	1137	1351	1171	1161	1204	1031
Additional Non-Coding Region	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
tRNA-Phe/12S rRNA	-	-	-	-	-	-	-
12S rRNA/tRNA-Val	-	-	-	-	-	-	-
tRNA-Val/16S rRNA	-	-	-	-	-	-	-
16S rRNA/tRNA-Leu (UUR)	-	-	-	-	-	-	-
tRNA-Leu (UUR)/NADH1	8	10	9	11	11	10	9
NADH1/tRNA-Ile	3	3	1	11	11	11	1
tRNA-Ile/tRNA-Gln *(L)*	6	5	9	14	14	13	11
tRNA-Gln *(L)*/tRNA-Met	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
tRNA-Met/NADH2	-	-	-	-	-	-	-
NADH2/tRNA-Trp	1	1	1	1	1	1	1
tRNA-Trp/tRNA-Ala *(L)*	1	1	1	1	1	1	-
tRNA-Ala *(L)*/tRNA-Asn *(L)*	3	3	1	2	2	2	4
tRNA-Asn *(L)*/tRNA-Cys *(L)*	3	3	2	3	3	3	1
tRNA-Cys *(L)*/tRNA-Tyr *(L)*	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
tRNA-Tyr *(L)*/COI	1	1	1	1	1	1	1
COI/tRNA-Ser (UCN) *(L)*	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap
tRNA-Ser (UCN) *(L)*/tRNA-Asp	2	2	2	2	2	2	3
tRNA-Asp/COII	1	1	1	1	1	1	1
COII/tRNA-Lys	1	1	1	1	1	1	1
tRNA-Lys/ATPase8	1	1	1	1	1	1	1
ATPase8/ATPase6	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap
ATPase6/COIII	2	2	2	2	2	2	2
COIII/tRNA-Gly	1	1	1	1	1	1	1
tRNA-Gly/NADH3	-	-	-	-	-	-	-
NADH3/tRNA-Arg	4	4	4	5	5	5	4
tRNA-Arg/NADH4L	-	-	-	-	-	-	-
NADH4L/NADH4	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap
NADH4/tRNA-His	4	4	10	4	4	4	4
tRNA-His/tRNA-Ser (AGY)	-	-	-	-	-	-	-
tRNA-Ser (AGY)/tRNA-Leu (CUN)	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
tRNA-Leu (CUN)/NADH5	-	-	-	-	-	-	-
NADH5/Cyt*b*	12	12	15	15	15	15	12
Cyt*b*/tRNA-Thr	7	7	4	8	8	8	5
tRNA-Thr/tRNA-Pro *(L)*	82	82	83	9	9	9	55
tRNA-Thr/CR	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
CR/tRNA-Pro *(L)* +	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
tRNA-Pro *(L)*/NADH6 *(L)*	23	24	30	18	18	17	13
CR/NADH6 *(L)*	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
NADH6 *(L)*/tRNA-Glu *(L)*	3	3	3	3	3	3	3
tRNA-Glu *(L)*/CR #	-	-	-	-	-	-	-
CR/tRNA-Phe #	-	-	-	-	-	-	-
Total Length of Genome	16711	16756	16980	16714	16704	16749	16591

Ostrich 2	Eastern Moa 1	Eastern Moa 2	Little Bush Moa	Giant Moa	Great Tinamou	Crested-Tinamou	Range (paleognaths)
1034	1501	>1100	>1150	1508	1103	1444	1031 - 1508
N/A	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	1350	**N/A; 1350**
-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	0
9	11	11	11	11	9	9	8 - 11
1	4	4	4	4	19	21	1 - 21
11	7	7	8	8	7	7	5 - 14
1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
-	-	-	-	-	-	-	0
1	1	1	1	1	1	1	1
-	1	1	1	1	1	1	0; 1
4	1	1	1	1	1	-	0 - 4
1	2	2	2	2	2	2	1 - 3
1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
1	1	1	1	1	1	1	1
6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap
3	2	2	2	2	2	2	2; 3
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap
2	2	2	2	2	2	2	2
1	1	1	1	1	1	1	1
-	-	-	-	-	-	-	0
4	4	4	4	4	3	4	3 - 5
-	-	-	-	-	-	-	0
4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap
4	4	4	4	4	5	5	4 - 10
-	-	-	-	-	-	-	0
1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
-	-	-	-	-	-	-	0
12	13	13	13	13	2	2	2 - 15
5	4	4	4	4	4	4	4 - 8
55	26	26	27	27	85	5	5 - 85
N/A	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
N/A	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	-	**N/A**; 0
13	14	14	14	14	12	**N/A**	12 - 30; **N/A**
N/A	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**3**	**N/A**; 3
3	-	-	-	-	-	-	0 - 3
-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	0
16595	17061	>16662	>16716	17070	16700	18305	16591 - 18305

□able □continued

	Chicken 1	Chicken 2	Quail	Brush-turkey	Magpie Goose	Duck	Goose
Control Region (CR)	1227	1231	1155	1120	1335	1066	1174
Additional Non-Coding Region	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
tRNA-Phe/12S rRNA	-	-	-	-	-	-	-
12S rRNA/tRNA-Val	-	-	-	-	-	-	-
tRNA-Val/16S rRNA	-	-	-	-	-	-	-
16S rRNA/tRNA-Leu (UUR)	-	-	-	-	-	-	-
tRNA-Leu (UUR)/NADH1	9	9	8	12	8	4	6
NADH1/tRNA-Ile	3	3	3	1	1	1	1
tRNA-Ile/tRNA-Gln *(L)*	5	5	5	9	8	8	7
tRNA-Gln *(L)*/tRNA-Met	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
tRNA-Met/NADH2	-	-	-	-	-	-	-
NADH2/tRNA-Trp	1	1	1	1	2	1	1
tRNA-Trp/tRNA-Ala *(L)*	6	6	5	6	8	3	5
tRNA-Ala *(L)*/tRNA-Asn *(L)*	3	3	2	2	2	2	1
tRNA-Asn *(L)*/tRNA-Cys *(L)*	1	1	-	-	2	-	3
tRNA-Cys *(L)*/tRNA-Tyr *(L)*	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
tRNA-Tyr *(L)*/COI	1	1	1	1	1	1	1
COI/tRNA-Ser (UCN) *(L)*	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap
tRNA-Ser (UCN) *(L)*/tRNA-Asp	2	2	2	3	2	2	2
tRNA-Asp/COII	1	1	1	1	1	1	1
COII/tRNA-Lys	4	4	4	4	4	4	4
tRNA-Lys/ATPase8	1	1	1	1	1	1	1
ATPase8/ATPase6	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap
ATPase6/COIII	2	2	2	2	2	2	2
COIII/tRNA-Gly	1	1	1	1	1	1	1
tRNA-Gly/NADH3	-	-	-	-	-	-	-
NADH3/tRNA-Arg	4	4	4	4	4	4	4
tRNA-Arg/NADH4L	-	-	-	-	-	-	-
NADH4L/NADH4	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap
NADH4/tRNA-His	1	1	1	1	1	1	1
tRNA-His/tRNA-Ser (AGY)	-	-	-	-	-	-	-
tRNA-Ser (AGY)/tRNA-Leu (CUN)	-	-	-	-	-	1 overlap	1 overlap
tRNA-Leu (CUN)/NADH5	-	-	-	-	-	-	-
NADH5/Cyt*b*	7	7	2	15	9	2	10
Cyt*b*/tRNA-Thr	6	6	6	2	5	5	5
tRNA-Thr/tRNA-Pro *(L)*	-	-	2	2	2	10	8
tRNA-Thr/CR	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
CR/tRNA-Pro *(L)* +	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
tRNA-Pro *(L)*/NADH6 *(L)*	9	9	8	12	10	13	13
CR/NADH6 *(L)*	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
NADH6 *(L)*/tRNA-Glu *(L)*	2	2	1	1	1	-	-
tRNA-Glu *(L)*/CR #	-	-	-	-	-	-	-
CR/tRNA-Phe #	-	-	-	-	-	-	-
Total Length of Genome	16775	16788	16697	16698	16870	16616	16737

□able □continued

Morepork	Kakapo	Rifleman	Broadbill	Flycatcher	Lyrebird	Indigobird	Rook
>662	>69	>503	1453	1440	2107	1295	1339
N/A	**N/A**	**N/A**	301	178	173	**N/A**	**N/A**
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
13	7	19	10	13	13	8	9
1	1	1	24	10	1	9	11
9	8	8	11	4	5	5	5
1 overlap	1 overlap	1 overlap	12	1 overlap	1 overlap	1 overlap	1 overlap
-	-	-	-	-	-	-	-
1	2	1	4	1	2	2	5
1	1	1	1	1	1	1	1
1	1	4	6	1	2	9	9
2	2	13	6	2	2	-	2
1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
1	9	1	1	1	1	1	1
6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap
4	3	3	2	2	4	5	5
2	4	7	3	10	11	10	8
4	4	4	4	4	4	14	4
1	1	1	1	1	1	1	1
7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap
2	2	2	12	9	14	8	10
1	1	1	1	1	1	1	1
-	-	-	-	-	-	-	-
4	2	5	7	7	6	4	4
1	1	1	1	1	1	1	1
4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap
1	1	13	1	1	4	1	1
2	-	-	-	-	-	-	-
4	1 overlap	1	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
-	-	2	-	-	-	-	-
10	13	16	12	14	11	11	13
4	3	8	7	5	6	6	6
11	6	6	**N/A**	**N/A**	**N/A**	12	11
N/A	**N/A**	**N/A**	-	-	-	**N/A**	**N/A**
N/A	**N/A**	**N/A**	-	-	-	**N/A**	**N/A**
15	12	23	8	15	11	29	12
N/A	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
3	2	2	-	2	-	1	1
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
>16223	>15599	>16069	17344	17171	17839	16895	16931

□able □continued

	Falcon	Buzzard	Turkey Vulture	Oyster-catcher	Turnstone	Gull
Control Region (CR)	1510	1672	1177	1240	1173	1155
Additional Non-Coding Region	950	1455	**N/A**	**N/A**	**N/A**	**N/A**
tRNA-Phe/12S rRNA	-	-	-	-	-	-
12S rRNA/tRNA-Val	-	-	-	-	-	-
tRNA-Val/16S rRNA	-	-	-	-	-	-
16S rRNA/tRNA-Leu (UUR)	-	-	-	-	-	-
tRNA-Leu (UUR)/NADH1	15	9	13	2	16	4
NADH1/tRNA-Ile	19	1	1	1	1	1
tRNA-Ile/tRNA-Gln *(L)*	9	13	6	9	8	9
tRNA-Gln *(L)*/tRNA-Met	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
tRNA-Met/NADH2	1	-	-	-	-	-
NADH2/tRNA-Trp	2	1	1	1	1	1
tRNA-Trp/tRNA-Ala *(L)*	10	1	1	1	1	1
tRNA-Ala *(L)*/tRNA-Asn *(L)*	10	2	2	2	3	2
tRNA-Asn *(L)*/tRNA-Cys *(L)*	2	2	10	3	2	2
tRNA-Cys *(L)*/tRNA-Tyr *(L)*	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
tRNA-Tyr *(L)*/COI	1	1	6	1	1	1
COI/tRNA-Ser (UCN) *(L)*	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap
tRNA-Ser (UCN) *(L)*/tRNA-Asp	2	4	2	2	2	2
tRNA-Asp/COII	6	2	2	1	1	1
COII/tRNA-Lys	4	4	4	4	4	4
tRNA-Lys/ATPase8	1	1	1	1	1	1
ATPase8/ATPase6	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap
ATPase6/COIII	2	2	2	2	2	2
COIII/tRNA-Gly	1	1	1	1	1	1
tRNA-Gly/NADH3	-	-	-	-	-	-
NADH3/tRNA-Arg	4	7	7	5	7	7
tRNA-Arg/NADH4L	1	1	1	1	1	1
NADH4L/NADH4	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap
NADH4/tRNA-His	1	1	1	1	1	1
tRNA-His/tRNA-Ser (AGY)	-	-	-	-	-	-
tRNA-Ser (AGY)/tRNA-Leu (CUN)	1 overlap	-	1 overlap	1 overlap	1 overlap	1 overlap
tRNA-Leu (CUN)/NADH5	-	-	-	-	-	-
NADH5/Cyt*b*	8	18	13	17	16	22
Cyt*b*/tRNA-Thr	5	5	1	7	4	7
tRNA-Thr/tRNA-Pro *(L)*	**N/A**	**N/A**	31	14	7	6
tRNA-Thr/CR	-	-	**N/A**	**N/A**	**N/A**	**N/A**
CR/tRNA-Pro *(L)* +	-	-	**N/A**	**N/A**	**N/A**	**N/A**
tRNA-Pro *(L)*/NADH6 *(L)*	18	9	12	26	11	14
CR/NADH6 *(L)*	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
NADH6 *(L)*/tRNA-Glu *(L)*	3	3	4	3	3	3
tRNA-Glu *(L)*/CR #	-	-	-	-	-	-
CR/tRNA-Phe #	-	-	-	-	-	-
Total Length of Genome	18068	18674	16779	16791	16725	16701

White Stork	Oriental Stork	Loon	Penguin	Albatross	Petrel	Range (neognaths)	Range (all)
1779	2053	1984	2040	1459	>812	1066 - 2107	1031 - 2107
N/A	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**; 173 - 1455	**N/A**; 173 - 1455
-	-	-	-	-	-	0	0
-	-	-	-	-	-	0	0
-	-	-	-	-	-	0	0
-	-	-	-	-	-	0	0
17	17	15	5	4	6	2 - 19	2 - 19
1	1	1	1	1	1	1 - 24	1 - 24
8	8	10	9	9	7	4 - 13	4 - 14
1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap; 12	1 overlap; 12
-	-	-	8	-	-	0; 1, 8	0; 1, 8
2	2	1	1	1	1	1 - 5	1 - 5
1	1	1	1	1	1	1 - 10	0 - 10
2	2	4	2	2	16	1 - 16	0 - 16
2	2	2	2	2	2	0 - 13	0 - 13
1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
1	1	1	1	1	1	1; 6, 9	1; 6, 9
6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap
2	2	3	4	2	2	2 - 5	2 - 5
1	1	2	2	1	1	1 - 11	1 - 11
4	4	4	4	4	4	4; 14	1 - 14
1	1	1	1	1	1	1	1
7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap
2	2	2	2	2	2	2; 8 - 14	2; 8 - 14
1	1	1	1	1	1	1	1
-	-	-	-	-	-	0	0
7	7	7	7	6	7	2 - 7	2 - 7
1	1	1	1	1	1	0; 1	0 - 1
4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap
1	1	1	2	1	1	1; 2, 4, 13	1 - 13
-	-	-	-	-	-	0; 2	0; 2
1 overlap	1 overlap	1 overlap	1 overlap	-	1 overlap	1 overlap - 4	1 overlap - 4
-	-	-	-	-	-	0; 2	0; 2
13	13	14	10	13	18	2 - 22	2 - 22
4	4	5	6	7	6	1 - 8	1 - 8
11	11	16	9	21	14	0 - 31; **N/A**	0 - 85; **N/A**
N/A	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**; 0	**N/A**; 0
N/A	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**; 0	**N/A**; 0
13	13	9	16	13	30	8 - 30	8 - 30; N/A
N/A	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**; 3
3	3	4	3	3	3	0 - 4	0 - 4
-	-	-	-	-	-	0	0
-	-	-	-	-	-	0	0
17347	17622	17573	17611	17026	>16414	16616 - 18674	16591 - 18674

□able □ Length (in nucleotides) of avian mt tRNA and rRNA genes

tRNAs (22)	Emu	Cassowary	Kiwi	Greater Rhea 1	Greater Rhea 2	Lesser Rhea	Ostrich 1
F, Phe	70	70	72	68	68	68	69
V, Val	73	71	73	71	71	72	68
L, Leu (UUR)	74	74	74	74	74	74	74
I, Ile	71	72	71	71	71	71	71
Q, Gln *(L)*	71	71	71	71	71	71	71
M, Met	71	71	69	69	69	69	69
W, Trp	75	75	77	71	71	71	74
A, Ala *(L)*	69	69	69	69	69	69	69
N, Asn *(L)*	73	73	73	73	73	73	73
C, Cys *(L)*	67	67	67	67	67	67	67
Y, Tyr *(L)*	71	71	71	71	71	71	71
S, Ser (UCN) *(L)*	74	74	74	73	73	73	74
D, Asp	69	69	69	69	69	70	69
K, Lys	70	70	71	68	68	68	68
G, Gly	69	69	69	69	69	69	69
R, Arg	68	68	68	68	68	68	69
H, His	70	70	70	70	70	70	70
S, Ser (AGY)	66	66	66	66	66	66	66
L, Leu (CUN)	71	71	71	71	71	71	71
T, Thr	70	70	70	70	70	70	70
P, Pro *(L)*	70	70	70	70	70	70	70
E, Glu *(L)*	68	69	68	69	69	69	68
rRNAs (2)							
12S	964	967	969	963	963	964	966
16S	1596	1592	1595	1583	1583	1585	1579

□able □continued

Ostrich 2	Eastern Moa 1	Eastern Moa 2	Little Bush Moa	Giant Moa	Great Tinamou	Crested-Tinamou	Range (paleognaths)
69	69	69	71	69	66	66	66 - 72
68	74	74	74	75	71	70	68 - 75
74	74	74	74	74	74	74	74
71	71	71	71	71	71	72	71; 72
71	71	71	71	71	71	71	71
69	69	69	69	69	69	69	69; 71
74	77	77	77	76	72	72	71 - 77
69	69	69	69	69	69	69	69
73	73	73	73	73	73	73	73
67	67	67	67	67	67	67	67
71	72	72	72	72	71	71	71; 72
74	73	73	73	73	73	73	73 - 74
69	69	69	69	69	69	69	69; 70
68	70	70	70	70	70	70	68 - 71
69	69	69	69	69	69	69	69
69	68	68	68	68	70	69	68 - 70
70	70	70	70	70	70	70	70
66	66	66	66	66	66	66	66
71	73	73	73	73	71	72	71 - 73
70	70	70	70	70	67	70	67; 70
70	70	70	70	70	70	73	70; 73
68	67	67	67	67	68	70	67 - 70
966	969	971	969	968	989	972	963 - 989
1580	1600	1600	1602	1601	1588	1585	1579 - 1602

□able □continued

tRNAs (22)	Chicken 1	Chicken 2	Quail	Brush-turkey	Magpie Goose	Duck	Goose
F, Phe	70	70	68	73	68	70	68
V, Val	73	73	71	74	72	71	71
L, Leu (UUR)	74	74	74	74	75	74	74
I, Ile	72	72	71	73	72	72	73
Q, Gln *(L)*	71	71	71	71	71	71	71
M, Met	69	69	69	69	69	69	69
W, Trp	76	76	76	78	79	76	73
A, Ala *(L)*	69	69	69	69	69	69	69
N, Asn *(L)*	73	73	73	72	73	73	73
C, Cys *(L)*	66	66	66	67	67	66	66
Y, Tyr *(L)*	71	71	71	72	71	71	71
S, Ser (UCN) *(L)*	75	75	75	75	74	73	73
D, Asp	69	69	69	69	69	69	69
K, Lys	68	68	68	69	68	68	69
G, Gly	69	69	69	69	69	69	69
R, Arg	68	68	69	68	68	70	71
H, His	69	69	69	70	69	69	69
S, Ser (AGY)	67	67	71	69	66	66	67
L, Leu (CUN)	71	71	71	71	71	71	71
T, Thr	69	69	70	69	68	69	68
P, Pro *(L)*	70	70	70	70	71	70	69
E, Glu *(L)*	68	68	68	68	68	68	68
rRNAs (2)							
12S	975	975	973	980	973	982	988
16S	1621	1626	1615	1615	1594	1604	1609

□able □continued

Morepork	Kakapo	Rifleman	Broadbill	Flycatcher	Lyrebird	Indigobird	Rook
67	65	>19	70	69	69	70	70
71	72	71	71	70	71	70	70
74	74	74	74	74	75	75	75
71	71	73	70	73	73	72	74
71	71	71	71	71	71	71	71
69	69	69	69	69	69	69	69
75	70	72	72	69	70	70	70
69	69	69	69	69	69	69	69
74	74	73	74	73	73	73	74
70	67	67	67	67	68	67	67
72	71	71	72	71	71	71	71
74	76	74	74	74	75	72	73
69	69	69	69	69	69	69	69
72	69	68	69	71	70	69	70
70	70	69	69	69	69	69	69
69	68	68	68	69	69	70	70
70	69	69	70	70	70	70	69
67	65	66	67	66	66	66	66
71	71	71	71	71	71	71	71
69	69	70	68	70	69	69	69
70	69	71	71	69	69	70	70
73	70	73	71	70	72	70	72
975	968	975	976	981	973	978	975
1582	1590	1618	1601	1580	1602	1600	1601

□able □continued

tRNAs (22)	Falcon	Buzzard	Turkey Vulture	Oyster-catcher	Turnstone	Gull
F, Phe	73	70	69	74	71	71
V, Val	73	71	72	72	71	72
L, Leu (UUR)	74	74	74	74	74	74
I, Ile	73	72	72	72	72	72
Q, Gln *(L)*	71	71	71	71	71	71
M, Met	72	69	69	69	69	69
W, Trp	77	72	72	72	73	72
A, Ala *(L)*	70	69	69	69	69	69
N, Asn *(L)*	73	73	73	73	73	73
C, Cys *(L)*	67	67	67	67	67	67
Y, Tyr *(L)*	71	71	72	71	71	71
S, Ser (UCN) *(L)*	74	74	75	74	74	74
D, Asp	69	69	69	69	69	69
K, Lys	71	68	71	68	71	70
G, Gly	69	69	69	69	69	69
R, Arg	69	70	69	69	69	69
H, His	71	70	70	70	69	69
S, Ser (AGY)	66	66	67	66	67	66
L, Leu (CUN)	71	71	71	71	71	71
T, Thr	69	70	69	70	70	70
P, Pro *(L)*	69	70	70	70	70	70
E, Glu *(L)*	71	71	73	72	71	72
rRNAs (2)						
12S	979	972	973	963	970	964
16S	1599	1598	1606	1589	1595	1597

□able □continued

White Stork	Oriental Stork	Loon	Penguin	Albatross	Petrel	Range (neognaths)	Range (all)
70	70	71	69	69	70	65 - 74	65 - 74
71	71	71	71	73	73	70 - 74	68 - 75
74	74	74	74	75	75	74 - 75	74; 75
72	72	72	73	72	72	70 - 74	70 - 74
71	71	71	71	71	71	71	71
69	69	69	70	69	69	69; 70, 72	69; 70 - 72
76	76	70	69	70	70	69 - 79	69 - 79
69	69	69	69	69	69	69; 70	69; 70
73	73	73	73	73	73	72 - 74	72; 73; 74
67	67	67	67	67	67	66 - 70	66; 67; 68, 70
71	71	71	71	72	72	71 - 72	71 - 72
74	74	74	76	74	74	72 - 76	72 - 76
69	69	74	69	69	69	69; 74	69; 70, 74
72	72	71	70	71	70	68 - 72	68 - 72
69	69	70	69	69	69	69; 70	69; 70
69	69	68	69	69	69	68 - 71	68 - 71
70	70	70	70	70	70	69 - 71	69 - 71
66	66	66	66	66	66	65 - 71	65 - 71
71	71	71	71	71	71	71	71; 72, 73
68	68	70	70	70	70	68 - 70	67 - 70
70	70	70	70	70	70	69 - 71	69 - 73
72	72	72	71	69	73	68 - 73	67 - 73
968	968	979	975	974	974	963 - 988	963 - 989
1608	1612	1609	1608	1599	1608	1580 - 1626	1579 - 1626