Genomic Support for a Moa–Tinamou Clade and Adaptive Morphological Convergence in Flightless Ratites

Allan J. Baker,*,[1,2] Oliver Haddrath,[1] John D. McPherson,[3] and Alison Cloutier[2]

[1]Department of Natural History, Royal Ontario Museum, Toronto, Ontario, Canada
[2]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada
[3]Ontario Institute for Cancer Research, Toronto, Ontario, Canada
*Corresponding author: E-mail: allanb@rom.on.ca.

Associate editor: Nicolas Vidal

Abstract

One of the most startling discoveries in avian molecular phylogenetics is that the volant tinamous are embedded in the flightless ratites, but this topology remains controversial because recent morphological phylogenies place tinamous as the closest relative of a monophyletic ratite clade. Here, we integrate new phylogenomic sequences from 1,448 nuclear DNA loci totaling almost 1 million bp from the extinct little bush moa, Chilean tinamou, and emu with available sequences from ostrich, elegant crested tinamou, four neognaths, and the green anole. Phylogenetic analysis using standard homogeneous models and heterogeneous models robust to common topological artifacts recovered compelling support for ratite paraphyly with the little bush moa closest to tinamous within ratites. Ratite paraphyly was further corroborated by eight independent CR1 retroposon insertions. Analysis of morphological characters reinterpreted on a 27-gene paleognath topology indicates that many characters are convergent in the ratites, probably as the result of adaptation to a cursorial life style.

Key words: phylogenomics, ratite paraphyly, CR1 retroposon insertions, little bush moa, ancient DNA, morphological convergence.

Introduction

Among the most contentious issues in resolving the avian Tree of Life is the phylogenetic relationships within the Paleognathae, an ancient clade that diverged from the remaining crown-group Neognathae at the base of the avian tree (Cracraft et al. 2004; Hackett et al. 2008; Mayr 2011). Paleognaths share a paleognathous bony palate, open ilioschiatic fenestra, and rhampothecal grooves (Cracraft 1974), and include the volant tinamous of Central and South America and the large flightless ratites, which are distributed across the southern landmasses except Antarctica. Ratites lack a keeled sternum, and comprise the extant African ostrich, South American rheas, and the Australasian emu, cassowaries and kiwi, in addition to the recently extinct Madagascan elephant birds and New Zealand moa. Within this clade of charismatic species, perhaps no group has captured the imagination more than the wingless moa (Aves: Dinornithiformes) of New Zealand following the first description of a moa femur by the renowned anatomist Sir Richard Owen (Owen 1839). The relatively recent extinction of the nine species of moa following the Polynesian colonization of New Zealand in 13th century, along with a wealth of discovered moa remains, has enabled studies not only of their morphology and paleontology, but also the use of ancient DNA (aDNA) to elucidate aspects of moa evolutionary biology and life history (Allentoft and Rawlence 2012).

Morphological and molecular systematists currently agree that paleognaths constitute a monophyletic clade, but phylogenetic relationships among paleognaths are disputed (fig. 1). Most recent morphological studies support ratite monophyly, with tinamous as the closest relative of the ratites (Livezey and Zusi 2007; Bourdon et al. 2009; Worthy and Scofield 2012). Conversely, analysis of characters of the tongue apparatus and cranial osteology supports ratite paraphyly (but not a moa–tinamou clade), and suggests that these characters might better retain phylogenetic signal, whereas postcranial skeletal characters might exhibit homoplasy due to convergence (Bock and Bühler 1990; Elzanowski 1995; Johnston 2011). Most earlier molecular studies of mitochondrial DNA also supported ratite monophyly (e.g., Haddrath and Baker 2001), but there were exceptions (e.g., Braun and Kimball 2002). However, studies using nuclear DNA, but lacking sequences from moa, recovered a paraphyletic clade that nested tinamous within the ratites and placed the ostrich as the closest relative of all other paleognaths (Hackett et al. 2008; Harshman et al. 2008; [N.B. We use the term "paraphyly" to refer to a nonmonophyletic ratite clade that includes some, but not all, descendants of the most recent common ancestor of ratites and tinamous. Harshman et al. (2008) instead characterized the ratites as "polyphyletic" because it is inferred that their uniting characters (e.g., a keel-less sternum) arose via convergence]). This novel topology was supported by a reanalysis of mitochondrial genomes that accounted for a base compositional bias at third codon positions (Phillips et al. 2010), and recovered a highly supported moa–tinamou clade (fig. 1). Most recently, separate analyses of 27 nuclear loci including moa sequences (Haddrath and Baker 2012), and 40 nuclear loci that lacked
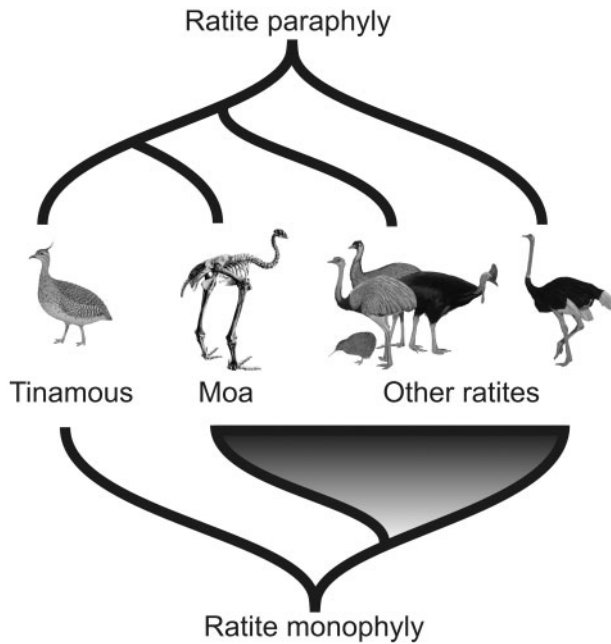
FIG. 1. Competing hypotheses of phylogenetic relationships in the Paleognathae. Ratite monophyly posits evolution from a flightless common ancestor. The ratite paraphyly hypothesis that places the volant tinamous as the closest relative of moa (Phillips et al. 2010; Haddrath and Baker 2012; Smith et al. 2013) instead posits either that ratites and tinamous evolved from a volant common ancestor and that flight has been lost multiple times in the ratites or that tinamous have regained flight. Bird drawings are modified from del Hoyo et al. (1992) with permission from Lynx Edicions, Barcelona, Spain.

moa sequences (Smith et al. 2013), both recovered tinamous nested within the ratites. Resolving paleognath relationships is essential to investigate the discord between morphological and molecular topologies, and for interpretations of phenotypic evolution and biogeography within this group. Ratite monophyly is consistent with evolution from a flightless common ancestor and a predominantly vicariant mode of speciation following the breakup of Gondwana, whereas ratite paraphyly requires that either tinamous regained flight from a flightless paleognath ancestor or, perhaps more likely, that multiple independent losses of flight have occurred across the ratites (Bock and Bühler 1990; Harshman et al. 2008; Phillips et al. 2010; Haddrath and Baker 2012; Smith et al. 2013).

High-throughput sequencing (HTS) can produce large-scale phylogenomic data sets that reflect genome-wide phylogenetic signal, and can also maximize the amount of nuclear sequence recovered from endogenous aDNA, which is typically degraded into short fragments (Millar et al. 2008; Rizzi et al. 2012; Huynen et al. 2012). A major advantage to using DNA sequences to test competing phylogenetic hypotheses is the large number of phylogenetically informative characters they provide, which are key to estimating species trees under concatenation or jointly with gene trees. Discordance between gene histories and species histories can arise when incomplete lineage sorting (ILS) occurs across short internal branches, even in deep time (Edwards et al. 2005; Degnan and Rosenberg 2009; Oliver 2013). Simulations using the multi-species coalescent suggest that phylogenomic discordance

can potentially mislead estimation of evolutionary relationships of taxa under concatenation because stochastic population processes are not taken into account (Oliver 2013). Conversely, concatenation is expected to yield a good estimate of species histories when ILS is rare or absent, as the most common gene tree topology should reflect the true history of species. Short internal branches associated with the diversification of neoavian orders deep in the Avian Tree of Life suggest that phylogenomic discordance arising through ILS could underpin the substantial difficulties encountered in reconstructing some higher-level avian relationships (Poe and Chubb 2004; Edwards et al. 2005; Hackett et al. 2008; Matzke et al. 2012; Oliver 2013). ILS could prove similarly problematic within the Paleognathae, where the order of deep branches is very difficult to resolve (Haddrath and Baker 2012). Therefore, a much larger number of gene histories needs to be sampled in the aDNA genome of an extinct moa and their putative closest relatives, the tinamous, to more stringently test whether tinamous are truly embedded within the ratites or whether this relationship is an artifact resulting from ILS. However, with large phylogenomic data sets, it is also imperative to ensure that systematic error from heterogeneity in base composition and exchange rates, as well as mutational saturation and long-branch attraction, do not overwhelm phylogenetic signal to produce incorrect, yet highly supported, topologies (Jeffroy et al. 2006; Philippe et al. 2011; Telford and Copley 2011).

Here, we combine high-throughput sequences from aDNA of the extinct little bush moa (*Anomalopteryx didiformis*), and DNA of the extant Chilean tinamou (*Nothoprocta perdicaria*) and emu (*Dromaius novaehollandiae*), with available phylogenomic sequences from two additional paleognaths as well as four representative neognath taxa and the green anole to serve as outgroups. We assembled the largest phylogenomic data set to date comprised of 854 ultraconserved elements (UCEs) and 594 single-copy protein-coding genes to more thoroughly test competing hypotheses of ratite paraphyly versus monophyly. Phylogenies constructed with Bayesian concatenation and maximum pseudolikelihood species tree approaches were checked by fitting heterogeneous models of evolution which are robust to most systematic biases (Lartillot and Philippe 2004; Foster 2004; Foster et al. 2009). Independent information was obtained from retroposon insertions to corroborate inferences drawn from sequence-based analyses. To investigate why molecular and morphological trees are at such odds in supporting different hypotheses of paleognath relationships, we reinterpreted morphological characters on the molecular topology to determine whether concerted convergence in morphology might have occurred.

Results

Support for Ratite Paraphyly from UCEs and Protein-Coding Genes

The total data set comprises 1,133,474 bp of aligned nuclear DNA. UCEs total 337,982 bp of sequence from 854 loci (supplementary data S1, Supplementary Material online), and
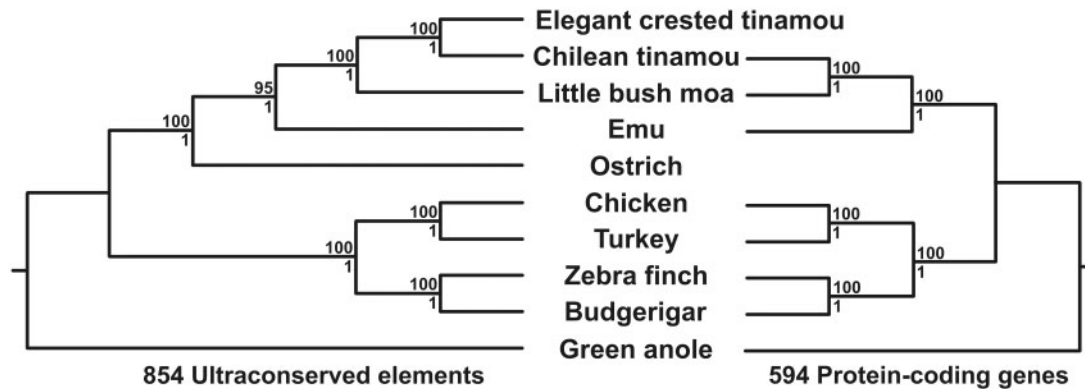
FIG. 2. Phylogenetic analyses of 854 UCE and 594 protein-coding gene data sets in MrBayes and MP-EST. Bayesian posterior probabilities are shown (below nodes) along with the MP-EST bootstrap percentages based on 1,000 replicates (above nodes).

protein-coding genes total 795,492 bp of sequence from 594 loci (supplementary data S2, Supplementary Material online). Alignment data are available from the Dryad Digital Repository: doi:10.5061/dryad.9g3cm. Notably, an HTS approach recovered sequence for 841 of 854 UCEs and 593 of 594 protein-coding genes from aDNA of the little bush moa. Moa sequence totaled 901,554 bp across all loci, or 79.5% of sites across total alignment lengths, with a mean depth of coverage of 10.85 ± 0.0008 SE and mean quality score of 36.46 ± 0.0024 SE.

Phylogenetic analyses of both UCEs and protein-coding genes under either concatenation in MrBayes or with the gene tree/species tree approach implemented in MP-EST recovered identical topologies, with the exception that only UCE sequences were available for the ostrich and elegant crested tinamou (fig. 2). In addition to the expected relationships for the four neognath representatives, this topology supports ratite paraphyly by embedding the tinamous within the ratites in a highly supported clade with the little bush moa. Successive nodes to this embedded moa + tinamou clade are formed with the emu, followed by the ostrich. Maximum support is observed at all nodes except that separating emu from (moa + tinamou) in the maximum pseudolikelihood species tree for UCEs, where multiscale bootstrap support falls to 95%.

Graphical exploration of observed transitions and transversions plotted against corrected genetic distance shows linear trends for UCEs and for first and second codon positions of protein-coding genes, indicating that they are not saturated with superimposed substitutions (supplementary fig. S1, Supplementary Material online). This conclusion is supported by Xia et al.'s (2003) test, where the entropy-based indices of substitution saturation (Iss) estimated for UCEs and first and second codon positions are all significantly smaller than their respective estimated critical values regardless of the symmetry of the tree (results not shown). Plots of transitions and transversions for third codon positions exhibit a slight deviation from linearity in comparisons between the ingroup and outgroup taxa suggesting that multiple substitutions have occurred between more distantly related taxa at these sites (supplementary fig. S1, Supplementary Material online).

However, the Iss estimate for third codon positions is 0.3750, which is significantly less than the critical value of 0.8577 for a symmetrical tree ($t_{[df = 129,112]} = 442.6$, $P_{[two-tailed]} < 0.00001$), and the critical value of 0.7837 for an extreme asymmetrical tree ($t_{[df = 129,112]} = 334.5$, $P_{[two-tailed]} < 0.00001$) and indicates that these sites have not experienced substitutional saturation to a degree that would obscure the underlying phylogenetic signal.

Compositional χ^2 tests using posterior predictive simulations showed statistically significant compositional nonstationarity for both UCEs and protein-coding genes (supplementary table S4, Supplementary Material online; both $P < 0.000001$). However, observed differences in mean GC content among species were associated with small to moderate effect sizes (supplementary tables S5–S7, Supplementary Material online; maximum effect size [Cohen's d] of 0.21 for UCEs and 0.43 for protein-coding genes). Nevertheless, phylogenies were built with heterogeneous models in P4 (Foster et al. 2009) and PhyloBayes (Lartillot and Philippe 2004; Lartillot et al. 2013) to confirm that the topologies recovered from homogeneous models in MrBayes and MP-EST did not simply reflect systematic biases in compositional or exchange rate nonstationarity. For UCEs, the best P4 heterogeneous model incorporated seven rate matrices and six compositional matrices (supplementary data S3, Supplementary Material online; model 7GTR_6C), with Bayes factor (BF) comparisons providing evidence for a strong improvement in model fit relative to the homogeneous 1GTR_1C model (2ln BF = 1,002.5). The CAT–GTR model in PhyloBayes further improved model fit, with 2ln BF = 21,355.1 relative to the best 7GTR_6C model in P4. PhyloBayes chains failed to stabilize for amino acid translations of protein-coding genes; thus, results are shown for P4 only. The best P4 model, with a single JTT rate matrix and six compositional matrices (1JTT_6C), shows strong evidence for improved model fit relative to the homogeneous 1JTT_1C model (2ln BF = 2,038.5).

There was no longer a significant departure (at $\alpha = 0.05$) from the null distribution of χ^2 values generated from posterior predictive simulations with the addition of at least two compositional matrices for UCEs and three compositional
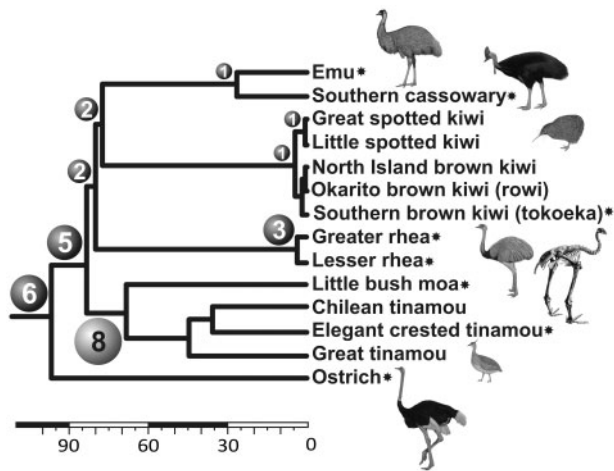
FIG. 3. Eight retroelement insertions shared by little bush moa and four tinamou species identified in the current study (light gray sphere with black numbering), mapped onto a published paleognath topology constructed from 27 nuclear genes and 21 retroelements (Haddrath and Baker 2012; published CR1 insertions are indicated by dark gray spheres with white numbering). Scale bar indicates time in millions of years. Asterisks mark species included in comparisons to morphological data. Bird drawings are modified from del Hoyo et al. (1992) with permission from Lynx Edicions, Barcelona, Spain.

matrices for protein-coding genes. All topologies built with heterogeneous models for both UCEs and protein-coding genes were identical to those constructed with homogeneous models. Maximal node support occurred throughout, except at the node separating the zebra finch and budgie in the PhyloBayes tree for UCEs where posterior probability (PP) was 0.77 (supplementary data S3, Supplementary Material online).

## Corroboration of a Moa–Tinamou Clade from CR1 Retroposon Insertions

In silico bioinformatic approaches identified 17 candidate retroposon insertions that were shared by the Chilean tinamou and little bush moa, but were absent from the emu. Experimental validation from polymerase chain reaction (PCR) amplification and sequencing across a wider sampling of paleognaths provided unequivocal evidence for eight CR1 insertions that support a moa–tinamou node (fig. 3). All insertions were shared by additional tinamou species, but were absent from all non-moa ratites (supplementary table S3, Supplementary Material online; alignments are deposited in the Dryad Digital Repository: doi:10.5061/dryad.9g3cm). These eight loci had orthologous flanking sequences, with CR1 insertions of the same size, subfamily, orientation, and degree of truncation, in addition to recognizable target site duplications (TSDs), indicating they are homologous insertion events. The remaining nine insertion events could not be unambiguously interpreted due to missing sequences, deletions, and/or unrecognizable TSDs; however, no information contradicting ratite paraphyly was observed (supplementary table S3, Supplementary Material online).

Searches to identify CR1 insertions supporting the two alternative hypotheses of relationships among moa, emu, and Chilean tinamou identified no insertions that were shared by the emu and tinamou, but were absent from moa. However, we identified one insertion (locus 7292) that was shared by emu and moa, but absent from the Chilean tinamou. Experimental validation confirmed the presence of this CR1 insertion in all investigated ratites, but its absence in the great tinamou and the Chilean tinamou (supplementary table S3, Supplementary Material online; Dryad Digital Repository: doi:10.5061/dryad.9g3cm). The probability that the observed pattern of CR1 insertions (e.g., [8 1 0]) arose by chance is $P = 0.0005$ (Waddell et al. 2001), providing strong support for the moa–tinamou clade.

Interpreting Morphological Characters on the Molecular Tree

The morphological characters described by Worthy and Scofield (2012) were clustered into cliques of mutually compatible characters relative to a molecular topology constructed from a 27-gene data set for eight paleognaths and an outgroup taxon (fig. 4A, part d; also see supplementary fig. S2, Supplementary Material online, for the 27-gene topology, and supplementary data S4, Supplementary Material online, for additional information regarding morphological characters). The largest clique (clique 1) groups 35 characters, or 27% of all morphological characters included in the analysis, and encompasses three predominant phylogenetic signals. In addition to 14 of 18 characters that support uncontested shallower nodes that are common to the morphological and molecular topologies (i.e., grouping emu and cassowary, or greater and lesser rhea; shown as green boxes in fig. 4A, part a), clique 1 encompasses all characters that support an inferred homoplasious grouping of moa with kiwi ($n = 7$, shown as gold links in fig. 4A, part f), and 11 of 14 characters where the tinamou and outgroup share a character state that differs from the ratites, in direct refutation of ratite paraphyly (drawn as blue links). Nodes that are unique to the molecular tree are supported by eight morphological characters (red boxes in fig. 4A, part b). Five of these characters unite moa and tinamou; of these, four (characters 26, 39, 45, and 114) co-occur in clique 4, whereas the fifth (character 96) is positioned within the closest relative of this clique. Inspection of character states across all 26 species scored for morphology by Worthy and Scofield (2012) revealed that two of these characters (96 and 114) would also group the elephant birds with moa and tinamou. Although this result is not inconsistent with the molecular topology, we cannot evaluate the potential for homoplasious signal within this larger clade.

Each of the seven largest cliques contains characters from across body compartments (fig. 4B). There is limited support that cranial characters are better able to recover nodes that are unique to the molecular topology. The three morphological characters that unambiguously support a moa–tinamou node are cranial characters, versus only 2 of 14 characters that directly refute paraphyly (Fisher's exact test, $P_{(two\text{-}tailed)} = 0.01$). Cliques 2 and 4 combined, with characters that support
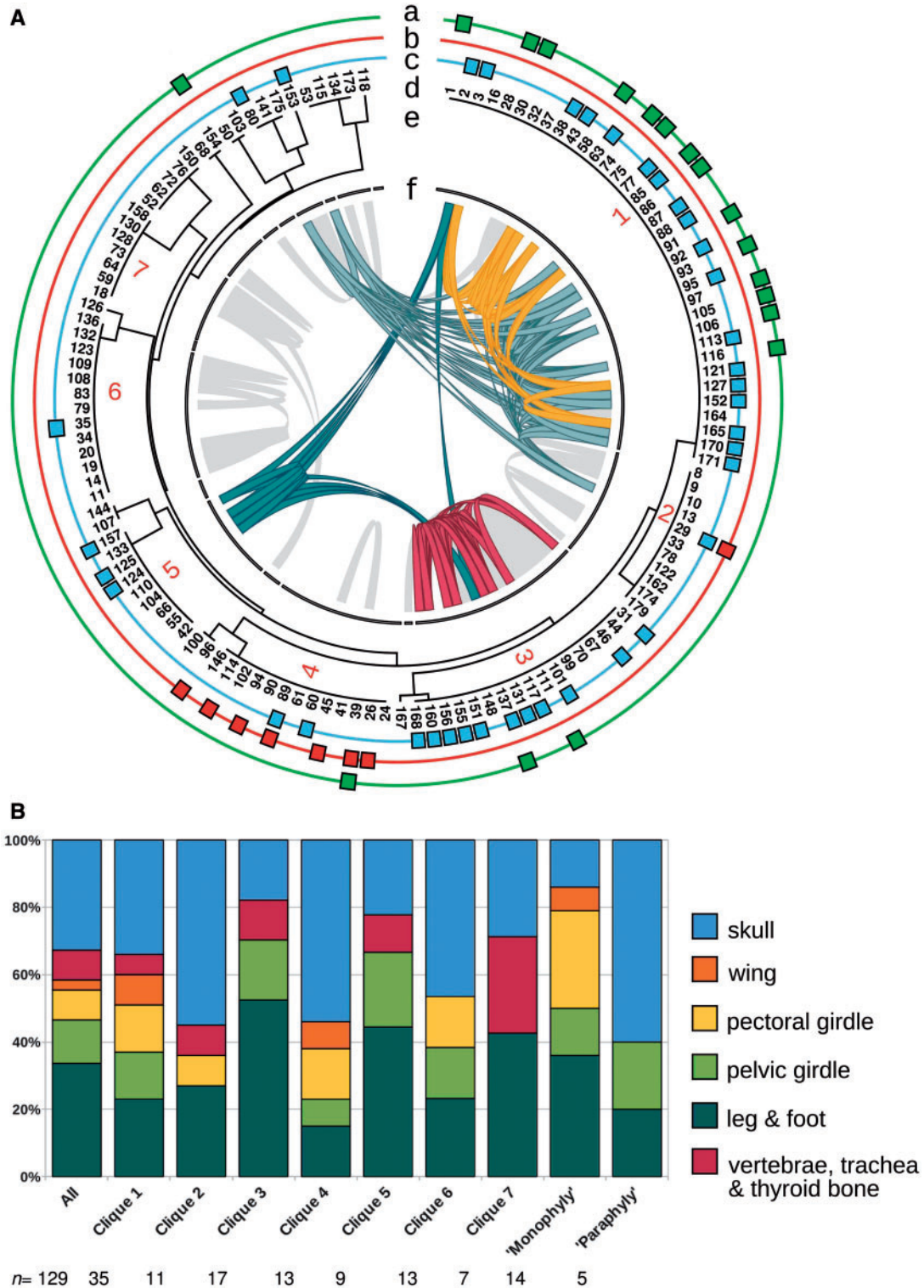
FIG. 4. Analysis of morphological characters. (A) Cliques of compatible morphological characters from UPGMA clustering of pairwise dissimilarity scores (d), with numbers assigned to cliques with n > 5 constituent members (e). Outer orbitals show characters that provide unambiguous support for nodes shared by the morphological and molecular phylogenies [(emu, cassowary); (rheas)] (a), for nodes that are unique to the molecular phylogeny [(moa, tinamou); (emu, cassowary, kiwi); (emu, cassowary, kiwi, rheas)] (b), or for nodes unique to the morphological phylogeny [(ratite monophyly); (moa, kiwi); (ostrich, rhea); (emu, cassowary, ostrich, rhea)] (c). Inner circle (f) depicts inferred homoplasies from plotting each morphological character on the 27-gene molecular topology. Outer ideograms in part (f) echo the position of each clique, and inner links join characters with the same inferred homoplasious signal (drawn in Circos [Krzywinski et al. 2009]). Characters grouping tinamou with the outgroup are drawn in blue, those placing kiwi with moa are in gold, ostrich with rheas are in teal, and (emu, cassowary, ostrich, rhea)/(kiwi, moa, tinamou, outgroup) are shown in red. Other inferred homoplasious groupings with at least two constituent members are drawn in light gray. (B) Proportion of characters from each anatomical category in the full morphological data set, in the seven largest cliques, and in support of ratite monophyly (grouping tinamou + outgroup) versus paraphyly (grouping moa + tinamou).

nodes found only in the molecular tree, are also more heavily weighted toward cranial characters relative to the remaining five largest cliques, which contain no support for nodes found only on the molecular tree (fig. 4B; cliques 2 + 4: 13 skull vs. 11 nonskull characters; cliques 1 + 3 + 5 + 6 + 7: 25 skull vs. 56 nonskull characters; $\chi^2_{(df = 1)} = 4.35$, $P = 0.037$). However, only 8 of the 24 characters (33%) that support any node in the molecular tree (including uncontested nodes shared with the morphological topology, green + red boxes shown in fig. 4A, parts a and b, respectively) are cranial characters, which is almost identical to the proportion of cranial characters in the total data set (42 of 129 characters, or 32.6%). Furthermore, the proportion of characters that are free from any inferred homoplasies does not differ for cranial and postcranial groupings (7 of 42 cranial characters vs. 13 of 87 postcranial; $\chi^2_{(df = 1)} = 0.007$, $P = 0.917$).

Discussion

Using the first large phylogenomic nuclear data set for paleognaths to include aDNA sequences from an extinct moa from New Zealand, we recover compelling support for the ratite paraphyly hypothesis originally proposed by Bock and Bühler (1990), and supported by molecular studies (Harshman et al. 2008; Phillips et al. 2010; Haddrath and Baker 2012; Smith et al. 2013). Congruent topologies that place the little bush moa as the closest relative of tinamous, nested within the ratites, are obtained from large, independent data sets of two different types of nuclear loci (UCEs and protein-coding genes), and are supported by both Bayesian concatenation and maximum pseudolikelihood species tree analytical approaches. Systematic biases that can affect tree topology, such as compositional and exchange rate heterogeneity or long-branch attraction, can likely be ruled out. Phylogenetic analyses with heterogeneous models of evolution (Lartillot and Philippe 2004; Foster 2004; Foster et al. 2009) more robust to these artifacts recovered the same tree topology with strong support. Long-branch attraction could potentially occur if sequences are saturated by superimposed parallel substitutions; however, we demonstrated that sufficient phylogenetic signal is retained in the analyzed data sets. Furthermore, translation of protein-coding genes and use of the site-heterogeneous CAT mixture model for UCEs, which is believed to be robust to long-branch attraction (Lartillot et al. 2007), should also safeguard against erroneous inference caused by mutational saturation.

Independent corroboration from eight retroposon insertions also resolutely supports moa as the closest relative of the tinamous, and five previously published insertions (Haddrath and Baker 2012) further support ratite paraphyly. These results agree with previous phylogenetic analyses of smaller data sets that included ancient mtDNA or nuclear DNA from moa that also recovered a moa–tinamou clade (Phillips et al. 2010; Haddrath and Baker 2012), and with nuclear gene phylogenies (Harshman et al. 2008; Smith et al. 2013) that lacked moa sequence but nevertheless recovered tinamous embedded in the ratites. We therefore reject the traditional hypothesis of ratite monophyly in which tinamous are placed as the closest relative of the ratites.

In addition to the much more powerful test of the ratite paraphyly hypothesis afforded by the large number of gene histories recovered from HTS of the little bush moa, emu, and Chilean tinamou, we gained strong genomic evidence against a role for ILS as an explanation for the observed relationship between moa and tinamous. ILS would result in gene tree heterogeneity and should also produce conflicting patterns in observed retroposon insertions (Edwards et al. 2005; Matzke et al. 2012). Although the single conflicting retroposon insertion found in all the ratite taxa we sampled but absent in two tinamou species could represent an instance of ILS, it is probably instead a secondary loss that occurred by near-perfect deletion of the CR1 element in the tinamou common ancestor (refer to Dryad Digital Repository: doi:10.5061/dryad.9g3cm retroelement locus 7292 for a multiple sequence alignment showing the boundaries of tinamou flanking sequence relative to the retroelement insertion observed in ratites). Although transposable element insertions are viewed as virtually homoplasy free, rare instances of parallel insertions and precise deletions are documented (e.g., Ray et al. 2006; Han et al. 2011) and underscore the utility of genome-level approaches to more rigorously evaluate support for alternative phylogenetic hypotheses. We caution that it was not possible to search for additional retroposon insertions that might occur in other ratite genomes not sampled here, and which could be affected by ILS across very short internal branches separating the major lineages of non-ostrich paleognaths (e.g., see fig. 3 and Haddrath and Baker 2012). However, the branch connecting the moa–tinamou clade to other ratites is reasonably long. Thus, the likelihood of ILS is expected to be very low at this node unless ancestral population size was very large because there would have been approximately 10–15 Myr (Haddrath and Baker 2012) for genetic drift to fix alternative alleles in either species. The next step to resolve phylogenetic relationships of major lineages across the paleognath tree requires the construction of similarly large phylogenomic data sets for other taxa not sampled here. Heterogeneity in gene tree topologies is expected to pose significant challenges in resolving some deep divergences in the avian tree (Poe and Chubb 2004; Edwards et al. 2005; Haddrath and Baker 2012), and a large number of gene histories and rare genomic events might therefore be required to reconstruct phylogenetic relationships among major paleognath lineages.

The robust support for ratite paraphyly from phylogenomic data contrasts with several recent morphological topologies that place tinamous as the closest relative of a monophyletic ratite clade (Livezey and Zusi 2007; Bourdon et al. 2009; Worthy and Scofield 2012), and suggests that multiple independent losses of flight have likely occurred within the ratites (Bock and Bühler 1990; Harshman et al. 2008; Phillips et al. 2010; Smith et al. 2013). Flightlessness, which is documented across 15 avian orders, is commonly associated with extensive modifications to the skeleto-muscular system, with paedomorphic reduction of the pectoral apparatus and peramorphosis leading to an enlarged or more robust pelvic girdle and hind limbs (Livezey 1995; Cubo and Arthur 2001). Our reinterpretation of morphological data

identified several characters that are consistent with the molecular topology, especially at shallower nodes that are shared with the morphological tree. However, we also identified many inferred homoplasies that are likely to have arisen through convergence associated with independent losses of flight and adaptation to a cursorial lifestyle within ratites. We note that a full interpretation of signal present in the morphological data set was not possible due to the large number of extinct forms surveyed (11 of 23 ingroup taxa). However, comparisons to the 27-gene molecular topology did include at least one representative from all extant paleognath lineages, in addition to the extinct little bush moa.

Beyond inferring homoplasies at individual characters, the morphological data set is clustered into cliques of mutually compatible characters that could reflect concerted convergence across suites of characters in response to shared selective pressures or developmental processes (Holland et al. 2010; Blanke et al. 2013). Similar to a previous analysis in cormorants and shags (Holland et al. 2010), cliques encompass characters from across body compartments including the skull, and direct functional links are not necessarily apparent among clique members. Cranial and tongue characters, which might be less affected by convergence due to the modularity of cranial evolution relative to postcranial structures (Bhullar et al. 2012), have previously recovered ratite paraphyly, although not a moa–tinamou clade (Bock and Bühler 1990; Johnston 2011). However, Cubo and Arthur (2001) found significant associations between peramorphosis of the pelvic apparatus and of the skull among lineages of flightless birds. The loss of flight might lessen selective pressures for a lightweight skull and associated musculature, and a concomitant increase in muscle insertion areas (Gussekloo and Cubo 2013). We found limited support that cranial characters can better recover nodes consistent with the molecular topology, but did not find that cranial characters in general are less affected by inferred homoplasies. These observations suggest that the use of skull characters alone might not offer a panacea to resolve the ongoing disconnect between paleognath phylogenies estimated from molecular and morphological data. However, investigating the underlying basis for shared phylogenetic signals within character cliques could lead to a better understanding of the functional and ontogenetic interrelationships of their constituent members.

Materials and Methods

Sample Information

Source information for all specimens associated with new sequence data is provided in supplementary table S1, Supplementary Material online.

HTS of Paleognaths

Phylogenomic sequence data were retrieved from draft genome assemblies built from 454 and Illumina paired-end reads for the emu (D. novaehollandiae), and Illumina paired-end reads for the Chilean tinamou (N. perdicaria; Baker AJ, unpublished data). Additionally, unassembled 101 bp paired-end reads were obtained from two lanes of an Illumina HiSeq 2500 platform for the extinct little bush moa (A. didiformis).

Ultraconserved Genomic Elements (UCEs)

Sequences for 854 previously published avian UCEs were obtained from the Dryad Digital Repository doi:10.5061/dryad.64dv0tg1 (Faircloth et al. 2012) for chicken (Gallus gallus), emu, ostrich (Struthio camelus), elegant crested tinamou (Eudromia elegans), and green anole (Anolis carolinensis [reptilian outgroup]; supplementary data S1, Supplementary Material online). Published emu sequences were compared with the draft genome assembly and degenerate bases were introduced at variable sites. BLAST (Altschul et al. 1997) searches were used to retrieve each UCE from the publicly available genome sequences for three additional species: turkey (Meleagris gallopavo, using chicken query sequences), zebra finch (Taeniopygia guttata), and budgerigar (Melopsittacus undulatus, both queried with the banded pitta, Pitta guajana [Faircloth et al. 2012]), as well as from the draft Chilean tinamou assembly (queried with elegant crested tinamou). Moa UCEs were assembled by mapping reads to emu reference sequences with Stampy v. 1.0.17 (Lunter and Goodson 2011), and moa–emu alignments were visualized with Integrative Genomics Viewer v. 2.2 (Robinson et al. 2011). Moa consensus sequences were constructed using SAMtools v. 0.1.18 (Li et al. 2009) and custom Perl scripts, and required a minimum coverage depth of 2 and Phred quality score $Q \geq 13$ at each position.

Protein-Coding Genes

Ensembl BioMart (Kinsella et al. 2011) was used to identify one-to-one orthologs with conserved exon–intron structure in the chicken, turkey, zebra finch, and green anole for which there also existed homology-based annotations in the draft emu and Chilean tinamou genomes ($n = 594$; supplementary data S2, Supplementary Material online). Computationally derived gene models for emu and tinamou were curated manually in Apollo Genome Annotation Curation Tool v. 1.11.6 (Lewis et al. 2002), and sequences were retrieved as cDNAs. Budgerigar sequences were identified with BLASTn searches of zebra finch queries against the set of predicted budgie cDNAs available in Ensembl Pre! (ftp://ftp.ensembl.org/pub/pre/fasta/cdna/melopsittacus_undulatus/, last accessed April 24, 2014). Moa consensus sequences were generated from alignment to emu references as described above.

Assessment of Mutational Saturation

Mutational saturation was assessed by graphical exploration of patterns of transitions and transversions as well as using the Xia et al. (2003) information entropy-based index of substitution saturation (Iss). Both approaches were performed in DAMBE v. 5.2.57 (Xia 2013). Observed transitional and transversional substitutions were plotted against corrected genetic distances estimated using a GTR model for the UCE data set and for each of the three codon positions in the protein-coding gene data set. The extent of substitutional saturation for each data partition was then formally tested by comparing

the Iss to Iss.c, the critical value at which saturation is sufficient to cause sequences to begin to fail to recover the true topology (Xia et al. 2003).

Isolation of CR1 Retroelements

Two in silico approaches were employed to isolate shared CR1 retroelement insertions that support a moa–tinamou clade observed in phylogenies reconstructed from UCEs and protein-coding genes. First, CR1 repeats were identified in emu and Chilean tinamou scaffolds with RepeatMasker v. 3.3.0 (Smit et al. 2010), and sequences were retrieved for each CR1 plus up to 500 bp of nonrepetitive flanking DNA in both directions. Tinamou CR1-containing regions were used as queries in BLASTn searches of the emu genome, and candidates were retained when both tinamou flanking sequences could be aligned with emu up to the CR1 boundary, but where emu lacked the intervening repeat. BLAST searches against the data set of moa sequencing reads flagged potential candidates when a moa read overlapped the CR1-flank boundary or when one member of a moa read pair fell within the flanking sequence and the second aligned within the CR1 repeat (in both cases, supporting a shared CR1 in tinamou and moa that is absent from the emu).

Second, we inverted the search strategy by repeat-masking the raw moa reads and searching for read pairs where one partner was annotated as a CR1 repeat and its partner was nonrepetitive. The nonrepetitive partner was used to query the tinamou draft genome, after which the CR1-containing read was used in a bl2seq search against the tinamou hit sequence extended by up to 2 kb in the expected direction based on moa read pair orientation. Candidate tinamou sequences that matched both members of a moa read pair were used to query the emu genome, and were retained when the corresponding genomic region in emu lacked a CR1 insertion.

Similar bioinformatic methods were used to search for CR1 insertions supporting the two alternative hypotheses (i.e., grouping emu and moa to the exclusion of tinamou, or emu and tinamou to the exclusion of moa). For the former hypothesis, candidate CR1-containing regions in emu that lacked an insert in tinamou were flagged if a moa read overlapped the CR1-flank boundary or if members of a moa read pair aligned within the flank and the CR1 repeat. For the latter hypothesis, candidate regions containing CR1 insertions in both the emu and tinamou were flagged if a single moa read was split across the insertion site and aligned to the flanking sequence on either side of the repeat (i.e., indicating that moa lacks a CR1 insertion that is shared by emu and tinamou).

Primers were designed in FastPCR v. 5.1.68 beta1 (Kalendar et al. 2009) within regions flanking each candidate CR1 insertion using consensus sequences from emu, tinamou, and moa. In vitro screening for the presence or absence of homologous insertions was performed for ostrich, Southern cassowary, lesser rhea, all five kiwi species, and three additional tinamou species (supplementary table S1, Supplementary Material online). PCR amplifications followed a touchdown protocol (Haddrath and Baker 2012), and all amplified products were recovered (Dean and Greenwald 1995), sequenced with BigDye v. 3.1 chemistry following standard protocols, and resolved on an ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA). Queries against the RepeatMasker database identified which amplified products contained CR1 insertions.

Phylogenomic Analysis

Sequences for phylogeny reconstruction were aligned with Muscle v. 3.8.31 (Edgar 2004) and inspected in BioEdit v. 7.1.1 (Hall 1999). Best-fitting models of nucleotide substitution for UCEs were chosen according to the corrected Akaike information criterion (AICc) in jModelTest v. 2.1.1 (Posada 2008), and single-locus maximum-likelihood trees were built in PhyML v. 20120412 (Guindon and Gascuel 2003) from 1,000 bootstrap replicates. For protein-coding genes, the best partitioning scheme was determined with PartitionFinder v. 1.0.1 (Lanfear et al. 2012). Gene trees were estimated from 1,000 bootstrap replicates in RaxML v. 7.2.8 (Stamatakis et al. 2005). For both UCEs and protein-coding genes, species trees were estimated from the set of individual gene trees with MP-EST v. 1.2 (Liu et al. 2010).

Analyses were also conducted on separate concatenated data sets for UCEs and protein-coding genes using Bayesian inference as implemented in MrBayes v. 3.1.2 (Huelsenbeck and Ronquist 2001). The concatenated UCE data set was partitioned to apply the best-fitting models of nucleotide substitution. The protein-coding gene data set was partitioned according to codon position with the best-fitting model for each partition determined using AICc in jModelTest. MrBayes runs were performed with four chains per run, set for 50 million generations and sampling every 5,000 generations. All other run parameters were set as default and a 25% burnin was applied.

MrBayes runs were performed on the CIPRES Science Gateway portal (Miller et al. 2010); all other phylogenetic analyses, including those detailed below, were conducted on the GPC supercluster at the SciNet HPC consortium (Loken et al. 2010).

Bayesian Phylogenetic Reconstructions Using Heterogeneous Models

Bayesian inference incorporating heterogeneous models of evolution was used to investigate the influence of nonstationarity in base composition and exchange rate upon the topologies recovered from MrBayes concatenated and MP-EST species tree approaches. For computational tractability, data sets were simplified through the removal of constant sites and singletons and by translation of protein-coding genes, which produced alignments of 10,424 bp for UCEs and 16,849 amino acid residues for protein-coding genes. PhyloBayes MPI v. 1.4 (Lartillot et al. 2013) was used to build topologies with CAT profile mixture models, which attempt to account for heterogeneity in the substitution process through the use of multiple site classes associated with unique substitution matrices, and with the total number of classes itself estimated as a free parameter. Parameter estimates and consensus trees

were derived from three independent runs per data set run for a minimum of 50,000 cycles using a CAT–GTR substitution scheme.

A second approach to modeling heterogeneity in the evolutionary process was performed in P4 (Foster et al. 2009), which incorporates tree-heterogeneous models that allow multiple compositional and/or rate matrices across the topology. Parameters for "base" homogeneous models (i.e., with single compositional and rate matrices) were estimated with jModelTest for UCEs and with ProtTest v. 3.3 (Darriba et al. 2011) for the amino acid data set. Consensus estimates were obtained from a minimum of three chains per model that were run for a minimum of 2 million generations each, with sampling every 100 generations and with rate and composition matrices modeled as free parameters and initially assigned randomly upon the starting tree. Adequacy of P4 models to account for compositional nonstationarity was assessed with compositional χ^2 tests where the observed χ^2 value was compared with a null distribution generated from posterior predictive simulations from the post burnin sample for each investigated model (Foster 2004).

Postrun diagnostics and examination of raw trace data were performed in Tracer v. 1.5 (Rambaut and Drummond 2009). Convergence among runs was assessed with the bpcomp subprogram in PhyloBayes and from the calculation of average standard deviation of split frequencies (ASDOSS) in P4. Log marginal likelihoods were calculated using eq. (16) of Newton and Raftery (1994) in P4 and were used in turn to compute BFs for model comparisons. Following Kass and Raftery (1995), values of 2ln BF >6 were interpreted as providing strong evidence for the rejection of one model in favor of an alternative.

Comparison of Molecular and Morphological Data

We used the morphological data set of Worthy and Scofield (2012) to investigate the discordance between topologies constructed from morphological and molecular data. Of the 26 taxa scored for morphology, published sequence data were available from 27 nuclear genes in nine species (including 7 of the 12 extant paleognaths scored for morphology): elegant crested tinamou, little bush moa, ostrich, emu, Southern cassowary (Casuarius casuarius), lesser rhea (Pterocnemia pennata), greater rhea (Rhea americana), Southern brown kiwi (Apteryx australis), and the outgroup magpie goose (Anseranas semipalmata) (Haddrath and Baker 2012). A molecular phylogeny was estimated in MrBayes from the concatenated data set of 31,013 bp of aligned DNA, with gene sequences consisting entirely of exonic nucleotides partitioned by codon position and sequences spanning introns partitioned by gene. All run parameters follow those for MrBayes analyses detailed above.

Excess and retention indices were calculated with PAUP* v. 4.0b10 (Swofford 2001) for each morphological character (N.B. only 129 of 179 morphological characters remained informative with reduced taxon sampling). Median values are reported from calculations using 1,000 trees drawn from the PP distribution from the Bayesian analysis of molecular data

after chains reached convergence. These indices assess the fit of each morphological character on the molecular topology, with high excess and low retention values, respectively, indicating poor agreement (Holland et al. 2010).

Cliques of mutually compatible morphological characters, which could be indicative of concerted convergence in response to similar selective regimes, were identified following Blanke et al. (2013). Pairwise excess indices, which measure the dissimilarity between characters as the additional number of tree steps relative to the minimum number needed to map individual characters onto the molecular topology, were calculated for each pair of morphological characters using PAUP* and custom Perl scripts. UPGMA clustering of the pairwise dissimilarity matrix in PAUP* produced character cliques of compatible members whose pairwise dissimilarity equals zero.

Supplementary Material

Supplementary data S1–S4, figures S1 and S2, and tables S1–S7 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

References

Allentoft ME, Rawlence NJ. 2012. Moa's ark or volant ghosts of Gondwana? Insights from nineteen years of ancient DNA research on the extinct moa (Aves: Dinornithiformes) of New Zealand. Ann Anat. 194:36–51.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Bhullar B-AS, Marugán-Lobón J, Racimo F, Bever GS, Rowe TB, Norell MA, Abzhanov A. 2012. Birds have paedomorphic dinosaur skulls. Nature 487:223–226.

Blanke A, Greve C, Wipfler B, Beutel RG, Holland BR, Misof B. 2013. The identification of concerted convergence in insect heads corroborates Palaeoptera. Syst Biol. 62:250–263.

Bock WJ, Bühler P. 1990. The evolution and biogeographical history of the palaeognathous birds. In: Schuchmann K-L, Schmidt-Keonig K, van den Elsen R, editors. Current topics in avian biology. Proceedings of the International Centennial Meeting of the Deutsche Ornithologen Gesellschaft; 1998. Garmisch-Partenkirchen (Germany): Verlag der Deutsche Ornithologen-Gesellschaft. p. 31–36.

Bourdon E, de Ricqles A, Cubo J. 2009. A new transantarctic relationship: morphological evidence for a Rheidae–Dromaiidae–Casuariidae clade (Aves, Palaeognathae, Ratitae). Zool J Linn Soc Lond. 156:641–663.

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

Cracraft J. 1974. Phylogeny and evolution of the ratite birds. *Ibis* 116: 494–521.

Cracraft J, Barker FK, Braun M, Harshman J, Dyke GJ, Feinstein J, Stanley S, Cibois A, Schikler P, Beresford P, et al. 2004. Phylogenetic relationships among modern birds (Neornithes): toward an avian tree of life. In: Cracraft J, Donoghue MJ, editors. Assembling the tree of life. New York: Oxford University Press. p. 468–489.

Cubo J, Arthur W. 2001. Patterns of correlated character evolution in flightless birds: a phylogenetic approach. *Evol Ecol.* 14:693–702.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27: 1164–1165.

Dean A, Greenwald J. 1995. Use of filtered pipette tips to elute DNA from agarose gels. *BioTechniques* 18:980.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol.* 24: 332–340.

del Hoyo J, Elliot A, Sargatal J. 1992. Handbook of the birds of the world. Barcelona: Lynx Edicions.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Edwards SV, Jennings WB, Shedlock AM. 2005. Phylogenetics of modern birds in the era of genomics. *Proc R Soc B Biol Sci.* 272:979–992.

Elzanowski A. 1995. Cretaceous birds and avian phylogeny. *Cour Forsch Inst Senckenberg.* 181:37–53.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol.* 61:717–726.

Foster PG. 2004. Modeling compositional heterogeneity. *Syst Biol.* 53: 485–495.

Foster PG, Cox CJ, Embley TM. 2009. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Philos Trans R Soc B Biol Sci.* 364:2197–2207.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52: 696–704.

Gussekloo SWS, Cubo J. 2013. Flightlessness affects cranial morphology in birds. *Zoology* 116:75–84.

Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han K-L, Harshman J, et al. 2008. A phylogenetic study of birds reveals their evolutionary history. *Science* 320:1763–1768.

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

Haddrath O, Baker AJ. 2012. Multiple nuclear genes and retroposons support vicariance and dispersal of the palaeognaths, and an early Cretaceous origin of modern birds. *Proc R Soc B Biol Sci.* 279: 4617–4625.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp.* 41:95–98.

Han K-L, Braun EL, Kimball RT, Reddy S, Bowie RCK, Braun MJ, Chojnowski JL, Hackett SJ, Harshman J, Huddleston CJ, et al. 2011. Are transposable element insertions homoplasy free? An examination using the avian tree of life. *Syst Biol.* 60:375–386.

Harshman J, Braun EL, Braun MJ, Huddleston CJ, Bowie RCK, Chojnowski JL, Hackett SJ, Han K-L, Kimball RT, Marks BD, et al. 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. *Proc Natl Acad Sci U S A.* 105:13462–13467.

Holland BR, Spencer HG, Worthy TH, Kennedy M. 2010. Identifying cliques of convergent characters: concerted evolution in the cormorants and shags. *Syst Biol.* 59:433–445.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.

Huynen L, Millar CD, Lambert DM. 2012. Resurrecting ancient animal genomes: the extinct moa and more. *BioEssays* 34:661–669.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? *Trends Genet.* 22:225–231.

Johnston P. 2011. New morphological evidence supports congruent phylogenies and Gondwana vicariance for palaeognathous birds. *Zool J Linn Soc.* 163:959–982.

Kalendar R, Lee D, Schulman AH. 2009. FastPCR software for PCR primer and probe design and repeat search. *Genes Genom Genomics.* 3: 1–14.

Kass RE, Raftery AE. 1995. Bayes factors. *J Am Stat Assoc.* 90:773–795.

Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, et al. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* 2011:bar030.

Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* 19:1639–1645.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol.* 29:1695–1701.

Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol.* 7:S4.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 21:1095–1109.

Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures profiles in a parallel environment. *Syst Biol.* 62:611–615.

Lewis SE, Searle SMJ, Harris N, Gibson M, Iyer V, Richter J, Wiel C, Bayraktaroglu L, Birney E, Crosby MA, et al. 2002. Apollo: a sequence annotation editor. *Genome Biol.* 3:RESEARCH0082.1–0082.14.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.

Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol Biol.* 10:302.

Livezey BC. 1995. Heterochrony and the evolution of avian flightlessness. In: McNamara KJ, editor. Evolutionary change and heterochrony. New York: Wiley. p. 169–193.

Livezey BC, Zusi RL. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zool J Linn Soc.* 149:1–95.

Loken C, Gruner D, Groer L, Peltier R, Bunn N, Craig M, Henriques T, Dempsey J, Yu C-H, Chen J, et al. 2010. SciNet: lessons learned from building a power-efficient top-20 system and data centre. *J Phys: Conf Ser.* 256:012026.

Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* 21: 936–939.

Matzke A, Churakov G, Berkes P, Arms EM, Kelsey D, Brosius J, Kriegs JO, Schmitz J. 2012. Retroposon insertion patterns of neoavian birds: strong evidence for an extensive incomplete lineage sorting era. *Mol Biol Evol.* 29:1497–1501.

Mayr G. 2011. Metaves, Mirandornithes, Strisores and other novelties—a critical review of the higher-level phylogeny of neornithine birds. *J Zool Syst Evol Res.* 49:58–76.

Millar CD, Huynen L, Subramanian S, Mohandesan E, Lambert DM. 2008. New developments in ancient genomics. *Trends Ecol Evol.* 23:386–393.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE); New Orleans (LA). Nov 14, 2010. Institute of Electrical and Electronics Engineers. p. 1–8.

Newton MA, Raftery AE. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *J R Stat Soc B Methodol.* 56:3–48.

MBE

Oliver JC. 2013. Microevolutionary processes generate phylogenomic discordance at ancient divergences. *Evolution* 67:1823–1830.

Owen R. 1839. On the bone of an unknown struthious bird from New Zealand. *Proc Zool Soc Lond.* 7:169–171.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 9:e1000602.

Phillips MJ, Gibb GC, Crimp EA, Penny D. 2010. Tinamous and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites. *Syst Biol.* 59:90–107.

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

Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 25:1253–1256.

Rambaut A, Drummond AJ. 2009. Tracer v1.5. [cited 2014 Apr 24]. Available from: http://tree.bio.ed.ac.uk/software/tracer/.

Ray DA, Xing J, Salem A-H, Batzer MA. 2006. SINEs of a nearly perfect character. *Syst Biol.* 55:928–935.

Rizzi E, Lari M, Gigli E, De Bellis G, Caramelli D. 2012. Ancient DNA studies: new perspectives on old samples. *Genet Sel Evol.* 44:21.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol.* 29:24–26.

Smit AFA, Hubley R, Green P. 2010. RepeatMasker Open-3.0. [cited 2014 Apr 24]. Available from: http://www.repeatmasker.org.

Smith JV, Braun EL, Kimball RT. 2013. Ratite nonmonophyly: independent evidence from 40 novel loci. *Syst Biol.* 62:35–49.

Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21:456–463.

Swofford DL 2001. *PAUP**. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.

Telford MJ, Copley RR. 2011. Improving animal phylogenies with genomic data. *Trends Genet.* 27:186–195.

Waddell PJ, Kishino H, Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Inform.* 12:141–154.

Worthy TH, Scofield RP. 2012. Twenty-first century advances in knowledge of the biology of moa (Aves: Dinornithiformes): a new morphological analysis and moa diagnoses revised. *N Z J Zool.* 39: 87–153.

Xia X. 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol.* 30: 1720–1728.

Xia X, Xie Z, Salemi M, Chen L, Wang Y. 2003. An index of substitution saturation and its application. *Mol Phylogenet Evol.* 26:1–7.