

Phylogenomic evidence for multiple losses of flight in ratite birds

John Harshman^{a,b,c}, Edward L. Braun^{c,d,e}, Michael J. Braun^{c,f,g}, Christopher J. Huddleston^f, Rauri C. K. Bowie^{a,h,i}, Jena L. Chojnowski^d, Shannon J. Hackett^a, Kin-Lan Han^{d,f,g}, Rebecca T. Kimball^d, Ben D. Marks^j, Kathleen J. Miglia^k, William S. Moore^k, Sushma Reddy^a, Frederick H. Sheldon^j, David W. Steadman^l, Scott J. Steppan^m, Christopher C. Witt^{j,n}, and Tamaki Yuri^{d,f}

^aZoology Department, Field Museum of Natural History, 1400 South Lakeshore Drive, Chicago, IL 60605; ^b4869 Pepperwood Way, San Jose, CA 95124; ^dDepartment of Zoology, University of Florida, Gainesville, FL 32611; ^fDepartment of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, 4210 Silver Hill Road, Suitland, MD 20746; ^gBehavior, Ecology, Evolution, and Systematics Program, University of Maryland, College Park, MD 20742; ^hMuseum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720; ⁱDepartment of Science and Technology/National Resource Foundation Centre of Excellence at the Percy FitzPatrick Institute, Department of Botany and Zoology, Stellenbosch University, Matieland 7602, South Africa; ^jMuseum of Natural Science, 119 Foster Hall, Louisiana State University, Baton Rouge, LA 70803; ^kDepartment of Biological Sciences, Wayne State University, 5047 Gullen Mall, Detroit, MI 48202; ^lFlorida Museum of Natural History, University of Florida, Gainesville, FL 32611; ^mDepartment of Biological Science, Florida State University, Tallahassee, FL 32306; and ⁿDepartment of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, NM 87131

Edited by Morris Goodman, Wayne State University School of Medicine, Detroit, MI, and approved July 17, 2008 (received for review April 2, 2008)

Ratites (ostriches, emus, rheas, cassowaries, and kiwis) are large, flightless birds that have long fascinated biologists. Their current distribution on isolated southern land masses is believed to reflect the breakup of the paleocontinent of Gondwana. The prevailing view is that ratites are monophyletic, with the flighted tinamous as their sister group, suggesting a single loss of flight in the common ancestry of ratites. However, phylogenetic analyses of 20 unlinked nuclear genes reveal a genome-wide signal that unequivocally places tinamous within ratites, making ratites polyphyletic and suggesting multiple losses of flight. Phenomena that can mislead phylogenetic analyses, including long branch attraction, base compositional bias, discordance between gene trees and species trees, and sequence alignment errors, have been eliminated as explanations for this result. The most plausible hypothesis requires at least three losses of flight and explains the many morphological and behavioral similarities among ratites by parallel or convergent evolution. Finally, this phylogeny demands fundamental reconsideration of proposals that relate ratite evolution to continental drift.

convergence | flightlessness | Paleognath | homoplasy | vicariance biogeography

Living birds are divided into two major groups, Palaeognathae and Neognathae (1, 2), a classification based originally on bony palate structure (3, 4). Palaeognathae also is traditionally divided into two groups, the flightless ratites (defined by absence of a keel on the sternum) and the volant tinamous. Although they include fewer than 1% of extant avian species, paleognaths have long been viewed as central to understanding the early evolution of birds. The many morphological and behavioral similarities of ratites suggest common ancestry, but some have proposed that they instead reflect convergent adaptation to a flightless, cursorial lifestyle. The distribution of ratites is also remarkable; ostriches live in Africa, rheas in South America, emus and cassowaries in Australasia, kiwis and moas (now extinct) in New Zealand, and elephant birds (also now extinct) in Madagascar. How did these flightless birds get to these far-flung southern landmasses?

Paleognath relationships have been controversial since the earliest days of evolutionary biology. When Huxley defined the paleognathous palate he also stated that extant ratites “are but waifs and strays of what was once a very large and important group” (3). Nevertheless, the notion that ratites have independent origins arose around the same time, when Owen suggested that they have closer affinities to various volant groups while being united by the “arrested development of wings unfitting

them for flight” (5). Ratite monophyly was debated throughout much of the last century, with Mayr and Amadon (6) stating in 1951 that the “present consensus is that the main groups of these birds are of independent origin.” DeBeer provided a developmental explanation for the similarities among ratites when he interpreted the paleognathous palate and other features of extant ratites as neotenic (7). Paleognath monophyly was questioned as late as the 1980s (8), but it has been confirmed by many recent morphological and molecular studies (9–13).

Most recent studies have also strongly supported ratite monophyly (9–12, 14), suggesting a single loss of flight in their common ancestor. This puzzled biogeographers for more than a century, because ratites would be unable to achieve their current distribution on southern land masses if their common ancestor was flightless. Continental drift provided a compelling solution. No longer was it necessary to imagine giant flightless birds crossing vast oceans; they could have rafted to their current distributions on fragments of the Earth’s crust (15). Although the proposed phyletic branching patterns for ratites do not correspond perfectly to the order of separation of land masses during the breakup of Gondwana, the convenient serendipity of continental drift as a mechanistic explanation for ratite distribution proved irresistible (10, 11, 14, 16), and it stands today as a textbook example of vicariance biogeography (17, 18).

Despite the current consensus, some continue to question ratite monophyly (13, 19) and the role of Gondwana in ratite distribution (20). It has long been recognized that adaptation to a flightless, cursorial lifestyle can result in morphological convergence or parallelism, especially in the postcranial skeleton (1, 19, 21). Such convergent adaptations might mislead phylogenetic inference based on morphology. In fact, one study based solely

Author contributions: J.H., E.L.B., M.J.B., S.J.H., R.T.K., W.S.M., F.H.S., and D.W.S. designed research; J.H., E.L.B., M.J.B., C.J.H., R.C.K.B., J.L.C., S.J.H., K.-L.H., R.T.K., B.D.M., K.J.M., S.R., S.J.S., C.C.W., and T.Y. performed research; J.H., E.L.B., M.J.B., and C.J.H. analyzed data; and J.H., E.L.B., and M.J.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: New DNA sequences are deposited in GenBank (accession nos. EU805776–EU805796, and EU822937). Alignments and trees have been deposited in TreeBase (study accession no. S2138).

J.H., E.L.B., and M.J.B. contributed equally to this work.

†To whom correspondence should be addressed at: Department of Zoology, P.O. Box 118525, University of Florida, Gainesville, FL 32611. E-mail: ebraun68@ufl.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0803242105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

Table 1. Bootstrap support for the crucial branch uniting a non-ostrich paleognath clade from ML and MP analyses of individual loci

Gene	ML		MP	
	Support	Conflict	Support	Conflict
<i>ALDOB</i>	64	—	63	—
<i>BDNF</i>	—	57*	—	91
<i>CLTC</i>	42	—	—	59
<i>CLTCL1[†]</i>	68	—	—	—
<i>CRYAA</i>	66	—	54	—
<i>EEF2</i>	99	—	88	—
<i>EGR1</i>	77	—	56	—
<i>FGB</i>	—	42	62	—
<i>GH1</i>	71	—	77	—
<i>HMG2</i>	81	—	47	—
<i>IRF2</i>	60	—	71	—
<i>MB</i>	97	—	88	—
<i>MUSK</i>	95	—	91	—
<i>MYC[‡]</i>	85	—	—	—
<i>NGF</i>	78	—	74	—
<i>NTF3</i>	83	—	52	—
<i>PCBD1</i>	92	—	76	—
<i>RHO</i>	78	—	65	—
<i>TGFB2[‡]</i>	—	52	—	—
<i>TPM1</i>	47	—	69	—
Number of loci	17 [‡]	3	15 [§]	2

*ML analysis of *BDNF* does not support paleognath monophyly, instead placing the tinamous within neognaths.

[†]These genes have multiple MP trees, some of which support and some of which conflict with ratite polyphyly.

[‡]Binomial test, $P = 8 \times 10^{-15}$ (pure birth model) or 3×10^{-12} (equiprobable model).

[§]Binomial test, $P = 9 \times 10^{-12}$ (pure birth model) or 2×10^{-9} (equiprobable model).

nation of avian *FGB* intron 7 indels (34) revealed that 1-bp indels are more than twice as likely to exhibit homoplasy as longer (≥ 5 -bp) indels. Thus, the likelihood of homoplasy in the 1-bp indel is much higher than the combined likelihood of homoplasy in the two longer indels that support ratite polyphyly.

Known Phylogenetic Artifacts Do Not Explain Ratite Polyphyly. Several distinct systematic biases that can mislead phylogenetic analysis have been characterized (35–38), and we determined whether any of these artifacts could explain ratite polyphyly.

Inconsistent phylogenetic estimation due to long-branch attraction (35) is a concern in phylogenomics. Superficially, it might appear problematic for paleognath phylogeny because tinamous have a high rate of molecular evolution (Table S4). Although the branch leading to tinamous is long (e.g., Fig. 1), it is not united with the other long branch on the tree, which leads to the outgroup. In fact, long-branch attraction would be expected to favor the conventional hypothesis of ratite monophyly rather than the surprising alternative of ratite polyphyly that we observed. We confirmed this expectation in two ways. First, a parametric bootstrap (Swofford–Olson–Waddell–Hillis, SOWH) test (39) rejected a null hypothesis of ratite monophyly ($P < 0.001$). Second, a series of 1,000 MP analyses in which the outgroup was replaced with a random sequence of similar base composition, simulating the longest possible branch (40), all rooted the tree within tinamous or along the branch leading to them. This confirmed our expectations and rejected long-branch attraction as an explanation for ratite polyphyly ($P < 0.001$).

Convergence in base composition can also create artifactual relationships (23). If the observed topology reflects composi-

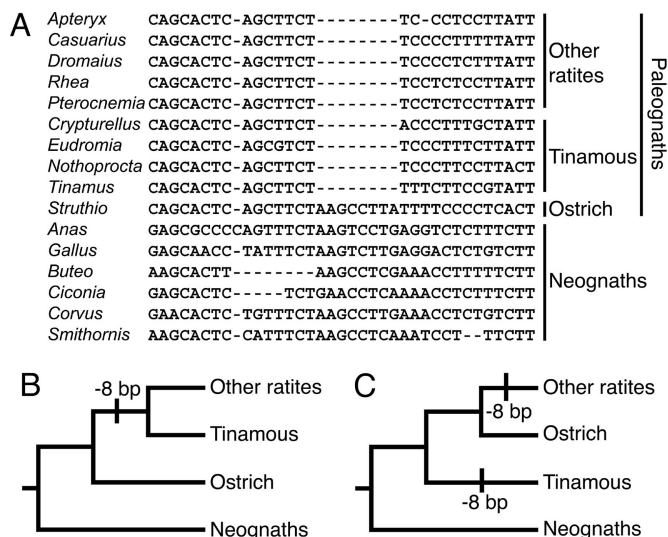


Fig. 3. An 8-bp deletion in *ALDOB* supports ratite polyphyly. (A) Alignment of the region around the informative deletion in *ALDOB* (positions 3213–3220). The ostrich shares its character state (+8 bp) with neognaths, whereas tinamous share the character state of all other ratites (–8 bp). (B) The distribution of character states can be mapped as a single deletion on the optimal topology found in this study. (C) The distribution requires at least two steps on the traditional topology (one possible reconstruction shown).

tional convergence, we would expect a tree derived from base-compositional information alone to have tinamous nested within ratites. When we clustered taxa using base-compositional distances, however, tinamous fell within ratites for few loci (Table S5). Instead, the most common grouping (10 of 20 loci) clustered ratites, a signal expected to reinforce the conventional topology. Furthermore, ratite polyphyly was supported both by analyses conducted after RY coding (Fig. S1A), which increases historical signal relative to compositional bias (37, 41), and ML analyses using a model allowing base compositional change (Table S6 and Fig. S1B). Base compositional convergence therefore cannot explain our results.

The one locus that failed to support paleognath monophyly was *BDNF*, which places tinamous within neognaths in ML analyses. However, *BDNF* has the greatest base-compositional variation and longest tinamou branch of all loci studied (Table S4). ML analyses of *BDNF* after RY coding do not place tinamous within neognaths (data not shown), suggesting that the anomalous results for that locus are artifactual.

Another potential problem occurs when individual gene trees differ from the species tree. Under some conditions, stochastic lineage sorting can produce a bias toward gene trees more symmetrical than the species tree (38), and the gene tree signal can predominate when many loci are analyzed (42). However, this is unlikely to occur unless the relevant branches are short relative to the coalescence time for the genes examined. The expected coalescence time for nuclear genes is twice the effective population size in generations. Because the relevant branch length and the ancestral paleognath population size are unknown, we assessed the potential impact of lineage sorting empirically. The gene-tree bias is weak (42), so it is unlikely to produce the consistency of single-gene trees we observed (Table 1). The aggregate probability of all trees with non-ostrich paleognath monophyly is 1/10, given the pure birth model for gene trees (42), so the probability of finding this result in 15 or more of 20 single-gene analyses is very small ($P = 9 \times 10^{-12}$). This calculation is conservative; it reflects the minimum number of loci supporting the relevant branch in single-gene MP analyses. More loci support non-ostrich paleognath monophyly using

etic. We have discovered a robust genome-wide signal that is not associated with any known phylogenetic artifact. We believe this phylogeny resolves a debate on ratite origins that began in the time of Huxley and Owen (3–5). Our phylogeny implies that the numerous striking similarities associated with flightlessness (1) had independent origins in various ratite lineages. Thus, the flightless ratites are living evidence of parallel evolutionary trajectories from flighted ancestors. The possibility that multiple, unique developmental genetic pathways underlie the ratite form should be tested in light of this new phylogenetic hypothesis. Finally, our phylogeny removes the need to postulate vicariance by continental drift to explain ratite distribution. Although that theory seemed to represent a concisence between evolutionary biology and geology, it was never completely consistent either with any published phylogeny or the existence of paleognath fossils in the Northern hemisphere (45, 57). Perhaps the impact of our phylogeny should be viewed as yet another example of the phenomenon that Huxley called “the great tragedy of science—the slaying of a beautiful theory by an ugly fact.”

Methods

Sequencing and Alignment of Nuclear Loci. We amplified and sequenced 20 nuclear loci (see *SI Methods* for detailed methods). Sequences (accession numbers listed in *Table S7*) were aligned manually (see *SI Methods*); ambiguously aligned regions and sparsely sampled sites (those not present in at least four birds and three paleognaths) were excluded from analyses.

Phylogenetic Analyses. Four nonpasserine neognaths were used as outgroups in most analyses of the 14 taxon, 23,902-bp dataset (e.g., Fig. 1). Crocodylians and passerines were included only in analyses designed to test the position of the root of the avian tree using the subset of sequences for which crocodylian and avian sequences could be aligned (4,668 bp; Fig. 2).

Optimal MP and ML trees using unpartitioned data were identified by using PAUP* 4.0b10 (58). MP analyses used branch-and-bound searches with equally weighted characters and assessed support using 1,000 bootstrap replicates. The appropriate nucleotide substitution model (*Table S8*) for ML analyses was determined by using the Akaike information criterion and Modeltest 3.6 (59) or the models appropriate for RY-coded data (see *SI Methods*). ML analyses used heuristic searches with 10 random addition sequence replicates and tree bisection and reconnection (TBR) branch swapping; support was assessed by using 100 bootstrap replicates.

Three ML analyses with the data partitioned by locus were conducted. First, a partitioned RAxML (60) analysis, using the general time-reversible (GTR)+I model with distinct parameter values for each partition and linked branch length parameters, was conducted and support assessed by using 100 bootstrap replicates. The other two analyses used PAUP* and nhPhyML (61), allowing use of a more diverse set of nucleotide substitution models. Branch lengths were unlinked and monophyly of well established clades was assumed *a priori*, focusing on a “plausible set” of 315 trees (105 arrangements of the five major paleognath groups and three arrangements within tinamous; Fig. S2). The log likelihoods of all partitions for each tree were summed after being calculated using PAUP* (*Table S9*) or nhPhyML (not shown). PAUP* analyses

used the best-fit model for each locus (*Table S8*) and nhPhyML analyses used the Galtier and Gouy (62) model, which allows the GC content to vary across the tree. In these analyses, support was assessed by using 1,000 resampling of estimated log-likelihoods (RELL) (63) bootstrap replicates.

Bayesian Markov chain Monte Carlo analyses were performed by using MrBayes 3.1.1 (64). We ran four chains for 10 million generations, sampling every 500 generations and discarding the first 500 trees, with other run parameters set at defaults. Partitioned Bayesian analyses used linked branch lengths.

The relationship between the probability that a clade is correct and bootstrap support is complex (65, 66), but the bootstrap is conservative under many circumstances. In contrast, Bayesian posterior probabilities can overestimate the probability that a clade is correct (e.g., 66). We also assessed branch (Bremer) support as well as hidden support and conflict for specific branches (31, 32).

Tests of Topologies, Base Composition, and Individual Gene Trees. We used two topology tests appropriate for comparisons of trees that were not specified *a priori* (67) to examine the position of the ostrich. The SH test (68) was performed by using the plausible set of 315 distinct trees (Fig. S2) because the test requires the inclusion of all trees that can be entertained as the true topology (67). The parametric bootstrap (SOWH) test (39) compares the difference in optimality scores for the empirical data on the optimal tree and a null hypothesis tree to a distribution generated by simulation. The null hypothesis topology was the most likely one with ratite monophyly (tree 19a of Fig. S2); 1,000 simulated datasets were analyzed by using MP and ML (*SI Methods*).

Base-compositional clustering was performed by using minimum evolution in PAUP* with a matrix of Euclidean distances between base-composition vectors for each taxon (see *SI Methods*). Topologies examined were limited to the plausible set (Fig. S2).

Individual gene trees may differ from the species tree. We tested whether the observed number of gene trees showing monophyly of non-ostrich paleognaths was unexpected given the null hypothesis of a completely polytomous species tree. We used binomial tests and two appropriate models of tree probability to assess this possibility (*SI Methods*).

Identification of Informative Indels. To identify low-homoplasy indels, all gaps in a 19-gene, 171-taxon dataset (25) were coded by using the simple gap coding method (69). These indels were mapped by using MP on the ML tree for the 171-taxon dataset (25) and the same tree rearranged so ratites were monophyletic. Indels mapping unambiguously on the branch of interest with a consistency index ≥ 0.5 were examined further, and ambiguously aligned regions were removed. *CLTCL1* was not included in the 171-taxon analysis (25), so it was examined independently by using a similar methodology.

ACKNOWLEDGMENTS. We thank D. L. Swofford, P. O. Lewis, and D. J. Zwickl for advice on phylogenetic analyses; D. J. Levey, K. Braun, and anonymous reviewers for helpful comments; and A. B. Thistle for a technical edit of the manuscript. For the loan of samples, we thank the Field Museum of Natural History, the University of Kansas, Louisiana State University, the United States National Museum of Natural History, L. D. Densmore, and the collectors (*Table S2*). This work was supported by the U.S. National Science Foundation Assembling the Tree of Life Program Grants DEB-0228675 (to S.J.H.); DEB-0228682 (to R.T.K., E.L.B., and D.W.S.); DEB-0228688 (to F.H.S.); and DEB-0228617 (to W.S.M.).

- Davies SJJF (2002) *Ratites and Tinamous: Tinamidae, Rheidae, Dromaiidae, Casuariidae, Apterygidae, Struthionidae* (Oxford Univ Press, Oxford).
- Harshman J (2007) in *Reproductive Biology and Phylogeny of Birds*, ed Jamieson BGM (Science Publishers, Enfield, NH), pp 1–35.
- Huxley TH (1867) On the classification of birds; and on the taxonomic value of the modifications of certain of the cranial bones observable in that class. *Proc Zool Soc London* 145:415–472.
- Pycraft WP (1900) Part II. On the morphology and phylogeny of the Palaeognathae (Ratitae and Crypturi) and Neognathae (Carinatae). *Trans Zool Soc London* 15:149–290.
- Owen R (1866) *The Anatomy of Vertebrates, Vol. 2: Birds and Mammals* (Longmans, Green, London).
- Mayr E, Amadon D (1951) A classification of recent birds. *Am Mus Novit* 1496: 1–42.
- DeBeer G (1956) The evolution of ratites. *Bull Br Mus (Nat Hist) Zool* 4:57–76.
- Olson SL (1985) The fossil record of birds. *Avian Biol* VIII:79–238.
- Lee K, Feinstein J, Cracraft J (1997) in *Avian Molecular Evolution and Systematics*, ed Mindell DP (Academic, San Diego), pp 173–211.
- Cooper A, et al. (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* 409:704–707.
- Haddrath O, Baker AJ (2001) Complete mitochondrial DNA genome sequences of extinct birds: Ratite phylogenetics and the vicariance biogeography hypothesis. *Proc R Soc London Ser B* 268:939–945.
- 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.
- Elzanowski A (1995) Cretaceous birds and avian phylogeny. *Cour Forsch Inst Senck* 181:37–53.
- Sibley CG, Ahlquist JA (1990) *Phylogeny and Classification of Birds* (Yale Univ Press, New Haven, CT).
- Cracraft J (1973) Continental drift, palaeoclimatology, and the evolution and biogeography of birds. *J Zool* 169:455–545.
- Cracraft J (2001) Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. *Proc R Soc London Ser B* 268:459–469.
- Futuyama DJ (2005) *Evolution* (Sinauer, Sunderland, MA).
- Gill F (2007) *Ornithology* (Freeman, New York).
- Bock WJ, Bühler P (1990) in *Proceedings of the International Centennial Meeting of the Deutsche Ornithologen-Gesellschaft*, eds van den Elzen R, Schuchmann K-L, Schmidt-Koenig K (Verlag der Deutschen Ornithologen-Gesellschaft, Bonn, Germany), pp 31–36.

