Supporting information for "Ratite Non-Monophyly: Independent Evidence From 40 Novel Loci" by J.V. Smith, E.L. Braun, and R.T. Kimball

Table of Contents:

Table S1. Specimens used in this study. Sequences from the chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*) were taken from whole genome sequences in Genbank.

Species	Common name	Instituition[a]	Voucher	Collector
Palaeognathae				
Dromaius novaehollandiae	Emu	LSUMNS	B5895	Captive
Rhea americana	Greater rhea	FLMNH	44923	Captive
Pterocnemia (*Rhea*) *pennata*[b]	Darwin's rhea	USNM	620827	Captive
Struthio camelus	Common ostrich	—	No Voucher[c]	Captive
Crypturellus soui	Little tinamou	USNM	586295	K. S. Bostwick
Tinamus guttatus	White-throated tinamou	FMNH	389673	B. D. Patterson
Galloanserae				
Chauna torquata	Southern screamer	USNM	614546	J. P. Angle
Crax alector	Black curassow	USNM	625104	C. M. Milensky

[a] FLMNH, Florida Museum of Natrual History, FMNH, Field Museum of Natural History, LSUMNS, Louisiana State University Museum of Natural Science, USNM, United States National Museum.
[b] *Pterocnemia pennata* is sometimes placed in the genus *Rhea*.
[c] The *Struthio camelus* sample used for this study was validated by sequence comparison for several loci with published data from LSUMNS B1526.

Table S2. Locus information using gene codes from Entrez Gene and primer sequences. In some cases, primers were re-designed to amplify all taxa; these primers are listed below the original primers. For CLOCK, separate primers were designed for introns 6 and 10.

Locus	Chromosome[a]	Forward Primer	Reverse Primer	Type
ACTB[b]	10	CCTGATGGTCAGGTCATCA	CAGCAATGCCAGGGTACAT	EPIC
ARNTL[c]	5	TGGTTCAGTTTCATGAACCCTTG	CCTGAAGCACRCTGTCCATGCT	EPIC
BMP5	3	GTCCACCAGTNMGGTATTAATC	CCTNCCACTGTANCAAANG	Anon[e]
		GAGGGGAAACTTGGAAYTTTGTGG	GCCAAATACCATACCWTGAACTGAC	
CALB1[d]	2	AGGGTGTCAARATGTGTGSGAAAGA	GTANAGCTTCCCTCCATCNGACAA	EPIC
CHMP5	2	CTAAGTAGGAATTGTCTTCATCAGC	GATGAAGACGATTTGGAAGC	EPIC
CIZ1	17	CCCTGAATCAGCCCTCAAATTCTACTGTTA	AATCTCCCCAAGTCGCTGCTG	EPIC
CLOCK (6)	4	CCAGAGGGGGAACATTCAGAA	TCCTTTGGGTCTATTGTTCCTCG	EPIC
CLOCK (10) [c]		CATGTGGATGATCTAGATAATCTGGC	GYAATGTGTTTGCAGCCAAATCCA	EPIC
CRAT	17	AGAAAGGCCTGGAGAGGAGAGC	GTCTTCAACCACCAGTCCGAGAG	EPIC
			CATGSTATCTGCTGCTGTTYGGTCC	
CSDE1[c]	2	CTGGTGCTGTAAGTGCTCGTAAC	CCAGGCTGTAAGGTTTCTAGGTCAC	EPIC
CSNK1E	2	GACTACGTCTTTGACTGGAACATGCT	ATCCTCAGGGTTTCGGGCTG	EPIC
DDX5	18	CATATAAATCATCAGCCATTCCTGG	GTTGGTGCCAGCACAAGAC	EPIC
EIF5	5	GTCCAGCAATGAGACACCTCCAC	CCAGTCATCATCGTCCTCCTCC	EPIC
ENO1	21	GGTGATGATCTGACTGTGACCAACC	CATCACACCCCAGCCATTGGAC	EPIC
ETS2	1	CAGTGGCTTCACAAAGGAACAGTGTC	CAAACATGCTGTTCAGTCCACAACC	EPIC
GAPDH[d]	1	TGCGGGTGCTGGCATTGC	TGCATGCCATGTGGACCAT	EPIC
GNB2L1	16	GATCACCTCCTGCTTCAGCTC	GGCCCCAGCATCAAGATCTG	EPIC
GRIA2	4	GGTTGGAGAACTTGTTTATGG	GGCTCATAAAGGGCTTTG	EPIC
HNRPA2B1	2	GCATTTCCTGTCTAGAGAGGGCTTTC	CATTTGATGACCATGATCCTGTGTGG	EPIC
Intergenic 1	2	GGATGCTCGCTCAGKAMTTTG	GTGGTTTAGCCTGGAGTTAAG	Anon[f]
Intergenic 2	Z	CACAACACTTGACTATGGC	GAYKCAGACACCRCATTATC	Anon[f]
Intergenic 3	8	GTTATCAAGTGATCTGTTTGCAGTC	GCAKCYGTGATGCCAGGRTG	Anon[f]
Intergenic 4	4	GCTGCACTYATAGKTRAGG	GACCAACTGAAAAGACTTG	Anon[f]
		AATGTGGGACTYAAGGAAGCTG	CAAYAGCCAAGGTCCTGATTCAG	
KCNQ5	3	CATGGACCGAAGAGGAGGCACT	CCAGAGAGCATCTGCATATGTGGAG	EPIC
NAT15[c]	14	ATCAGAGGGGTTCTCAAAGATGG	AGAGAAGGCTCTGGGCTTGTCGGTA	EPIC
NUSAP1	5	GGYTGTTGAGAGRACAGAAGG	GTTTKGGCAGGTRCTGGC	Anon[h]
PALLD	4	CTGCACWGGAAGGCTKATG	GTCTGAWGATGTCTSAGCTGTG	Anon[h]
		TCCTGGYCAGCCTCATRTTAGAAG		
PARK7[c]	21	GCAGGCCTRRCTGGAAAAGARCC	TTCTGAGCTCCWAGRTTACC	EPIC

PAXIP1 [c]	2	CCCTCAGACACTGGATTAYGAATCAT	CCAAGGATTCCGAAGCAGTAAG	EPIC
PER2 [c]	9	CATCTTCAYCCAAATGACAGACC	CCTGATTGGTGAATAGTCAAAAGG	EPIC
PHB	27	TGGGGYTGGCTGTNGCRGGTGGAGT	CAGGGGATGAGGAAGTGGGTRCCTTC	EPIC
PSMA2	2	GTATAGTGGTATGGGTCCAGATTAC	GCTGTTGGAATGGGCTCATGATAAAC	EPIC
PUM2	3	GGAGACCTTRTTGGACATATTGTTG	CACMGCTTCAATGAGACACCTC	Anon[h]
SEPT2	15	CTTGGCGATCACAGGGACAATG	GAACAGGCGCCACATTATAGACAATAG	EPIC
SFRS3	Z	GCTGTGTATTTGGTCTATTCAGAG	CAGGTGGCAAATGTAAAGATGTG	EPIC
SLC25A21	5	CTGCTAAAWAGAATCCNGG	GRAGTCATTRTGCACCCTC	Anon[e]
TCP1	3	CCCNGATCGCAAAAATCTGAAATG	CGAAGAATAGTAATTGCWGCTTCTGTTGC	EPIC
TTN	7	GGCAGYATTAAGGAAACGCAC	GAGTAYGTAGACCAAMCCATACG	Anon[g]
TXNDC12 [c]	8	GGAAACCCCAGCTACAAGTATTTC	GGCCTCCTTCATCCCTTGG	EPIC
VDAC2	6	GTCAACGTCACAACCTAGG	CAACTTTCTCACCAAATACAGG	EPIC
VIM [c]	2	GACCGTGGAAACTAGAGATGGAC	GTCATCGTGATGCTGGGAAGTTTC	EPIC

[a] Chromosome number refers to the position in the chicken genome.
[b] Previously published primer from Waltari and Edwards (2002).
[c] Previously published primer from Kimball et al. (2009).
[d] Previously published primer from Cox et al. (2007).
[e] Anonymous locus -Large Intron
[f] Anonymous locus- Intergenic region
[g] Anonymous locus- exon only
[h] Anonymous locus -intron and exon

REFERENCES:

Cox W.A., Kimball, R.T., Braun, E.L. 2007. Phylogenetic position of the New World quail (Odontophoridae): Eight nuclear loci and three mitochondrial regions contradict morphology and the Sibley-Ahlquist tapestry. The Auk. 124: 71-84.

Kimball R.T., Braun E.L., Barker F.K., Bowie R.C.K., Braun M.J., Chojnowski J.L., Hackett S.J., Han K.L., Harshman J., Heimer-Torres V., Holznagel W., Huddleston C.J., Marks B.D., Miglia K.J., Moore W.S., Reddy S., Sheldon F.H., Smith J.V., Witt C.C., Yuri T. 2009. A well-tested set of primers to amplify regions spread across the avian genome. Mol. Phylo. Evol. 50:654-660.

Waltari, E., Edwards, S.V. 2002. The evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. Am. Nat. 160: 539-552.

Table S3. Results of ML analyses for each locus[a] (topology numbers from Table 1).

Locus	Best-fit Model[b]	Total Sites	Variable Sites	Topology	Root[c]
EPIC	GTR+Γ+I			**1**	**Ostrich**
ACTB	HKY+Γ	360	230	**2**	**Ostrich**
ARNTL	TVM+I	635	332	**1**	**Ostrich**
CALB1	K81uf+I	546	244	**2**	**Ostrich**
CHMP5	TVM+Γ	475	246	7	Rhea
CIZ1	HKY+I	383	156	**2**	**Ostrich**
CLOCK	TVM+Γ	1512	721	**1**	**Ostrich**
CRAT	TrN+Γ	515	249	**1**	**Ostrich**
CSDE1	TVM+I	375	216	**1**	**Ostrich**
CSNK1E	TVM+I	647	322	**3**	**Ostrich**
DDX5	TVM+I	494	333	8	Rhea
EIF5	TVM+Γ	580	307	10[d]	Sym
ENO1	K81uf+I	440	255	**1**	**Ostrich**
ETS2	TVM+Γ	435	173	**3**	**Ostrich**
GAPDH	HKY	441	271	14	Tinamou
GNB2L1	GTR+I	295	192	4[e]	Emu
GRIA2	GTR+I	510	267	**3**	**Ostrich**
HNRPA2B1	TVM+I	506	261	—[f]	—
KCNQ5	K81uf+I	797	218	**1**	**Ostrich**
NAT15	TVM+Γ	1138	568	9[g]	Rhea
PARK7	GTR+I	762	415	**2**	**Ostrich**
PAXIP1	TVM	469	217	8	Rhea
PER2	HKY+I	749	408	**3**	**Ostrich**
PHB	TVM+Γ	489	202	—[f]	—
PSMA2	TVM+Γ	516	250	**3**	**Ostrich**
SEPT2	GTR+I	505	280	**2**	**Ostrich**
SFRS3	HKY+I	459	235	**2**	**Ostrich**
TCP1	HKY+I	230	144	**1**[g]	**Ostrich**
TXNDC12	TVM	463	245	—[h]	—
VDAC2	TVM+Γ	484	280	8	Rhea
VIM	HKY+Γ	563	361	**1**	**Ostrich**
Anonymous	GTR+Γ+I			**1**	**Ostrich**
BMP5	HKY+Γ	405	62	—[f]	—
Intergenic 1	HKY+Γ	702	350	**3**	**Ostrich**
Intergenic 2	K80+Γ	532	315	**1**	**Ostrich**
Intergenic 3	K80+Γ	580	277	**3**	**Ostrich**
Intergenic 4	TVMef+Γ	411	233	15	Tinamou
NUSAP1	TIM+Γ	421	214	—[i]	—
PALLD	K81uf+I	659	331	—[h]	—
PUM2	GTR+Γ	534	282	—[f]	**Ostrich**
SLC25A21	GTR+Γ	660	310	5	Emu
TTN	TrN+Γ	541	102	**1**	**Ostrich**

[a] Continued on the next page along with all footnotes.

Table S3. Continued.

Locus	Best-fit Model	Total Sites	Variable Sites	Topology[b]	Root[c]
Published	TVM+Γ+I			**1**	**Ostrich**
ALDOB	GTR+1	2962	1137	**3**	**Ostrich**
BDNF	TVM+Γ+I	688	134	**1**	**Ostrich**
CLTC	TVM+Γ+I	2175	1019	7	Rhea
CLTCL1	HKY+1	810	311	**2**	**Ostrich**
CRYAA	HKY+Γ+I	1363	779	**1**	**Ostrich**
EEF2	TVM+Γ	2223	908	**3**	**Ostrich**
EGR1	GTR+Γ	1723	374	**3**	**Ostrich**
FGB	TVM+I	3041	1483	**3**	**Ostrich**
GH1	HKY+Γ	1434	723	**3**	**Ostrich**
HMGN2	GTR+Γ+I	2100	1053	13	Tinamou
IRF2	HKY+Γ	653	268	**3**	**Ostrich**
MB	TrN+I	958	313	**3**	**Ostrich**
MUSK	TVM+I	721	399	**3**	**Ostrich**
MYC	TVM+Γ+I	1256	348	**1**	**Ostrich**
NGF	TVM+I	749	260	—[j]	—
NTF3	TrN+Γ	731	176	**2**	**Ostrich**
PCBD1	TrN+Γ	1228	668	**3**	**Ostrich**
RHO	TVM+I	1979	922	**2**	**Ostrich**
TGFB2	TVM+I	731	356	12	Sym
TPM	TIM+Γ	491	124	7	Rhea
Mitochondrion	GTR+Γ+I	15393	3408	13	Tinamou

[b] The best-fitting model based upon the AIC. Models are from Modeltest 3.7.

[c] The 15 plausible topologies (Table 1) can be placed into five groups, four of which have one lineage sister to other paleognaths and one of which is symmetrical. These groups are: ostrich sister = 1-3; emu sister = 4-6; rheas sister = 7-9; symmetrical = 10-12; tinamous sister = 14-15. Topologies with ostrich sister to other paleognaths are also presented in bold.

[d] The extended majority rule bootstrap consensus tree for EIF5 corresponded to topology 9 (rheas sister to other paleognaths; 41% support).

[e] The extended majority rule bootstrap consensus tree for GNB2L1 corresponded to topology 2 (ostrich sister to other paleognaths; 49% support).

[f] Topologies with a polytomy are indicated using a dash. The relevant topologies corresponded to: HNRPA2B1 (Ostrich,Rheas,(Emu,Tinamous)); PHB (Emu,Rheas,(Ostrich,Tinamous)); BMP5 — major lineages unresolved; PUM2 (Ostrich,(Emu,Rheas,Tinamous)).

[g] Extended majority rule bootstrap consensus trees for NAT15 and TCP1 conflict with the ML topology (topology 11 for NAT15 and topology 9 for TCP1). In both cases the bootstrap support for the conflicting groups was <50%.

[h] Topologies with a missing lineage are indicated using a dash. The relevant topologies for both TXNDC12 and PALLD correspond to (Ostrich,(Rheas,Tinamous)).

[i] The strict consensus of three ML trees for NUSAP1 is (Emu,Tinamous,(Ostrich,Rheas)); the extended majority rule bootstrap consensus tree for NUSAP1 corresponded to topology 15 (tinamous sister to other paleognaths; 21% support).

[i] The ML topology for NGF was (Passerines,Tinamous,(Ostrich,(Galloanserae,(Emu,Rheas)))); this topology does not support paleognath monophyly. Note, however, that the relevant branch (separating passerines and tinamous from other taxa) only has 51% bootstrap support.

Table S4. Joint posterior probabilities for each topology (discordance prior α = 3).

Locus	Topology 1 (40 loci) [b]	Topology 2 (40 loci) [b]	Topology 1 (61 loci) [c]	Topology 2 (61 loci) [c]
EPIC				
ACTB	0.261	0.739	0.211	0.789
ARNTL	0.83	0.17	0.797	0.203
CALB1	0.784	0.216	0.732	0.268
CHMP5	0.995	0.005	0.993	0.007
CIZ1	0.122	0.878	0.094	0.906
CLOCK	0.806	0.194	0.774	0.226
CRAT	0.034	0.966	0.025	0.975
CSDE1	0.981	0.019	0.976	0.024
CSNK1E	0.469	0.531	0.344	0.656
DDX5[d]	0.283	0.697	0.013	0.987
EIF5[d]	0.202	0.796	0.123	0.877
ENO1	0.666	0.334	0.617	0.383
ETS2	0.571	0.429	0.527	0.473
GAPDH	0.676	0.324	0.553	0.447
GNB2L1	0.018	0.982	0.01	0.99
GRIA2	0.511	0.489	0.446	0.554
HNRPA2B1	0.477	0.522	0.407	0.593
KCNQ5	0.839	0.161	0.801	0.199
NAT15	0.973	0.027	0.98	0.02
PARK7	0.39	0.61	0.335	0.665
PAXIP1[d]	0.249	0.729	0.287	0.713
PER2	0.695	0.305	0.65	0.35
PHB	0.795	0.205	0.612	0.388
PSMA2	0.264	0.736	0.227	0.773
SEPT2	0.404	0.596	0.349	0.651
SFRS3	0.094	0.906	0.069	0.931
TCP1	0.844	0.156	0.762	0.238
TXNDC12	0.776	0.224	0.733	0.267
VDAC2[d]	0.885	0.111	0.877	0.123
VIM	0.98	0.02	0.977	0.023
Anonymous				
BMP5	0.56	0.44	0.512	0.488
Intergenic 1	0.365	0.635	0.321	0.679
Intergenic 2	0.821	0.179	0.789	0.211
Intergenic 3	0.583	0.417	0.565	0.435
Intergenic 4	0.939	0.061	0.954	0.046
NUSAP1	0.644	0.356	0.436	0.564
PALLD	0.864	0.136	0.83	0.17
PUM2	0.483	0.517	0.428	0.572
SLC25A21	0.575	0.425	0.55	0.45
TTN	0.963	0.037	0.955	0.045

[a] Continued on the next page along with all footnotes.

Table S4. Continued.

Locus	Topology 1 (61 loci) [c]	Topology 2 (61 loci) [c]
Published		
ALDOB	0.334	0.666
BDNF	0.537	0.463
CLTC	0.982	0.018
CLTCL1	0.544	0.456
CRYAA	0.79	0.21
EEF2	0.79	0.21
EGR1	0.199	0.801
FGB	0.792	0.208
GH1	0.188	0.812
HMGN2	0.748	0.252
IRF2	0.102	0.898
MB	0.45	0.55
MUSK	0.316	0.684
MYC	0.668	0.332
NGF	0.964	0.036
NTF3	0.004	0.996
PCBD1	0.098	0.902
RHO	0.4	0.6
TGFB2	0.596	0.404
TPM	0.668	0.332
Mitochondrion	1	0

[b] Results of the BUCKy analysis the 40 novel loci and a gene tree discordance prior (α) of 3 (the α prior was based upon the 20 published loci and mitochondrial genome; see text for details).

[c] Results of the BUCKy analysis the total evidence analyses (60 nuclear loci and mitochondrial genomes). For consistency, an α prior of 3 was used. However, these joint posteriors were not sensitive to the value of the α prior (runs were conducted using $\alpha = 10$ and $\alpha = 50$ and almost identical values were obtained).

[d] The joint posteriors for topologies 1 and 2 was < 0.999 for DDX, EIF5, PAXIP1, and VDAC2 for the 40 gene analysis (although the joint posterior for these gene tree topologies was >0.95). The joint posterior for gene tree topologies1 and 2 was >0.999 in all other cases.

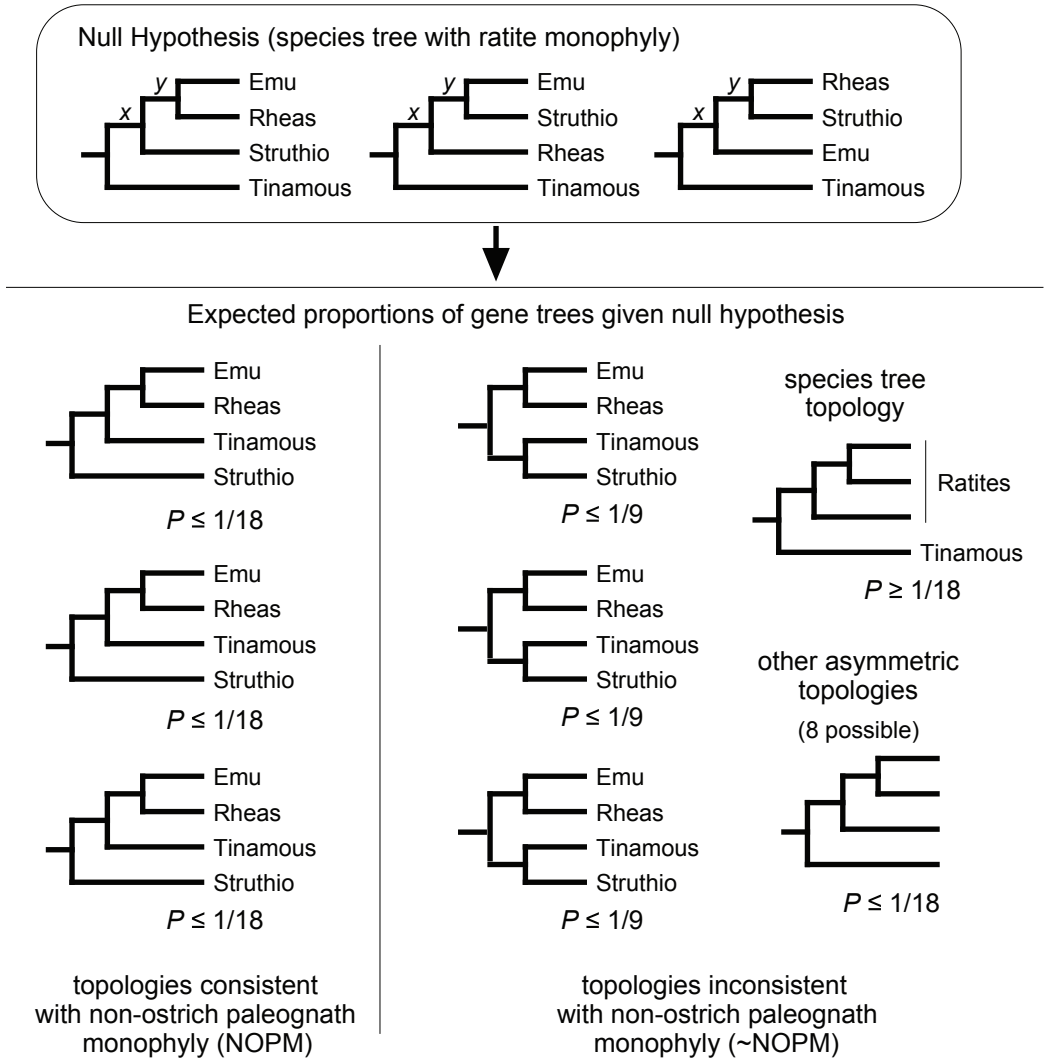

Figure S1. Expected proportions of gene tree topologies used for χ^2 test given a null hypothesis that the species tree lacks non-ostrich paleognath monophyly (~NOPM). The species trees consistent with the traditional hypothesis of ratite monophyly are shown as an example. If the species tree lacks NOPM the probability of observing a specific gene tree topology consistent with NOPM is $\leq$ 1/18 (the exact probability for specific topologies depends upon the species tree branch lengths *x* and *y*; Degnan and Rosenberg 2006). Thus, it is clear that gene trees with NOPM are not favored when the species trees lacks the non-ostrich paleognath clade, even when *x* and *y* are very short. In fact, when *x* and *y* are very short (i.e., when the species tree is in the anomaly zone) NOPM gene trees are not expected to be the most common even when the species tree has a non-ostrich paleognath clade. Thus, finding that the number of gene tree topologies consistent with NOPM is significantly greater than 1/6 is most consistent with a species tree that has NOPM (i.e., the species tree has a non-ostrich paleognath clade).

REFERENCES:

Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genetics. 2:e68.