

APPENDIX: Identification of the ‘Magnificent Seven’ and ‘Indicator Clades’

The goal of this study is to evaluate a set of hypotheses to explain the differences between the Jarvis TENT and Prum trees (Fig. 1). We used two approaches to test these hypotheses. First, we examined the presence of two different sets of indicator clades. Second, we clustered trees based on their topologies. This appendix focuses on the first approach.

The idea of using indicator clades that can be defined *a priori* has been used in a number of phylogenetic studies of Neoaves to assess convergence of different datasets on a single topology (e.g., Gibb et al. 2007; Pratt et al. 2009). This same approach can be used to assess whether different taxon samples or data types converge on specific topologies (i.e., a topology resembling the Jarvis TENT on the one hand or the Prum tree on the other). However, before the indicator clade approach can be used one must assess the degrees of freedom available for Neoaves phylogeny. We assert that the base of Neoaves has ten lineages in play, the ‘magnificent seven’ superordinal clades and three ‘orphan orders’ (Supplementary Fig. S1a).

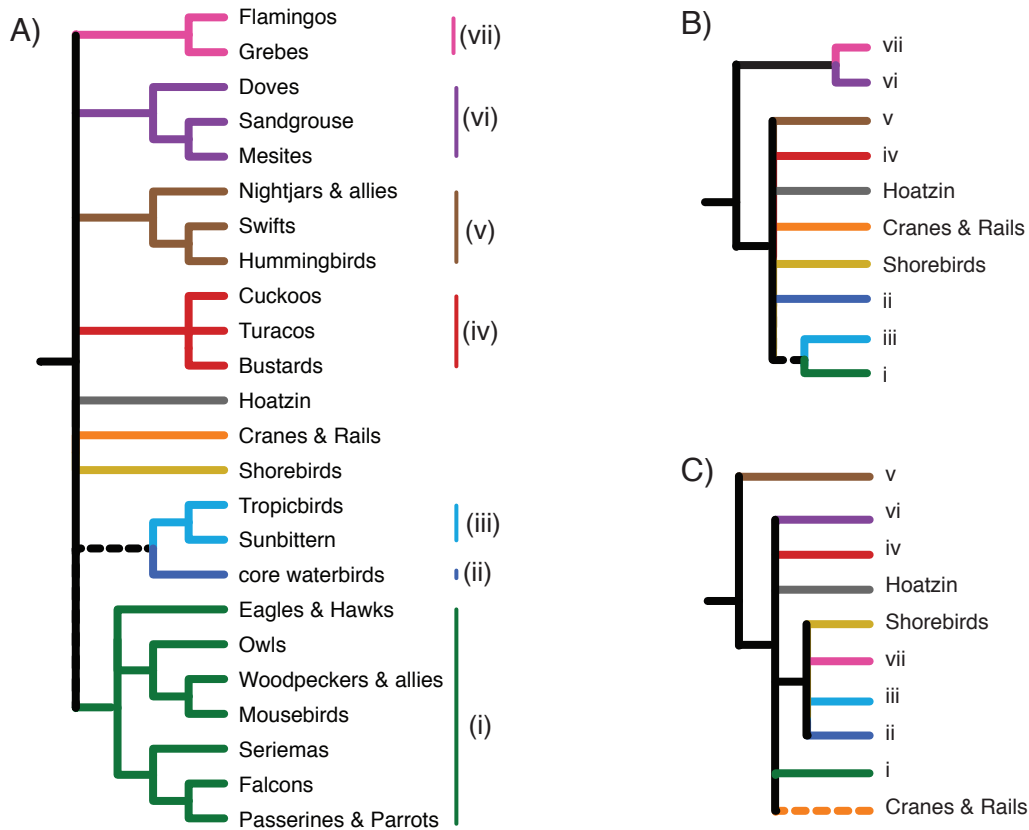


FIGURE S1. The magnificent seven and the indicator clades (J1/J2 or P1/P2). (A) A strict consensus of the Jarvis TENT and the Prum tree suggest that deep Neoaves phylogeny can be viewed as a rooted nine-taxon tree. We expand the set of plausible trees for Neoaves to those consistent with the rooted ten-taxon tree produced by collapsing the dashed branch uniting clades ii and iii. This reflects the observation that the ii + iii clade is not present in the Jarvis intron and UCE trees. (B) We view trees with clades J1 (vi + vii) and J2 (all members of Neoaves except J1) as ‘Jarvis-like’ trees. The Jarvis-like trees correspond to 0.13% of all plausible trees. (C) We view trees with clades P1 (all Neoaves except v) and P2 (ii, iii, vii, and shorebirds) or P2<sub>J</sub> (P2 + cranes & rails) as ‘Prum-like’ trees. The Prum-like trees correspond to 0.08% of all plausible trees.

The magnificent seven correspond to the superordinal clades present in a strict consensus of the Jarvis TENT and the Prum tree. All of the magnificent seven clades also received 100% bootstrap support in the Jarvis TENT. Several previous studies also support monophyly of at least some members of the magnificent seven. Hackett et al. (2008) only recovered five of the magnificent seven (clades i, ii, v, vi, and vii), Kimball et al. (2013) recovered six (clades i, ii, iii, v, vi, and vii), and McCormack et al. (2013) recovered all seven (more accurately, it recovered a tree consistent with all of the magnificent seven since it did not have sufficient taxon sampling to rigorously assess their monophyly).

The Jarvis TENT includes two superordinal clades with 100% bootstrap support that are not found in the Prum tree (Fig. 1a and Supplementary Fig. S1b). We have designated these clades J1 (clades vi and vii) and J2 (clades i, ii, iii, iv, and v) and we consider them to be indicators of a Jarvis-like topology. These clades appear in a number of additional trees, all dominated by non-coding data:

1. **The Jarvis intron tree** (Fig. 1c and Fig. 4a of Jarvis et al. 2014): This tree contains **clades J1 and J2**; it also includes **all of the magnificent seven**. The data used to generate the intron tree is the largest subset of the TENT data (it is 46% of the TENT data).
2. **The Jarvis UCE tree** (Fig. 4b of Jarvis et al. 2014): This tree contains **clade J2**; it also includes **all of the magnificent seven**. The data used to generate the UCE tree is a subset of the TENT data (it is 22% of the TENT data), but it is completely independent of the intron data.
3. **The Jarvis ‘whole genome tree’** (Fig. S4d of Jarvis et al. (2014): This tree contains **clades J1 and J2**; it also includes **all of the magnificent seven**. This data used to generate this tree includes all data types (both coding and non-coding) because it comprises the 322-million bp that could be aligned to the chicken genome using MULTIZ (Blanchette et al. 2004). The MULTIZ method used to generate breaks up putatively orthologous regions into short segments Jarvis et al. (2014) suggest that multiple sequence alignment within those segments is likely to be inferior to the TENT alignments. Despite these limitations it provides another line of evidence that corroborates J1 and J2.

The Prum tree actually shared two interesting features with several of the Jarvis coding exon trees (Fig. 1b and Supplementary Fig. S1c). Specifically, the Prum tree placed clade v sister to all other of members of Neoaves (clade P1) and a clade that Prum et al. (2015) named Aequorlitorhithes that resembled the ‘aquatic/semiaquatic’ group noted by Jarvis et al. (2014). For simplicity, we call Aequorlitorhithes ‘clade P2’ and the Jarvis aquatic/semiaquatic group clade P2J. The Jarvis et al. (2014) trees used for this comparison were:

1. **The Jarvis c12 exon tree** (Fig. 1d and Fig. 4c of Jarvis et al. 2014): This tree contains **clade P1 and P2<sub>J</sub>**; it also includes **five of the magnificent seven** (clades i, ii, iii, v, and vii). The data used to generate this tree is the coding subset of the TENT data (it is 32% of the TENT data) and it excludes third codon positions.
2. **The Jarvis low-GC variation exon tree** (Fig. 6a of Jarvis et al. 2014): This tree contains **clade P1 and P2<sub>J</sub>**; it also includes **five of the magnificent seven** (clades i, ii, iii, v, and vii). The data used to generate this tree corresponds to the 830 protein coding genes with the most limited GC-content variation (10% of the loci included in the c12 tree). However, this data matrix includes third codon positions so it is not simply a subset of the c12 exon data.

The base of Neoaves is a rooted ten-taxon tree if we assume the magnificent seven are very likely to be monophyletic but we remain agnostic regarding relationships among the magnificent seven and the three orphan orders (i.e., Supplementary Fig. S1a). There are almost 34.5 million rooted ten-taxon trees (Felsenstein 1978) and we view this as the size of the set of plausible trees for Neoaves. The number of plausible trees with either set the indicator clades (i.e., J1 and J2 or P1 and P2/P2<sub>J</sub>) is very small compared to the total number of plausible trees:

1. **Number of trees that contain J1 and J2:** There are 135,135 arrangements of the eight taxa in clade J2 (Felsenstein 1978). Only one arrangement of clade J2 and the two remaining taxa is consistent with J1 monophyly. Thus, there are 135,135 trees that divide Neoaves into clades J1 and J2 (0.39% of the plausible trees).
2. **Number of trees that contain P1 and P2:** There are nine lineages in clade P1 and four of those lineages must themselves form a clade (clade P2). Thus, clade P1 can be viewed as a rooted six-taxon tree (see Fig. S1c) and there are 945 possible six-taxon trees. There are 15 possible arrangements of the P2 taxa. Thus, there are 14,175 possible trees consistent with both P1 and P2 clades (0.04% of the plausible trees).
3. **Number of trees that contain P1 and P2 or P2<sub>J</sub>:** The match between the Prum tree and the Jarvis et al. (2014) exon trees is imperfect because the latter include P2<sub>J</sub>, not P2. P2<sub>J</sub> comprises the four lineages in P2 plus cranes & rails (Gruiformes; dashed line in Fig. S1c). There are *a priori* reasons to view Gruiformes as a potential rogue taxon that can fall anywhere in clade P1, including within clade P2 (Prum et al. 2015). If we simply prune Gruiformes from the tree there will still 15 possible arrangements of the P2 taxa but there will only be 105 possible arrangements of taxa for the remaining P1 taxa. This is a total of 1,575 trees. However, the total number of plausible trees is lower (only 2,027,025) since Gruiformes is excluded from consideration. Ultimately 0.08% of the plausible trees are compatible with clades P1 and P2/P2<sub>J</sub>.

This discussion suggests that the probability that *analyses of a novel data matrix* (e.g., the EB2 data) will contain either set of indicator clades by chance alone is very low.

The significance of observing **clade J3<sub>N</sub>** (dashed in part B of Fig. S1) is debatable. The two Jarvis trees that are based only on non-coding data (the intron and UCE trees) include clade J3<sub>N</sub>. However, 5.9% of the plausible ten-taxon trees are consistent with the presence of any specific two-taxon clade (like clade J3<sub>N</sub>). A slightly larger percentage (7.7%) of the trees compatible with J1 and J2 monophyly also contain clade J3<sub>N</sub>. Thus, as indicated in the main text, recovery of clade J3<sub>N</sub> in trees based on non-coding data is provocative but it is not necessarily significant.

All calculations assume that the magnificent seven can be treated as monophyletic and that the only questions that remain are the relationships among those clades and the three ‘orphan’ orders (shorebirds, cranes & rails, and the hoatzin) (see Supplementary Fig. S1). We note that the EB2 tree, like the Hackett et al. (2008) and Kimball et al. (2013) trees, did not support clade iv. However, failure to recover any of the magnificent seven will not bias analyses toward either set of indicator clades. Indeed, it to only places more lineages ‘in play’ and decreases the likelihood of finding a Jarvis-like (i.e., clades J1 and J2 present) or Prum-like tree (i.e., clades P1 and P2/P2<sub>J</sub> present) tree by chance alone.

## REFERENCES

- Blanchette M., Kent W.J., Riemer C., Elnitski L., Smit A.F.A., Roskin K.M., Baertsch R., Rosenbloom K., Clawson H., Green E.D., Haussler D., Miller W. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*, 2004;14:708-715.
- Felsenstein J. The number of evolutionary trees. *Syst Zool*, 1978;27:27-33.
- Gibb G.C., Kardailsky O., Kimball R.T., Braun E.L., Penny, D. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. *Mol Biol Evol*, 2007;24:269-280.
- Hackett S.J., Kimball R.T., Reddy S., Bowie R.C.K., Braun E.L., Braun M.J., Chojnowski J.L., Cox W.A., Han K.L., Harshman J., Huddleston C.J., Marks B.D., Miglia K.J., Moore W.S., Sheldon F.H., Steadman D.W., Witt C.C., Yuri T. A phylogenomic study of birds reveals their evolutionary history. *Science*, 2008;320:1763-1768.
- Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., da Fonseca R.R., Li J.W., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V., Subramanian S., Gabaldon T., Capella-Gutierrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W.C., Ray D., Green R.E., Bruford M.W., Zhan X.J., Dixon A., Li S.B., Li N., Huang Y.H., Derryberry E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C.V., Lovell P.V., Wirthlin M., Schneider M.P.C., Prosdocimi F., Samaniego J.A., Velazquez A.M.V., Alfaro-Nunez A., Campos P.F., Petersen B., Slicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell A.C., Shapiro B., Xiong Z.J., Zeng Y.L., Liu S.P., Li Z.Y., Liu B.H., Wu K., Xiao J., Yinqi X., Zheng Q.M., Zhang Y., Yang H.M., Wang J., Smeds L., Rheindt F.E., Braun M., Fjeldsa J., Orlando L., Barker F.K., Jonsson K.A., Johnson W., Koepfli K.P., O'Brien S., Haussler D., Ryder O.A., Rahbek C., Willerslev E., Graves G.R., Glenn T.C., McCormack J., Burt D., Ellegren H., Alstrom P., Edwards S.V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T., Jun W., Gilbert M.T.P., Zhang G.J. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 2014;346:1320-1331.
- Kimball R.T., Wang N., Heimer-McGinn V., Ferguson C., Braun E.L. Identifying localized biases in large datasets: A case study using the avian tree of life. *Mol Phylogenet Evol*, 2013;69:1021-1032.
- McCormack J.E., Harvey M.G., Faircloth B.C., Crawford N.G., Glenn T.C., Brumfield R.T. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS ONE* 2013;8:e54848.
- Pratt R.C., Gibb G.C., Morgan-Richards M., Phillips M.J., Hendy M.D., Penny D. Toward resolving deep Neoaves phylogeny: data, signal enhancement, and priors. *Mol Biol Evol*, 2009;26:313-326.
- Prum R.O., Berv J.S., Dornburg A., Field D.J., Townsend J.P., Lemmon E.M., Lemmon A.R. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, 2015;526:569-573.
- Smith J.V., Braun E.L., Kimball R.T. Ratite nonmonophyly: Independent evidence from 40 novel loci. *Syst Biol*, 2013;62:35-49.
- Wang N., Braun E.L., Kimball R.T. Testing hypotheses about the sister group of the passeriformes using an independent 30-locus data set. *Mol Biol Evol*, 2012;29:737-750.