IBIS
BOU international journal of avian science

Ibis (2017), doi: 10.1111/ibi.12464

The phylogeny of the world's bulbuls (Pycnonotidae) inferred using a supermatrix approach

SUBIR B. SHAKYA* & FREDERICK H. SHELDON
Museum of Natural Science and Department of Biological Sciences, Louisiana State University, 119 Foster Hall, Baton Rouge, LA 70803, USA

The bulbuls comprise an ecologically important group of frugivorous, seed-dispersing birds found in Asia and Africa. Although several studies have examined the phylogenetic relationships of subsets of bulbul species, a comprehensive phylogeny of the family Pycnonotidae has hitherto been lacking. We used publicly available sequences generated from previous phylogenetic studies, augmented by new sequences from several unstudied taxa, to create a supermatrix from which to infer the phylogeny of the family. In all, we compared 121 of the 130 bulbul species. Our tree supports the monophyly of the family and comprises an exclusively African and a predominantly Asian clade. Several genera were found not to be monophyletic and we suggest taxonomic changes to provide a more accurate classification based on phylogeny.

Keywords: biogeography, bulbul, phylogeny, Pycnonotidae, supermatrix, taxonomy.

Bulbuls comprise a prominent group of passerines in the tropics and subtropics of the Old World. They are generalists, feeding on a wide variety of fruits and arthropods, and they play a particularly important ecological role as seed dispersers (e.g. Corlett 1998). They are also sometimes common garden birds and are known for their melodious songs. Several species are kept in captivity and some, such as the Straw-headed Bulbul *Pycnonotus zeylanicus*, have suffered severe population declines due to their popularity as cage-birds (Eaton *et al.* 2015). Bulbuls have been introduced in many parts of the world, with species such as the Red-whiskered Bulbul *Pycnonotus jocosus* flourishing in such disparate places such as Florida, Hawaii and coastal Australia.

Currently, 30 genera and 130 species of bulbuls are recognized (Dickinson & Christidis 2014). A diagnostic characteristic of this family is a thin sheet of bone in the operculate nostrils (Fishpool & Tobias 2005). Most species also have drab or dark-coloured fluffy plumage, with characteristically colourful patches of feathers, especially on the throat and abdomen. The family is distributed throughout Africa, the Middle East and most of southern and south-eastern Asia but is absent (barring introductions) from the Americas, Europe and east of Wallacea.

Several molecular studies have examined the phylogenetic relationships of species within the Pycnonotidae (Pasquet *et al.* 2001, Moyle & Marks 2006, Johansson *et al.* 2007, Oliveros & Moyle 2010, Zuccon & Ericson 2010, Fuchs *et al.* 2015). Pasquet *et al.* (2001) used two mitochondrial markers, 12S and 16S, to compare 27 species and demonstrated that the family is divisible into two fundamental clades, one from Africa and another from Asia. Moyle and Marks (2006) compared 57 species and used both mitochondrial and nuclear genes to determine relationships within the family. Like Pasquet *et al.* (2001), they recovered African and Asian clades and noted that a few members of the Asian clade had subsequently dispersed into Africa, including Madagascar (Warren *et al.* 2005, Moyle & Marks 2006). Johansson *et al.* (2007) compared numerous African members of the family using three nuclear genes and ended up splitting several African genera. Most recently, molecular studies have concentrated on phylogenetic relationships within specific groups, such as the Philippine bulbuls (Oliveros & Moyle 2010),

*Corresponding author.
Email: sshaky6@lsu.edu

© 2017 British Ornithologists' Union

the Sundaic bulbuls (Dejtaradol *et al.* 2016), and the genera *Alophoixus* (Fuchs *et al.* 2015) and *Bleda* (Huntley & Voelker 2016). Oliveros and Moyle (2010), for example, redefined the limits of several Asian taxa to encompass monophyletic groups. Despite these advances, the classification of bulbul genera is still in flux. This is especially true in the largest bulbul genus, *Pycnonotus* (41 species), which is almost certainly polyphyletic, although little has been done to revise its classification (Fishpool & Tobias 2005).

Recent studies at higher phylogenetic levels have caused several putative members of the Pycnonotidae to be assigned to other families. Five taxa in the genus *Phyllastrephus* are now considered members of the endemic Malagasy family Bernieridae (Cibois *et al.* 2001, 2010, Gill & Donsker 2015). Similarly, three species of the genus *Nicator* are now generally placed in their own family Nicatoridae (Beresford *et al.* 2005, Johansson *et al.* 2008, Fregin *et al.* 2012). The phylogenetic position of Malia *Malia grata* is still in doubt; one possibility is that it belongs in the Locustellidae (Oliveros *et al.* 2012). It turns out that most of these former 'bulbuls' lack operculate nostrils, strengthening the usefulness of this character in defining the group. Nevertheless, at least one true pycnonotid, the Black-collared Bulbul *Neolestes torquatus*, lacks it (Fishpool & Tobias 2005).

The recent, substantial restructuring of African and Asian bulbul classification demonstrates how molecular studies have helped improve our understanding of bulbul relationships, but it also indicates how much more remains to be done. Limited taxon sampling has severely undermined efforts at phylogenetic reconstruction. So far, most molecular studies of the Pycnonotidae have either sampled broadly at the intergeneric level or focused on specific genera or biogeographical groups. A comprehensive phylogenetic reconstruction of the entire family is needed. Fortunately, DNA sequences from all previous molecular studies are available to serve as a basis for such a study. In addition, tissue samples have become available for a few bulbul taxa that were not included in earlier studies, and further gaps in sampling may be filled using toe-pad DNA from traditional museum specimens. Given these sources of data, it is now possible to infer phylogenetic relationships among a much broader group of bulbuls than ever before, using a supermatrix approach.

Supermatrices constructed from DNA sequences available in public databases, such as GenBank, have been successfully used to reconstruct phylogeny. (e.g. de Queiroz & Gatesy 2007, Thomson & Shaffer 2010). The strengths and weaknesses of the supermatrix approach have been well considered, especially the problems posed by large amounts of missing sequence data (Lemmon *et al.* 2009, Roure *et al.* 2013, Hosner *et al.* 2016, Streicher *et al.* 2016). Proponents of supermatrices recognize that this as an important issue but suggest it is a minor problem given the advantages of greater taxonomic coverage. Moreover, one particularly useful feature of supermatrices is that they can always be enhanced as future studies produce more sequences and improve sampling.

We use DNA sequences from previous phylogenetic projects coupled with new sequences derived from fresh tissues and toe-pads of missing species to generate a more comprehensive tree of relationships within the Pycnonotidae. We compare sequences of 121 of the 130 currently recognized species of bulbul (Dickinson & Christidis 2014), together with individuals of several recently recognized species (i.e. subspecies that have been raised to species level). These comparisons substantially improve our understanding of phylogenetic relationships within the Pycnonotidae and thus help with the formulation of a more accurate classification. The comparisons also provide a view of the efficacy, strengths and weakness of the supermatrix approach to phylogenetic reconstruction.

METHODS

All available sequences of bulbuls (and a few related species) were downloaded from GenBank (http://www.ncbi.nlm.nih.gov/genbank). For each species, we manually selected a single sequence representing each locus, and each sequence-type was binned into a specific folder based on its locus. Whenever possible, we chose sequences from the same individual and used the minimum number of individuals per species. If sequences were from multiple individuals, we made sure they were members of the same subspecies group. For all sequences, we also checked tissue and specimen numbers to ensure correct species assignment of any taxon that has been split recently. We followed Dickinson and Christidis (2014) for most species-level assignments. We also included sequences of taxa designated as subspecies by

© 2017 British Ornithologists' Union

Dickinson and Christidis (2014) that have recently been raised to species status by other taxonomists. Four species, three of which were formerly considered bulbuls, were included as outgroups: Malia *Malia grata*, Long-billed Bernieria *Bernieria madagascariensis*, Yellow-spotted Nicator *Nicator chloris* and Barn Swallow *Hirundo rustica*. Sequences in bins were aligned using MUSCLE (Edgar 2004) and manually checked for anomalous sequences or misidentified loci. The trailing ends of each alignment were trimmed to provide maximum sequence overlap (i.e. each nucleotide site was present in at least 50% of the taxa).

In addition, we sequenced loci of bulbul species not in GenBank (Supporting Information Table S1). Total genomic DNA was extracted from frozen or alcohol-preserved tissue or blood samples using a DNEasy Blood and Tissue Kit (Qiagen, Hilden, Germany) and the manufacturer's protocol. PCR amplifications were performed in 25-μL reactions using Taq DNA Polymerase (New England BioLabs Inc., Ipswich, MA, USA). The primer pairs, L5215 (Hackett 1996) and HTrpC (STRI), L10755 and H11151 (Chesser 1999), and MB-2R and MB-3F (Kimball *et al.* 2009) were used to amplify the loci NADH dehydrogenase subunit 2 (ND2), NADH dehydrogenase subunit 3 (ND3) and Myoglobin-intron 3 (MB-I3), respectively. Thermocycler runs were set to 34 cycles with denaturing temperatures of 95 °C, annealing temperature varying based on the primer pair used, and an extension temperature of 72 °C. The PCR products were visualized in 1% agarose gel stained with SYBR Safe DNA Gel Stain (Invitrogen, Carlsbad, CA, USA). The PCR products were sequenced by Beckman Coulter Genomics (Danvers, MA, USA).

DNA from toepads was extracted using well-established ancient-DNA protocols (Sheldon *et al.* 2012). ND2 was amplified using primers designed for Black-headed Bulbul *Pycnonotus atriceps* (Chua *et al.* 2015). Because several amplification efforts using these primer pairs failed, we used forward primers L5215 (Hackett 1996), and *P. atriceps* F2, F3, F4 and F5 (Chua *et al.* 2015), and reverse primers *Copsychus malabaricus* R1, R2, R3, R4 (Chua *et al.* 2015), and HTrpC (STRI) to amplify the missing fragments.

Including GenBank and newly generated sequences we had seven mitochondrial genes and seven nuclear loci (Table 1). These were aligned using MUSCLE (Edgar 2004). Maximum-likelihood (ML) tree searches were conducted for each locus using RAxML v8 (Stamatakis 2014) to see whether the sequences produced reasonable bulbul trees (i.e. did not include anomalous or paralogous loci). After this check, the loci were concatenated into a single supermatrix using a custom Python script. Substitution models were estimated using PARTITIONFINDER v1.1.1 (Lanfear *et al.* 2012) and partitioned by locus. Mitochondrial loci were further partitioned by codon position. We implemented a greedy search algorithm, specified blocks as described above and used the BIC criterion to find the best partitioning scheme.

ML searches were run on the concatenated supermatrix using RAxML v8 (Stamatakis 2014) implemented through the CIPRES Science Gateway (Miller *et al.* 2010). The supermatrix was partitioned based on the best scheme obtained from PARTITIONFINDER and the GTRGAMMA model was used for each partition. A total of 1000 non-parametric bootstrap replicates were run to obtain the

Table 1. Genetic loci included in the supermatrix analyses. The column-headed 'Taxa' lists the total number of individuals used in the study for each locus. Loci marked with an asterisk were used in the reduced supermatrix.

Loci	Abbreviation	Taxa	Consensus length (bp)
Mitochondrially encoded 12S RNA	12S	34	860
Mitochondrially encoded 16S RNA	16S	35	523
ATP synthase subunit 6	ATP6	17	684
Cytochrome c oxidase subunit I	COI	37	652
Cytochrome b*	CYTB	58	1143
NADH dehydrogenase subunit 2*	ND2	121	1041
NADH dehydrogenase subunit 3*	ND3	106	351
Fibrinogen beta chain intron 5*	FIB-I5	64	588
Fibrinogen beta chain intron 7*	FIB-I7	79	1016
Glyceraldehyde-3-phosphodehydrogenase intron 11	GAPDH-I11	16	347
Myoglobin intron 2*	MB-I2	80	730
Ornithine decarboxylase intron 6 and 7*	ODC-I6/I7	70	801
Recombination activating protein 1	RAG1	17	1467
Transforming growth factor beta 2 intron 5*	TGFβ2-I5	50	596

© 2017 British Ornithologists' Union

branch support values on the best tree generated from RAxML. Bayesian tree reconstruction was performed using MrBayes v3 (Huelsenbeck & Ronquist 2001, Ronquist & Huelsenbeck 2003) implemented through the CIPRES Science Gateway (Miller *et al.* 2010), using the same partitioning scheme as the ML search. Two runs each with four chains were performed for 10 000 000 generations, sampling every 1000 generations. Burn-in of 250 000 was discarded and convergence of runs was checked in Tracer v1.6 (Rambaut *et al.* 2014).

Because of missing and short sequences, and sequences containing numerous unresolved sites, the complete matrix of sequences produced trees with many unresolved nodes. To improve resolution, we maximized species sequence coverage by reducing the number of loci. A smaller matrix was constructed using loci that occurred in at least one-third of the taxa. This reduced supermatrix consisted of three mitochondrial genes and five nuclear loci (Table 1). We also identified 'rogue taxa' (i.e. taxa that move around disproportionately during tree reconstructions, thereby reducing branch resolution) in ML tree searches of the complete supermatrix using *RogueNaRok* (Aberer *et al.* 2013). Several of these 'rogue taxa' were pruned from the dataset before compiling the reduced matrix. Fortunately, some of the 'rogue taxa' were also taxa that already had been pruned from the large matrix because of sequence-quality problems. Using the reduced matrix, we repeated ML and Bayesian tree searches.

RESULTS

Together, the GenBank sequences and the 58 newly generated sequences amounted to 790 sequences representing eight loci: seven mitochondrial genes and seven nuclear gene segments (Tables 1 and S1). In all, 125 species were compared, including 121 species of bulbul representing 26 of 27 genera in the family (*Acritillas* was missing from our study). The sequences used in the analysis are listed in Table S1, and the sequences used per taxon are shown in Figure 1. Taxon coverage for the different loci ranged from 13% (for Glyceraldehyde-3-phosphodehydrogenase intron 11) to 97% (for ND2). The total matrix contained 55.2% missing sequence data.

Gene trees constructed for each locus did not show any unexpected relationships that might have indicated paralogous, misidentified or other sequence problems. The concatenated, eight-locus supermatrix consisted of 10 802 nucleotide sites. Trees generated using ML and Bayesian approaches were highly consistent but most branches in the trees were not well supported (Fig. 1). Although most taxa appear to evolve at similar rates, a few taxa lay on unusually long branches.

In the reduced supermatrix, we included only those sequences occurring in more than one-third of the taxa: three mitochondrial genes (ND2, ND3 and cytb) and five nuclear loci (Fib5, Fib7, MB, ODC and TGF). We also removed four rogue species identified by *RogueNaRok*: Prigogine's Greenbul *Chlorocichla prigoginei*, Joyful Greenbul *Chlorocichla laetissima*, White-bearded Greenbul *Criniger ndussumensis* and Yellow-eared Bulbul *Pycnonotus penicillatus*. Compared with the full matrix, the reduced matrix had a coverage per gene ranging from 40% of the taxa (transforming growth factor (TGF) β-I5) to 97% of the taxa (for ND2). The reduced matrix contained 37% missing sequences. Trees generated from the reduced matrix were very similar to ones generated from the complete supermatrix but had much stronger nodal support (Fig. 2). All phylogenetic inferences below are based on the reduced matrix tree.

Phylogenetic analysis indicates that the Pycnonotidae is monophyletic and the overall tree structure was similar to that found by Moyle and Marks (2006), especially with respect to the existence of African and Asian clades (Fig. 2). Relationships among the African species were similar to those found by Johansson *et al.* (2007). For example, *Calyptocichla serinus*, *Stelgidillas gracilirostris* and *Andropadus importunus* formed a clade that was sister to the rest of the African clade. Similarly, in agreement with Zuccon and Ericson (2010), we also found *Neolestes torquatus* to be a member of the African bulbul clade. Among the other African bulbuls, several groupings are clear. Most African genera were monophyletic. The only exceptions were *Chlorocichla* and *Arizelocichla; Chlorocichla simplex* is sister to *Baeopogon*, not to the other *Chlorocichla* species, and *Arizelocichla montana* is sister to a clade containing *Chlorocichla* and *Baeopogon*. Among the largely Asian clade, two major subdivisions were well supported, one including most of *Pycnonotus* and *Spizixos*, and the other including all remaining Asian genera. As in most previous studies (Moyle & Marks 2006, Oliveros & Moyle 2010), we found

© 2017 British Ornithologists' Union

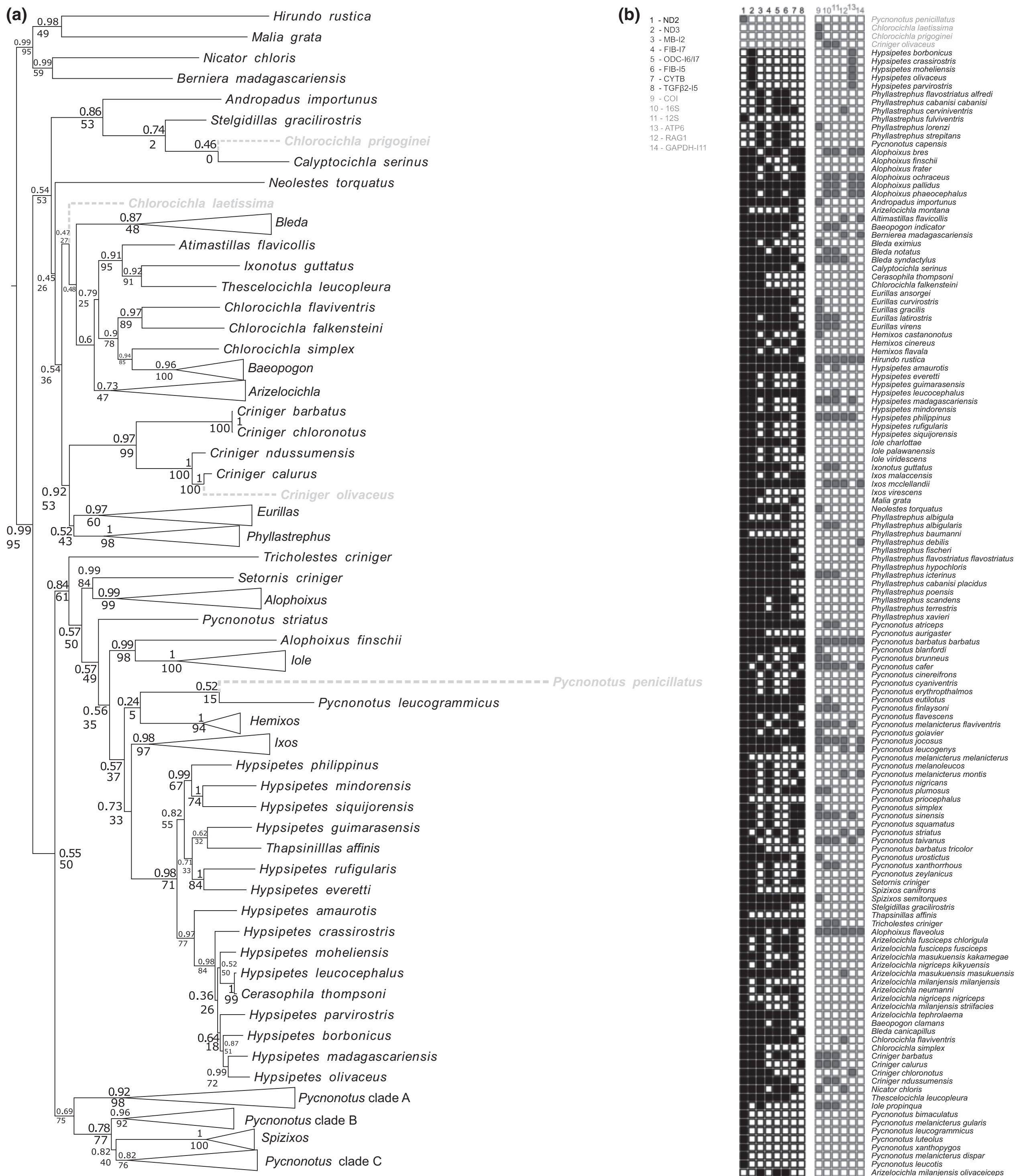

Figure 1. (a) Bayesian tree generated from the complete supermatrix with posterior probability and bootstrap support indicated by numbers above and below branches, respectively. Several monophyletic groups are collapsed into triangles for easier visualization. Grey dotted branches respresent 'rogue taxa' identified by *RogueNaRok* and subsequently pruned from the complete supermatrix to make the reduced supermatrix. (b) Data completeness matrix showing presence/absence of loci per taxon used in our analyses. Filled boxes indicate sequences used. Grey boxes and labels indicate loci or taxa pruned from our complete supermatrix to make the reduced supermatrix. See Table S1 for GenBank accession numbers of each locus.

© 2017 British Ornithologists' Union

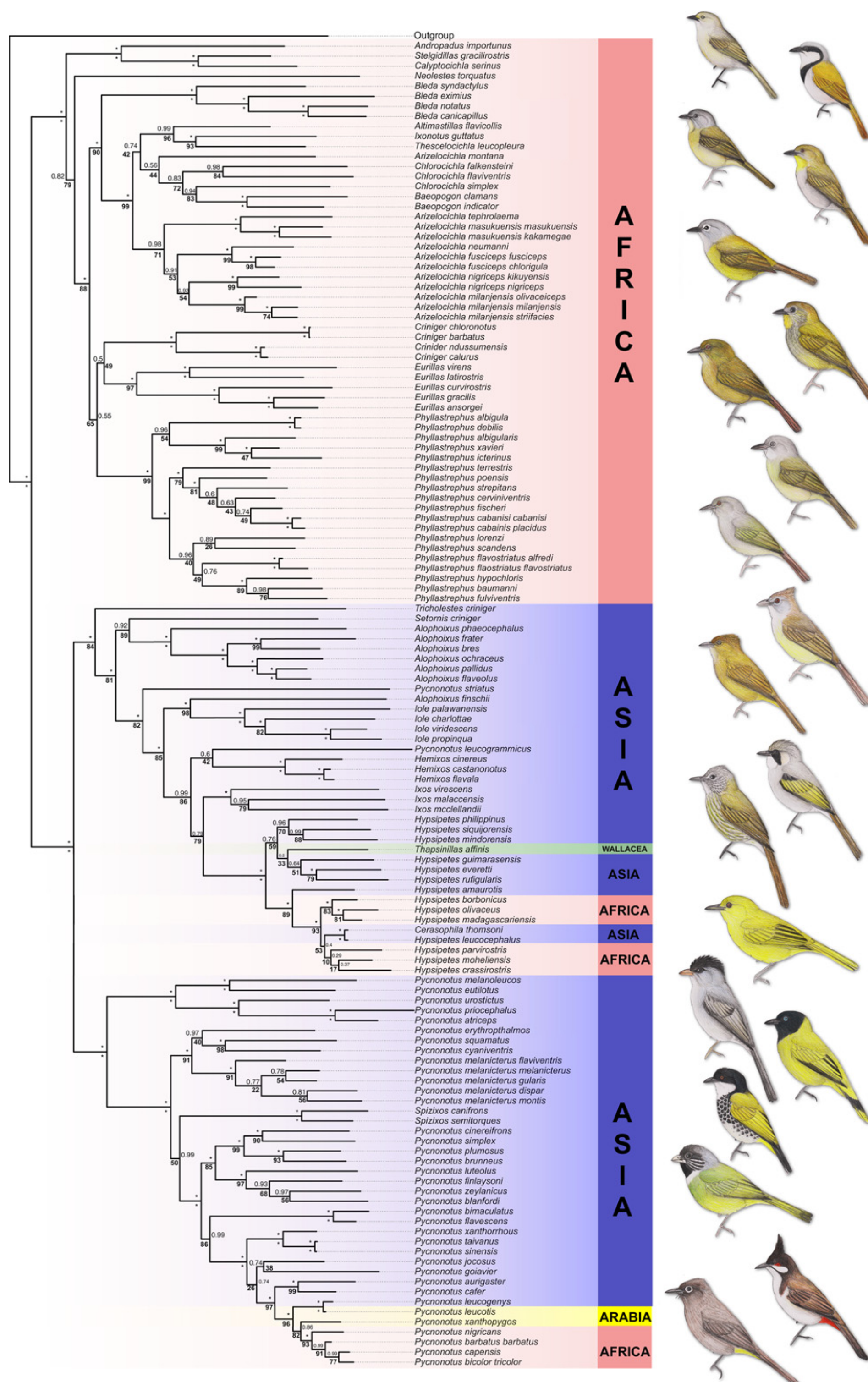

Figure 2. Phylogenetic tree generated from the reduced supermatrix. Posterior probability and bootstrap support are indicated by numbers above and below branches, respectively. Asterisks represent a posterior probability of 1.0 or bootstrap support of 100%. Continental regions where species occur are shown to the right of the taxon name. Illustrations of exemplar taxa are shown to the right (illustrated by Subir B. Shakya).

© 2017 British Ornithologists' Union

Pycnonotus to be polyphyletic. Most *Pycnonotus* taxa formed a clade with *Spizixos* (finchbills) embedded within it.

Several species that have never been compared using molecular methods were included in our study, and we summarize their relationships here. Within *Chlorocichla*, *Chlorocichla falkensteini* was sister to *Chlorocichla flaviventris*, but their congener *Chlorocichla simplex* was sister to *Baeopogon*, albeit with low support. In *Bleda*, *Bleda notatus* was sister to *Bleda canicapillus*, which in turn was sister to *Bleda eximius* and then *Bleda syndactylus*. *Criniger chloronotus* was sister to *Criniger barbatus*, and *Criniger calurus* was sister to *Criniger chloronotus*. Two species of *Phyllastrephus* that have not been included in previous molecular studies, *Phyllastrephus fulviventris* and *Phyllastrephus baumanni*, were sister species and together formed the sister group of *Phyllastrephus hypochloris*. Among the Asian species, *Alophoixus finschii* was not closely related to other *Alophoixus* species, but instead was sister to *Iole*. Additionally, *Iole viridescens* was sister to *Iole propinqua*. *Ixos virescens* was sister to a clade containing *Ixos malaccensis* and *Ixos mcclellandii*. *Hypsipetes* was not monophyletic, with *Thapsinillas affinis* embedded within it, closest to *Hypsipetes guimarasensis*. *Cerasophila thompsoni* was also embedded within *Hypsipetes* and was sister to *Hypsipetes leucocephalus*. Among *Pycnonotus* species, *Pycnonotus striatus* was sister to the clade containing *Iole*, *Hemixos*, *Ixos* and *Hypsipetes*. *Pycnonotus leucogrammicus* was sister to *Hemixos*, albeit with low support. *Pycnonotus priocephlaus* was sister to *Pycnonotus atriceps*. *Pycnonotus squamatus* was sister to *Pycnonotus cyaniventris*. *Spizixos canifrons* was sister to *Spizixos semitorques*. Among the subspecies of *Pycnonotus barbatus*, *Pycnonotus barbatus tricolor* was sister to *Pycnonotus capensis*, and this clade was sister to *Pycnonotus barbatus barbatus*.

DISCUSSION

Supermatrix issues

We used newly generated and GenBank sequences to build a supermatrix to infer the phylogeny of the Pycnonotidae. Initially, our approach was to compile a matrix of all loci from previous studies and to supplement these data with new sequences, either from fresh tissues that had recently become available (e.g. *Pycnonotus squamatus*, *Alophoixus finschii* and *Spizixos semitorques*) or from toe-pads of particularly difficult-to-obtain taxa (e.g. *P. leucogrammicus* and *Thapsinillas affinis*). However, when all old and new data were compiled into a matrix, the resulting tree was poorly resolved (Fig. 1). Its lack of resolution could have been due to missing loci, small sequence sizes, pseudogenes or a variety of other issues. Removal of loci that were represented in less than one-third of the sampled taxa greatly improved resolution of the tree (Fig. 2).

There are several disadvantages of building supermatrices using sequences from public databases such as GenBank. These are primarily inconsistency in sampled loci and low sequence quality (e.g. sequences that consist only of subsections of commonly used loci, that include pseudogenes or have incorrectly identified bases), although the information accompanying sequences from public databases is not necessarily correct. Regarding the first issue, while centralized repositories assemble a variety of loci of variable quality, no standardized set of loci is used in all phylogenetic studies. Thus, some of the available sequences, such as ND2, are widely represented in phylogenetic studies (and so are available for many taxa), whereas others are under-represented (only available for a few taxa). If commonly used sequences are available for all or a majority of the taxa, then the supermatrix approach usually produces a well-resolved tree (Wiens 2003, Wiens & Morrill 2011, Jiang *et al.* 2014). Thorough coverage of one or two genes, however, was not a problem in our study; the problem was under-representation as exemplified by *Chlorocichla prigoginei* and *C. laetissima*. The only gene available for these species was cytochrome oxidase I (COI). Although well known as the barcode gene (Austerlitz *et al.* 2009), COI is not commonly used in phylogenetic studies of birds. Moreover, the COI sequences from these two species were short (< 300 bp). Thus, when these species were included in the supermatrix, they appeared in several different African clades during tree construction, causing low support at several nodes. *RogueNaRak* identified these two species as the top 'rogue taxa' and their removal from the supermatrix greatly improved nodal support values.

The second issue, that of low-quality sequences, is usually the result of sequences generated from degraded DNA in toe-pads or mistaken amplification and sequencing of pseudogenes. Low

© 2017 British Ornithologists' Union

sequence quality can affect nodal support as adversely as inconsistent sequence sampling. For example, the *Pycnonotus penicillatus* sequence we used was generated from a toe-pad of a 60-year-old specimen. The sequence was short (300 bp) and contained several unidentified nucleotides. Its removal greatly improved support of several nodes in the Asian clade. Other clades with low nodal support also contained taxa represented by short sequences (< 200 bp). These included the Indian Ocean *Hypsipetes* group (*Hypsipetes parvirostris*, *Hypsipetes moheliensis* and *Hypsipetes crassirostris*) and the *Pycnonotus melanicterus* complex (*Pycnonotus melanicterus melanicterus*, *Pycnonotus melanicterus flaviventris*, *Pycnonotus melanicterus gularis*, *Pycnonotus melanicterus montis* and *Pycnonotus melanicterus dispar*).

A final consideration in using supermatrices constructed from public databases is the influence of locus sampling on branch lengths. Missing data is known to affect branch lengths (Darriba *et al.* 2016). Similarly, sequences of a locus compiled for a single species may, in fact, be derived from different individuals representing geographically disparate populations. Branch lengths of trees generated from such heterogeneous datasets do not accurately reflect the evolutionary history of a species, but rather represent sort of an average for the composite 'species'. Although resulting trees may reflect the appropriate branching position of a composite species, the branch length of that 'species' must not be considered a reliable parameter in analyses that require accurate tree dimensions.

Biogeography

As in previous studies, we recovered two distinct groups of bulbuls: one restricted to Africa and the other consisting mainly of Asian taxa. Two Asian genera, *Hypsipetes* and *Pycnonotus*, have expanded into Africa quite recently (Fig. 2). *Hypsipetes* appears to have reached the islands of east Africa, including Madagascar, by island-hopping across the Indian Ocean (Warren *et al.* 2005). Its route is evidenced by several Indian Ocean populations of Black Bulbuls (Fig. 3). If we interpret our tree literally, this cross-ocean dispersal happened either in two waves or in a single event followed by reinvasion of mainland Asia. The other dispersal event, by *Pycnonotus* bulbuls, seems to have occurred via the Arabian Peninsula.

Hypsipetes appears to be a highly vagile group, with a predisposition for colonizing remote islands. With the exception of two mainland forms, *H. leucocephalus* (including *H. leucocephalus ganeesa*) and *Cerasophila thompsoni* (which is embedded in *Hypsipetes*), all other members of the genus are endemic to islands (Fig. 3). They have reached (and are the only bulbul species present on) the islands of east Africa, including Comoros, Madagascar, Seychelles, Reunion, Aldabra and Moheli. They have also colonized islands in the Philippines and Japan. Intriguingly, our comparisons demonstrated that an enigmatic island genus, *Thapsinillas*, endemic to the Moluccas, in fact belongs in *Hypsipetes*.

Taxonomy

By including African taxa not previously compared by Johansson *et al.* (2007), we discovered potential relationships that help to improve bulbul classification. However, the availability of only short sequences (100–200 bp) for some key species, such as C. *prigoginei* and C. *laetissima*, might have caused their inaccurate placement in our tree. Without better sampling, and longer and reliable sequences for taxa in the African clade, it is premature for us to suggest taxonomic changes in this group. Such improvements will have to wait for thorough investigations of individual groups, such as that by Huntley and Voelker (2016), who compared sequences of many individuals in all *Bleda* species and suggested raising *Bleda notatus ugandae* to species status.

Most of our taxonomically useful discoveries pertain to Asian taxa. We compared 35 species of *Pycnonotus* and five subspecies of *P. melanicterus* that recently have been raised to full species (Rasmussen & Anderton 2005). *Pycnonotus* and *Spizixos* form a large monophyletic group. However, three species of '*Pycnonotus*' are not members of this clade. The Striped Bulbul *Pycnonotus striatus* of the Himalayas is sister to a large clade comprising *Alophoixus finschii*, *Iole*, *Hemixos*, *Ixos*, *Thapsinillas*, *Hypsipetes* and *Cerasophila* bulbuls. We recommend resurrecting the genus *Alcurus* Hodgson, 1844 for this species and calling it *Alcurus striatus*. The Cream-striped Bulbul *P. leucogrammicus* is sister to *Hemixos* and could warrant its own genus on morphological grounds (Fishpool & Tobias 2005). However, to emphasize its phylogenetic position we suggest applying the generic name *Hemixos* to it. We were not able to

© 2017 British Ornithologists' Union

Figure 3. Ranges of *Hypsipetes* species (including *Cerasophila* and *Thapsinillas*), showing their tendency to inhabit islands. Ranges were estimated from accounts in Fishpool and Tobias (2005), Dickinson and Christidis (2014), and Gill and Donsker (2015).

obtain good quality sequences for the Yellow-eared Bulbul *P. penicillatus*, and it was excluded from the final matrix. However, on the basis of its short DNA fragment, this species does not appear to belong in *Pycnonotus*.

At the highest level, *Pycnonotus* may be divided into three groups (indicated by letters in Fig. 1). The members of Clade A were previously revised by Dickinson and Christidis (2014), who placed *P. atriceps*, *Pycnonotus priocephalus* and *Pycnonotus fuscoflavens* (not sampled in our study) in *Brachypodius*; *Pycnonotus urostictus* in *Poliolophus*; *Pycnonotus eutilotus* in *Euptilotus* and *Pycnonotus melanoleucos* in *Microtarsus*. For Clade B, Fishpool and Tobias (2005) suggested placing *P. melanicterus* in *Rubigula* and splitting *P. melanicterus* into five species, *P. melanicterus*, *P. montis*, *P. flaviventris*, *P. gularis* and *P. dispar* (also see Rasmussen & Anderton 2005, Collar & Pilgrim 2007). Fishpool and Tobias (2005) also suggest placing *P. squamatus* and *P. cyaniventris* in *Ixidea*. Our study indicated that *P. erythropthalmos* also belongs in *Ixidea*. Only the species in Clade C should remain in *Pycnonotus*. Within this clade,

© 2017 British Ornithologists' Union

we found that *P. barbatus tricolor* is sister to *P. capensis* not *P. b. barbatus*. *P. b. tricolor*, along with *P. b. somaliensis* and *P. b. dodsoni*, which were not included in our study, are sometimes raised to full species (Fishpool & Tobias 2005, Turner & Pearson 2015). Such an arrangement, at least for *P. b. tricolor*, seems warranted based on our study.

Provided these generic name changes are adopted for *Pycnonotus*, the name *Spizixos* may be retained for the two species of finchbill. The remarkable, long-term misplacement of *Spizixos* is because of the oddly shaped bill of this species, which is adapted to seed-eating. Such plastic characters are not expected to be useful in phylogenetic reconstruction (McCracken & Sheldon 1998).

Finsch's Bulbul *Alophoixus finschii* and several similar Asian bulbuls (e.g. *A. bres*, *A. phaeocephalus*, *A. frater*, *A. ochraceus* and *A. pallidus*) were originally placed in the genus *Criniger*. However, Pasquet *et al.* (2001) showed that African *Criniger* species formed a group distinct from Asian taxa. The Asian forms, including *A. finschii*, were then reclassified as *Alophoixus*. *Alophoixus finschii* was moved to this genus even though it does not resemble the other *Alophoixus* bulbuls. Our study suggests that *A. finschii* is most closely related to the similar-looking *Iole* bulbuls and should be called *Iole finschii*.

The only bulbul found east of Wallace's line is the Golden Bulbul *Thapsinillas affinis*. This bird has a messy taxonomic history, having at various times been synonymized with *Criniger*, *Alophoixus* and *Ixos* (Fishpool & Tobias 2005, Collar *et al.* 2013). Our study found it to be a member of *Hypsipetes*. *Hypsipetes* species have colonized many islands in the Indian Ocean (Fig. 3), so discovering that the island-dwelling Golden Bulbul of Wallacea is a part of the *Hypsipetes* clade is not surprising. Another species, White-headed Bulbul *Cerasophila thompsoni* from Indochina, is also a member of *Hypsipetes*.

Unfortunately, we were not able to include Yellow-browed Bulbul *Acritillas indica* of South India and Sri Lanka in our study. *Acritillas* is a monotypic genus previously considered allied with *Criniger* and *Iole*, but its unique morphology and nest structure led Dickinson and Gregory (2002) to move it to its own genus. The phylogenetic position of this species remains unknown.

We thank colleagues at Louisiana State University Museum of Natural Science (LSUMNS), University of Kansas Biodiversity Institute and Natural History Museum (KU), Yale Peabody Museum (YPM), Field Museum of Natural History (FMNH), and California Academy of Sciences (CAS) for providing tissues and toe-pads. We thank Dr Martin Päckert and Sontaya Manawatthana for providing sequences from their studies. We also acknowledge the help of Dr Rebecca Kimball and two anonymous reviewers for insightful comments on improving the manuscript. Funding for lab work at LSU was provided by the NSF grant DEB 1241059.

REFERENCES

Aberer, A.J., Krompass, D. & Stamatakis, A. 2013. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Syst. Biol.* **62**: 162–166.

Austerlitz, F., David, O., Schaeffer, B., Bleakley, K., Olteanu, M., Leblois, R., Veuille, M. & Laredo, C. 2009. DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. *BMC Bioinformatics* **10** (Suppl 14): S10.

Beresford, P., Barker, F.K., Ryan, P.G. & Crowe, T.M. 2005. African endemics span the tree of songbirds (Passeri): molecular systematics of several evolutionary 'enigmas'. *Proc. R. Soc. B Biol. Sci.* **272**: 849–858.

Chesser, R. 1999. Molecular systematics of the Rhinocryptid genus *Pteroptochos*. *Condor* **101**: 439–446.

Chua, V.L., Phillipps, Q., Lim, H.C., Taylor, S.S., Gawin, D.F., Rahman, A., Moyle, R.G. & Sheldon, F.H. 2015. Phylogeography of three endemic birds of Maratua Island, a potential archive of Bornean biogeography. *Raffles Bull. Zool.* **63**: 259–269.

Cibois, A., Slikas, B., Schulenberg, T.S. & Pasquet, E. 2001. An endemic radiation of Malagasy songbirds is revealed by mitochondrial DNA sequence data. *Evolution* **55**: 1198–1206.

Cibois, A., David, N., Gregory, S.M.S. & Pasquet, E. 2010. Bernieridae (Aves: Passeriformes): a family-group name for the Malagasy Sylvoid radiation. *Zootaxa* **2554**: 65–68.

Collar, N.J. & Pilgrim, J.D. 2007. Species-level changes proposed for Asian birds, 2005–2006. *BirdingASIA* **8**: 14–30.

Collar, N.J., Eaton, J.A. & Hutchinson, R.O. 2013. Species limits in the Golden Bulbul *Alophoixus (Thapsinillas) affinis* complex. *Forktail* **29**: 19–24.

Corlett, R.T. 1998. Frugivory and seed dispersal by vertebrates in the Oriental (Indomalayan) region. *Biol. Rev.* **73**: 413–448.

Darriba, D., Weiß, M. & Stamatakis, A. 2016. Prediction of missing sequences and branch lengths in phylogenomic data. *Bioinformatics* **32**: 1331–1337.

Dejtaradol, A., Renner, S.C., Karapan, S., Bates, P.J.J., Moyle, R.G. & Päckert, M. 2016. Indochinese-Sundaic faunal transition and phylogeographical divides north of the Isthmus of Kra in Southeast Asian Bulbuls (Aves: Pycnonotidae). *J. Biogeogr.* **43**: 471–483.

© 2017 British Ornithologists' Union

Dickinson, E.C. & Christidis, L. 2014. *The Howard and Moore Complete Checklist of the Birds of the World 4th edn, Vol. 2. Passerines*. Eastbourne, UK: Aves Press.

Dickinson, E.C. & Gregory, S. 2002. Systematic notes on Asian birds. 24. On the priority of the name *Hypsipetes* Vigors, 1831, and the division of the broad genus of that name. *Zool. Verh.* **340**: 75–92.

Eaton, J.A., Shepherd, C.R., Rheindt, F.E., Harris, J.B.C., van Balen, B., Wilcove, D.S. & Collar, N.J. 2015. Trade-driven extinctions and near-extinctions of avian taxa in Sundaic Indonesia. *Forktail* **31**: 1–12.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Res.* **32**: 1792–1797.

Fishpool, L.D.C. & Tobias, J.A. 2005. Family Pycnonotidae. In del Hoyo, J., Elliott, A. & Sargatal, J. (eds) *Handbook of the Birds of the World*, Vol. 10: 124–251. Barcelona: Lynx Edicions/Birdlife International.

Fregin, S., Haase, M., Olsson, U. & Alström, P. 2012. New insights into family relationships within the avian superfamily Sylvioidea (Passeriformes) based on seven molecular markers. *BMC Evol. Biol.* **12**: 157.

Fuchs, J., Ericson, P.G.P., Bonillo, C., Couloux, A. & Pasquet, E. 2015. The complex phylogeography of the Indo-Malayan *Alophoixus* Bulbuls with the description of a putative new ring species complex. *Mol. Ecol.* **24**: 5460–5471.

Gill, F. & Donsker, D. 2015. *IOC World Bird List (v5.4)*. doi: 10.14344/IOC.ML.5.4

Hackett, S. 1996. Molecular phylogenetics and biogeography of Tanagers in the genus *Ramphocelus* (Aves). *Mol. Phylogenet. Evol.* **5**: 368–382.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L. & Kimball, R.T. 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the Landfowl (Aves: Galliformes). *Mol. Biol. Evol.* **33**: 1110–1125.

Huelsenbeck, J.P. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**: 754–755.

Huntley, J.W. & Voelker, G. 2016. Cryptic diversity in Afro-tropical lowland forests: the systematics and biogeography of the avian genus *Bleda*. *Mol. Phylogenet. Evol.* **99**: 297–308.

Jiang, W., Chen, S.-Y., Wang, H., Li, D.-Z. & Wiens, J.J. 2014. Should genes with missing data be excluded from phylogenetic analyses? *Mol. Phylogenet. Evol.* **80**: 308–318.

Johansson, U.S., Fjeldså, J., Lokugalappatti, L.G.S. & Bowie, R.C.K. 2007. A nuclear DNA phylogeny and proposed taxonomic revision of African Greenbuls (Aves, Passeriformes, Pycnonotidae). *Zool. Scr.* **36**: 417–427.

Johansson, U.S., Fjeldså, J. & Bowie, R.C.K. 2008. Phylogenetic relationships within Passerida (Aves: Passeriformes): a review and a new molecular phylogeny based on three nuclear intron markers. *Mol. Phylogenet. Evol.* **48**: 858–876.

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., Witti, C.C. & Yuri, T. 2009. A well-tested set of primers to amplify regions spread across the avian genome. *Mol. Phylogenet. Evol.* **50**: 654–660.

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

Lemmon, A.R., Brown, J.M., Stanger-Hall, K. & Lemmon, E.M. 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst. Biol.* **58**: 130–145.

McCracken, K.G. & Sheldon, F.H. 1998. Molecular and osteological heron phylogenies: sources of incongruence. *Auk* **115**: 127–141.

Miller, M.A., Pfeiffer, W. & Schwarz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proc. Gatew. Comput. Environ. Work. (GCE), 14 November 2010, New Orleans, LA: 1–8.

Moyle, R.G. & Marks, B.D. 2006. Phylogenetic relationships of the Bulbuls (Aves: Pycnonotidae) based on mitochondrial and nuclear DNA sequence data. *Mol. Phylogenet. Evol.* **40**: 687–695.

Oliveros, C.H. & Moyle, R.G. 2010. Origin and diversification of Philippine Bulbuls. *Mol. Phylogenet. Evol.* **54**: 822–832.

Oliveros, C.H., Reddy, S. & Moyle, R.G. 2012. The phylogenetic position of some Philippine ‘babblers’ spans the muscicapoid and sylvioid bird radiations. *Mol. Phylogenet. Evol.* **65**: 799–804.

Pasquet, E., Han, L.-X., Khobkhet, O. & Cibois, A. 2001. Towards a molecular systematics of the genus Criniger, and a preliminary phylogeny of the Bulbuls (Aves, Passeriformes, Pycnonotidae). *Zoosystema* **23**: 857–863.

de Queiroz, A. & Gatesy, J. 2007. The supermatrix approach to systematics. *Trends Ecol. Evol.* **22**: 34–41.

Rambaut, A., Suchard, M.A., Xie, D. & Drummond, A.J. 2014. *Tracer v1.6*. http://beast.bio.ed.ac.uk/Tracer

Rasmussen, P. & Anderton, J.C. 2005. *Birds of South Asia: The Ripley Guide*. Washington, DC: Smithsonian Institute.

Ronquist, F. & Huelsenbeck, J.P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Roure, B., Baurain, D. & Philippe, H. 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Mol. Biol. Evol.* **30**: 197–214.

Sheldon, F.H., Oliveros, C.H., Taylor, S.S., McKay, B., Lim, H.C., Abdul Rahman, M., Mays, H. & Moyle, R.G. 2012. Molecular phylogeny and insular biogeography of the lowland tailorbirds of Southeast Asia (Cisticolidae: Orthotomus). *Mol. Phylogenet. Evol.* **65**: 54–63.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis for large phylogenetic trees. *Bioinformatics* **30**: 1312–1313.

Streicher, J.W., Schulte, J.A. & Wiens, J.J. 2016. How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. *Syst. Biol.* **65**: 128–145.

Thomson, R.C. & Shaffer, H.B. 2010. Sparse supermatrices for phylogenetic inference: taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Syst. Biol.* **59**: 42–58.

Turner, D. & Pearson, D. 2015. Systematic and taxonomic issues concerning some East African bird species, notably those where treatment varies between authors. *Scopus* **34**: 1–23.

© 2017 British Ornithologists’ Union

Warren, B.H., Bermingham, E., Prys-Jones, R.P. & Thebaud, C. 2005. Tracking island colonization history and phenotypic shifts in Indian Ocean Bulbuls (Hypsipetes: Pycnonotidae). *Biol. J. Linn. Soc.* **85**: 271–287.
Wiens, J.J. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. *Syst. Biol.* **52**: 528–538.
Wiens, J.J. & Morrill, M.C. 2011. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Syst. Biol.* **60**: 719–731.
Zuccon, D. & Ericson, P.E.R.G.P. 2010. The phylogenetic position of the Black-Collared Bulbul *Neolestes torquatus*. *Ibis* **152**: 386–392.

Received 5 May 2016;
revision accepted 26 January 2017.
Associate Editor: Farah Ishtiaq.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. GenBank accession numbers for all taxa included in supermatrix analyses. Accession numbers for new sequences generated for this study are presented in italics. All tissues used in this study and previous studies are listed in the column 'Tissue No.'. Superscript numbers in front of each GenBank accession number correspond to the respective numbers in front of tissue numbers from which they were sequenced.

© 2017 British Ornithologists' Union