SPECIAL FEATURE: Species Limits and Taxonomy in Birds

Performance of a points-based scoring system for assessing species limits in birds

Joseph A. Tobias,[1,*] Paul F. Donald,[2] Rob W. Martin,[2] Stuart H. M. Butchart,[2,3] and Nigel J. Collar[2]

[1] Department of Life Sciences, Imperial College London, Silwood Park, Ascot, Berkshire, UK
[2] BirdLife International, Cambridge, UK
[3] Department of Zoology, University of Cambridge, Cambridge, UK
* Corresponding author: j.tobias@imperial.ac.uk

ABSTRACT

Species are fundamental to biology, conservation, and environmental legislation; yet, there is often disagreement on how and where species limits should be drawn. Even sophisticated molecular methods have limitations, particularly in the context of geographically isolated lineages or inadequate sampling of loci. With extinction rates rising, methods are needed to assess species limits rapidly but robustly. Tobias et al. devised a points-based system to compare phenotypic divergence between taxa against the level of divergence in sympatric species, establishing a threshold to guide taxonomic assessments at a global scale. The method has received a mixed reception. To evaluate its performance, we identified 397 novel taxonomic splits from 328 parent taxa made by application of the criteria (in 2014–2016) and searched for subsequent publications investigating the same taxa with molecular and/or phenotypic data. Only 71 (18%) novel splits from 60 parent taxa have since been investigated by independent studies, suggesting that publication of splits underpinned by the criteria in 2014–2016 accelerated taxonomic decisions by at least 33 years. In the evaluated cases, independent analyses explicitly or implicitly supported species status in 62 (87.3%) of 71 splits, with the level of support increasing to 97.2% when excluding subsequent studies limited only to molecular data, and reaching 100% when the points-based criteria were applied using recommended sample sizes. Despite the fact that the training set used to calibrate the criteria was heavily weighted toward passerines, splits of passerines and non-passerines received equally strong support from independent research. We conclude that the method provides a useful tool for quantifying phenotypic divergence and fast-tracking robust taxonomic decisions at a global scale.

Keywords: avian systematics, species limits, taxonomy

LAY SUMMARY

- A scoring system based on quantitative criteria was developed for classifying bird species and applied to the global avifauna in 2014–2016.
- We assess the performance of the criteria by searching for independent taxonomic assessments published subsequently.
- A minimum of 87% of novel taxonomic splits proposed by the criteria are supported by independent research, increasing to 97–100% when focusing only on integrative analyses based on genotypic and phenotypic information, or when the criteria were applied using more robust samples of individuals.
- The proportion of novel splits assessed by subsequent independent studies suggests that the application of the criteria in 2014–2016 accelerated taxonomic decisions by at least 33 years.
- We conclude that the criteria offer a useful tool for fast-tracking robust taxonomic decisions, although they do not remove the need for verification by more sophisticated analyses.

Desempeño de un sistema de puntuación para evaluar los límites de las especies en las aves

RESUMEN

Las especies son fundamentales en biología, conservación y legislación ambiental, pero usualmente hay desacuerdo en cómo y dónde deben trazarse los límites de las especies. Incluso los métodos moleculares sofisticados tienen limitaciones, particularmente en el contexto de linajes geográficamente aislados o de muestreo inadecuado de loci. Con el aumento de las tasas de extinción, se necesitan métodos para evaluar de modo rápido y robusto los límites de las especies. Tobias et al. idearon un sistema basado en puntos para comparar la divergencia fenotípica entre taxones con el nivel

de divergencia en especies simpátricas, estableciendo un umbral para guiar las evaluaciones taxonómicas a una escala global. El método ha recibido una recepción variada. Para evaluar su desempeño, identificamos 397 nuevas divisiones taxonómicas a partir de 328 taxones de origen, realizadas por la aplicación de los criterios (en 2014–2016), y buscamos publicaciones subsecuentes que investigaron los mismos taxones con datos moleculares y/o fenotípicos. Solo 71 (18%) de estas nuevas divisiones han sido evaluadas desde entonces por estudios independientes, sugiriendo que los criterios aceleraron las decisiones taxonómicas para esta muestra en por lo menos 33 años en total. En los casos evaluados, los análisis independientes apoyaron explícita o implícitamente el estatus de especie en 62 (87.3%) de las 71 divisiones, con el nivel de apoyo aumentando hasta 97.2% cuando se excluyeron los estudios subsecuentes que se limitaron solo a los datos moleculares, y alcanzaron el 100% cuando los criterios basados en puntos fueron aplicados usando los tamaños de muestra recomendados. A pesar del hecho de que el set de entrenamiento usado para calibrar los criterios estaba muy inclinado hacia los paseriformes, las divisiones de los paseriformes y no-paseriformes recibieron un apoyo igualmente fuerte por parte de las investigaciones independientes. Concluimos que el método representa una herramienta útil para cuantificar la divergencia fenotípica y brinda decisiones taxonómicas rápidas y sólidas a escala global.

Palabras clave: límites de las especies, sistemática de aves, taxonomía

INTRODUCTION

Species taxonomy underpins much of biological research, with the establishment of stable and globally standardized species limits being particularly critical for macroecology, macroevolution, and the setting of conservation priorities. Alas, taxonomic stability and standardization have proved elusive for all major taxa, especially birds. The waning and waxing of the number of bird species recognized by global authorities over the past century have reflected, first, a major fluctuation in the accepted definition of species and, second, a rapid expansion in knowledge and data relating to the biological characteristics of avian taxa around the world. From a high of 18,939 (Sharpe 1899–1909), the number of bird species recognized in published world checklists fell in under 40 years to a mere 8,616 (Mayr 1946), but now totals between 10,175 (Christidis et al. 2018) and 11,158 (Handbook of the Birds of the World [HBW]/BirdLife International 2020) species.

The upward trend in the number of recognized bird species in recent decades does not reflect the discovery of new species so much as the redrawing of species boundaries based on new information and new ways of processing old information. Rates of taxonomic change have been driven in part by legions of birdwatchers and sound-recordists collecting data on distributions and vocal signals, as well as by museum researchers examining larger samples of preserved material. An even more significant role has been played by molecular biologists and systematists uncovering the evolutionary relationships among taxa and developing new tools to examine species boundaries. Nevertheless, the interpretation and use of the information, new and old, has been far from consistent. Different world and regional lists have used different methods to gauge the validity of the myriad taxonomic judgments that derive from the continuing cascade of new information, resulting in a divergence of listings that has been characterized as "taxonomic anarchy" (Garnett and Christidis 2017).

To some extent, this is a matter of preferred "species concept," the multiple forms of which are themselves the most salient evidence of "anarchy" in taxonomy (for ornithology, see Haffer 1992, 1997). Species concepts come in various guises, but in the context of avian taxonomy, the choice is often between the phylogenetic species concept (PSC), which espouses monophyly as its key criterion, and the biological species concept (BSC), which makes reproductive incompatibility its central tenet (e.g., Winker et al. 2007). Although both approaches have advantages and limitations, a wholesale switch to the PSC is currently viewed as problematic from the perspective of standardizing species lists (Collar 1997, 2018, Johnson et al. 1999), not least because PSC-based assessments may double or treble the number of recognized bird species (Barrowclough et al. 2016). Therefore, world and regional lists still adhere, at least nominally, to the BSC, building upward and outward from the base provided by Peters and successors (1931–1986). Even so, the subjectivity involved in deciding the rank of allopatric taxa under the BSC remains a serious challenge, leading to ever-increasing disparities between the various world lists of bird species (Garnett and Christidis 2017).

The problem of "subjectivity" of the BSC in assigning rank to allopatric taxa was treated as unfounded ("allopatric populations can be assessed objectively") in a robust rejection of the PSC by Johnson et al. (1999), whose argument, however, rested entirely on emerging technological capacities to analyze vocalizations, displays, and genetic sequences. Their paper offered no way forward in matters of morphological (plumage and size) differences, and failed to reflect on the complexities of acoustic analysis in the context of learned or innate vocalizations, or of genetic analysis where outcomes are obscured by incomplete sampling and rapidly changing methodologies. Its concluding formulation of a "comprehensive biologic [sic] species concept" could commonly be applied to taxa ranked as subspecies.
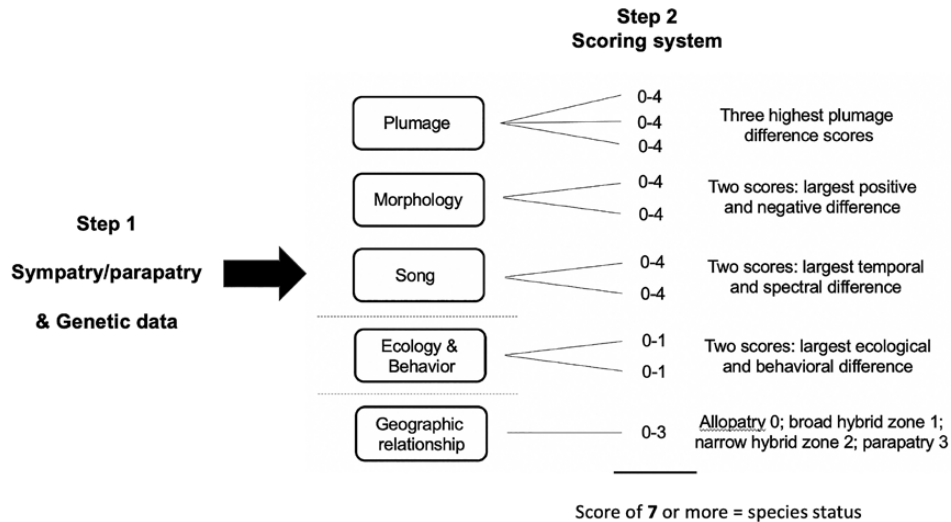
FIGURE 1. Diagram of the two-step application of the taxonomic criteria proposed by Tobias et al. (2010). Step 1 gives priority to genetic data in cases where taxa are in confirmed contact (e.g., sympatric or parapatric). If molecular evidence indicates substantial divergence across multiple loci or low levels of gene flow between such taxa, they are treated as species regardless of their level of phenotypic divergence. If molecular evidence is inconclusive, or the taxa are allopatric, their level of divergence is scored in Step 2. No individual character can score more than 4, with 4 reflecting "exceptional divergence" in a major character (particularly those involved in reproductive isolation, e.g., mating display, song, or ornament). To limit multiple counting of correlated traits, scores are capped to 2 orthogonal morphometric and vocal characters, defined by thresholds in effect size (Cohen's *d*) calculated from measurements taken from >10 individuals/songs.

Shortly afterwards, Helbig et al. (2002) proposed that species rank be assigned to allopatric taxa that are "fully diagnosable in each of several discrete or continuously varying characters" and where "the sum of the character differences corresponds to or exceeds the level of divergence seen in related species that coexist in sympatry." They further proposed that "allospecies" rank—without clarifying what this represents—be assigned to taxa that are fully diagnosable by at least one character and in which the level of divergence is equivalent to that found in related sympatric species. The difficulties with this formulation lie in the non-specificity of "several," the potential triviality of the diagnostic characters identified, the predictable difficulty in many cases of finding sufficiently closely related sympatric species with which to make comparisons, and problems in gauging levels of divergence when represented by different kinds of characters. Thus, they could be used to assign species rank to virtually any diagnosable subspecies. Perhaps for this reason, this formulation has been little used, even by its own authors, who in multiple subsequent decisions on allopatric taxa in their "taxonomic recommendations for British birds" (Ibis volumes 144–158) did not apply the comparison with levels of difference in "related sympatric species."

Recognizing the weaknesses in these formulations but respecting the intention behind them, Tobias et al. (2010) proposed what were intended to be more robust and more explicit criteria by allowing for *strength* of character as well as number. These were based on an exercise in which 58 pairs of closely related and morphologically similar sympatric or parapatric bird species from all continents and latitudes (albeit largely focused on passerines) were scored for the strength and number of their phenotypic and ecological differences. The level of divergence in these universally accepted species was established as a preliminary benchmark (or "yardstick") to assess species status, a concept with a long history in ornithology (Mayr 1969, Isler et al. 1998, Helbig et al. 2002) and championed more widely as a solution to runaway taxonomic inflation under the PSC (Isaac et al. 2004, Meiri and Mace 2007). A simple method was developed for converting quantitative measurements into scores, and a variety of limits were placed on scoring to increase objectivity and avoid double counting (Figure 1). This study suggested that a total score of 7 can serve as a general threshold for the recognition of species rank in birds.

In a recent world checklist (del Hoyo and Collar 2014–2016; henceforth "the Checklist"), these methods (henceforth "the criteria") were applied in cases where the elevation of distinctive-looking (and/or distinctive-sounding) subspecies to species appeared a possible outcome. (The Checklist also incorporated numerous revisions following other sources, including many molecular studies, where the criteria were often used as a means of validation: Burfield et al. 2017.) This list, which is updated annually using the same approach (see http://datazone.birdlife.org/species/taxonomy), now underpins the taxonomy for birds on the IUCN Red List, and has widespread policy impact

(e.g., being adopted by various multilateral environmental agreements: Burfield et al. 2017). The criteria have experienced a relatively negative reception in some quarters, yet in the decade since they appeared in print no peer-reviewed paper has to our knowledge subjected them to direct analytical scrutiny.

We therefore assess quantitatively how well the criteria have performed in matching and predicting taxonomic proposals, by comparing the results of the application of the criteria during the preparation of the 2014‒2016 Checklist with independent research that has been published subsequently, based on analyses of both molecular and phenotypic data. We also take the opportunity to address criticisms and common misunderstandings of the criteria.

METHODS

We identified all taxa that were recognized as species in the Checklist (del Hoyo and Collar 2014–2016) that were split on the basis of the criteria and not previously given species rank in either the 16 volumes of HBW (del Hoyo et al. 1992‒2011) or the 2013 version of the BirdLife International taxonomic checklist (BirdLife International 2013). We then excluded all newly described species and any species already recognized by the IOC (International Ornithological Congress) list (https://www.worldbirdnames.org/new/) in versions prior to the publication of the Checklist—with reference to v4.1 (January 2014) for non-passerines and v6.1 (January 2016) for passerines—or by Howard and Moore v4.1 (2014). To avoid the potential bias of sampling splits most likely to be supported, we also excluded several hundred taxa split in the Checklist which had previously (since the 1950s but prior to the Checklist) been proposed by other authors for treatment as species. This left 397 novel splits (excluding nominate subspecies) from 328 "parent" taxa that were proposed by the Checklist based on the criteria.

To identify which of these had been subsequently and independently assessed by others, we compiled a library of papers relating to the taxonomy of the world's birds published from 2014 to October 2020, including a comprehensive survey from 2014 to 2019 provided by J. L. Copete of Lynx Edicions of some 120 journals that published over 500 papers and articles relevant to bird taxonomy. This survey was supplemented by J. V. Remsen's regular circulation of new titles and by our own regular searches. The resulting inventory was thus likely to have been close to a comprehensive list of the avian taxonomic literature postdating the Checklist up to and including 2019, with a less comprehensive collection of further studies published up to October 2020. From this we then sought, through a rapid review of the assembled material, to identify those taxonomic studies by other authors that simultaneously or subsequently, and independently, assessed the parent

taxa that had been split by application of the criteria in the Checklist. Such studies included both explicit assessments of species limits, either at the level of individual species or across higher taxonomic levels, and studies that presented taxonomic data, for example, in the form of phylogenies, without making explicit recommendations regarding species limits. To retrospectively assess the extent to which our methods are likely to have missed relevant papers, we used a two-stage search process. First, we randomly selected 50 novel splits from different parent taxa for which we had not found a subsequent independent study, and searched on the scientific name of the parent species, and the term "taxonomy," in Google Scholar. We then searched the list of results using the species name of the split, and examined all remaining results for evidence that the taxon or taxa had been assessed. We then calibrated the results of this process by repeating it exactly for 20 randomly selected splits from different parent taxa for which we had identified subsequent studies.

We scored the outcome of each such study according to whether it provided evidence that (1) did not support the split, (2) implicitly supported the split, for example, by showing a phylogeny or indicating significant vocal or morphometric differences that support the split but without making a direct taxonomic declaration in its favor, or (3) explicitly supported the split by presenting evidence in its favor and using that evidence to propose that the same taxonomic division be made. Note that "implicit" support is not necessarily any weaker than "explicit" support, and merely indicates that the authors refrained from any taxonomic judgement. Previous studies have criticized the criteria on the basis that the threshold of 7 for species status was calibrated on a sample heavily biased toward passerines, and therefore may not be relevant to non-passerines (Remsen 2016). To assess whether the criteria performed differently in passerines and non-passerines, we quantified support for these taxonomic groupings separately.

Full application of the criteria involves the use of measurements from at least 10 individuals when calculating the effect size of morphometric or vocal differences between taxa. However, when applying the criteria, assessments of vocal differences or morphological measurements were often based on samples of fewer than 10 individuals per taxon, owing to a shortage of available material. To assess whether splits were more likely to be supported if they were based on more robust sample sizes, we scored sampling (or "robustness") as Low = split resulting from scores assigned on the basis of a combination of plumage, ecology, geography, morphology (but with <10 specimens per taxon measured), and voice (with <10 individuals typically sampled); Medium = split resulting from scores assigned on the basis of a combination of fully sampled plumage, ecology, geography, and morphology

(with ≥10 specimens per taxon measured) but with under-sampled vocal data (<10 individuals sampled); or High = split resulting from scores assigned on the basis of full sampling of all traits, ecology, and geography, including vocal sampling (with ≥10 specimens per taxon measured). Note that some species scored over 7 on the basis of plumage characters alone and in such cases, vocal traits were not needed and therefore not sampled. We classified these cases as High robustness because no parameters were based on <10 individuals per taxon. The criteria for High robustness meet the sample sizes recommended by Tobias et al. (2010).

RESULTS

We identified 328 parent taxa for which application of the criteria led to the split of one or more previously unrecognized species in the Checklist, totaling 397 splits. Of these, we found subsequent studies of 60 (18.3%), which among them accounted for 71 splits (17.9%) made in the Checklist through application of the criteria (Table 1). Searches in Google Scholar for a random sample of 50 of the 268 parent taxa for which we found no subsequent study returned only one paper relevant to our purposes, and that was published outside our search period (in December 2020). In contrast, when applied to a random sample of 20 parent taxa for which we identified subsequent studies, the same search protocol successfully located those studies in 17 cases, usually as the first item in the list of search results. The 3 cases that were missed either used a different genus name (two instances), or the same genus name was abbreviated to a single letter where the target taxon was mentioned in the paper, so the search term was not located. Taking the probability of a paper being missed by the search protocol as $1 - (17/20) = 0.15$ (with exact binomial 95% CL of 0.05–0.36), and assuming our sample of 71 was 90% complete, we estimate the probability of returning no search results from a random sample of 50 of the 268 parent taxa for which no studies was found as $0.15^{((71 \times (100/90)) - 71) \times (50/268)} = 0.06$ (95% CL: 0.012–0.22). Thus, we can be 78–99% confident that our sample contains 90% or more of relevant studies. Furthermore, we see no reason why any studies that we missed should be systematically different from those we found in terms of their support or otherwise for splits made using the criteria.

Of these 71 splits, 9 (of 6 parent taxa) received no support from subsequent research (although in 2 cases, they were identified as separate "conservation units"), 19 (of 16 parent taxa) received implicit support from subsequent research, and 43 (of 38 parent taxa) received explicit support (Table 1, Figure 2). Thus, the overall concordance rate was 87.3% (62 of 71 splits), with no difference between passerines and non-passerines in the extent to which splits made by application of the criteria received subsequent independent support (Figure 2). However, in only 13 cases were the original splits based on the sample size of 10 or more individuals per taxon recommended under the criteria, owing to paucity of accessible specimens or recordings. In the case of these 13 splits, all (100%) were implicitly or explicitly supported by subsequent independent research (Figure 2). In addition, there was a significant association between the degree of support and the methods used in subsequent analyses (genetic data only, phenotypic data only, or integrated genotypic/phenotypic data: Fisher's exact test, $P < 0.001$; Figure 2). This was because for 8 of the 9 splits that received no support, subsequent evaluation had been based on analyses with genetic data only. When excluding cases where subsequent studies did not use genetic methods, the concordance rate rose to 95%, and when excluding genetics-only studies it rose again to 97.2%. Finally, concordance was 100% among studies that used an "integrative taxonomy" approach, combining genetic, phenotypic, and other metrics (Figure 2).

Of the 196 taxa that were scored against the criteria and not split on the basis of the resulting scores, we found only 6 that were independently assessed subsequently; in 3 cases, the decision not to split was supported and in 3 cases, the subsequent analyses proposed splits. Although sample sizes were small, the proportion of splits based on the criteria that received subsequent independent support (explicit or implicit) was significantly higher than the proportion of non-splits that received subsequent support (62/71 vs 3/6; Fisher's exact test, $P < 0.02$), suggesting that the criteria may be conservative in their likelihood of producing splits with respect to the wider taxonomic literature. Furthermore, the proportion of criteria-based splits that went on to be independently assessed (60/328 parent taxa; 18.3%) was significantly higher than the proportion of criteria-based non-splits that received subsequent independent taxonomic attention (6/196; 3.1%; Fisher's exact test, $P < 0.0001$), suggesting that research attention was more focused toward taxa split by application of the criteria.

During the preparation of the Checklist, we also applied the criteria to 145 taxa proposed by other authorities to be treated as species largely or wholly on the basis of molecular evidence. Of these, 120 (82.8%) qualified as species by application of the criteria and were treated as such in the Checklist, while 25 (17.2%) were treated as subspecies because, on the available evidence (acknowledging that analysis of additional data, especially acoustic, might alter these conclusions), they scored less than 7 using the criteria. It is notable that none of the 25 rejected cases had sample sizes classed as "high."

A key rationale for the development of the criteria was operational speed. Taking 2015 as the average

TABLE 1. List of splits made using scoring criteria that have subsequently been independently assessed. Robustness of criteria application is scored largely in respect of sample size (High: >10 individuals sampled) and whether data on vocalizations were available (see Methods). The method(s) used in the subsequent independent assessment are scored: A = genetic, B = phenotypic, C = vocalizations, D = other. Thus, a study marked 'A,C' used a combination of genetic and vocal data. Degree of support indicates the extent to which the subsequent independent assessment supported the split made using the scoring criteria (see Methods).

Parent taxon	Split	Robustness of criteria application	Method in subsequent assessment	Degree of support	Source
Acridotheres melanopterus	*Acridotheres tricolor*	Medium	A	None[a]	Sadanandan et al. 2020
Acridotheres melanopterus	*Acridotheres tertius*	Medium	A	None[a]	Sadanandan et al. 2020
Alcedo cyanopectus	*Ceyx nigrirostris*	Low	A	Implicit	Andersen et al. 2018
Actenoides monachus	*Actenoides capucinus*	Medium	A	None	Andersen et al. 2018
Amazona autumnalis	*Amazona lilacina*	Medium	B	Explicit	Donegan et al. 2016
Amazona festiva	*Amazona bodini*	High	B	Explicit	Donegan et al. 2016
Anthus lutescens	*Anthus peruvianus*	Medium	A,C	Explicit	van Els and Norambuena 2018
Arremon taciturnus	*Arremon axillaris*	Medium	B,C	Explicit	Buainain et al. 2017
Aulacorhynchus prasinus	*A. cyanolaemus*	Medium	B,D	None[b]	Winker 2016
Bambusicola thoracica	*Bambusicola sonorivox*	High	A,C	Explicit[c]	Hung et al. 2014
Basileuterus culicivorus	*Basileuterus cabanisi*	High	C	Explicit	Freeman and Montgomery 2017
Basileuterus luteoviridis	*Myiothlypis striaticeps*	Low	C	Explicit	Freeman and Montgomery 2017
Bleda notatus	*Bleda ugandae*	Medium	A	Explicit[c]	Huntley and Voelker 2016
Bowdleria punctata	*Poodytes caudatus*	Medium	A	Implicit	Alström et al. 2018a
Brachypteryx montana	*Brachypteryx erythrogyna*	Medium	A	Implicit	Kyriazis et al. 2018
Brachypteryx montana	*Brachypteryx poliogyna*	Medium	A	Implicit	Kyriazis et al. 2018
Brachypteryx montana	*Brachypteryx cruralis*	Low	A,B,C,D	Explicit	Alström et al. 2018b
Brachypteryx montana	*Brachypteryx goodfellowi*	Medium	A,B,C,D	Explicit	Alström et al. 2018b
Brachypteryx montana	*Brachypteryx sinensis*	Medium	A,B,C,D	Explicit	Alström et al. 2018b
Ceyx melanurus	*Ceyx mindanensis*	Low	A	Implicit	Andersen et al. 2018
Charadrius alexandrinus	*Charadrius dealbatus*	High	A,B,D	Explicit	Sadanandan et al. 2019, Wang et al. 2019
Charadrius obscurus	*Charadrius aquilonius*	Medium	A	None[c]	Barth et al. 2013
Cittura cyanotis	*Cittura sanghirensis*	High	A	Implicit	Andersen et al. 2018
Colaptes auratus	*Colaptes mexicanoides*	Medium	A	Implicit	Manthey et al. 2017
Coracias benghalensis	*Coracias affinis*	Medium	A	Explicit	Johansson et al. 2018
Cyornis tickelliae	*Cyornis sumatranus*	Low	C	Explicit	Gwee et al. 2019
Dinopium benghalense	*Dinopium psarodes*	Medium	A	Explicit	Fernando et al. 2016
Edolisoma tenuirostre	*Edolisoma grayi*	Medium	A	None	Pedersen et al. 2018
Edolisoma tenuirostre	*Edolisoma obiense*	Medium	A	None	Pedersen et al. 2018
Euscarthmus meloryphus	*Euscarthmus fulviceps*	Medium	B,C	Explicit	Franz et al. 2020
Forpus xanthopterygius	*Forpus spengeli*	Medium	B	Explicit[d]	Bocalini and Silveira 2015
Francolinus castaneicollis	*Pternistis atrifrons*	Medium	A,B,C	Explicit[c]	Töpfer et al. 2014
Francolinus psilolaemus	*Scleroptila elgonensis*	Medium	A,B,C	Explicit	Hunter et al. 2019, Turner et al. 2020
Goura scheepmakeri	*Goura sclaterii*	Medium	A	Explicit	Bruxaux et al. 2018
Gracupica contra	*Gracupica jalla*	High	A	Explicit	Baveja et al. 2020
Grallaria quitensis	*Grallaria alticola*	Low	C	Explicit	Freeman and Montgomery 2017
Grallaricula ferrugineipectus	*Grallaricula leymebambae*	Low	A,B,C	Explicit	van Doren et al. 2018
Halcyon smyrnensis	*Halcyon gularis*	High	A	Implicit	Andersen et al. 2018
Junco phaeonotus	*Junco bairdi*	Medium	A,B	Implicit[c]	Friis et al. 2016
Lacedo pulchella	*Lacedo melanops*	Medium	A	Implicit	Andersen et al. 2018
Macronous flavicollis	*Mixornis prillwitzi*	Medium	C	Implicit	Cros and Rheindt 2017
Malacoptila striata	*Malacoptila minor*	Medium	A	Explicit	Ferreira et al. 2017
Melozone leucotis	*Melozone occipitalis*	Medium	B,C	Explicit	Sandoval et al. 2017
Merops viridis	*Merops americanus*	Medium	A	Explicit	Huang et al. 2017

TABLE 1. Continued

Parent taxon	Split	Robustness of criteria application	Method in subsequent assessment	Degree of support	Source
Mulleripicus funebris	*Mulleripicus fuliginosus*	High	A	Explicit	Shakya et al. 2017
Ninox squamipila	*Ninox hantu*	High	A,C	Explicit	Gwee et al. 2017
Oriolus melanotis	*Oriolus finschi*	Low	A	Implicit[c]	Jønsson et al. 2016
Oriolus cruentus	*Oriolus consanguineus*	Medium	A	Implicit	Jønsson et al. 2019a
Paramythia montium	*Paramythia olivacea*	Medium	A	Implicit	Jønsson et al. 2019b
Pica pica	*Pica asirensis*	Low	A	Explicit	Kryukov et al. 2017
Pica pica	*Pica mauritanica*	Medium	A,C	Explicit	Kryukov et al. 2017
Pomatorhinus erythrocnemis	*Erythrogenys gravivox*	High	A	Explicit	Dai et al. 2019
Pomatorhinus erythrocnemis	*Erythrogenys swinhoei*	High	A	Explicit	Dai et al. 2019
Psittacara wagleri	*Psittacara frontatus*	High	B	Explicit	Donegan et al. 2016
Ptilinopus porphyraceus	*Ptilinopus hernsheimi*	Medium	C	Explicit	Hayes et al. 2016
Pycnonotus flavescens	*Pycnonotus leucops*	Medium	A	Implicit[c]	Dejtaradol et al. 2016
Pyrocephalus rubinus	*Pyrocephalus dubius*	Medium	A,B,C	Explicit[c]	Carmi et al. 2016
Pyrocephalus rubinus	*Pyrocephalus nanus*	Medium	A,B,C	Explicit[c]	Carmi et al. 2016
Pyrrhura melanura	*Pyrrhura pacifica*	Low	B	Explicit	Donegan et al. 2016
Sittasomus griseicapillus	*Sittasomus griseus*	Low	C	Explicit	Freeman and Montgomery 2017
Sporophila torqueola	*Sporophila morelleti*	Medium	A,B	Explicit	Mason et al. 2018
Stephanoxis lalandi	*Stephanoxis loddigesii*	High	B	Explicit[c]	Cavarzere et al. 2014
Thamnistes anabatinus	*Thamnistes aequatorialis*	Low	C	Implicit[e]	Isler and Whitney 2017
Thryothorus euophrys	*Pheugopedius schulenbergi*	Low	C	Explicit	Freeman and Montgomery 2017
Turdinus crispifrons	*Gypsophila calcicola*	Low	A,B,C	Explicit[f]	Gwee et al. 2020
Zosterops cinereus	*Zosterops ponapensis*	Medium	B,C	Explicit[c]	Hayes et al. 2016
Zosterops poliogastrus	*Zosterops eurycricotus*	Low	A[g]	Implicit	Pearson and Turner 2017
Zosterops poliogastrus	*Zosterops mbuluensis*	Low	A[g]	Implicit	Pearson and Turner 2017
Zosterops poliogastrus	*Zosterops winifredae*	Low	A[g]	Implicit	Pearson and Turner 2017
Zosterops poliogastrus	*Zosterops kaffensis*	Low	A[g]	None	Pearson and Turner 2017
Zosterops poliogastrus	*Zosterops kulalensis*	Low	A[g]	None	Pearson and Turner 2017

[a] Authors suggest that the forms should be treated as separate conservation units.
[b] Winker (2016) otherwise concurred with all species limits as defined under the criteria in the *A. prasinus* group.
[c] Paper published concurrently or immediately prior to Checklist but only seen subsequently.
[d] Split supported by Donegan et al. (2016), although they speculated that *spengeli* might be a subspecies of *F. passerines*.
[e] Isler and Whitney (2017) selected only *T. rufescens* to split; the criteria split *rufescens*, *aequatorialis*, and *gularis*, Isler and Whitney (2017) conceded *aequatorialis* is vocally distinct from the 4 trans-Andean taxa.
[f] Gwee et al. (2020) also split *annamensis* but this was not scored using the criteria owing to lack of material.
[g] Authors reviewed molecular evidence from earlier studies that had made no taxonomic recommendations.

publication date of the 2 Checklists, in the following 6 years up to 2020 inclusive, 60 of our restricted sample of 328 parent taxa that were split by application of the scoring method were subsequently independently analyzed. On this basis, if we simplistically assume the same rate of taxonomic investigation and concordance (i.e. 60/6 = 10 parent taxa split per year), it would take until 2047 for the remaining 268 parent taxa to be subject to independent scrutiny. Use of the criteria to underpin splits in the Checklist therefore brought forward taxonomic revisions by at least 33 years, with the true figure potentially likely to be substantially larger considering the increased research attention on splits in the Checklist (see above), and the hundreds of other previously proposed splits that were assessed but not included in our sample (see Methods).

DISCUSSION

The key finding of this review is that the criteria of Tobias et al. (2010) produce decisions on biological species limits in birds that conform well with decisions subsequently reached by other researchers using various other methods. Splits of species first proposed by application of the criteria were 10 times more likely to receive support from subsequent independent study than not. The context in which the criteria matched least well with the conclusions of independent analyses was when those subsequent studies were
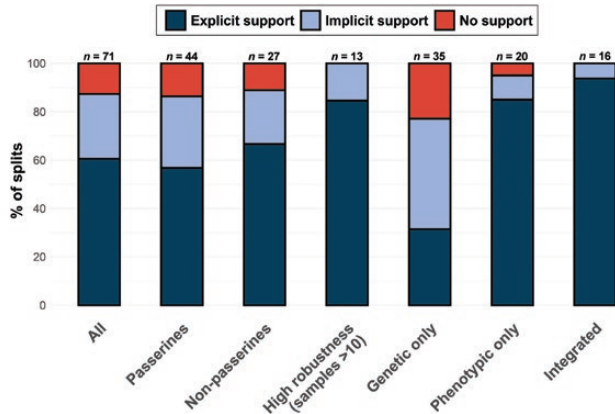
FIGURE 2. Degree of concordance (%) between taxonomic splits made by application of the criteria of Tobias et al. (2010) and subsequent, independent taxonomic research. Levels of explicit and implicit support are high for all splits that have been subsequently assessed (All). In addition, despite the criteria being calibrated with a sample heavily biased toward passerines, support was high for splits made in both passerines and non-passerines, separately. In cases where the criteria were applied using >10 individuals for morphometric analyses, in line with "High" robustness (see Methods) and following recommendations of Tobias et al. (2010), independent support for the split was 100%. Support was lower in studies focusing only on molecular data ("Genetic only"), much higher in studies focusing only on phenotypic data, and 100% in studies combining a range of data sources (Integrated). Sample sizes (number of splits) are given above the bars.

based solely on molecular data. However, this lower level of support can be partly explained by the higher proportion of genetics-only studies adopting an entirely different species concept (PSC), sometimes leading to taxonomic recommendations at odds with the BSC. Concordance with the taxonomic conclusions or implications of previous studies was 100% when the criteria were applied fully (see Figure 1), or when evaluated with integrative studies based on a range of data sources (Figure 2). This of course does not imply that the criteria can achieve total concordance with other treatments, particularly given numerous potential sources of error and bias underlying the scores, as well as the inherent subjectivity of all taxonomic classification systems (Tobias et al. 2010). Nonetheless, the degree of agreement among studies is highly encouraging, and suggests that the criteria can deliver robust taxonomic assessments in a short time frame.

Our observation that performance improves with more extensive sampling of individuals in morphometric and vocal analyses is likely to be caused by 2 main factors. First, larger samples reduce error and uncertainty in effect sizes, and second, they also reduce the problem of bias caused by individual variation or contextual differences (e.g., songs sampled in different seasons or settings). Another sampling issue justifiably raised in a previous

review of the criteria (Remsen 2016) was that the species pairs used to derive the scoring threshold were dominated by passerines. This bias in sampling would obviously be problematical if it meant the criteria worked less well for non-passerines. However, we found no evidence of this effect, because splits of non-passerines were equally likely to receive subsequent independent support.

The criteria are designed for fast-tracking taxonomic decisions and are therefore relatively basic by design, as discussed by Tobias et al. (2010). We do not believe that there is anything "magic" about the score of 7, as defined in Figure 1, merely that it reflects a level of differentiation between taxa that correlates with species limits assigned through other methods. There is a tendency among ornithologists to assume that highly sophisticated genomic analyses must be able to delimit species more accurately than any method relying largely on phenotypic divergence, but this assumption is risky in the case of allopatric taxa for which monophyly can simply reflect population structure (arising from reduced gene flow between spatially isolated populations of the same species; Tobias et al. 2010, 2020). The criteria counter this problem by defining a minimum threshold of phenotypic divergence associated with sympatric or parapatric species pairs. Our results suggest that even in their current form the criteria offer a useful tool for making largely accurate taxonomic classifications across numerous species in a short time frame.

Promise and Pitfalls of Genetic Evidence for Reproductive Isolation

In the decade since their inception, the criteria have received plenty of criticism, often on the basis of misunderstandings about their goals and methods (Collar et al. 2016). A common misconception is that the approach taken is anti-molecular, and molecular biologists have been understandably sensitive to the fact that genetic information was not incorporated as a quantifiable component into the criteria (e.g., Collinson et al. 2017). However, far from ignoring genetic evidence, Tobias et al. (2010) explicitly prioritized genetic data when populations are in contact (Figure 1) and emphasized that the criteria are designed to accommodate genetic distances as soon as their relevance to species limits under the BSC are better understood.

The reason genetic measures were not integrated at the outset is because simplified molecular metrics—along the lines of mtDNA divergence—are difficult to interpret in the context of species limits (Winker et al. 2007). The reasons for this problem are widely reported, including heterogeneous rates of molecular evolution and gene flow among currently or historically parapatric taxa, which can pose serious difficulties for species delimitation (Tobias et al. 2020, Jiao and Yang 2021). On the one hand, incomplete lineage sorting can explain low or zero divergence at

target loci (Joseph et al. 2009). On the other hand, when secondary contact occurs before reproductive isolation is complete, hybridization can lead to anything from zero gene flow to complete merger of gene pools (e.g., Kearns et al. 2018), or genome-wide introgression with species limits maintained by very few "barrier loci," often on the sex chromosomes (e.g., Toews et al. 2016). In this context, introgression of the mitogenome can periodically reset mtDNA divergence to zero during the process of speciation with gene flow, even accounting for unexpectedly low mtDNA divergence among well-established species (Irwin et al. 2009, Rheindt and Edwards 2011, Tobias et al. 2020, Miller et al. 2021).

While it is often assumed that genetic information provides greater reliability in taxonomic decisions, any system over-reliant on molecular evidence will also suffer instability (Cadena and Zapata 2021). For example, Martens et al. (2008) described a new species—Alpine Leaf-warbler (*Phylloscopus occisinensis*)—on the basis of divergent mtDNA. However, this was later identified as a case of "deep mitochondrial divergence" within populations of *Phylloscopus affinis* potentially caused by hybridization with a now-extinct congener, which left the "ghost of introgression past" in the genes of the colonizing population (Zhang et al. 2019). When assessed under points-based criteria for the Checklist, *P. occisinensis* did not qualify for species status because it was undiagnosable using vocal or plumage characters. By contrast, all other major taxonomic checklists adopted the split of *P. occisinensis*, and will now need to correct the error.

Other cases of instability caused by deference to genetic evidence stem from the ever-changing landscape of molecular methods and evolutionary modeling. As methods and datasets improve, reversals of earlier judgments can diminish confidence in work published only a few years before. Contradictory assessments of the genetic differences between Common Swifts (*Apus apus*) and Pallid Swifts (*Apus pallidus*) (Päckert et al. 2012, Pellegrino et al. 2017) and of the validity of *Heliangelus zusii* (Kirchman et al. 2010, Pérez-Emán et al. 2018) are cases in point. Overall, there is an emerging view that the use of mtDNA alone, the mainstay of earlier genetic studies of birds, is insufficient to determine taxonomic relationships reliably: Drovetski et al. (2018) "caution against the out-of-hand dismissal of traditional taxonomy in cases when mtDNA appears to contradict it, regardless of how strong the support of geographically coherent clades in the mtDNA gene tree might be."

None of this is intended to downplay the crucial contribution of molecular evidence to systematic revision and species delimitation in birds. Both genotypic and phenotypic evidence have major strengths and limitations, and they clearly need to be considered together

as complementary aspects of "integrative taxonomy" (Winker 2009, Padial et al. 2010, Cadena and Zapata 2021). Many studies have taken up the challenge of bringing multiple lines of evidence to bear on taxonomic judgments (e.g., Alström et al. 2008, 2018a, Cadena and Cuervo 2010). These are the gold standard in terms of integrative taxonomy, although some cases appear to give heavier weighting to genetics. Hosner et al. (2018), for example, invoked "operational criteria" for deciding species rank that involve the congruence of "(1) well-supported monophyly of geographic clades, (2) significant genetic differentiation, as identified by a coalescent model, and (3) fixed plumage and morphological differences." In this formulation, taxonomy is integrative but not exactly balanced: as long as the first 2 genetic criteria are met, morphological differences, however tiny, can trigger species rank. Even without such differences, the first 2 criteria identify "cryptic lineages" that *might* be species, although the authors admit "their genetic distinctiveness could be an artifact of strong population structure." If so, however, why should this possibility not equally extend to splits that are only marginally distinct in morphology?

In a related study of Bornean birds, Moyle et al. (2017) found significant genetic differences between lowland and upland representatives of 3 taxa with apparent "elevational parapatry," arguing that these consequently merited recognition at species level, and concluding with the remark that "any species concept that attempts to predict interbreeding potential simply on the basis of perceived morphological differences (Tobias et al. 2010) likely underestimates species diversity." While this may be the case, we think the criteria would only overlook a very small proportion of parapatric cryptic species acceptable under the BSC simply because the score for parapatry is high (3), and when added to consistent vocal differences, often triggers species status (Figure 1). Given how regularly this simple fact is overlooked, we suspect that some who have dismissed the criteria may have never attempted to apply them fully in practice.

A classic example of over-hasty dismissal is provided by a recent study establishing *Catharus maculatus* as a species distinct from *C. dryas* (Halley et al. 2017). The authors argued that their split would fail under the "yardstick" criteria used in previous studies, yet also reported that the 2 taxa "are 100% diagnosable in genetic, vocal, morphometric, and plumage characters." These findings suggest that *C. maculatus* and *Catharus dryas* would be treated as species under 3 recent permutations of yardstick criteria developed for birds (Isler et al. 1998, Helbig et al. 2002, Tobias et al. 2010). Indeed, when the Tobias et al. (2010) criteria are applied to the data from Halley et al. (2017), *C. maculatus* emerges with a score of >10, well above the threshold set for species rank. In some of these cases, we

suspect that misinterpretation may arise from an assumption that any taxon not split in the Checklist failed to meet the 7-point threshold, whereas a lack of splitting often simply means that the case was not investigated or the relevant data were not available. The case of *C. maculatus*, for example, was not scored for the Checklist and so does not appear in Table 1 of this paper.

LIMITATIONS AND FUTURE DIRECTIONS

Species delimitation is to some extent inherently arbitrary, and disagreement about the conclusions of any taxonomic system is therefore unavoidable (Hey et al. 2003, Winker et al. 2007). With regard to taxonomic changes in the Checklist, a recurring point of contention involves the assignment of species status to taxa joined by hybrid zones (e.g., Donegan et al. 2015). The criteria are designed to classify such cases as species even when lineages hybridize freely across a broad zone, as long as lineages appear to retain highly divergent and stable phenotypes on either side of the zone, as explained with reference to examples in Collar et al. (2016). We consider our approach to be supported by evidence that species limits can be maintained in genetically near-identical species by relatively few barrier loci (i.e. islands of genomic differentiation), such as those coding for highly diagnostic plumage differences in *Vermivora* warblers (Toews et al. 2016). Important questions remain about how such cases should be treated taxonomically, with one alternative approach being demotion to subspecies status. In other cases, new information about the level of phenotypic divergence, or the extent of genetic introgression, or indeed the width of the hybrid zone in relation to the overall range of taxa, may support remerging of taxa split by the criteria (Céspedes-Arias et al. 2021). However, our rationale for conferring species status to hybridizing taxa with high scores of phenotypic divergence remains unchanged.

Another widely repeated criticism is that the criteria depend on "subjective" assessments of character differences (e.g., Martens and Bahr 2016). For instance, Hosner et al. (2018) claimed that the criteria are "subject to individual interpretation, and often result in conflicting limits drawn from differing data sources," although they provided no evidence to support the latter point and, as they noted, their own proposed criteria require "individual interpretation" to decide the taxonomic status of cryptic lineages. To some extent, this echoes ongoing species concept debates, with phylogenetic systematists criticizing the BSC for relying on subjective assessments, despite equivalent levels of subjectivity inherent in deciding species limits under the PSC (Johnson et al. 1999, Winker et al. 2007, Tobias et al. 2010). Nonetheless, we acknowledge that some degree of subjectivity is impossible to eliminate from the criteria,

and that this issue may be accentuated because most of the scores in the first round of taxonomic evaluations for the Checklist were made by a single observer (N.J.C.). Further steps should be taken to refine the process in this regard, including averaging across scores from multiple observers whenever possible and using online data to calibrate scoring among observers (see our Data depository statement). Another solution to the problem of subjectivity is transparency. Not only is the scoring of character differences under the criteria done using explicit guidelines, but the scores for individual characters are reported along with any underlying quantitative data, including samples of vocal and morphological measurements used to generate effect sizes. To ensure that the scores can be checked and challenged where necessary, these datasets are publicly available for examination (see Data depository in Acknowledgments).

Some published suggestions relating to the criteria are constructive. The use and treatment of effect sizes have been debated (Donegan 2018), including the suggestion that the approach be abandoned entirely because of problems associated with delimiting species on the basis of central tendency in phenotypic data (Cadena et al. 2018). In addition, it has been suggested that the procedure used to assign and add scores violates elements of measurement theory (for discussion of which, see Houle et al. 2011). We agree that these issues warrant attention, and that alternative models and procedures should be examined in cases where larger phenotypic datasets are available. However, we also caution against over-complicating an approach designed for rapid application and which seems to work remarkably well in its current format. Thankfully, few systems are as taxonomically intractable as the *Geospiza* finches investigated by Cadena et al. (2018).

The scoring system and associated threshold value should be applied to a wider range of accepted sympatric or parapatric species, particularly non-passerines, to refine and recalibrate the system. Furthermore, additional work might reveal that a wider range of phenotypic characters could usefully be included in the system; spectral reflectance of feathers, tarsal scutellation, osteology, egg color, plumage of downy young, and even smell have all been used or proposed as ways to assess taxonomic status. Finally, even in its current form, the method can provide a useful framework for quantifying phenotypic divergence among lineages in studies testing evolutionary hypotheses (e.g., Campbell et al. 2016).

CONCLUSIONS

Ideally, taxonomic decisions should be based entirely on painstaking research combining multiple lines of genetic and phenotypic evidence. However, with many threats intensifying and bird populations rapidly declining

worldwide, there is a strong argument for seeking a more rapid approach—at least as a preliminary assessment—to fast-track taxonomic decisions before we lose many cryptic bird species forever (Lees and Pimm 2015, Remsen 2016). Our findings suggest that Tobias et al.'s (2010) criteria, as applied in the HBW/BirdLife International Checklists, provide a reasonably reliable method for achieving this goal. Indeed, the level of concordance is well over 90% when discounting genetics-only studies or implementing the points-based system using recommended sample sizes. Some will think that a potential 10% error rate is too high, but there is much to be gained from a pragmatic approach achieving over 90% concordance and then allowing science to correct a relatively small number of mistakes. The criteria can always be criticized for being too "quick and dirty," but our results suggest that they produce taxonomic decisions with high consistency in a fraction of the time, and also improve the knowledge base by increasing the focus of subsequent research on likely splits. At the very least, it is clear that they offer a useful tool for proactively investigating hitherto neglected cases and reactively evaluating taxonomic changes proposed by other analytical methods.

ACKNOWLEDGMENTS

LITERATURE CITED

Alström, P., P. C. Rasmussen, U. Olsson, and P. Sundberg (2008). Species delimitation based on multiple criteria: The Spotted Bush Warbler *Bradypterus thoracicus* complex (Aves: Megaluridae). Biological Journal of the Linnean Society 154:291–307.

Alström, P., A. Cibois, M. Irestedt, D. Zuccon, M. Gelang, J. Fjeldså, M. J. Andersen, R. G. Moyle, E. Pasquet, and U. Olsson (2018a). Comprehensive molecular phylogeny of the grassbirds and allies (Locustellidae) reveals extensive non-monophyly of traditional genera, and a proposal for a new classification. Molecular Phylogenetics and Evolution 127:367–375.

Alström, P., P. C. Rasmussen, C. Xia, M. Gelang, Y. Liu, G. Chen, H. Zhao, Y. Hao, C. Zhao, J. Zhao, et al. (2018b). Taxonomy of the White-browed (*Brachypteryx montana*) complex on mainland Asia and Taiwan: An integrative approach supports recognition of three instead of one species. Avian Research 9:34.

Andersen, M. J., J. M. McCullough, W. M. Mauck, B. Tilston Smith, and R. G. Moyle (2018). A phylogeny of kingfishers reveals an Indomalayan origin and elevated rates of diversification on oceanic islands. Journal of Biogeography 45:269–281.

Barrowclough, G. F., J. Cracraft, J. Klicka, and R. M. Zink (2016). How many kinds of birds are there and why does it matter? PLoS One 11:e0166307.

Barth, J. M., M. Matschiner, and B. C. Robertson (2013). Phylogenetic position and subspecies divergence of the endangered New Zealand Dotterel (*Charadrius obscurus*). PLoS One 8:e78068.

Baveja, P., K. M. Garg, B. Chattopadhyay, K. R. Sadanandan, D. M. Prawiradilaga, P. Yuda, J. G. H. Lee, and F. E. Rheindt (2020). Using historical genome-wide DNA to unravel the confused taxonomy in a songbird lineage that is extinct in the wild. Evolutionary Applications. doi:10.1111/eva.13149

BirdLife International (2013). The BirdLife checklist of the birds of the world, with conservation status and taxonomic sources, version 6. http://www.birdlife.org/datazone/userfiles/file/Species/Taxonomy/BirdLife_Checklist_Version_6.zip

Bocalini, F., and L. F. Silveira (2015). Morphological variability and taxonomy of the Blue-winged Parrotlet *Forpus xanthopterygius* (Psittacidae). Revista Brasileira de Ornitologia 23:64–75.

Bruxaux, J., M. Gabrielli, H. Ashari, R. Prŷs-Jones, L. Joseph, B. Milá, G. Besnard, and C. Thébaud (2018). Recovering the evolutionary history of crowned pigeons (Columbidae: *Goura*): Implications for the biogeography and conservation of New Guinean lowland birds. Molecular Phylogenetics and Evolution 120:248–258.

Buainain, N., C. P. de Assis, and M. A. Raposo (2017). Geographic variation and taxonomy of the *Arremon taciturnus* (Hermann, 1783) species complex (Aves: Passerellidae). Journal of Ornithology 158:631–650.

Burfield, I. R., S. H. M. Butchart, and N. J. Collar (2017). BirdLife, conservation and taxonomy. Bird Conservation International 27:1–5.

Cadena, C. D., and A. M. Cuervo (2010). Molecules, ecology, morphology and songs in concert: How many species is *Arremon torquatus* (Aves: Emberizidae)? Biological Journal of the Linnean Society 99:152–176.

Cadena, C. D., and F. Zapata (2021). The genomic revolution and species delimitation in birds (and other organisms): Why phenotypes should not be overlooked. Ornithology 138:1–18. doi:10.1093/ornithology/ukaa069

Cadena, C. D., F. Zapata, and I. Jiménez (2018). Issues and perspectives in species delimitation using phenotypic data:

Atlantean evolution in Darwin's finches. Systematic Biology 67:181–194.

Campbell, K. K., T. Braile, and K. Winker (2016). Integration of genetic and phenotypic data in 48 lineages of Philippine birds shows heterogeneous divergence processes and numerous cryptic species. PLoS One 11:e0159325.

Carmi, O., C. C. Witt, A. Jaramillo, and J. P. Dumbacher (2016). Phylogeography of the Vermilion Flycatcher species complex: Multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galápagos-endemic bird species. Molecular Phylogenetics and Evolution 102:152–173.

Cavarzere, V., L. F. Silveira, M. F. de Vasconcelos, R. Grantsau, and F. C. Straube (2014). Taxonomy and biogeography of *Stephanoxis* Simon, 1897 (Aves: Trochilidae). Papéis Avulsos de Zoologia 54:69–79.

Céspedes-Arias, L. N., A. M. Cuervo, E. Bonaccorso, M. Castro-Farias, A. Mendoza-Santacruz, J. L. Pérez-Emán, C. C. Witt, and C. D. Cadena (2021). Extensive hybridization between two Andean warbler species with shallow divergence in mtDNA. Ornithology 138:1–28.

Christidis, L. (Editor). (2018). The Howard and Moore Complete Checklist of the Birds of the World, v 4.1. https://www.howardandmoore.org

Collar, N. J. (1997). Taxonomy and conservation: Chicken and egg. Bulletin of the British Ornithologists' Club 117:122–136.

Collar, N. J. (2018). Taxonomy as tyranny. The Ibis 160:481–484.

Collar, N. J., L. D. C. Fishpool, J. del Hoyo, J. D. Pilgrim, N. Seddon, C. N. Spottiswoode, and J. A. Tobias (2016). Toward a scoring system for species delimitation: A response to Remsen. Journal of Field Ornithology 87:104–115.

Collinson, J. M., P. Dufour, A. A. Hamza, Y. Lawrie, M. Elliott, C. Barlow, and P. A. Crochet (2017). When morphology is not reflected by molecular phylogeny: The case of three 'orange-billed terns' *Thalasseus maximus*, *Thalasseus bergii* and *Thalasseus bengalensis* (Charadriiformes: Laridae). Biological Journal of the Linnean Society 20:1–7.

Cros, E., and F. E. Rheindt (2017). Massive bioacoustic analysis suggests introgression across Pleistocene land bridges in *Mixornis* tit-babblers. Journal of Ornithology 158:407–419.

Dai, C., F. Dong, and X. Yang (2019). Morphotypes or distinct species? A multilocus assessment of two East Asian scimitar babblers (Aves, Timaliidae). Zoologica Scripta 49:265–279.

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

Donegan, T. M. (2018). What is a species? A new universal method to measure differentiation and assess the taxonomic rank of allopatric populations, using continuous variables. ZooKeys 757:1–67.

Donegan, T., A. Quevedo, J. C. Verhelst, O. Cortés-Herrera, T. Ellery, and P. Salaman (2015). Revision of the status of bird species occurring or reported in Colombia, with discussion of BirdLife International's new taxonomy. Conservación Colombiana 23:3–48.

Donegan, T. M., J. C. Verhelst, T. Ellery, O. Cortés-Herrera, and P. Salaman (2016). Revision of the status of bird species occurring or reported in Colombia 2016 and assessment of

BirdLife International's new parrot taxonomy. Conservación Colombiana 24:12–36.

van Doren, B. M., B. G. Freeman, N. Aristizabal, M. Alvarez-R, J. Pérez-Emán, A. M. Cuervo, and G. A. Bravo (2018). Species limits in the Rusty-breasted Antpitta (*Grallaricula ferrugineipectus*) complex. The Wilson Journal of Ornithology 130:152–167.

Drovetski, S. V., A. B. Reeves, Y. A. Red'kin, I. V. Fadeev, E. A. Koblik, V. N. Sotnikov, and G. Voelker (2018). Multi-locus reassessment of a striking discord between mtDNA gene trees and taxonomy across two congeneric species complexes. Molecular Phylogenetics and Evolution 120:43–52.

van Els, P., and H. V. Norambuena (2018). A revision of species limits in Neotropical pipits *Anthus* based on multilocus genetic and vocal data. The Ibis 160:158–172.

Fernando, S. P., D. E. Irwin, and S. S. Seneviratne (2016). Phenotypic and genetic analysis support distinct species status of the Red-backed Woodpecker (Lesser Sri Lanka Flameback: *Dinopium psarodes*) of Sri Lanka. The Auk: Ornithological Advances 133:497–511.

Ferreira, M., A. Aleixo, C. C. Ribas, and M. P. D. Santos (2017). Biogeography of the Neotropical genus *Malacoptila* (Aves: Bucconidae): The influence of the Andean orogeny, Amazonian drainage evolution and palaeoclimate. Journal of Biogeography 44:748–759.

Franz, I., D. J. Alvares, and M. Borges-Martins (2020). Species limits in the Tawny-crowned Pygmy-tyrant *Euscarthmus meloryphus* complex (Aves: Passeriformes: Tyrannidae). Zootaxa 4809:475–495.

Freeman, B. G., and G. A. Montgomery (2017). Using song playback experiments to measure species recognition between geographically isolated populations: A comparison with acoustic trait analyses. The Auk: Ornithological Advances 134:857–870.

Friis, G., P. Aleixandre, R. Rodríguez-Estrella, A. G. Navarro-Sigüenza, and B. Milá (2016). Rapid postglacial diversification and long-term stasis within the songbird genus *Junco*: Phylogeographic and phylogenomic evidence. Molecular Ecology 25:6175–6195.

Garnett, S. T., and L. Christidis (2017). Taxonomy anarchy hampers conservation. Nature 546:25–27.

Gwee, C. Y., L. Christidis, J. A. Eaton, J. A. Norman, C. R. Trainor, P. Verbelen, and F. E. Rheindt (2017). Bioacoustic and multi-locus DNA data of *Ninox* owls support high incidence of extinction and recolonisation on small, low-lying islands across Wallacea. Molecular Phylogenetics and Evolution 109:246–258.

Gwee, C. Y., J. A. Eaton, K. M. Garg, P. Alström, S. van Balen, R. O. Hutchinson, D. M. Prawiradilaga, M. H. Le, and F. E. Rheindt (2019). Cryptic diversity in *Cyornis* (Aves: Muscicapidae) jungle-flycatchers flagged by simple bioacoustic approaches. Zoological Journal of the Linnean Society 186:725–741.

Gwee, C. Y., Q. L. Lee, S. P. Mahood, H. Le Manh, R. Tizard, K. Eiamampai, P. D. Round, and F. E. Rheindt (2020). The interplay of colour and bioacoustic traits in the differentiation of a Southeast Asian songbird complex. Molecular Ecology 30:297–309.

Haffer, J. (1992). The history of species concepts and species limits in ornithology. Bulletin of the British Ornithologists' Club 112A:107–158.

Haffer, J. (1997). Species concepts and species limits in ornithology. In Handbook of the Birds of the World (J. del Hoyo, A. Elliott, and J. Sargatal, Editors). Lynx Edicions, Barcelona, Spain. Vol. 4, pp. 11–24.

Halley, M. R., J. C. Klicka, P. R. Sesink Clee, and J. D. Weckstein (2017). Restoring the species status of *Catharus maculatus* (Aves: Turdidae), a secretive Andean thrush, with a critique of the yardstick approach to species delimitation. Zootaxa 4276:387–404.

Handbook of the Birds of the World and BirdLife International (2020). Handbook of the Birds of the World and BirdLife International Digital checklist of the birds of the world. Version 5. http://datazone.birdlife.org/userfiles/file/Species/Taxonomy/HBW-BirdLife_Checklist_v5_Dec20.zip

Hayes, F. E., H. D. Pratt, and C. J. Cianchini (2016). The avifauna of Kosrae, Micronesia: History, status, and taxonomy. Pacific Science 70:91–127.

Helbig, A. J., A. G. Knox, D. T. Parkin, G. Sangster, and M. Collinson (2002). Guidelines for assigning species rank. The Ibis 144:518–525.

Hey, J., R. S. Waples, M. L. Arnold, R. K. Butlin, and R. G. Harrison (2003). Understanding and confronting species uncertainty in biology and conservation. Trends in Ecology & Evolution 18:597–603.

Hosner, P. A., L. C. Campillo, M. J. Andersen, L. A. Sánchez-González, C. H. Oliveros, R. C. Urriza, and R. G. Moyle (2018). An integrative species delimitation approach reveals finescale endemism and substantial unrecognized avian diversity in the Philippine Archipelago. Conservation Genetics 19:1153–1168.

Houle, D., C. Pélabon, G. P. Wagner, and T. F. Hansen (2011). Measurement and meaning in biology. The Quarterly Review of Biology 86:3–34.

del Hoyo, J., A. Elliott, J. Sargatal, and D. A. Christie (1992–2011). Handbook of the Birds of the World. Vols. 1–16. Lynx Edicions, Barcelona, Spain.

del Hoyo, J., and N. J. Collar (2014–2016). The HBW–BirdLife International Illustrated Checklist of the Birds of the World. Vols. 1–2. Lynx Edicions, Barcelona, Spain.

Huang, Z., F. Tu, and D. Ke (2017). Complete mitochondrial genome of Blue-throated Bee-eater *Merops viridis* (Coraciiformes: Meropidae) with its taxonomic consideration. Pakistan Journal of Zoology 49:79–84.

Hung, C. M., H. Y. Hung, C. F. Yeh, Y. Q. Fu, D. Chen, F. Lei, C. T. Yao, C. J. Yao, X. J. Yang, Y. T. Lai, and S. H. Li (2014). Species delimitation in the Chinese Bamboo Partridge *Bambusicola thoracica* (Phasianidae; Aves). Zoologica Scripta 43:562–575.

Hunter, N., M. Mills, and C. Cohen (2019). Elgon Francolin *Scleroptila elgonensis* should be treated as a species distinct from Moorland Francolin *S. psilolaema*. Bulletin of the African Bird Club 26:92–100.

Huntley, J. W., and G. Voelker (2016). Cryptic diversity in Afro-tropical lowland forests: The systematics and biogeography of the avian genus *Bleda*. Molecular Phylogenetics and Evolution 99:297–308.

Irwin, D. E., A. S. Rubtsov, and E. N. Panov, (2009). Mitochondrial introgression and replacement between Yellowhammers (*Emberiza citrinella*) and Pine Buntings (*Emberiza leucocephalos*) (Aves: Passeriformes). Biological Journal of the Linnean Society 98:422–438.

Isaac, N. J., J. Mallet, and G. M. Mace (2004). Taxonomic inflation: Its influence on macroecology and conservation. Trends in Ecology & Evolution 19:464–469.

Isler, M. L., and B. M. Whitney (2017). Species limits in the genus *Thamnistes* (Aves: Passeriformes: Thamnophilidae): An evaluation based on vocalizations. Zootaxa 4291:192–200.

Isler, M. L., P. R. Isler, and B. M. Whitney (1998). Use of vocalizations to establish species limits in antbirds (Passeriformes: Thamnophilidae). The Auk 115:577–590.

Jiao, X., and Z. Yang (2021). Defining species when there is gene flow. Systematic Biology 70:108–119.

Johansson, U. S., M. Irestedt, Y. Qu, and P. G. P. Ericson (2018). Phylogenetic relationships of rollers (Coraciidae) based on complete mitochondrial genomes and fifteen nuclear genes. Molecular Phylogenetics and Evolution 126:17–22.

Johnson, N. K., J. V. Remsen, and C. Cicero (1999). Resolution of the debate over species concepts in ornithology: A new comprehensive biologic species concept. In Proceedings of the 22nd International Ornithological Congress (N. J. Adams and R. H. Slotow, Editors), BirdLife South Africa, Durban, South Africa. pp. 1470–1482.

Jønsson, K. A., K. Delhey, G. Sangster, P. G. P. Ericson, and M. Irestedt (2016). The evolution of mimicry of friarbirds by orioles (Aves: Passeriformes) in Australo-Pacific archipelagos. Proceedings of the Royal Society B 283:20160409.

Jønsson, K. A., M. P. K. Blom, P. Z. Marki, L. Joseph, G. Sangster, P. G. P. Ericson, and M. Irestedt (2019a). Complete subspecies-level phylogeny of the Oriolidae (Aves: Passeriformes): Out of Australasia and return. Molecular Phylogenetics and Evolution 137:200–209.

Jønsson, K. A., A. H. Reeve, M. P. K. Blom, M. Irestedt, and P. Z. Marki (2019b). Unrecognised (species) diversity in New Guinean passerine birds. Emu - Austral Ornithology 119:233–241.

Joseph, L., G. J. Adcock, C. Linde, K. E. Omland, R. Heinsohn, R. T. Chesser, and D. Roshier (2009). A tangled tale of two teal: Population history of the grey *Anas gracilis* and Chestnut Teal *A. castanea* of Australia. Journal of Avian Biology 40:430–439.

Kearns, A. M., M. Restani, I. Szabo, A. Schrøder-Nielsen, J. A. Kim, H. M. Richardson, J. M. Marzluff, R. C. Fleischer, A. Johnsen, and K. E. Omland (2018). Genomic evidence of speciation reversal in ravens. Nature Communications 9:906.

Kirchman, J. J., C. C. Witt, J. A. McGuire, and G. R. Graves (2010). DNA from a 100-year-old holotype confirms the validity of a potentially extinct hummingbird species. Biology Letters 6:112–115.

Kryukov, A. P., L. N. Spiridonova, S. Mori, V. Y. Arkhipov, Y. A. Red'kin, O. A. Goroshko, E. G. Lobkov, and E. Haring (2017). Deep phylogeographic breaks in Magpie *Pica pica* across the Holarctic: Concordance with bioacoustics and phenotypes. Zoological Science 34:185–200.

Kyriazis, C. C., B. Alam, M. Wjodyla, S. Hackett, P. Hosner, H. L. Mays, L. R. Heaney, and S. Reddy (2018). Colonization and diversification of the White-browed Shortwing (Aves: Muscicapidae: *Brachypteryx montana*) in the Philippines. Molecular Phylogenetics and Evolution 121:121–131.

Lees, A. L., and S. L. Pimm (2015). Species, extinct before we know them? Current Biology 25:R177–R180.

Manthey, J. D., M. Geiger, and R. G. Moyle (2017). Relationships of morphological groups in the Northern Flicker superspecies

complex (*Colaptes auratus* & *C. chrysoides*). Systematics and Biodiversity 15:183–191.

Martens, J., and N. Bahr (2016). Dokumentation neuer Vogel-Taxa, 10—Bericht für 2014. Vogelwarte 54:195–230.

Martens, J., Y. Sun, and M. Päckert (2008). Intraspecific differentiation of Sino-Himalayan bushdwelling *Phylloscopus* leaf warblers, with description of two new taxa (*P. fuscatus*, *P. fuligiventer*, *P. affinis*, *P. armandii*, *P. subaffinis*). Vertebrate Zoology 58:233–265.

Mason, N. A., A. Olvera-Vital, I. J. Lovette, and A. G. Navarro-Sigüenza (2018). Hidden endemism, deep polyphyly, and repeated dispersal across the Isthmus of Tehuantepec: Diversification of the White-collared Seedeater complex (Thraupidae: *Sporophila torqueola*). Ecology and Evolution 8:1867–1881.

Mayr, E. (1946). The number of species of birds. The Auk 63:64–69.

Mayr, E. (1969). Principles of Systematic Zoology. McGraw-Hill, New York, NY, USA.

Meiri, S., and G. M. Mace (2007). New taxonomy and the origin of species. PLoS Biology 5:e194.

Miller, M. J., E. Bermingham, B. L. Turner, J. C. Touchon, A. B. Johnson, and K. Winker (2021). Demographic consequences of foraging ecology explain genetic diversification in Neotropical bird species. Ecology Letters 24:563–571.

Moyle, R. G., J. D. Manthey, P. A. Hosner, M. Rahman, M. Lakim, and F. H. Sheldon (2017). A genome-wide assessment of stages of elevational parapatry in Bornean passerine birds reveals no introgression: Implications for processes and patterns of speciation. PeerJ 5:e3335.

Päckert, M., J. Martens, M. Wink, A. Feigl, and D. T. Tietze (2012). Molecular phylogeny of Old World swifts (Aves: Apodiformes, Apodidae, *Apus* and *Tachymarptis*) based on mitochondrial and nuclear markers. Molecular Phylogenetics and Evolution 63:606–616.

Padial, J. M., A. Miralles, I. De la Riva, and M. Vences (2010). The integrative future of taxonomy. Frontiers in Zoology 7:16.

Pearson, D. J., and D. A. Turner (2017). A taxonomic review of the genus *Zosterops* in East Africa, with a revised list of species occurring in Kenya, Uganda and Tanzania. Scopus 37:1–13.

Pedersen, M. P., M. Irestedt, L. Joseph, C. Rahbek, and K. A. Jønsson (2018). Phylogeography of a 'great speciator' (Aves: *Edolisoma tenuirostre*) reveals complex dispersal and diversification dynamics across the Indo-Pacific. Journal of Biogeography 45:826–837.

Pellegrino, I., M. Cucco, J. A. Harvey, F. Liberatore, M. Pavia, G. Voelker, and G. Boano (2017). So similar and yet so different: Taxonomic status of Pallid Swift *Apus pallidus* and Common Swift *Apus apus*. Bird Study 64:344–352.

Pérez-Emán, J. L., J. P. Ferreira, N. Gutiérrez-Pinto, A. M. Cuervo, L. N. Céspedes, C. C. Witt, and C. D. Cadena (2018). An extinct hummingbird species that never was: A cautionary tale about sampling issues in molecular phylogenetics. Zootaxa 4442:491–497.

Peters, J. L., and successors (1931–1986). Check-list birds of the world. Vols. 1–15. Harvard University Press/Museum of Comparative Zoology, Cambridge, MA, USA.

Remsen, Jr, J. V. (2016). A 'rapid assessment program' for assigning species rank? Journal of Field Ornithology 87:110–115.

Rheindt, F. E., and S. V. Edwards (2011). Genetic introgression: An integral but neglected component of speciation in birds. The Auk 128:620–632.

Sadanandan, K. R., C. Küpper, G. W. Low, C. T. Yao, Y. Li, T. Xu, F. E. Rheindt, and S. Wu (2019). Population divergence and gene flow in two East Asian shorebirds on the verge of speciation. Scientific Reports 9:8546.

Sadanandan, K. R., G. W. Low, S. Sridharan, C. Y. Gwee, E. Y. X. Ng, P. Yuda, D. M. Prawiradilaga, J. G. H. Lee, A. Tritto, and F. E. Rheindt (2020). The conservation value of admixed phenotypes in a critically endangered species complex. Scientific Reports 10:15549.

Sandoval, L., P. P. Bitton, A. D. Demko, S. M. Doucet, and D. J. Mennill (2017). Phenotypic variation and vocal divergence reveals a species complex in White-eared Ground-sparrows (Cabanis) (Aves: Passerellidae). Zootaxa 4291:155–170.

Shakya, S. B., J. Fuchs, J. M. Pons, and F. H. Sheldon (2017). Tapping the woodpecker tree for evolutionary insight. Molecular Phylogenetics and Evolution 116:182–191.

Sharpe, R. B. (1899–1909). A Hand-List of the Genera and Species of Birds. Trustees of the British Museum, London, UK.

Tobias, J. A., N. Seddon, C. N. Spottiswoode, J. D. Pilgrim, L. D. C. Fishpool, and N. J. Collar (2010). Quantitative criteria for species delimitation. The Ibis 152:724–746.

Tobias, J. A., J. Ottenburghs, and A. Pigot (2020). Avian diversity: Speciation, macroevolution and ecological function. Annual Review of Ecology, Evolution and Systematics 51:533–560.

Toews, D. P., S. A. Taylor, R. Vallender, A. Brelsford, B. G. Butcher, P. W. Messer, and I. J. Lovette (2016). Plumage genes and little else distinguish the genomes of hybridizing warblers. Current Biology 26:2313–2318.

Töpfer, T., L. Podsiadlowski, and K. Gedeon (2014). Rediscovery of the Black-fronted Francolin *Pternistis* (*castaneicollis*) *atrifrons* (Conover, 1930) (Aves: Galliformes: Phasianidae) with notes on biology, taxonomy and conservation. Vertebrate Zoology 64:261–271.

Turner, D. A., F. Dowsett-Lemaire, and R. J. Turner (2020). Further support for the specific distinctness of Elgon Francolin *Scleroptila elgonensis*. Bulletin of the African Bird Club 27:80–83.

Wang, X., P. Que, G. Heckel, J. Hu, X. Zhang, C. Y. Chiang, N. Zhang, Q. Huang, S. Liu, J. Martinez, et al. (2019). Genetic, phenotypic and ecological differentiation suggests incipient speciation in two *Charadrius* plovers along the Chinese coast. BMC Evolutionary Biology 19:135.

Winker, K. (2009). Reuniting phenotype and genotype in biodiversity research. BioScience 59:657–665.

Winker, K. (2016). An examination of species limits in the *Aulacorhynchus* "*prasinus*" toucanet complex (Aves: Ramphastidae). PeerJ 4:e2381.

Winker, K., D. A. Rocque, T. M. Braile, and C. L. Pruett (2007). Vainly beating the air: Species-concept debates need not impede progress in science or conservation. In Festschrift for Ned K. Johnson: Geographic Variation and Evolution in Birds (C. Cicero and J. V. Remsen, Jr. Editors). Ornithological Monograph 63:30–44.

Zhang, D., L. Tang, Y. Cheng, Y. Hao, Y. Xiong, G. Song, Y. Qu, F. E. Rheindt, P. Alström, C. Jia, and F. Lei (2019). 'Ghost introgression' as a cause of deep mitochondrial divergence in a bird species complex. Molecular Biology and Evolution 36:2375–2386.