



Commentary

Is it a species?

KEVIN WINKER*

University of Alaska Museum, 907 Yukon Drive,
Fairbanks, AK 99775, USA

Plumbing the depths of biodiversity, and describing and studying the wondrous variation among Earth's living creatures require that they be classified and placed on the tree of life. But the fundamental assessment of species-level diversity has yet to be fully completed, even in a comparatively well-studied group such as birds. When confronted with remarkable animals such as the Kagu *Rhynchotus jubatus* or the Shoebill *Balaeniceps rex*, which each represents an entire (monotypic) avian family and is morphologically highly distinct, we can all agree that they represent distinct species. But the evolutionary process of differentiation begins at small scales and progresses gradually through rather subjective stages, to which we have given labels such as subspecies, semi-species, species, superspecies, subgenera, genera, sub-families, families and so forth. Determining when two lineages have diverged sufficiently to call them full species has been a long-standing problem in biology (Darwin 1859, Mayr 1982). This problem is most difficult when the populations in question do not occur together but exist apart, in allopatry.

The definition of species has changed through time, and at present there are numerous competing concepts and definitions (Mayr 1982, Mayden 1997). But the problem of discerning what is a species and what is not is inherent in any definition, as doing so represents the human imposition of categories on various stages of the continuous process of differentiation. In the case of speciation, there is no widely accepted easy or obvious threshold indicating when it has been completed, short of two lineages coming into contact and not freely interbreeding. However, many lineages that may be species do not provide such a convenient test, leaving us with the need to infer taxonomic status based on other criteria. The answer to the question 'Is it a species?' can affect a diverse array of people, from scientists, birders, wildlife managers and conservationists, to developers and politicians.

In this issue, Tobias *et al.* (2010) take a major step forward in developing a pragmatic approach to determining species limits in birds. The system they develop follows

the classic methodology of inferring species limits between allopatric forms by contrasting them with congeneric species (or other close relatives) that are sympatric or parapatric (i.e. the forms come into contact to a greater or lesser extent). If the form(s) in question are as different from each other as congeners that demonstrate reproductive isolation, then by inference we can classify them as full species. If not, then they are probably subspecies, assuming that sufficient differences exist to posit the initial question of species limits. Although this methodology has deep historical roots, it has usually not been placed into a quantitative framework and it has not been extended to fit an entire organismal class (Mayr 1969, Haffer 2007). What is perhaps most refreshing is that the system that Tobias *et al.* develop focuses on phenotype, and it actually might work in a majority of cases.

Because of disagreement on species concepts, this issue is dealt with right off. The authors briefly summarize three of these concepts, two that have emerged from cladistics (the phylogenetic and monophyletic species concepts, PSC and MSC, respectively), and the biological species concept (BSC). After circumscribing some of the debate and choosing the BSC as their operative framework, they effectively set the debate aside and get on with the job. This is a necessary approach to separating the philosophical question 'What is a species?' from the practical question of whether a particular lineage is a species. No progress would occur if we had to have a definitive answer to the philosophical question, because it has been debated for centuries. The authors acknowledge that all three concepts have commonalities, placing them in the family of the general lineage species concept (de Queiroz 2007), but one might observe that some of the most intractable issues occur in families. Therefore, choosing a definition and rigorously addressing the practical problem is critical (Winker & Haig 2010). No species concept is perfect, all have some inherent subjectivity, and in birds the biological species concept remains dominant. This is true in areas as diverse as the ever-growing number of 'Birds of...' treatises and in national and international law (e.g. US Endangered Species Act and the Convention on International Trade in Endangered Species). Why this is so might be summed up with two observations. First, the application of cladistic principles to population-level phenomena (e.g. PSC) is problematic (Avise 2000). The answers one obtains in terms of species delimitation are likely to vary depending on what characters are used, and with increasing numbers of characters one can cladistically diagnose lineages with increasing resolution (to the level of the individual); the strength of these methods lies elsewhere. Secondly, we as humans consider ourselves, *Homo sapiens*, a biological species; not only is this unlikely to change, but it provides a common framework for our understanding of biodiversity. These are my own observations. The authors themselves are

*Email: kevin.winker@alaska.edu

not so forceful in their adoption of the BSC, but their effort would have had substantially less utility if they had chosen another species concept.

In developing their criteria, Tobias *et al.* focus on phenotypic evidence and set aside genetic data. They do this for two reasons: (i) genetic data are not yet sufficiently available to be widely applied in the manner done here with phenotype; and (ii) there is no widespread agreement on how genetic data can or should be used to delimit avian species. They point out, however, that the approach they develop will work with any type of quantitative data. With the increasing importance and availability of genetic data, and with the widespread desire to have some easy genetic yardstick of differentiation to apply to species delimitation, it was critical to prop open this doorway. As much as we would like to see a clear, simple and correct genetic solution to this long-standing problem, it remains unlikely that aspects of genetic differentiation alone will provide such a master key (Price 2008). However, in conjunction with phenotype, genetic evidence is playing an increasingly important role in species delimitation. The practical effect of setting genetic evidence aside here is that the methods proposed are applicable by scientists everywhere, with minimal technological investment.

Tobias *et al.* develop their phenotypically based taxonomic method using 58 species pairs (from 29 families) that come into contact and demonstrate sufficient reproductive isolation to be widely accepted as biological species. They then measured or categorized differences between lineage pairs in multiple characters chosen from four character suites: morphometrics, plumage (or bare parts) colour or pattern, song and behaviour or ecology. The basis of their quantitative approach is to categorize degrees of difference in each character as minor, medium, major or exceptional, and then to assign a corresponding numerical score.

In the method developed here, it is necessary to meld together truly quantitative measures with qualitative (subjective) categorizations of phenotypic divergence. This is effectively the only practical way to be able quickly to put together datasets on character size differences (e.g. wing length or song duration) with those of character quality differences (e.g. plumage colour or pattern). For their qualitative characters, Tobias *et al.* provide written guidelines for categorization. For their quantitative characters, they make an important departure from *P*-values, which we all know to be affected by sample size and which at times indicate significant differences that are biologically trivial. Instead, they use effect sizes upon which to base their categories of degrees of divergence (minor to exceptional). As used here, effect sizes provide a standardized measure of the magnitude of the difference between two lineages in the measured traits. Although the use of effect size in taxonomy can be traced back to the 1940s (Mayr 1969, Patten 2010),

it has only recently gained popularity, probably because of its utility in enabling direct comparisons among heterogeneous datasets (such as vocalizations and morphometrics, as used here), and among different studies (Whitlock & Schluter 2009). Tobias *et al.* use effect sizes in an atypical way, however, choosing to interpret them in terms of categorical levels of divergence, thus ultimately giving them scores of 1–4 depending on their magnitude.

But what should be measured? Ideally, in species delimitation we want to focus on a relatively small set of phenotypic attributes (and genetic, if possible) that are widely definitive of species limits. It is a biologically complex problem, though, and in some ways analogous to determining the causes of complex disease in humans; simple answers tend not to work (Buchanan *et al.* 2006). All characters or character suites are not equally important in all lineages undergoing speciation. Early results in whitefish (*Coregonus* spp.), for example, suggest that genes associated with metabolism are more important in divergence than genes associated with morphology and behaviour (Whiteley *et al.* 2008).

Using what knowledge we have about avian speciation, it is reasonable to select from the character suites that Tobias *et al.* use, and in their system, 'character shopping' within these suites is both possible and expected. The problem of cultural evolution in learned song, which is not a small one, is dealt with by focusing on vocal attributes thought to have a high level of heritability, such as peak frequency and pace. An indication that this approach is successful is that within Passeriformes, 33 species pairs of oscine (learned song) and 17 pairs of suboscine (endogenous song) did not show significant differences in mean unsigned effect size for any of the seven vocal characters used. Vocal characters did have more variation than morphometric ones, but this might be affected by the smaller sample sizes used.

Scores are summed across all characters measured to indicate how far a taxon pair has diverged, and based on these empirical datasets the authors then derived a more general method. The desire to have criteria that work across the entire class Aves requires some tailoring to come up with a universal yardstick that puts about 95% of the focal species pairs into species status. To counter overemphasis of the trivial, the authors capped the number and nature of characters that can be used. They also limited the scoring potential on ecological and behavioural characters. Weighting characters has a long history in taxonomy, and *a posteriori* character-weighting such as that used here can be very effective (Mayr 1964). Although geographical relationships were also scored (0–3, based on allopatric, broadly hybridizing, narrowly hybridizing or parapatric situations), this was snipped off in the tailoring.

Originally, the methodology of inferring the taxonomic status of allopatric forms by studying the degrees

of difference among closely related lineages that are clearly species was not intended to work across an entire class. It is a big step indeed to expand the approach to cover all of Aves and to arrive at a 'one-size-fits-all' threshold number (7.0) indicating that speciation has occurred. The seductive allure of the number seven may not be fully appreciated, but it has almost mystical properties, from days of the week, to pillars of wisdom and luck in gambling, to real psychological attributes (Miller 1956).

A test of the system was made among 23 pairs of European subspecies, though without using vocal data. Two of these taxon pairs would qualify for elevation to full species under their criteria. It is also inferred, but not demonstrated, that application of the method will produce many changes in tropical bird taxonomy. The real test of the system developed will be, as the authors recognize, among additional species and subspecies that are widely accepted.

What if the answer is that a lineage is not a full species? That answer does not mean that the lineage is unimportant; good subspecies, for example, also have evolutionary and conservation importance (Winker & Haig 2010). But we can expect a failure rate of at least 5% with this method, which is likely to include both false positives and false negatives. If applied to hybridizing populations as the authors propose, examining only pure phenotypes, this error rate is likely to be higher because it unduly diminishes the importance of gene flow and the degrees of reproductive isolation achieved in secondary contact. The breadth of hybrid zones was part of the geographical scoring, but that is a poor surrogate (and was discarded in the method's development), and phenotype is not always a dependable indicator of gene flow. Hybridization is a particularly thorny issue in birds, given its frequency, and one can argue that under any species concept, sampling in zones of potential gene flow between lineages is important. Although the authors spent effort on this issue, it remains a loose end.

As important as the tools of taxonomy are, too much focus on them can divert us from the more fundamental question of what exactly is occurring evolutionarily to cause sufficient divergence for two lineages to be species. How are phenotypic and genotypic changes favoured, disfavoured, and accrued such that speciation occurs and passes our subjective thresholds? And how different might this process be in different lineages? The interplay of characters being used on the one hand as taxonomic criteria and on the other hand to help us understand these evolutionary processes has in this study produced two important insights. The first, noted above, was that a focus on conserved elements in song can help us set aside the difficult problem of cultural evolution in taxa that have song learning. The second is that morphological divergence and vocal divergence are negatively correlated among the diverse taxa sampled – not unexpected,

but a nice result to see in developing a method intended to work across all birds.

Bird collections are at the heart of this continually developing frontier of taxonomic research. Both traditional and non-traditional (e.g. tissues, song) collections will clearly be fundamental for continued advances. We need to look to the health of these resources if we are to achieve robust success, particularly as we move into this more quantitative era and sample sizes and power analyses become increasingly important (Winker 1996, 2009, Stoeckle & Winker 2009). Furthermore, we need to recognize that recent samples are important; significant morphometric and genetic changes can occur within decades with environmental change (e.g. Pergams & Lacy 2007).

While this approach has broad utility, it will need to be used with care. Missing data, sample sizes and sample distribution, data quality and sources, and other handicaps to the taxonomist are perhaps even more important to consider when applying a relatively simple method so broadly. I am not convinced of the use of illustrations to score characters, and I look forward to continued development and publication (with larger sample sizes) of the dependable and undependable attributes of learned vocalizations in species delimitation. The authors are also realistic in providing cautions.

This system will not make species delimitation wholly objective, but it is a substantial step forward. Despite the subjectivity involved in selecting which taxa and characters to include, in choosing categorical bins of the magnitudes of differences between two lineages, in capping variable inclusion, and in choosing a single speciation threshold indicator, there is a great advantage to be had in standardizing evaluations of differences across multiple character suites. Critics might focus on the method's subjectivity and the fact that its quantitative basis rests upon somewhat arbitrary magnitudes of difference. However, by demonstrating repeatability, by rendering what have historically been idiosyncratic judgment calls (albeit usually by experts) into a quantitative framework, and by insisting that multiple diverse character suites have been sampled, the authors firmly place a degree of consistency and transparency upon taxonomic decisions, and this will be widely welcomed. A method like this can certainly clean up and straighten out the road a bit in the twisty, brushy areas of taxonomy, opening the way for more detailed study as time and resources permit.

Tobias *et al.* recognize that their method serves as a baseline that will be modified. They view likely improvements as coming from better measurements, more reliance directly on effect sizes, and the possible inclusion of genetic data (though they emphasize that retention of the method's simplicity is important for data-poor situations). I am fully supportive of the goals and advances made in this study, and look forward to seeing

this method put widely to the test. One threshold may not work satisfactorily across such a diverse group, but the method is very well suited for pursuing 'local optima', e.g. character sets and threshold values tailored to specific subsets of taxa, more as the classic method was originally conceived. It seems likely that speciation and its cues differ sufficiently among lineages that such focused pursuits will also be rewarding, and this framework is well suited to such applications. We can also look forward to the rapidly expanding frontiers of genomics and phenomics (Houle 2010) to provide a better understanding of the process of differentiation and to improve our focus on the genotypic and phenotypic characteristics most important in speciation.

REFERENCES

- Avise, J.** 2000. Cladists in wonderland. *Evolution* **54**: 1828–1832.
- Buchanan, A.V., Weiss, K.M. & Fullerton, S.M.** 2006. Dissecting complex disease: the quest for the Philosopher's Stone? *Int. J. Epidemiol.* **35**: 562–571.
- Darwin, C.** 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.
- Haffer, J.** 2007. *Ornithology, Evolution, and Philosophy: The Life and Science of Ernst Mayr 1904–2005*. Berlin: Springer.
- Houle, D.** 2010. Numbering the hairs on our heads: the shared challenge and promise of phenomics. *Proc. Natl Acad. Sci. USA* **107**(suppl. 1): 1793–1799.
- Mayden, R.L.** 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. In Claridge, M.F., et al. (eds) *Species: The Units of biodiversity*: 381–424. New York: Chapman & Hall.
- Mayr, E.** 1964. The new systematics. In Leone, C.A. (ed.) *Taxonomic Biochemistry and Serology*: 13–32. New York: The Ronald Press Company.
- Mayr, E.** 1969. *Principles of Systematic Zoology*. New York: McGraw-Hill Book Company.
- Mayr, E.** 1982. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Cambridge, MA: Belknap Press.
- Miller, G.A.** 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* **101**: 343–352.
- Patten, M.** 2010. Null expectations in subspecies diagnosis. *Ornithol. Monogr.* **67**: 35–41.
- Pergams, O.R.W. & Lacy, R.C.** 2007. Rapid morphological and genetic change in Chicago-area *Peromyscus*. *Mol. Ecol.* **17**: 450–463.
- Price, T.** 2008. *Speciation in Birds*. Greenwood, CO: Roberts and Company Publishers.
- de Queiroz, K.** 2007. Species concepts and species delimitation. *Syst. Biol.* **56**: 879–886.
- Stoeckle, M. & Winker, K.** 2009. A global snapshot of avian tissue collections: state of the enterprise. *Auk* **126**: 684–687.
- Tobias, J.A., Seddon, N., Spottiswoode, C.N., Pilgrim, J.D., Fishpool, L.D.C. & Collar, N.J.** 2010. Quantitative criteria for species delimitation. *Ibis* **152**: 724–746.
- Whiteley, A.R., Derome, N., Rogers, S.M., St-Cyr, J., Laroche, J., Labbe, A., Nolte, A., Renaut, S., Jeukens, J. & Bernatchez, L.** 2008. The phenomics and expression quantitative trait locus mapping of brain transcriptomes regulating adaptive divergence in lake whitefish species pairs (*Coregonus* spp.). *Genetics* **180**: 147–164.
- Whitlock, M.C. & Schluter, D.** 2009. *The Analysis of Biological Data*. Greenwood Village, CO: Roberts and Company Publishers.
- Winker, K.** 1996. The crumbling infrastructure of biodiversity: the avian example. *Conserv. Biol.* **10**: 703–707.
- Winker, K.** 2009. Reuniting genotype and phenotype in biodiversity research. *Bioscience* **59**: 657–665.
- Winker, K. & Haig, S.M.** (eds). 2010. Avian subspecies. *Ornithol. Monogr.* **67**: i–viii, 1–200.