Point of View

Syst. Biol. 61(1):165–169, 2012

DOI:10.1093/sysbio/syr030
Advance Access publication on April 11, 2011

Determining Species Boundaries in a World Full of Rarity: Singletons, Species Delimitation Methods

GWYNNE S. LIM[1], MICHAEL BALKE[2], AND RUDOLF MEIER[1,3,*]

[1]*Department of Biological Sciences, National University of Singapore, Science Drive 4, Singapore 117543, Singapore;* [2]*Zoologische Staatsammlung, Muenchhausenstrasse 21, 81247 Munich, Germany; and* [3]*University Scholars Programme, National University of Singapore, Science Drive 4, Singapore 117543, Singapore;*
[*]*Correspondence to be sent to: Department of Biological Sciences and University Scholars Programme, National University of Singapore, Science Drive 4, Singapore 117543, Singapore; E-mail:meier@nus.edu.sg.*

Received 14 May 2010; reviews returned 29 July 2010; accepted 7 January 2011
Associate Editor: Mark Fishbein

Singletons—species only known from a single specimen—and uniques—species that have only been collected once—are very common in biodiversity samples. Recent reviews suggest that in tropical arthropod samples, 30% of all species are represented by only one specimen (Bickel 1999; Novotny and Basset 2000; Coddington et al. 2009), with additional sampling helping little with eliminating rarity. Usually, such sampling only converts some of the singleton species to doubletons, with new singleton species being discovered in the process (Scharff et al. 2003; Coddington et al. 2009). Here, we first demonstrate that rare species are similarly common in specimen samples used for taxonomic research before we argue that the phenomenon of rarity has been insufficiently considered by the new quantitative techniques for species delimitation.

Addressing this disconnect between theory and reality is pressing given that the last decade has seen a renewed interest in methods for species identification and delimitation (Sites and Marshall 2004; O'Meara 2010). Much of this interest has been fuelled by the availability of DNA sequences (Meier 2008). However, many newly proposed techniques implicitly or explicitly assume that all populations and species can be well sampled. But what is the value of these techniques if many species have only been collected once and/or are only known from one specimen? Here, we argue that all existing techniques need to be modified to accommodate the commonness of rarity and that all future techniques should be explicit about how rare species can be discovered and treated.

Rare Species are Common in Taxonomic Treatises

In order to determine whether rarity is also a common phenomenon in specimen samples that are used for species descriptions, we conducted a survey of the taxonomic literature. An initial screen involving the monograph series of the American Museum of Natural History (AMNH) mostly yielded arthropod data, which we then complemented with data from *Systematic Botany* for plants and *Zootaxa* for nonarthropod animals. For each taxon, we initially surveyed the most recent 10 papers with species descriptions (Plants, Porifera, Platyhelminthes, Nematoda, and Mollusca), but for some taxa, additional publications had to be evaluated in order to obtain data for at least 20 new species (Coelenterata, Annelida, Tardigrada, Echinodermata, Pisces, Amphibia, Reptilia, and Mammalia). Note that for Bryozoa, Protozoa, Nemertea, and Aves fewer than 20 species are described in *Zootaxa*.

The invertebrate papers in the AMNH publications (2000–2010: $N = 25$) covered 695 new species (Coleoptera, Heteroptera, Hymenoptera, Araneae, Scorpiones, and Arhynchobdellida) and revealed that one in six newly described species is only known from one specimen (singleton: $123/695 = 17.7\%$). One in four species is described from only one locality (uniques: $191/695 = 27.5\%$). For vertebrates, the proportions of singletons (19%) and uniques (35%) are even higher ($N = 89$ species), whereas the remaining invertebrate taxa have a lower proportion of singletons (13%), but a very high proportion of uniques (57%; $N = 261$ species). Only plant species are rarely described based on singletons (8%) or unique specimens (17%; $N = 66$ species). Overall, the proportion of rare species in taxonomy is thus similar to what has been reported for biodiversity samples where the proportions are usually around 30% (Coddington et al. 2009). Given that some systematists prefer not to describe species based on one specimen, it is not surprising that the proportion of rare species in collections is even higher (Petersen and Meier 2003; Meier and Dikow 2004). This means that rarity is not only a prevalent phenomenon in "species discovery" studies but also in research that is based on specimen samples that accumulated over decades and are stored in many natural history museums.

Despite centuries of taxonomic work, systematists have only recently intensified efforts to move taxonomy from a nonquantitative science to a numerical, analytical

one. This interest increased when DNA sequences became available because such data are arguably more amenable to quantification given that many morphological characters are difficult to quantify (e.g., genitalia shape). In order to delimit species, many of the new methods designed for DNA sequences aim to detect a change in the quality of the evolutionary signal that may reflect a shift from intraspecific variation to interspecific isolation. Here, we briefly present the different species delineation approaches and discuss if and how rarity is accommodated.

DNA Barcoding

DNA barcoding was initially presented as a method for species identification (Hebert et al. 2003), but it now also encompasses species delimitation and/or prediction thus blurring the distinction between DNA barcoding and DNA taxonomy (Meier et al. 2006; Vogler and Monaghan 2007; Meier 2008). Either tree-based on cluster-based techniques are used for analyzing the DNA barcode data (Meier 2008). We would argue that tree-based techniques alone cannot reliably identify singleton species because recognizing a particular branch as representing such a species requires a threshold value for branch lengths and/or pairwise distances that distinguish intra- from interspecific variability (Chang et al. 2008; Gustafsson et al. 2009; Trontelj et al. 2009; Vieites et al. 2009). Such a threshold is also needed when DNA barcode data are analyzed using clustering approaches and the species delimitations are then entirely based on distances. One technical problem with clustering is that the pairwise distances for three or more sequences do not have to be equal so that the application of a strict threshold is often impossible (Meier et al. 2006). One solution is the use of algorithms that yield clusters where each sequence has at least one other sequence with a distance below the threshold, but where some distances are allowed to exceed this threshold (Meier 2008). However, regardless of which clustering technique is used, a rare species can in principle be recognized by threshold-based techniques as long as its sequence is so distinct that it does not group with sequences from other species.

The main problem with distance-based approaches lies elsewhere. There is no theoretical reason to believe that there would be a satisfactory universal threshold given that speciation is not a clockwise process (Meier et al. 2008). Not surprisingly, the search for an appropriate threshold has thus yielded many different values and the results have been criticized as arbitrary. For example, Hebert et al. (2003) initially suggested 3%, whereas the Barcode of Life Data Systems use 1% for identification purposes (Ratnasingham and Hebert 2007). Others have proposed relative thresholds such as 10x the intraspecific variability which, however, presupposes knowledge of intra- versus interspecific boundaries (Hebert et al. 2004). Therefore, although DNA-barcoding type analyses can accommodate rare species, this advantage comes at the expense of having to use distance thresholds that are hard to justify (Huang et al. 2008; Memon et al. 2006; Meier et al. 2008).

Population Aggregation and Cladistic Haplotype Analyses

Population aggregation analysis (PAA) was proposed for the delimitation of phylogenetic species (Davis and Nixon 1992). It starts with identifying individuals belonging to a population. The attributes of these individuals are collected and used to characterize the population. Invariable attributes are classified as "characters" and variable attributes as "traits." Different populations are assigned to the same species if they lack distinguishing characters. The results are species that are "the smallest aggregation of (sexual) populations or (asexual) lineages diagnosable by a unique combination of character states" (Davis and Nixon 1992). A modification of PAA is the cladistic haplotype analysis (CHA; Brower 1999) that differs from PAA in that it is tree-based and explicitly designed for molecular data. It was proposed because the use of "unique combinations of character states" to delimit species has the disadvantage that, for example, a reversal in a species defining DNA sequence character means that the individual does no longer have the "unique combination" and loses species membership. In CHA, the species membership of such specimens is maintained if it can be shown that the new haplotype is derived from one that had the unique combination of character states that characterized the species.

Regardless of whether PAA or CHA is used, it is not clear how a singleton species can be accommodated in the initial step. The distinction between character and trait is important for PAA and CHA, but it requires more than one specimen in order to test for the fixation of an attribute and its classification as a trait or character. Furthermore, even for nonsingleton samples, PAA and CHA require prior knowledge of population limits, which are notoriously difficult to obtain (Laamanen et al. 2003) and require multiple specimens. By definition, such dense taxon samples are unattainable for rare species (Wiens and Servedio 2000).

Coalescence

Coalescence methods for species delimitations were originally developed to reconstruct recent speciation events and to identify species that have not yet attained reciprocal monophyly. This is accomplished by modeling lineage sorting probabilistically (Carstens and Knowles 2007; Knowles and Carstens 2007). Not all methods and algorithms are explicit about how densely species have to be sampled in order for these methods to be successful, but some of the more commonly used software for coalescence analysis including "BEST" (Brumfield et al. 2008), "COAL," or "Brownie" (reviewed in O'Meara 2010) assume sampling frequencies

of ~5 individuals per species. Otherwise, an inadequate representation of intraspecific variability will lead to incorrect inferences. However, our survey of the biodiversity and taxonomic literature reveals that such sampling is unattainable for ca. 30% of all species, that is, the failure to account for rarity in coalescence analyses is likely to yield incorrect results for a large proportion of the species diversity.

One of the most commonly used species delimitation methods, and one that can be used on single-locus data sets, is the general mixed Yule coalescent (GMYC) analysis, first presented by Pons et al. (2006) and used in many subsequent studies (e.g., Papadopoulou et al. 2008; Bode et al. 2009; Monaghan et al. 2009; Yassin et al. 2009). It explicitly delineates intra- from interspecific branching in a repeatable manner by determining the coalescence point that marks the transition between population- and species-level relationships. Singleton species can be accommodated by determining whether they split from their sister branches before the threshold break that denotes a lineage transition (see Papadopoulou et al. 2008 for a specific example). Originally GMYC determined whether a particular clade should be considered a species by determining an age at which speciation occurs; in effect, a line was drawn across a tree with branch lengths fitted under the assumption of a molecular clock. Recent modifications of GMYC (Monaghan et al. 2009; Powell et al. 2011) allow for multiple thresholds and transition points can vary among different lineages. However, rate substitution models provide divergent results based on the treatment of data and choice of substitution models (Ho et al. 2008; Brandley et al. 2011). In addition, Lohse (2009) pointed out that the method is sensitive to undersampling because such sampling affects the estimation of the speciation point. Papadopoulou et al. (2009) replied that practically all systematics methods are unable to deal with the problems of undersampling, but this also means that in reality GMYC is not designed for a biota where ca. 30% of all species are so rare that they are only represented by one or a few specimens from one locality.

Ecological Niche Modeling

A relatively new addition to the techniques used in species delimitation and predicting species ranges is "ecological niche modeling." Different algorithms have been developed for identifying environmental variables that are associated with the known distribution of a species. These variables are then used to define the abiotic conditions within which populations of the species can be maintained. The same data can be used to predict the full species ranges (Raxworthy et al. 2007). However, according to our survey of the taxonomic literature, ca. 30% of all species are so rare that these techniques are not applicable given that ecological niche modeling requires large sample sizes in order to yield robust results (Stockwell and Peterson 2002; Wisz et al. 2008; Jimenez-Valverde et al. 2009). It thus appears that this approach is yet another example for a technique that suffers from disconnect between theoretical assumptions and the commonness of rarity.

The Systematic Challenge of Singletons

The species delimitation literature shows a surprising lack of awareness for the commonness of rarity. We conducted a keyword search ("single individual," "individual specimen," "single specimen," "single voucher," "singleton," "doubleton," "rarity", and "rare species") of significant papers describing new methods for species delimitation and or papers applying these techniques to large data sets. We find that few explicitly discuss rarity. These papers either utilize DNA barcoding (Hajibabaei et al. 2006; Smith et al. 2006, 2007, 2008) or GMYC analyses (Pons et al. 2006; Monaghan et al. 2009), whereas rarity is not discussed in character-based species delimitation methods (see Davis and Nixon 1992; Wiens and Penkrot 2002) and papers discussing coalescence methods (Knowles and Carstens 2007). DNA barcoding can accommodate singletons, but only by using hard-to-defend distance thresholds; GMYC models can accommodate singletons but yield skewed results when too many are included. These models are also making problematic assumptions about a "speciation time window" that is applied across a tree. Other techniques assume knowledge about population boundaries (PAA and CHA) and/or sampling completeness that appear unrealistic (PAA, coalescence). But techniques that are directly or indirectly built on the premise of having appropriately sampled intraspecific variability can only be a start; they need to be complemented with explicit techniques for identifying rare species for which such estimates cannot be obtained.

Rarity in Taxonomy: Past and Future

Rarity is not a new phenomenon and many singleton species are described as can be seen from our literature survey. Understanding why these species are being described based on few specimens may help with developing formalized quantitative species delimitation methods that can accommodate rarity. It appears that taxonomists describe a singleton species if it has a particularly distinct suite of morphological characters that renders it highly unlikely that it belongs to an already described species; that is, an implicit probability argument is used to justify the description. Ultimately, the argument is often based on the known intraspecific variability of closely related well-sampled species. If a species falls outside of this "normal" range, many taxonomists describe the species even if it is only known from one specimen. This approach has been explicitly rejected by some authors (e.g., Dayrat 2005), who argue that species should never be described based on one specimen because the intraspecific variability cannot be properly assessed. However, given the commonness of rarity, a strict adherence to this rule would prevent

the description of a very significant proportion of the species-level diversity.

Similar arguments for the recognition of new species based on one DNA sequence are sometimes also found in studies that use trees and/or coalescence methods for species delimitation. Here, putatively rare species are sometimes identified based on particularly discrete branches using ad hoc techniques, but most authors are only willing to assign species status if there is additional information supporting this decision (Gustafsson et al. 2009; Trontelj et al. 2009; Vieites et al. 2009; Tan et al. 2010). One can argue that this approach is justified because in studies using coalescence much of the evidence for species limits comes from coalescence points, which are by definition lacking for rare species and thus need to be replaced by other data. However, it appears to us that what is needed are more formalized approaches to recognizing singleton species based on other evidence and/or the known intraspecific variability of well-sampled species (Memon et al. 2006; Petersen et al. 2007). Fortunately, there are several approaches in statistics for recognizing "outliers" (Millar and Hamilton 1999; Fraley and Raftery 2002) or comparing a single measurement with the mean and variance of a population. It appears to us that future species delimitation techniques should apply these approaches to species recognition. Likely candidates are iterative techniques where singleton species are identified and removed prior to finding coalescence points for well-sampled species (Pagel et al. 2004).

Ecologists have long been aware and fascinated by the phenomenon of rarity in biodiversity samples (Magurran and Henderson 2003; Scharff et al. 2003; Cunningham and Lindenmayer 2005; Mao and Colwell 2005; Chao et al. 2009). This interest is not surprising given that rare species are particularly important from the points of view of conservation, ecology, and evolutionary biology. Such species are frequently also the focus for policy makers (Prendergast et al. 1993; Soltis and Gitzendanner 1999). Here, we argue that systematists should similarly embrace rare species and stop treating them as an anomaly. Instead, systematists should assume that a significant proportion of all species is rare. This applies to invertebrates and vertebrates alike while the proportion of species described based on singletons is smaller in plants. Future methods for delimiting species thus need to accommodate rarity, and we recommend that taxonomists treat singletons using ad hoc methods until these new techniques become available.

Funding

The research was supported by a grant from the Ministry of Education (R377-000-040-112).

Acknowledgments

The authors gratefully acknowledge J. Pons for his insight and comments.

References

Bickel D.J. 1999. What museum collections can reveal about species accumulation, richness, and rarity: an example from the Diptera. Mosman (Australia): Royal Zoological Society of New South Wales.

Bode S.N.S., Adolfsson S., Lamatsch D.K., Martins M.J.F., Schmit O., Vandekerkhove J., Mezquita F., Namiotko T., Rossetti G., Schön I., Butlin R.K., Martens K. 2009. Exceptional cryptic diversity and multiple origins of parthenogenesis in a freshwater ostracod. Mol. Phylogenet. Evol. 54:542–552.

Brandley M.C., Wang Y., Guo X., Montes de Oca A.N., Fería-Ortz M., Hikida T., Ota, H. 2011. Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of Plestiodon (Eumeces) lizards. Syst. Biol. 60:3–15.

Brower A.V.Z. 1999. Delimitation of phylogenetic species with DNA sequences: a critique of Davis and Nixon's population aggregation analysis. Syst. Biol. 48:199–213.

Brumfield R.T., Liu L., Lum D.E., Edwards S.V. 2008. Comparison of species tree methods for reconstructing the phylogeny of bearded Manakins (Aves: Pipridae, Manacus) from multilocus sequence data. Syst. Biol. 57:719–731.

Carstens B.C., Knowles L.L. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Syst. Biol. 56:400–411.

Chang C.H., Lin S.M., Chen J.H. 2008. Molecular systematics and phylogeography of the gigantic earthworms of the *Metaphire formosae* species group (Clitellata, Megascolecidae). Mol. Phylogenet. Evol. 49:958–968.

Chao A., Colwell R.K., Lin C.W., Gotelli N.J. 2009. Sufficient sampling for asymptotic minimum species richness estimators. Ecology. 90:1125–1133.

Coddington J.A., Agnarsson I., Miller J.A., Kuntner M.E., Hormiga G. 2009. Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. J. Anim. Ecol. 78:573–584.

Cunningham R.B., Lindenmayer D.B. 2005. Modeling count data of rare species: some statistical issues. Ecology. 86:1135–1142.

Davis J.I., Nixon K.C. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. Syst. Biol. 41:421–435.

Dayrat B. 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85:407–415.

Fraley C., Raftery A.E. 2002. Model-based clustering, discriminant analysis, and density estimation. J. Am. Stat. Assoc. 97:611–631.

Gustafsson D.R., Price D.A., Erseus C. 2009. Genetic variation in the popular lab worm *Lumbriculus variegatus* (Annelida: Clitellata: Lumbriculidae) reveals cryptic speciation. Mol. Phylogenet. Evol. 51:182–189.

Hajibabaei M., Janzen D.H., Burns J.M., Hallwachs W., Hebert P.D.N. 2006. DNA barcodes distinguish species of tropical Lepidoptera. Proc. Natl. Acad. Sci. U.S.A. 103:968–971.

Hebert P., Stoeckle M., Zemlak T., Francis C. 2004. Identification of birds through DNA barcodes. PLoS Biol. 2:E312.

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B. 270:313–321.

Ho S.Y.W., Saarma U., Barnett R., Haile J., Shapiro B. 2008. The effect of inappropriate calibration: three case studies in molecular ecology. PLoS ONE 3:e1615.

Huang D.W., Meier R., Todd P.A., Chou L.M. 2008. Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. J. Mol. Evol. 66:167–174.

Jimenez-Valverde A., Lobo J.M., Hortal J. 2009. The effect of prevalence and its interaction with sample size on the reliability of species distribution models. Community Ecol. 10:196–205.

Knowles L.L., Carstens B.C. 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56:887–895.

Laamanen T.R., Petersen F.T., Meier R. 2003. Kelp flies and species concepts—the case of *Coelopa frigida* (Fabricius, 1805) and *C. nebularum* Aldrich, 1929 (Diptera: Coelopidae). J. Zool. Syst. Evol. Res. 41:127–136.

Lohse K. 2009. Can mtDNA barcodes be used to delimit species? A response to Pons et al. (2006). Syst. Biol. 58:439–442.

Magurran A.E., Henderson P.A. 2003. Explaining the excess of rare species in natural species abundance distributions. Nature. 422:714–716.
Mao C.X., Colwell R.K. 2005. Estimation of species richness: mixture models, the role of rare species, and inferential challenges. Ecology. 86:1143–1153.
Meier R. 2008. DNA sequences in taxonomy: opportunities and challenges. In: Wheeler Q., editor. The new taxonomy. Boca Raton (FL): CRC Press.
Meier R., Dikow T. 2004. Significance of specimen databases from taxonomic revisions for estimating and mapping the global species diversity of invertebrates and repatriating reliable specimen data. Cons. Biol. 18:478–488.
Meier R., Shiyang K., Vaidya G., Ng P.K.L. 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst. Biol. 55:715–728.
Meier R., Zhang G., Ali F. 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the "Barcoding Gap" and leads to misidentification. Syst. Biol. 57:809–813.
Memon N., Meier R., Manan A., Su K.F.Y. 2006. On the use of DNA sequences for determining the species limits of a polymorphic new species in the stink bug genus *Halys* (Heteroptera: Pentatomidae) from Pakistan. Syst. Entomol. 31:703–710.
Millar A.M.L., Hamilton D.C. 1999. Modern outlier detection methods and their effect on subsequent inference. J. Stat. Comput. Simul. 64:125–150.
Monaghan M.T., Wild R., Elliot M., Fujisawa T., Balke M., Inward D.J.G., Lees D.C., Ranaivosolo R., Eggleton P., Barraclough T.G., Vogler A.P. 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. Syst. Biol. 58:298–311.
Novotny V., Basset Y. 2000. Rare species in communities of tropical insect herbivores: pondering the mystery of singletons. Oikos. 89:564–572.
O'Meara B.C. 2010. New heuristic methods for joint species delimitation and species tree inference. Syst. Biol. 59:59–73.
Pagel M., Meade A., Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. Syst. Biol. 53:673–684.
Papadopoulou A., Bergsten J., Fujisawa T., Monaghan M.T., Barraclough T.G., Vogler A.P. 2008. Speciation and DNA barcodes: testing the effects of dispersal on the formation of discrete sequence clusters. Philos. Trans. R. Soc. Lond. B. 363:2987–2996.
Papadopoulou A., Monaghan M.T., Barraclough T.G., Vogler A.P. 2009. Sampling error does not invalidate the Yule-Coalescent model for species delimitation. A response to Lohse (2009). Syst. Biol. 58: 442–444.
Petersen F.T., Damgaard J., Meier R. 2007. DNA taxonomy: how many DNA sequences are needed for solving a taxonomic problem? The case of two parapatric species of louse flies (Diptera: Hippoboscidae: *Ornithomya* Latreille, 1802). Arthr. Syst. Phyl. 65: 119–125.
Petersen F.T., Meier R. 2003. Testing species-richness estimation methods on single sample collection data using the Danish Diptera. Biodivers. Conserv. 12:667–686.
Pons J., Barraclough T.G., Gomez-Zurita J., Cardoso A., Duran D.P., Hazell S., Kamoun S., Sumlin W.D., Vogler A.P. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55:595–609.
Powell J.R., Monaghan M.T., Opik M., Rillig, M.C. 2011. Evolutionary criteria outperform operational approaches in producing ecologically relevant fungal species inventories. Mol. Ecol. 20: 655–666.
Prendergast J.R., Quinn R.M., Lawton J.H., Eversham B.C., Gibbons D.W. 1993. Rare species, the coincidence of diversity hotspots and conservation strategies. Nature. 365:335–337.
Ratnasingham S., Hebert P.D.N. 2007. BOLD: the Bbarcode of life data system (www.barcodinglife.org). Mol. Ecol. Notes. 7: 355–364.
Raxworthy C.J., Ingram C.M., Rabibisoa N., Pearson R.G. 2007. Applications of ecological niche modeling for species delimitation: a review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. Syst. Biol. 56:907–923.
Scharff N., Coddington J.A., Griswold C.E., Hormiga G., Bjorn P.D. 2003. When to quit? Estimating spider species richness in a northern European deciduous forest. J. Arachnol. 31:246–273.
Sites J.W. Jr., Marshall J.C. 2004. Operational criteria for delimiting species. Annu. Rev. Ecol. Syst. 35:199–229.
Smith M.A., Rodriguez J.J., Whitfield J.B., Deans A.R., Janzen D.H., Hallwachs W., Hebert P.D.N. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc. Natl. Acad. Sci. U.S.A. 105:12359–12364.
Smith M.A., Wood D.M., Janzen D.H., Hallwachs W., Hebert P.D.N. 2007. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proc. Natl. Acad. Sci. U.S.A. 104:4967–4972.
Smith M.A., Woodley N.E., Janzen D.H., Hallwachs W., Hebert P.D.N. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). Proc. Natl. Acad. Sci. U.S.A. 103: 3657–3662.
Soltis P.S., Gitzendanner M.A. 1999. Molecular systematics and the conservation of rare species. Cons. Biol. 13:471–483.
Stockwell D.R.B., Peterson A.T. 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148:1–13.
Tan D.S.H., Ang Y., Lim G.S., Ismail M.R.B., Meier R. 2010. From 'cryptic species' to integrative taxonomy: an iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of *Sepsis pyrrhosoma* (Sepsidae: Diptera). Zool. Script. 39:51–61.
Trontelj P., Douady C.J., Fiser C., Gibert J., Goricki S., Lefebure T., Sket B., Zaksek V. 2009. A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts? Freshwater Biol. 54:727–744.
Vieites D.R., Wollenberg K.C., Andreone F., Kohler J., Glaw F., Vences M. 2009. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. Proc. Natl. Acad. Sci. U.S.A. 106:8267–8272.
Vogler A.P., Monaghan M.T. 2007. Recent advances in DNA taxonomy. J. Zool. Syst. Evol. Res. 45:1–10.
Wiens J.J., Penkrot T.A. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Syst. Biol. 51:69–91.
Wiens J.J., Servedio M.R. 2000. Species delimitation in systematics: inferring diagnostic differences between species. Proc. R. Soc. Lond. B. 267:631–636.
Wisz M.S., Hijmans R.J., Li J., Peterson A.T., Graham C.H., Guisan A. 2008. Effects of sample size on the performance of species distribution models. Divers. Distrib. 14:763–773.
Yassin A., Amédégnato C., Cruaud C., Veuille M. 2009. Molecular taxonomy and species delimitation in Andean *Schistocerca* (Orthoptera: Acrididae). Mol. Phylogenet. Evol. 53:404–411.