Quantifying Uncertainty in Estimation of Tropical Arthropod Species Richness

Andrew J. Hamilton,[1,*] Yves Basset,[2] Kurt K. Benke,[3] Peter S. Grimbacher,[1] Scott E. Miller,[4] Vojtech Novotný,[5] G. Allan Samuelson,[6] Nigel E. Stork,[1] George D. Weiblen,[7] and Jian D. L. Yen[1]

1. Department of Resource Management and Geography, University of Melbourne, Burnley Campus, 500 Yarra Boulevard, Richmond, Victoria 3121, Australia; 2. Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Ancón, Panama; 3. Department of Primary Industries, 32 Lincoln Square North, Carlton, Parkville Centre, Victoria 3052, Australia; 4. National Museum of Natural History and National Zoological Park, Smithsonian Institution, Washington, DC 20013; 5. Biology Centre, Czech Academy of Sciences and Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic; 6. Bishop Museum, Honolulu, Hawaii 96817; 7. Department of Plant Biology, University of Minnesota, St. Paul, Minnesota 55108

Submitted November 29, 2009; Accepted March 8, 2010; Electronically published May 10, 2010

ABSTRACT: There is a bewildering range of estimates for the number of arthropods on Earth. Several measures are based on extrapolation from species specialized to tropical rain forest, each using specific assumptions and justifications. These approaches have not provided any sound measure of uncertainty associated with richness estimates. We present two models that account for parameter uncertainty by replacing point estimates with probability distributions. The models predict medians of 3.7 million and 2.5 million tropical arthropod species globally, with 90% confidence intervals of [2.0, 7.4] million and [1.1, 5.4] million, respectively. Estimates of 30 million or greater are predicted to have <0.00001 probability. Sensitivity analyses identified uncertainty in the proportion of canopy arthropod species that are beetles as the most influential parameter, although uncertainties associated with three other parameters were also important. Using the median estimates suggests that in spite of 250 years of taxonomy and around 855,000 species of arthropods already described, approximately 70% await description.

Keywords: Coleoptera, host specificity, Latin hypercube sampling, global richness estimate, stochastic model, tropical rain forest insects.

Introduction

The number of named distinct species of eukaryotic species on Earth is thought to be around 1.9 million, with approximately half of these (~1.1 million) being arthropods, predominantly insects (Chapman 2009). However, there is much greater disagreement on what the total number of species, including unnamed ones, might be (May 1988, 2000; Stork 1988, 1993; Nielsen and Mound 2000; Novotný et al. 2002). The problem of estimating global species richness has been approached from various angles, including body-size versus richness models, food-web structure, species' reporting rates, extrapolation from known faunas and regions, and taxonomic expert opinion (May 1988, 2000; Gaston 1991; Stork 1993; Nielsen and Mound 2000). But particular attention has been given to the empirical method proposed by Erwin (1982), in which the number of beetle species associated with an individual tropical rain forest tree species is used as the basis for extrapolating a tropical/global estimate.

This is for several reasons. First, focusing on tropical arthropods is logical because they make the largest contribution to global species richness through described species (May 2000) and species awaiting formal description but known from museum collections (Nielsen and Mound 2000). Within the arthropods the beetles are a commonly used surrogate group because they are functionally diverse and are the most species-rich order, with about one-quarter of all species on Earth thought to be beetles (Ødegaard 2000; Hunt et al. 2007). Second, previous estimates for arthropod species richness had been educated guesses, with no transparent and testable underlying method. In contrast, Erwin's model has several testable assumptions, evidenced by the many studies contributing data to model revisions (see Ødegaard 2000). Third, the original values used by Erwin (1988) produced values of 30 million to 100 million (Erwin 1982, 1988), 2 orders of magnitude higher than previous estimates, creating widespread scientific and public interest (Wilson 1999). Subsequent revisions applying different values to Erwin's general model have yielded estimates in the order of 5–10 million (Thomas 1990; Stork 1993; Ødegaard 2000; Novotný et al. 2002). But the uncertainty in parameter estimation and, therefore, that associated with the final estimation of how many species there are on Earth has never been formally included or analyzed in such models. This means that it is impossible to determine the precision of a particular estimate.

Ødegaard (2000) made a useful step in this direction by carrying out separate calculations based on minimum,

* Corresponding author; e-mail: andrewjh@unimelb.edu.au.

Am. Nat. 2010. Vol. 176, pp. 000–000.
DOI: 10.1086/652998

maximum, and working figures. This approach used quartiles for most (but not all) parameters to produce three separate estimates of species richness, but these lacked an associated measure of variance.

A more complete representation of uncertainty could be obtained by using stochastic models, that is, models in which the parameters are represented with probability distributions rather than point estimates. The aim of this note is to incorporate uncertainty into Erwin's model parameters by using the most comprehensive tropical arthropod data set available and to see how this translates into uncertainty in estimation of global tropical arthropod species richness.

Methods

Tropical arthropod species richness was estimated using two separate but related models. In both models, most parameters are represented by probability distributions. Latin hypercube sampling was applied to each distribution (McKay et al. 1979; Vose 2000). This involved stratifying a probability distribution into y ($j = 1, \ldots, y$) intervals of equal probability and drawing one realization from each interval, resulting in y ($=500{,}000$) iterations of the model. Model A is structurally analogous to those used by others (Thomas 1990; Stork 1993; Ødegaard 2000) and is given as

$$N_{\mathrm{A}j} = \left(\frac{x_j c_j}{p_{\mathrm{a}j} p_{\mathrm{c}j}}\right) n_{\mathrm{t}j}, \tag{1}$$

where, for the jth iteration, x_j is the average effective specialization of herbivorous beetle species across all tree species, c_j is a correction factor for nonherbivorous beetle species, p_{aj} is the proportion of canopy arthropod species that are beetles, p_{cj} is the proportion of all arthropod species found in the canopy, and n_{tj} is the number of tropical tree species. Each iteration of the model produced an estimate of tropical arthropod species richness, $N_{\mathrm{A}j}$, with the distribution shown in figure 1*A* representing the complete set. Simulations were performed using @Risk (Palisade, Newfield, NY) as an add-in to Microsoft Excel. Sensitivity analyses of the influence of parameter uncertainty on output uncertainty were conducted by determining Spearman's rank correlations. None of the distributions were held constant in sensitivity analyses.

To estimate x_j, we first need to consider n_k, the number of herbivorous canopy beetle species on tree species k ($k = 1, 2, \ldots, l$), and f_k, the proportion of the beetle species effectively specialized on that species. Both n_k and f_k were determined from a large data set of leaf-chewing canopy beetles on $l = 56$ tree species in New Guinea (Novotný et al. 2002). The latter was calculated following May (1990) as

$$f_k = \sum_{i=1}^{l} \left(\frac{1}{i}\right) p_k(i), \tag{2}$$

where $p_k(i)$ is the proportion of leaf-chewing canopy beetles present on tree species k that were found on a total of i different tree species (including species k). A distribution for x_j was then obtained by producing 500,000 nonparametric bootstrap estimates of $n_k f_k$. Note that in order to maintain the requisite connection between the number of beetle species and the proportion effectively specialized on a specific tree species, it was necessary to take random samples from the distribution of x_j rather than from the parent parameters n and f.

All other parameters were simply represented by continuous uniform distributions, that is, ranges. We used the literature survey of Ødegaard (2000) to define the following minima and maxima: $c = 1.79$–2.70, $p_{\mathrm{a}} = 0.25$–0.66, $p_{\mathrm{c}} = 0.18$–0.33. Further searching of the literature did not reveal any subsequent parameter estimates that lay beyond these ranges. The lower and upper bounds of the uniform distribution of n_{t} (43,000 and 50,000) were taken from expert estimates given by Erwin (1982) and Fine and Ree (2006).

Model B is a modified version of model A, the major difference being that it is based on the number of plant genera in New Guinea and the contribution of these to global tree richness. The arthropod species richness estimate from the jth iteration, $N_{\mathrm{B}j}$, was determined as

$$N_{\mathrm{B}j} = \frac{y_j c_j n_{\mathrm{g}}}{p_{\mathrm{a}j} p_{\mathrm{c}j} p_{\mathrm{n}}}, \tag{3}$$

where n_{g} is the number of tree genera in New Guinea (1,872; Höft and Wau Ecology Institute 1992), p_{n} is the ratio of New Guinean to tropical genera (0.05; Sekhran and Miller 1996), and y_j is the average effective specialization of herbivorous beetle species across all tree genera in the jth iteration. In the three cases where we had multiple tree species per genus, a representative host and its associated beetle species was created through random sampling of the candidate tree species, with a separate randomization being used for each beetle species. Given that some New Guinean tree genera are markedly more species rich than others, the likelihood that a specific genus would be chosen in any given iteration was weighted by using a nonuniform discrete distribution for y_j. Relative weightings were assigned according to the number of species known to occur in each genus (Höft and Wau Ecology Institute 1992). Unfortunately, no information on the un-

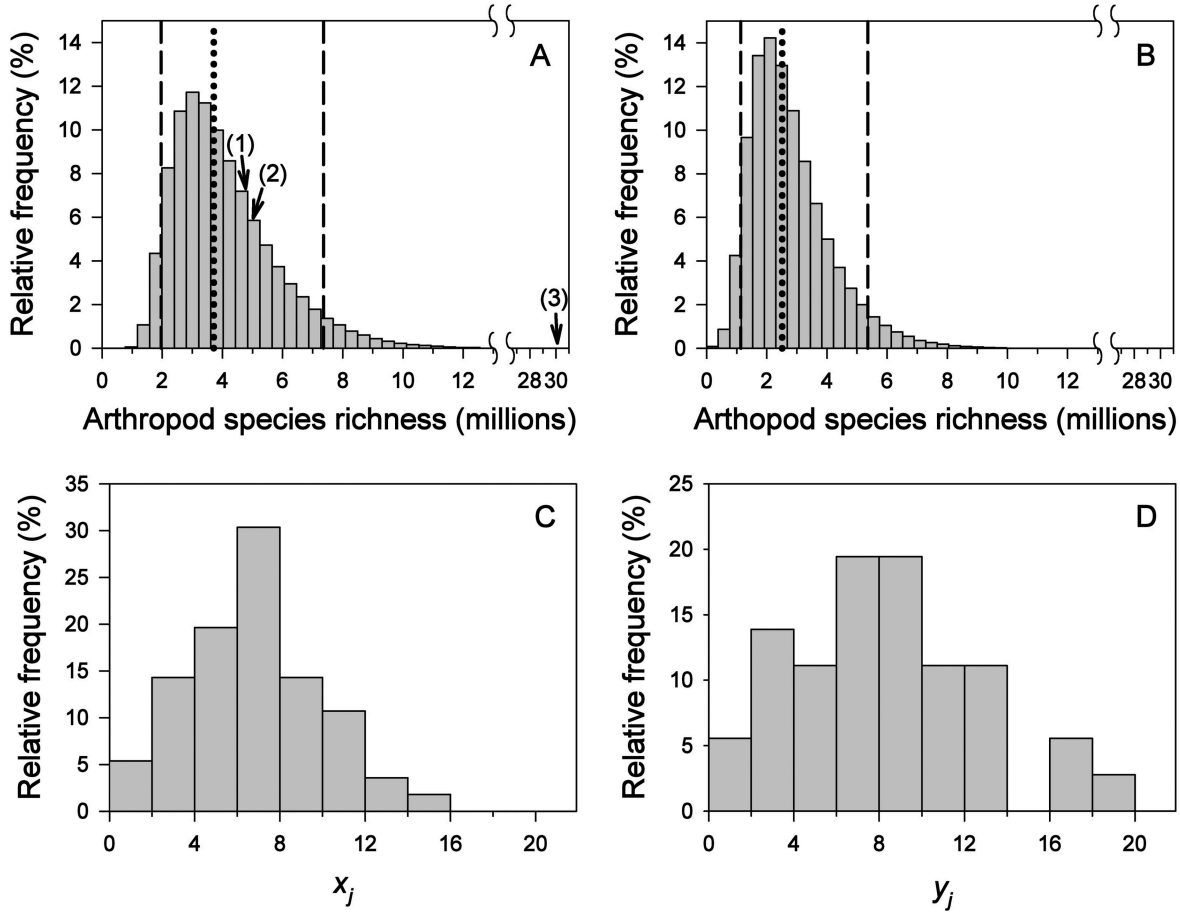

Figure 1: *A*, *B*, Probability distributions for estimates of tropical arthropod species richness arising from models A and B, respectively. The dotted vertical line denotes the median, and the left and right dashed lines mark the 5% and 95% confidence limits. Arrows indicate point estimates made by others (*1*, Ødegaard 2000; *2*, Stork 1993; *3*, Erwin 1982). Note break in *X*-axis. *C*, *D*, Number of beetle species effectively specialized on 56 New Guinean rain forest tree species (x_j) and 36 genera (y_j).

certainty associated with the estimate of p_n was available, and the point estimate itself was a rough approximation based on New Guinea's contribution of about 5% to the global richness of birds and flowering plants (Sekhran and Miller 1996). Given the overwhelming dominance of tropical plant and bird species to the global total and the roughness of the estimate in the first place, we have used the value of 0.05 as our best guess for the ratio of New Guinean to tropical genera. Of course, if the value of 0.05 is closer to the true ratio for New Guinean to global genera rather than for tropical genera, then N_B actually represents global (not just tropical) arthropod species richness. In any case, the failure to account for uncertainty in this ratio means that uncertainty in the final prediction would be underestimated because it is not propagated beyond the New Guinean estimate of arthropod species richness. Confidence intervals for all output distributions were determined using the percentile method (Buckland 1984).

Results

Both models produced right-skewed distributions for global species richness, with medians for models A and B of 3.7 million and 2.5 million species (fig. 1*A*, 1*B*) and 90% confidence intervals of [2.0, 7.4] and [1.1, 5.4] million, respectively. They suggest probabilities of <0.00001 for estimates of ≥30 million (i.e., <0.001% chance).

A sensitivity analysis showed that uncertainty in the proportion of canopy arthropod species that are beetles had the greatest impact on uncertainty in species estimation (fig. 2). However, other parameters (the number of beetle species effectively specialized on a tree species or

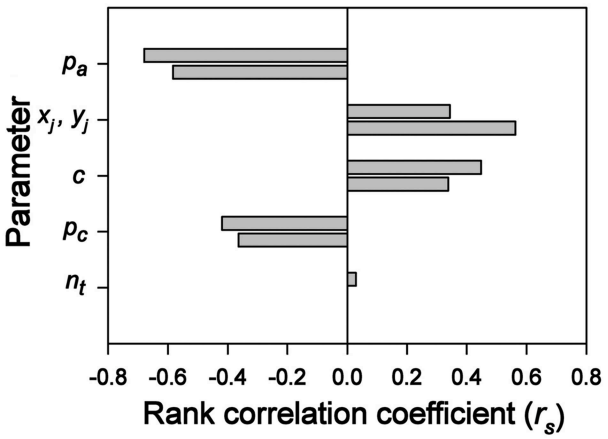

Figure 2: Sensitivity analyses of models A (*upper bar of each pair*) and B (*lower bar*) showing the Spearman rank correlation (r_s) between input and output values. p_a = proportion of canopy arthropod species that are beetles; x_j, y_j (models A and B, respectively) = number of beetle species effectively specialized on a species or genus of tree; c = correction factor for nonherbivorous species; p_c = proportion of all arthropod species that are found in the canopy; and n_t = number of tropical tree species (model A only).

genus, the proportion of arthropods that are found in the canopy, and the proportion of beetles that are nonherbivores) also had influence.

Discussion

Our models (see also the appendix) lend little support to hyperestimates of tropical arthropod species richness, with both models suggesting probabilities of <0.00001 for estimates of 30 million or greater. The median predictions from both models of 3.7 million and 2.5 million tropical arthropod species are much closer to those propounded by others (Gaston 1991; Stork 1993; May 2000; Ødegaard 2000). Perhaps what is more significant is that the 90% confidence intervals around these medians are quite broad. These confidence intervals are also asymmetrical, as the prediction distributions display positive skewness, which is born out of the shape of the effective specialization distributions and the multiplicative structure of the models. The sensitivity of the outputs to the effective specialization parameters, among others, highlights the risks associated with extrapolating from insect richness estimates for one or a few tree species or genera.

The appropriate distributional form for parameters c, p_a, p_c, and n_t was unknown, which is why we erred on the conservative side and used the continuous uniform, which assigns equal probability throughout the range. An alternative, and less conservative, approach would have been to use a distribution of more natural shape, that is, one with tails and a central peak. It was appropriate, therefore, to investigate the effect that such a conservative assumption had on the model output. To this end, we reran the models, substituting the Pert distribution for each continuous uniform distribution. The Pert is widely used in expert opinion modeling and is specified by three parameters: the minimum, the most likely, and the maximum. The minima and maxima for the four parameters equaled the minima and maxima for the continuous uniform distributions (i.e., lowest and highest published values), and the most likely value was assumed to be the midpoint in each case. Changing these distributions had negligible influence on the estimation of global arthropod species richness, producing medians and 90% confidence intervals of 3.6 [2.3, 5.9] million and 2.5 [1.3, 4.4] million species for models A and B, respectively.

The limitations of extrapolating from field surveys of a particular taxon in a specific part of the world to estimates of global species richness have been well documented (Stork 1993; May 2000; Ødegaard 2000). The models and richness estimates we present are also subject to many of these caveats, but they have the advantage of explicitly accounting for uncertainty. The use of data from only one part of the tropics to extrapolate to other regions is particularly relevant to the number of phytophagous canopy beetles effectively specialized on a tree species, x_j, because of the influence x_j has on global species richness estimations. Although our effective specialization data are derived from New Guinea only, the 90% confidence interval of our distribution of x_j, [1.96, 12.11], actually encompassed the entire range of previously published estimates from studies across the tropics (2.8–5.8; Ødegaard 2000), and the distribution for New Guinea is probably a reasonable representation of the tropics, given that the region hosts about a third of all tropical tree species (Höft and Wau Ecology Institute 1992; Fine and Ree 2006). We note that the maxima and minima of the other parameters represent studies from a variety of tropical locations (Brazil, Panama, Venezuela, and Sulawesi), with additional studies from many other places falling within these ranges.

On face value, our model involves extrapolating from plant-phytophage relationships. This could be perceived as a potential problem if most of the undescribed species are nonphytophagous (Stork 2007; Stork et al. 2008). We argue that our model accounts for this potential problem for several reasons. First, specialization among nonphytophagous insects still exists, owing to the interactions nonphytophagous insect species have with other species or resources (e.g., parasitoids and their hosts [Smith et al. 2008]). Second, nonphytophagous insects are likely to show similar biogeographic patterns to plants and other organisms (at broad spatial scales averaging over many

taxonomic groups) because most species probably respond to similar underlying environmental gradients or historical events (Gaston 2000; Moritz et al. 2001). We acknowledge that it would be preferential to have separate models for phytophages and nonphytophages, however, because the patterns and mechanisms explaining nonphytophagous insect diversity are much more poorly documented and understood than for phytophages (Kitching 2006), but data to perform uncertainty modeling are largely unavailable.

One factor that our models did not account for is that the phytophagous beetle fauna associated with a host plant may change across the geographic range of the plant (May 1990; Novotný et al. 2005; Gering et al. 2007). Conversely, a specific beetle species may specialize on different hosts in different parts of its range (May 1990; Thomas 1990; Novotný et al. 2007). Ødegaard (2000) attempted to incorporate these effects in his models, but the two correction factors effectively canceled each other out ($\times 2.1$, $\div 2.5$). Because so few data exist for the tropics (Lewinsohn and Roslin 2008), we chose not to include them.

A significant component of the uncertainty associated with x_j and y_j is likely to be aleatory (i.e., irreducible) because it mostly represents natural variation in specialization on different host tree species (fig. 1*C*, 1*D*), but at least some portion will be epistemic (i.e., reducible, representing poor knowledge) since our sampling of the host trees would have been incomplete (see Lewinsohn and Roslin 2008). Separating these two forms of uncertainty will be crucial to further refinement of global species richness estimates, as only epistemic uncertainty can be addressed through further research. The modeling technique used here is just one of a suite of statistical tools used across a wide variety of disciplines to account for uncertainty (Burgman 2005). Application of such methods to existing data sets could give further insights into the profound question of how many species exist.

Since the adoption of the Linnaean system of species descriptions 250 years ago, approximately 855,000 species of arthropods have been described (May 2000). If we use the median global species richness estimates from the present study (3.7 million and 2.5 million), 66%–77% of arthropod species are yet to be described. This represents an enormous amount of work for taxonomists that will take hundreds of years to complete at the current rate that species are described, taxonomists are trained, and funding is allocated for invertebrate taxonomy (Wheeler 2004; Rodman 2007), even with the application of new techniques such as DNA bar coding (Janzen et al. 2009).

Acknowledgments

We thank B. Brook, A. Chao, and R. Colwell for providing insightful comments on earlier versions of this manuscript and N. Karavarsamis for advice on confidence interval estimation. The host specificity studies in New Guinea were supported by the National Science Foundation, Christensen Fund, Grant Agency of the Czech Republic, Czech Academy of Sciences, Swedish Natural Science Research Council, Czech Ministry of Education, Otto Kinne Foundation, Darwin Initiative, International Centre of Insect Physiology and Ecology, and Bishop Museum. Parataxonomists in New Guinea are thanked for their assistance and are listed in Novotný et al. (2002).

APPENDIX

Additional Model

In addition to the two models described in the article, a third model, based on canopy lepidopterans collected in the same field study, was constructed. It corresponds to model B but makes the assumption that all caterpillar species can be found in the canopy and that all species are herbivorous. Thus, the number of tropical arthropod species for the lepidopteran model, N_{L}, is expressed as

$$N_{\mathrm{L}j} = \frac{z_{kj} n_{\mathrm{g}}}{p_{\mathrm{b}} p_{\mathrm{n}}}, \tag{A1}$$

where z_k is the number of lepidopteran species effectively specialized on the kth genus and p_{b} is the proportion of all arthropod species that are lepidopterans (cf. p_{a}, which relates to the proportion in the canopy). A point estimate of 0.10 (Novotný et al. 2002) was assigned to p_{b}, and z_{kj}, the only random variable in this simple model, was determined as per y_{kj}. This model yielded a slightly higher median prediction, 8.5 million species, than the beetle-based models, but uncertainty could not be adequately assessed here given the use of point estimates for all but one parameter.

Literature Cited

Buckland, S. T. 1984. Monte Carlo confidence intervals. Biometrics 40:811–817.

Burgman, M. 2005. Risks and decisions for conservation and environmental management. Cambridge University Press, Cambridge.

Chapman, A. D. 2009 Numbers of living species in Australia and the world. 2nd ed. Australian Biological Resources Study, Canberra. http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/pubs/nlsaw-2nd-complete.pdf.

Erwin, T. L. 1982. Tropical forests: their richness in Coleoptera and other arthropod species. Coleopterists Bulletin 36:74–75.

———. 1988. The tropical forest canopy: the heart of biotic diversity. Pages 123–129 *in* E. O. Wilson and F. M. Peter, eds. Biodiversity. National Academy, Washington, DC.

Fine, P. V. A., and R. H. Ree. 2006. Evidence for a time-integrated

species-area effect on the latitudinal gradient in tree diversity. American Naturalist 168:796–804.

Gaston, K. J. 1991. The magnitude of global insect species richness. Conservation Biology 5:283–296.

———. 2000. Global patterns in biodiversity. Nature 405:220–227.

Gering, J. C., K. A. DeRennaux, and T. O. Crist. 2007. Scale dependence of effective specialization: its analysis and implications for estimates of global insect species richness. Diversity and Distribution 13:115–125.

Höft, R., and Wau Ecology Institute. 1992. Plants of New Guinea and the Solomon Islands: dictionary of the genera and families of flowering plants and ferns. Wau Ecology Institute, Wau, Papua New Guinea.

Hunt, T., J. Bergsten, Z. Levkanicova, A. Papadopoulou, O. S. John, R. Wild, P. M. Hammond, et al. 2007. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. Science 318:1913–1916.

Janzen, D. H., W. Hallwachs, P. Blandin, J. M. Burns, J.-M. Cadio, I. Chacon, T. Dapkey, et al. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. Molecular Ecology Resources 9:1–26.

Kitching, R. L. 2006. Crafting the pieces of the diversity jigsaw puzzle. Science 313:1055–1057.

Lewinsohn, T. M., and T. Roslin. 2008. Four ways towards tropical herbivore megadiversity. Ecology Letters 11:398–416.

May, R. M. 1988. How many species are there on Earth? Science 241:1441–1449.

———. 1990. How many species? Philosophical Transactions of the Royal Society B: Biological Sciences 330:293–304.

———. 2000. The dimensions of life on Earth. Pages 30–45 *in* P. H. Raven and T. Williams, eds. Nature and human society: the quest for a sustainable world. National Academy, Washington, DC.

McKay, M. D., R. J. Beckman, and W. J. Conover. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21:239–245.

Moritz, C., K. S. Richardson, S. Ferrier, G. Monteith, J. Stanisic, S. E. Williams, and T. Whiffin. 2001. Biogeographical concordance and efficiency of taxon indicators for establishing conservation priority in a tropical rainforest biota. Proceedings of the Royal Society B: Biological Sciences 268:1875–1881.

Nielsen, E. S., and L. A. Mound. 2000. Global diversity of insects: the problems of estimating numbers. Pages 212–222 *in* P. H. Raven and T. Williams, eds. Nature and human society: the quest for a sustainable world. National Academy, Washington, DC.

Novotný, V., Y. Basset, S. E. Miller, G. D. Weiblen, B. Bremer, L. Cizek, and P. Drozd. 2002. Low host specificity of herbivorous insects in a tropical forest. Nature 416:841–844.

Novotný, V., S. E. Miller, Y. Basset, L. Cizek, K. Darrow, B. Kaupa, J. Kua, and G. D. Weiblen. 2005. An altitudinal comparison of caterpillar (Lepidoptera) assemblages on *Ficus* trees in Papua New Guinea. Journal of Biogeography 32:1303–1314.

Novotný, V., S. E. Miller, J. Hulcr, R. A. I. Drew, Y. Basset, M. Janda, G. P. Setliff, et al. 2007. Low beta diversity of herbivorous insects in tropical forests. Nature 448:692–698.

Ødegaard, F. 2000. How many species of arthropods? Erwin's estimate revised. Biological Journal of the Linnean Society 71:583–597.

Rodman, J. E. 2007. Reflections on PEET, the Partnerships for Enhancing Expertise in Taxonomy. Zootaxa 1668:41–46.

Sekhran, N., and S. E. Miller. 1996. Papua New Guinea country study on biological diversity. Papua New Guinea Department of Environment and Conservation, Waigani, Papua New Guinea.

Smith, M. A., J. J. Rodriguez, J. B. Whitfield, A. R. Deans, D. H. Janzen, W. Hallwachs, and P. D. N. Hebert. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proceedings of the National Academy of Sciences of the USA 105:12359–12364.

Stork, N. E. 1988. Insect diversity: facts, fiction and speculation. Biological Journal of the Linnean Society 35:321–337.

———. 1993. How many species are there? Biodiversity and Conservation 2:215–232.

———. 2007. World of insects. Nature 448:657–658.

Stork, N. E., P. S. Grimbacher, R. Storey, R. G. Oberprieler, C. A. M. Reid, and A. Slipinski. 2008. What determines whether a species of insect is described? evidence from a study of tropical forest beetles. Insect Conservation and Diversity 1:114–119.

Thomas, C. D. 1990. Fewer species. Nature 347:237.

Vose, D. 2000. Risk analysis: a quantitative guide. 2nd ed. Wiley, Chichester.

Wheeler, Q. D. 2004. Taxonomic triage and the poverty of phylogeny. Philosophical Transactions of the Royal Society B: Biological Sciences 359:571–583.

Wilson, E. O. 1999. The diversity of life. Harvard University Press, Cambridge, MA.

Associate Editor: Axel G. Rossberg
Editor: Donald L. DeAngelis