Molecular Phylogenetics and Evolution
Vol. 17, No. 1, October, pp. 129–132, 2000
doi:10.1006/mpev.2000.0737, available online at http://www.idealibrary.com on IDEAL®

COMMENTARY

The Importance of Being Earnest: What, if Anything, Constitutes a "Specimen Examined?"

Luis A. Ruedas,[1] Jorge Salazar-Bravo, Jerry W. Dragoo, and Terry L. Yates

Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico 87131-1091

Received June 14, 1999

In the probing theatrical opus "Jumpers," one of the characters created by playwright Tom Stoppard (1972) asks the question: "Is God?" The question is thus phrased, the character explains, because to phrase it in any other way would presuppose an assumption of doubt on the part of the questioner, when in fact an unbiased mind and question is what is required within the framework of philosophical inquiry. We ask a distinct question in a similar way: Is a specimen examined? The purpose of this line of inquiry is twofold: on the one hand, we would like to convince potential authors of the utility of a section in their paper detailing the specimens that were used in their study. On the other hand, we wish to stimulate editors and reviewers more consistently to require such a "specimens examined" section, particularly in papers that deal with the systematics and taxonomy of living organisms.

Increasingly, growing numbers of investigators are using molecular techniques to arrive at answers either to long-standing questions or to questions that heretofore were unthinkable. Concomitant with the growing number of sequence-based exploratory work is a growing body of work which cannot be confirmed or duplicated and may even be outright wrong. This is because the specimen used by the investigator cannot be reexamined because, in fact, it does not exist as such, having been discarded. When such work becomes incorporated into the body of scientific literature, errors increasingly become magnified as such work, having been published, is cited as fact rather than the fiction it very well might be. We believe in repeatability as a cornerstone of the scientific method and herein explore the *sine qua non* necessity of incorporating a legitimate "Specimens Examined" section into every paper published using molecular sequence data. Further, it is imperative to deposit in museum collections any and all specimens used in generating molecular data for publication. We would like to explicitly state at the outset that in no way do we wish to impugn the scientific integrity or accomplishments of any of the individuals whom we cite, either *pro* or *contra* our thesis that a legitimate specimens examined section should be a *sine qua non* requirement for all papers published in this venue or in any journal publishing sequence-based results. Only 15 of 56 papers that we scored in the past four numbers of this journal (27%) had a legitimate specimens examined section, with museum numbers for each voucher, and names of museums where the specimens used in the study could be examined. The remaining 41 papers on which we could have picked fell far short.

One of Kuhn's (1962) main theses might be paraphrased as stating that strict adherence to tradition should be viewed within the scientific establishment as an impediment to progress. We do not dispute that thesis; however, it often is true that there can be value in long-standing traditions. Unquestioning allegiance to tradition results in lack of progress but maintaining certain traditions after thorough evaluation of their value should not be misconstrued as unquestioning allegiance. "Normal" science advances as accumulated knowledge is reinterpreted in light of novel developments. That is the wherefore of the requirement—and accepted standard practice—of citing and acknowledging prior work upon which we are building, as a manner of reassuring the audience of the quality of the data being built upon and to give the reader the opportunity to verify our premises. It is, in fact, unthinkable to submit for publication any serious work without a fully cited references section.

One tradition of museum science which we believe should be retained is that of a thorough and meticulous section in every specimen-based paper detailing the specimens examined, wherein the primary data (the specimens themselves) could be reviewed and possibly reexamined should any question arise subsequently regarding their nature, provenance, taxonomic identification, or any other question: we are, after all, seek-

[1] To whom correspondence should be addressed. Fax: (505) 277-0304. E-mail: lruedas@sevilleta.unm.edu.

1055-7903/00 $35.00

ing to "have fairly true genealogical trees of each great kingdom of Nature." We submit that in the context of scientific work on living organisms based within an evolutionary framework, a properly referenced voucher specimen is as critical to the credibility of the work as are cited (but often unread) references.

Reckless disregard for taxonomic identification and the consequent possibility of nonrepeatability can be found even within the pages of this journal. For Volume 10 (1998) and Number 1 of Volume 11 (1999), we counted 15 papers in which there was no list of specimens examined (either tabular or in text), no mention of deposition of sequences in a genetic database, no citation of sequences from other studies, and, most egregiously, no list of specimens placed or deposited in a museum or university collection. Sixteen papers listed specimens examined, but these were not deposited anywhere subsequent to (or in parallel with) the study and thus could not be verified. Two papers used data from previously published papers, and 7 referenced materials to genetic databases (EMBL or GenBank). While the latter practice may be construed as acceptable by some and, indeed, when fully undertaken, constitutes a space-saving measure for the journal, for the most part, the information contained in these databases which is useful for further taxonomic determinations is generally no better than the electrons upon which it is printed. All sequence depositors should be as thorough in this respect as Zink *et al.* (1998), who state: "precise locality information, specimen information and voucher numbers, and sequence data are given in GenBank." The truth, however, is that most specimen data in GenBank are not congruent with potential repeatability of experiments.

Phylogenetic relationships among carnivores based on the complete sequence of the mitochondrial cytochrome *b* gene were reported upon by Ledje and Árnason (1996). These authors compared the sequences that they obtained to other partial sequences of carnivores available in the literature or from genetic data banks. They would express concern when their sequence for a particular species was not within 95% sequence similarity of those previously reported for the same species. Comparisons were undertaken with data from Lento *et al.* (1995), Vrana *et al.* (1994), Zhang and Ryder (1993), and Dragoo *et al.* (1993). Ledje and Árnason's (1996) sequences were similar within ≤5% of previously published sequences with the exception of two papers. Vrana *et al.* (1994) reported a sequence for the walrus that was >10% divergent from that of Ledje and Árnason (1996). Ledje and Árnason (1996) also found divergences (with respect to Vrana *et al.*, 1994) of ≥20% for the giant panda (but identical to the same species as reported by Zhang and Ryder, 1993), 14% for the red fox, 15% for the striped skunk (but within 3% of the same species as reported by Dragoo *et al.*, 1993), and >20% for the domestic cat. In point of fact, the cytochrome *b* sequence for the domestic cat reported by Vrana *et al.* (1994) actually is known to be derived from mink (Ledje and Árnason, 1996). Ledje and Árnason (1996) also reported a 6% difference between their spotted skunk sequence and that obtained by Dragoo *et al.* (1993); they suggested that a high level of polymorphism in the spotted skunk was responsible for this variance, as differences all were in synonymous substitutions. We can verify the identity of the specimen used in the study of Dragoo *et al.* (1993) as one of us (J.W.D.) actually was sprayed by the animal, which now resides in the Texas Cooperative Wildlife Collections! However, when questioned as to the specific locality from which their spotted skunk was derived, Árnason (pers. comm.) reported that it had been obtained from another researcher's freezer. It is not a bad thing to borrow specimens from other researchers; however, as shown in this example, specific localities and voucher specimens should be available somewhere in order to verify taxonomic identities.

It may be thought that deposition of specimens in vouchering collections (i.e., museums) may even be superfluous in some cases. For instance, is it really necessary to deposit a domestic dog and its ancillary materials in a collection? Kim *et al.* (1998) recently provided the complete nucleotide sequence of the mitochondrial genome of the domestic dog. However, there is no mention of deposition of skeletal or soft tissue remains in a collection for the specimen examined. Is this a serious error? It could be: a recent paper by Vilá *et al.* (1997) demonstrated that dogs and wolves showed paraphyletic relationships when examined using analysis of mitochondrial data. Without a specimen to examine, it would be hard to say where Kim *et al.*'s dog would fit morphologically in the picture painted by Vilá *et al.* (1997). It should further be noted that, thanks to explicitly stated specimen numbers, it was possible for Gardner (1998) to correctly identify an erroneously identified specimen used by Vilá *et al.* (1997). The answer to our, albeit, rhetorical question is therefore a resounding Yes: even in such apparently trivial cases as those involving a domestic dog, it is important to properly deposit vouchers in collections.

Situations involving other apparently well-known mammals also are rife. Early and Mouton (1973) karyotyped the swamp rabbit, *Sylvilagus aquaticus*, determining that the diploid number ($2n$) was 42. A subsequent report by Robinson *et al.* (1983) stated that the $2n$ was in fact 38. In neither of these reports was there a voucher for *S. aquaticus* (the answer is $2n = 38$; Ruedas and Elder, 1994). Despite the lack of a voucher, cells established and stored frozen by Robinson *et al.* (1983) were used in ensuing sequence studies by Halanych and Robinson (1997). Subsequent studies have shown that the latter authors were correct in their taxonomic assignation; in the case of Early and Mouton (1973), however, although the specimen reput-

edly was identified by a competent mammalian taxonomist, it could not be recalled who identified the specimen, which subsequently was lost (Ruedas and Elder, 1994). In fact, based on diploid number alone, it was a common Eastern cottontail, *Sylvilagus floridanus*, a species sympatric with *S. aquaticus* through the range of the latter and quite easily distinguished from *S. floridanus* under habitual circumstances.

Other papers examined specimens currently in zoos, including names or other identifying features of specimens living in zoos at the time of the study. Such is the case of Hall *et al.* (1998), who examine relationships among gibbon subgenera. For the most part, however, zoo animals do not end up deposited in museum collections (unless there is strong interest for such an eventuality on the part of a museum curator). Furthermore, even zoo animals are at times misidentified. In the case of primates in particular, there is evidence that such zoo staples as spider monkeys (*Ateles paniscus*) and certain howler monkeys (*Alouatta* sp.) may in fact be constituted by distinct, noninterbreeding species. Some authors consider *A. paniscus* to be constituted by at least four species; likewise, some authors consider that there exist more than the currently recognized six species of *Alouatta* (e.g., Froehlich and Froehlich, 1987). In cases such as these, it becomes of paramount importance not only to eventually voucher the specimens used in the study (and deposit soft tissues and DNA in ancillary collections) but further to note the geographic provenance of the stock from which were derived the specimens used in the study.

In instances potentially involving public health, it should be even more important to deposit materials, no matter what their biological origin. In our laboratory we have been addressing the relationship of *Calomys callosus* and Machupo virus, the agent responsible for Bolivian Hæmorrhagic Fever. We have examined several species and specimens of *Calomys* to address this relationship. In 1998, Engel *et al.* stated that *Calomys* was a polyphyletic lineage and the *C. callosus* was not even in the tribe Phyllotini. This came as a shock to us as *C. callosus* is the quintessential *Calomys* and we have numerous data from a number of distinct genes pointing to monophyly in *Calomys* (we are in the process of finalizing a manuscript on this issue). However, Engel *et al.* (1998) examined only three specimens representing three species of *Calomys* in their study. We attempted to verify the identity of the most genetically divergent specimen of *C. callosus* used in that study and found that the identification number that they provided indeed matched a skin and skull voucher specimen at the Sam Noble Oklahoma Museum of Natural History. However, further investigation revealed that identification numbers circumscribing that of the *Calomys* specimen represented species of *Oryzomys*. It is therefore possible that between collecting the specimen and depositing it in a museum, the specimen's frozen tissue tube was mislabeled: we cannot know with certainty, although it would be relatively trivial to verify this. Sequencing a second specimen of *C. callosus* would have shown the error; instead, the data were published without verification. Wrong data are worse than no data. The hypothesis that *Calomys* is monophyletic and at the base of the radiation of the Phyllotini is based on strong morphological evidence; when a result so far at variance with current evidence is obtained, it should be verified. Engel *et al.*'s (1998) hypothesis should have been verified or at least noted in text as potentially controversial or problematic. There were other inconsistencies in that data set as well, but we were concerned mostly with the *Calomys* issue. In this particular instance, because *Calomys* species are hosts for many different hæmorrhagic fevers, systematics is not merely an academic exercise: rather, it is a matter of human life and death.

CONCLUDING REMARKS

Based on the foregoing instances, we therefore strongly urge and recommend that papers addressing the phylogenetic relationships of a particular taxon or set of taxa and taking advantage of molecular or other techniques should contain a "specimens examined" section explicitly detailing the materials examined. Such materials should be deposited for posterity in a longstanding, legitimate collection dedicated to the storage of said materials. The specimens examined section minimally should include: (i) current scientific name of taxon or taxa; individual specimen identifier number (i.e., collector number, museum or collection catalogue number, lot number, etc.); (ii) name of collection wherein is housed the specimen(s) thus identified; (iii) exact location of geographic origin of specimen (i.e., precise collecting locality); and (iv) accession number of sequences obtained and used.

REFERENCES

Dragoo, J. W., Bradley, R. D., Honeycutt, R. L., and Templeton, J. W. (1993). Phylogenetic relationships among the skunks: A molecular perspective. *J. Mammal. Evol.* **1:** 225–267.

Early, E. M., and Mouton, M. (1973). Constitutive heterochromatin patterns (C-bands) of the laboratory rabbit, *Oryctolagus cuniculus*, and the wild cottontail rabbit, *Sylvilagus aquaticus*. *Mammal. Chromosomes Newslett.* **14:** 44–45.

Engel, S. R., Hogan, K. M., Taylor, J. F., and Davis, S. K. (1998). Molecular systematics and paleobiogeography of the South American sigmodontine rodents. *Mol. Biol. Evol.* **15:** 35–49.

Froehlich, J. W., and Froehlich, P. H. (1987). The status of Panama's endemic howling monkeys. *Primate Conserv.* **8:** 58–62.

Gardner, A. L. (1998). Red wolf specimen identity. *Cons. Biol.* **12:** 499.

Halanych, K. M., and Robinson, T. J. (1997). Phylogenetic relation-

ships of cottontails (*Sylvilagus,* Lagomorpha): Congruence of 12s rDNA and cytogenetic data. *Mol. Phylogenet. Evol.* **7:** 294–302.

Hall, L. M., Jones, D. S., and Wood, B. A. (1998). Evolution of the gibbon subgenera inferred from cytochrome *b* DNA sequence data. *Mol. Phylogenet. Evol.* **10:** 323–332.

Kim, K. S., Lee, S. E., Jeong, A. W., and Ha, J. H. (1998). The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome. *Mol. Phylogenet. Evol.* **10:** 210–220.

Kuhn, T. S. (1962). "The Structure of Scientific Revolutions," Univ. of Chicago Press, Chicago.

Ledje, C., and Árnason, Ú. (1996). Phylogenetic analyses of complete cytochrome *b* genes of the order Carnivora with particular emphasis on the Caniformia. *J. Mol. Evol.* **42:** 135–144.

Lento, G. M., Hickson, R. E., Chambers, G. K., and Penny, D. (1995). Use of spectral analysis to test hypotheses on the origin of pinnipeds. *Mol. Biol. Evol.* **12:** 28–52.

Robinson, T. J., Elder, F. F. B., and Chapman, J. A. (1983). Evolution of chromosomal variation in cottontails, genus *Sylvilagus* (Mamma-lia: Lagomorpha): *S. aquaticus, S. floridanus,* and *S. transitionalis. Cytogenet. Cell. Genet.* **35:** 216–222.

Ruedas, L. A., and Elder, F. F. B. (1994). On the putative existence of a diploid number 42 cytotype of *Sylvilagus aquaticus* (Bachman, 1837) (Lagomorpha: Leporidae). *Mammalia* **58:** 145–148.

Stoppard, T. (1972). "Jumpers," Faber & Faber, London.

Vilá, C., Savolainen, P., Maldonado, J. E., Amorim, I. R., Rice, J. E., Honeycutt, R. L., Crandall, K. A., Lundeberg, J., and Wayne, R. K. (1997). Multiple and ancient origins of the domestic dog. *Science* **276:** 1687–1689.

Vrana, P. B., Milinkovitch, M. C., Powell, J. R., and Wheeler, W. C. (1994). Higher level relationships of the arctoid Carnivora based on sequence data and "total evidence." *Mol. Phylogenet. Evol.* **3:** 47–58.

Zhang, Y., and Ryder, O. A. (1993). Mitochondrial DNA sequence evolution in the Arctoidea. *Proc. Natl. Acad. Sci. USA* **90:** 9557–9561.

Zink, R. M., Weller, S. J., and Blackwell, R. C. (1998). Molecular phylogenetics of the avian genus *Pipilo* and a biogeographic argument for taxonomic uncertainty. *Mol. Phylogenet. Evol.* **10:** 191–201.