

DNA barcoding and the renaissance of taxonomy

Scott E. Miller*

Office of the Under Secretary for Science, Smithsonian Institution, Washington, DC 20013-7012

How limited (or not) is the dietary range of an animal species? How does that host specificity plot across major lineages of organisms? How does that affect the structure of ecological communities? These are major questions in ecology, especially for invertebrates, where they have major implications for academic questions, such as how many species of insects there are, and practical applications in fields such as biological control (1). A multidisciplinary team representing field ecology, molecular genetics, and morphological taxonomy presents data in this issue of PNAS (2) that challenge our traditional assumptions of parasite diversity and host specificity in a system of tachinid flies that parasitize moth and butterfly caterpillars in Costa Rica (Fig. 1).

The work of Smith *et al.* (2) is a display of integrated taxonomy, demonstrating how DNA barcoding is a valuable addition to the taxonomic tool box. It integrates ecological, genetic, and morphological data to provide a robust analysis and also tests the primary genetic data (mitochondrial COI sequences) against other sequences (nuclear 28S and ITS1). This analysis is rooted in 29 years of inventory of $\approx 400,000$ wild-caught caterpillars in northwestern Costa Rica, the largest data set ever assembled on the basic biology of caterpillars, their host plants, and their parasites in a wildland setting. In their study, Smith *et al.* investigated 2,134 flies belonging to what appeared to be the 16 most generalist of the reared tachinid morphospecies. The combined evidence indicated that most of the flies were actually specialists and that only a few should be considered generalist species. A much smaller study of the biocontrol tachinid *Sturmiopsis parasitica* in Africa (3) suggests that this situation is not unique to Costa Rica.

Smith *et al.* (2) might best be read in parallel with Hajibabaei *et al.* (4), who reviewed DNA barcoding results for 521 morphologically defined species of moths and butterflies from the same inventory, many of which are hosts for the tachinids discussed by Smith *et al.* In contrast to the tachinid study, the moth species are relatively well known, and the barcode data agreed with morphological data in separating species. In only 13 cases did the barcode data suggest possible separations within a species; these separations were subse-



Fig. 1. Puparium (dark amber) of *Patelloa xanthura* DHJ02 (DHJPAR0003513), a truly generalist tachinid fly confirmed as such by ref. 2. The fly larva emerged from the newly killed caterpillar of *Parides iphidamas* (Papilionidae) a few hours before this image was taken (caterpillar voucher 05-SRNP-4432, August 19, 2005, Area de Conservacion Guanacaste, Costa Rica). (Photo courtesy of D. H. Janzen.)

quently confirmed by reanalysis of morphological and ecological data. In cases like these, barcode data led to more careful analysis of morphological data within a species. Subtle patterns of variation emerged within a purported species, supporting the presence of multiple species that were suggested by DNA and ecological data.

Although Smith *et al.* (2) provide an example of the power of integrated taxonomy, it is worth noting that their study is not a formal taxonomic publication. Although they have related each of their putative species to available species names (appendixes 1 and 4 in ref. 2), the formal taxonomic work remains to be done. The new and redefined species they propose need to be related to formally established species names by comparison with the morphology and DNA of type specimens in museums (if they can be located) or with newly collected material from the original “type localities” (as the best alternative if the type specimens are lost). In addition, the new species will need formally published names and descriptions. This is tedious, time-consuming work owing to the lack of new tools and procedures for taxonomic research. The results to date will stimulate further research and are well worth publishing in their present form.

The Taxonomic Enterprise

The taxonomic enterprise is entering a renaissance for many reasons, including new tools like DNA barcoding. Traditional taxonomists realize the need for a

more streamlined system of data gathering, analysis, and publication. Researchers in related fields (e.g., ecology, evolutionary biology, and agriculture) recognize that taxonomic names provide the basic vocabulary and information framework for biodiversity. Taxonomy as a field has always been tightly knit at the level of individual scientists but loosely organized at an institutional level. In recent years, a series of cooperating intergovernmental activities and nonprofit organizations have been created to build a more integrated taxonomic enterprise. These include the Global Biodiversity Information Facility (a global portal for taxonomic and specimen information and high-quality digital specimen images especially of “type specimens;” www.gbif.org), the Biodiversity Heritage Library (an effort to digitize and index the taxonomic literature; www.bhl.si.edu), the Global Taxonomy Initiative and Global Strategy for Plant Conservation of the Convention on Biological Diversity (www.biodiv.org), BioNet International (www.bionet-intl.org), and the Consortium for the Barcode of Life (www.barcoding.si.edu). A consortium is now forming to take the much-needed next step of presenting an online “Encyclopedia of Life” by merging diverse information into “species pages” with a standard format (5). All of these advances, plus the parallel advances in “smart mobile phones” such as the Blackberry, provide the opportunity to realize E. O. Wilson’s vision of a mobile biodiversity laboratory (6). These efforts are producing a surge in the availability on the Internet of four broad classes of information: taxonomic names, specimen data and images, literature, and attributes of species and specimens.

DNA barcoding has emerged at a critical time for taxonomy. Economic development and increased international commerce are leading to higher extinction rates and the introduction of invasive and pest species. As a result, local, national, and international user communities are demanding more and faster species identification services and better information about their biodiversity than ever before. At the same time, taxon-

Author contributions: S.E.M. wrote the paper.

Conflict of interest statement: S.E.M. is chair of the Consortium for the Barcode of Life.

See companion article on page 4967.

*E-mail: millers@si.edu.

omy as a field faces huge funding challenges and a dwindling professional workforce. Barcoding is emerging as a cost-effective standard for rapid species identification. It has the potential to accelerate our discovery of new species, improve the quality of taxonomic information, and make this information readily available to nontaxonomists and researchers outside of major collection centers. This is especially critical for tropical developing countries, which tend to have more species but fewer taxonomic resources (7).

What Is a Species?

Much of the controversy generated by DNA barcoding really involves two longstanding debates in taxonomy over analytical methods and species concepts (8–10). Given the historical depth of these debates, it is not reasonable to expect barcoding to provide the ultimate answers immediately, but barcoding studies are contributing to the debate. DNA barcoding, like previous applications of new categories of characters, creates the challenge of integrating new data into an established knowledge framework. Integrating multiple character systems adds to the robustness of species recognition, as demonstrated by studies such as ref. 2. Other studies (e.g., ref. 11) have shown that different approaches to species recognition can produce very similar results, adding to our confidence in a more integrative approach to taxonomy and species identification.

The debate also continues over the philosophical basis for species concepts and how to operationalize these concepts into a system for recognizing spe-

cies, despite several recent syntheses of the topic (e.g., refs. 12 and 13). Inevitably, DNA barcoding has been drawn into this debate. Taxonomists have always looked for discontinuous character variation that could signal divergence between species. This led naturally to discussions of threshold values for inter-specific divergence. The debate was essentially the same for morphological differences and genetic distance between species. Finding discontinuous variation in one character is typically the first step leading to further analysis of

Barcoding is emerging as a cost-effective standard for rapid species identification.

putative species, but few scientists would accept a single character as the basis for defining species boundaries (although in the case of some character-poor organisms, such as soil nematodes, there may be no other choice; ref. 8).

Value of Taxonomy

Biocontrol provides an example of why the seemingly arcane business of taxonomy is important and also why the renaissance in taxonomy and DNA barcoding should be encouraged. A new assessment by the Consultative Group on International Agriculture (CGIAR) (14) estimates that CGIAR agricultural research in sub-Saharan Africa has had a \$17 billion economic impact. Of this

impact, 80% is attributed to four biological control projects, the success of which all hinged on resolution of significant taxonomic problems. The identities of both the pests and their parasites required extensive and complex study to enable successful introductions of biocontrol agents.

Recent studies indicate that many pest insects that were previously considered widespread species are probably complexes of cryptic species with slight differences in their ecology. Understanding these differences will provide keys to managing the pest species and saving resources by relegating some of the species to nonpest status. Mistakes in species concepts or identifications are widespread in the history of biocontrol and account for many of the failures of introductions of biocontrol agents (15). In other cases, inappropriate choices of biocontrol agents cause great impact on nontarget species. For example, the introduced tachinid fly *Compsilura concinnata* has severely reduced populations of large native silk moths in northeastern North America (16).

An essential part of improving taxonomic data is the preservation of voucher specimens. I use the term specimen broadly here, because the appropriate vouchers may be cultures, tissue lines, or even images, depending on the organisms and the traits involved. The literature is replete with discoveries that could not be validated because of a lack of vouchers (17). The Smith *et al.* (2) study could not have been done without the retention of all of the reared flies as vouchers for the past three decades, first for their morphological iterative reanalysis, then for their backup of the ecological records, and, finally, for DNA barcoding.

1. Novotny V, Drozd P, Miller SE, Kulfan M, Janda M, Basset Y, Weiblen GD (2006) *Science* 313:1115–1118.
2. Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PDN (2007) *Proc Natl Acad Sci USA* 104:4967–4972.
3. Dittrich G, Mitchell A, Conlong D (2006) *Ann Soc Entomol France* 42:325–329.
4. Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) *Proc Natl Acad Sci USA* 103:968–971.
5. Wilson EO (2003) *Trends Ecol Evol* 18:77–80.
6. Wilson EO (2000) *Conserv Biol* 14:1–3.
7. Miller SE, Rogo LM (2002) *Cimbebasia* 17:197–218.
8. Cognato AI, Caesar RM, Blaxter M, Vogler AP (2006) *Front Ecol Environ* 4:268–273.
9. Vogler AP, Monaghan MT (2007) *J Zool Syst Evol Res* 45:1–10.
10. Waugh J (2007) *BioEssays* 29:188–197.
11. Dettman JR, Jacobson DJ, Turner E, Pringle A, Taylor JW (2003) *Evolution (Lawrence, Kans)* 57:2721–2741.
12. Agapow P-M, Bininda-Emonds ORP, Crandall KA, Gittleman JL, Mace GM, Marshall JC, Purvis A (2004) *Q Rev Biol* 79:161–179.
13. Sites JW, Jr, Marshall JC (2004) *Annu Rev Ecol Evol Syst* 35:199–227.
14. Maredia MK, Raitzer DA (2006) *CGIAR and NARS Partner Research in Sub-Saharan Africa: Evidence of Impact to Date* (Consultative Group on International Agricultural Research Science Council Secretariat, Rome).
15. Crawley MJ (1989) *Annu Rev Entomol* 34:531–562.
16. Boettner GH, Elkinton JS, Boettner CJ (2000) *Conserv Biol* 14:1798–1806.
17. Ruedas LA, Salazar-Bravo J, Dragoo JW, Yates TL (2000) *Mol Phylogenet Evol* 17:129–132.