



OXFORD JOURNALS  
OXFORD UNIVERSITY PRESS

## Society of Systematic Biologists

---

The Promise of DNA Barcoding for Taxonomy

Author(s): Paul D. N. Hebert and T. Ryan Gregory

Reviewed work(s):

Source: *Systematic Biology*, Vol. 54, No. 5 (Oct., 2005), pp. 852-859

Published by: [Oxford University Press](#) for the [Society of Systematic Biologists](#)

Stable URL: <http://www.jstor.org/stable/20061289>

Accessed: 16/12/2011 19:38

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at  
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*Society of Systematic Biologists* and *Oxford University Press* are collaborating with JSTOR to digitize, preserve and extend access to *Systematic Biology*.

<http://www.jstor.org>

*Syst. Biol.* 54(5):852–859, 2005  
 Copyright © Society of Systematic Biologists  
 ISSN: 1063-5157 print / 1076-836X online  
 DOI: 10.1080/10635150500354886

## The Promise of DNA Barcoding for Taxonomy

PAUL D. N. HEBERT AND T. RYAN GREGORY

*Department of Integrative Biology, Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, N1G 2W1, Canada;  
 E-mail: phebert@uoguelph.ca (P.D.N.H.)*

DNA barcoding is a novel system designed to provide rapid, accurate, and automatable species identifications by using short, standardized gene regions as internal species tags. As a consequence, it will make the Linnaean taxonomic system more accessible, with benefits to ecologists, conservationists, and the diversity of agencies charged with the control of pests, invasive species, and food safety. More broadly, DNA barcoding allows a day to be envisioned when every curious mind, from professional biologists to schoolchildren, will have easy access to the names and biological attributes of any species on the planet. In addition to assigning specimens to known species, DNA barcoding will accelerate the pace of species discovery by allowing taxonomists to rapidly sort specimens and by highlighting divergent taxa that may represent new species. By augmenting their capabilities in these ways, DNA barcoding offers taxonomists the opportunity to greatly expand, and eventually complete, a global inventory of life's diversity.

Despite the potential benefits of DNA barcoding to both the practitioners and users of taxonomy, it has been controversial in some scientific circles (Wheeler, 2004; Will and Rubinoff, 2004; Ebach and Holdredge, 2005; Will et al., 2005). A few have even characterized DNA barcoding as being "anti-taxonomy," arguing that its implementation will signal the death of a system 250 years in the making. We feel that this opposition stems from misconceptions about the DNA barcoding effort. As such, we welcome this opportunity to clarify both the rationale and potential impacts of DNA barcoding. In responding to this set of questions, we emphasize the multiple positive impacts of this approach for taxonomy and biodiversity science.

### QUESTIONS

1. *Given two billion US dollars (the amount a comprehensive program of DNA barcoding is estimated to cost [Whitfield, 2003]), how would you spend this money to benefit taxonomic and biodiversity research, and what would be the legacy of these data?*

This question ignores an inescapable reality; there is no prospect of a single \$2 billion infusion of support for any biodiversity research program. Such a level of investment may ultimately be achieved—but, if so, it will reflect a staged and geographically dispersed process of positive funding decisions that will depend heavily on both scientific progress and societal demand for species-level identifications. The small amount of funding so far

directed to DNA barcoding has yielded a rich harvest of scientific insights. This fact has led new organizations to provide the support needed to explore the scalability of these results across the animal kingdom. The early and positive results from this second wave of investigations have now motivated larger research groups to coalesce. In fact, the first alliances with a global reach have been assembled to lead the development of barcode sequence libraries for all birds and fishes. Segments of much more species-rich groups, such as plants and lepidopterans, are in the earlier stages of this process ([www.barcoding.si.edu](http://www.barcoding.si.edu)). These research groups may, in the longer term, form the nuclear units needed for the barcode initiative to move into the "big science" domain where success depends upon the coupling of a clearly enunciated, socially significant research agenda with strong international research alliances.

It is also important to note that the quest for large-scale support for DNA barcoding is not being carried out at the expense of taxonomic funding. Indeed, it is clear that any successful campaign to generate this support will result in a substantial infusion of funding for institutions and individuals engaged in taxonomic research. Overall, the costs associated with DNA sequencing will represent a small component of DNA barcoding efforts; the majority of funding will be employed for global collection efforts, for the curation of resultant specimens, and for developing online databases containing detailed information about them. Moreover, it bears pointing out that the funding already acquired has come from large foundations and government agencies and programs with no tradition of supporting taxonomic research, and that in some cases DNA barcoding proposals have competed directly with medical and comparative genomics projects rather than any related to taxonomic research. Viewed from this perspective, any large-scale DNA barcoding effort will represent a substantial boon, both financially and scientifically, to biodiversity and taxonomic research. It will certainly leave a lasting legacy in the form of a comprehensive, widely accessible system for the identification of species.

2. *Globally, alpha taxonomic research (the discovery and description of new species) is in crisis. Is DNA barcoding an expedient solution to this problem or will it expedite its decline?*

In our view the decline of alpha taxonomy is not a consequence of the growing use of molecular methods,

as has sometimes been suggested (Wheeler, 2004). In fact, we expect DNA barcoding to aid the resurgence of taxonomy. DNA barcoding programs will certainly direct new funding into the collection and cataloguing of specimens. They will, as well, aid taxonomic investigations by helping to reveal cryptic species (Hebert et al., 2004a, 2004b), by connecting sexes and life stages (Beskansky et al., 2003), and by clarifying problems of synonymy that now consume much taxonomic effort (Alroy, 2002). The novelty and scientific promise of DNA barcoding will additionally draw public interest to taxonomic and biodiversity issues, encouraging young researchers to enter the discipline and both academic departments and biomanagement agencies to hire them.

We are confident that DNA barcoding will play an increasingly important role as a taxonomic screening tool because of its ability to rapidly reveal the genetic discontinuities that ordinarily separate distinct species (e.g., Janzen et al., 2005; Smith et al., 2005). Its application in this fashion will allow an inversion of standard taxonomic approaches that operate in an a priori fashion—seeking the morphological discontinuities that signal reproductive isolation among unsorted assemblages of organisms. By contrast, DNA barcoding allows a more efficient a posteriori approach where predefined, genetically divergent groups are examined for trait variation. In this sense, DNA barcoding will clearly be a powerful enabler of alpha taxonomy.

3. *Overlapping character variation between and within species is well documented for many character systems. Why is this any more or less of a problem for DNA barcoding?*

Overlap in the variation of single characters is not problematic for any taxonomic system, be it morphological or molecular, so long as multiple characters are employed for taxon diagnosis. One common misunderstanding of DNA barcoding is that it is based on a single character, namely “one DNA sequence.” In fact, the 648-bp cytochrome *c* oxidase subunit I (*cox1* or *COI*) gene region used as the DNA barcode standard for members of the animal kingdom represents a complex composite character involving hundreds of independently varying components. Some of these component characters are invariant and therefore not all 648 bp are informative within a given taxonomic assemblage, but most are variable. For example, we have found variation at 512 of the 648 sites in a large set of lepidopteran barcodes (9715 sequences from 2215 species and 1047 genera). This means that even within a single insect order, DNA barcodes integrate the patterning of similarities and differences among hundreds of characters. In a certain sense, it is like the patterning generated by the scales on a moth’s wing—each scale is of almost no significance, but the composite character of wing coloration pattern is highly informative.

DNA barcoding using a single gene region does not assure complete taxonomic resolution, but it does promise proximity. Based on past results for varied animal groups, DNA barcoding will deliver species-level

resolution in 95% to 97% of cases (Hebert et al., 2004b; Janzen et al., 2005; Ward et al., 2005). When it fails, it will narrow the options to a small number of congeneric taxa (which, in many cases, could be resolved fully with additional genetic or other data). This impressive performance reflects two important, and perhaps unexpected, observations: the rarity of mitochondrial sequence sharing among species and the dearth of deep barcode divergences within species. Constrained intraspecific variation in diverse animal groups is a key early finding of the DNA barcode effort; one that merits deeper scientific investigation. Certainly, *cox1* shows far less variation within species than some early critics had projected would be the case (e.g., Mallet and Willmott, 2003), and this may reflect the impact of selective sweeps related to the coevolution of nuclear and mitochondrial genomes. Importantly, for the use of barcodes as species-level identifiers, barcode differences appear to accumulate quickly, making it possible to distinguish all but the youngest of sister species.

4. *Many taxonomists already practice DNA barcoding informally when delimiting and discovering species. Is this wrong, and what data are sufficient to demonstrate that a series of specimens represents a new species with traditional or barcoding methods?*

We recognize both the general utility of genetic data in taxonomic studies and the strong concordance in taxonomic signals from different genes. However, we emphasize that there is no such thing as “informal DNA barcoding.” A DNA barcode is not just any DNA sequence—it is a rigorously standardized sequence of a minimum length and quality from an agreed-upon gene, deposited in a major sequence database, and attached to a voucher specimen whose origins and current status are recorded. In fact, it has already been established that only those *cox1* sequences that meet these strict criteria will be designated as DNA barcodes by the National Center for Biotechnology Information’s GenBank (NCBI, GenBank; www.ncbi.nlm.nih.gov/Genbank), the European Molecular Biology Laboratory (EMBL; www.embl.org), and the DNA Data Bank of Japan (DDBJ; www.ddbj.nig.ac.jp).

There is an important distinction between “describing” and “delimiting” species, but a conflation of the two has created uneasiness about the use of DNA barcodes as the foundation of future taxonomic descriptions. We emphasize that DNA barcoding seeks merely to aid in *delimiting* species—to highlight genetically distinct groups exhibiting levels of sequence divergence suggestive of species status. By contrast, DNA barcodes—by themselves—are never sufficient to *describe* new species. At some stage, clearly divergent DNA barcodes, in combination with other information, will be used as the basis for providing a new Linnaean name (Smith et al., 2005) and, as with any taxonomic hypothesis, this would be subject to ongoing reevaluation. For example, in a recent survey of North American birds, the threshold for delineating probable new species was arbitrarily set at 10 × the average within-species variation of the entire

barcode data set. This led to the revelation of four presumptive new species (Hebert et al., 2004b), but decisions regarding the formal recognition of these taxa are left, appropriately, to the ornithological community (notably, existing morphological and behavioral information supports these new hypotheses). The synergy between DNA barcoding and studies of morphological/ecological diversity is further illustrated by the case of the skipper butterfly, *Astrartes fulgurator*, in which a combined morphological, natural history and barcoding approach revealed a complex of 10 species in one small area of Costa Rica. Importantly, several of these species showed a relatively small barcode divergence, but a coupling of this information with records on larval host plants and morphology illuminated the full diversity of the complex (Hebert et al., 2004a).

5. The proposed barcoding genes can fail to recover accurate species trees. Does this matter for DNA barcoding?

We emphasize that DNA barcodes do not aim to recover phylogenetic relationships; they seek instead to identify known species and to aid the discovery of new ones. Despite this fact, some opponents have claimed that DNA barcoding fails as a taxonomic approach because it does not always recover accurate species trees (e.g., Will and Rubinoff, 2004). It is important in this regard to emphasize that current taxonomic placements need to be viewed as hypotheses not facts. Consider a primary example offered by Will and Rubinoff (2004) in their critique of DNA barcoding: namely that DNA barcodes suggest a very close affinity between the moth *Simyra henrici* and certain species of *Acrionicta* (Hebert et al., 2003). Will and Rubinoff (2004:48) argue that this placement makes it “impossible [for DNA barcoding] to recover any taxonomic information below the suprageneric level, not even genus membership.” However, rather than reflecting a failure of DNA barcoding, we believe this case illustrates the power of the approach to illuminate taxonomic assignments in need of scrutiny. The traditional placement of *S. henrici* to a distinct genus reflects the fact that its adults have pale yellow forewings, showing striking divergence from the grey/black forewings of *Acrionicta* species. Yet larval morphology, adult forewing patterns, ecological niche, and genital anatomy all suggest that *S. henrici* has close affinities to *Acrionicta oblinata* (D. Wagner, personal communication), a conclusion reinforced by DNA barcodes (Fig. 1). Its distinctive forewing coloration likely reflects the fact that larvae of *S. henrici* feed on grasses, as opposed to the tree-feeding habits of typical *Acrionicta* species. A rapid shift in wing color has presumably been driven by natural selection to aid substrate matching

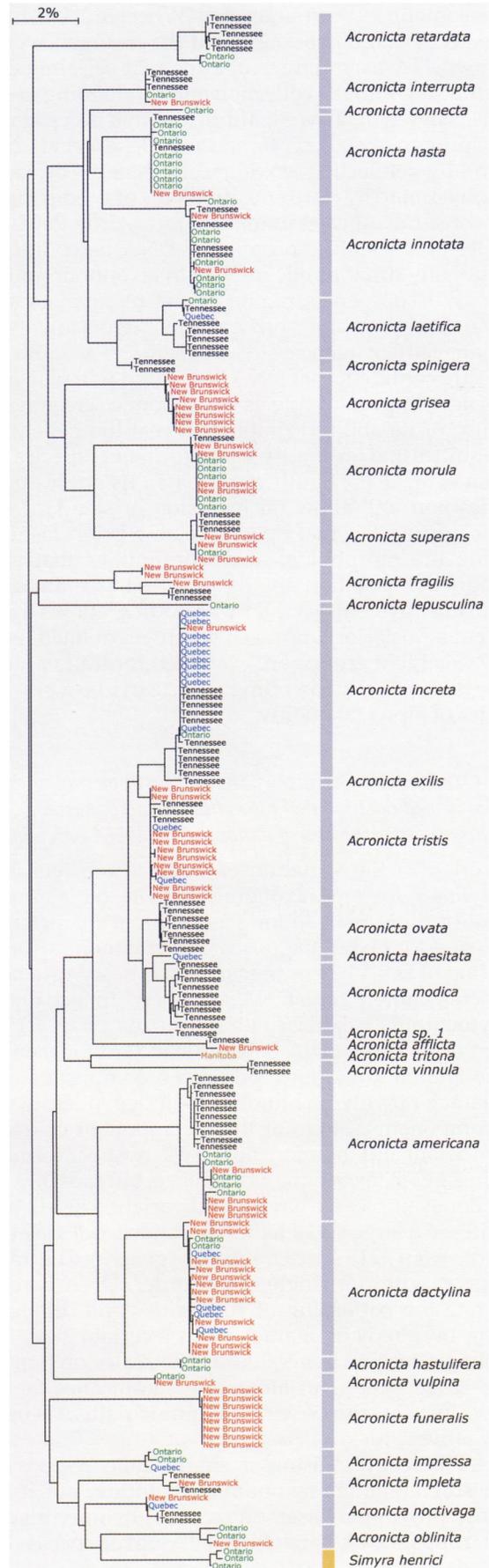


FIGURE 1. A taxon identification tree generated through neighbor-joining analysis of K2P distances showing patterns of *cox1* sequence divergence for 31 species of *Acrionicta* and 1 species of *Simyra*. Specimens from different provinces (Canada) or states (USA) are shown in different colors.

during adult life. Thus, this particular example reveals not only the ability of DNA barcoding to refine existing taxonomic hypotheses, but also to provide new insights into evolutionary trajectories (see Janzen et al., 2005, for other examples involving tropical lepidopterans).

6. *Some species are not mitochondrially monophyletic, sharing polymorphisms with unrelated taxa. How will this affect identifications using a barcoding approach?*

Although the horizontal exchange of mitochondria between taxonomically divergent organisms might theoretically occur, evidence for it not been found among the thousands of animal species that have now been barcoded. Shared mitochondrial sequences (and hence barcodes) have been observed, but only among closely related species and presumably as a result of ongoing hybridization. The taxonomic impacts of such sharing are far from catastrophic—they restrict identifications to a small complex of congeners. Shared barcodes do not represent a substantive taxonomic problem because they are uncommon and their impacts are parochial.

It is worth emphasizing that critical tests of mitochondrial sequence sharing between species are difficult to execute. Many studies presume that discrepancies between identifications made using morphological traits and DNA barcodes signal flaws in the barcode data. Before critically accepting such conclusions, stronger validation of the morphology-based assignments is needed. For example, Wahlberg et al. (2003) reported conflicts between morphology and mitochondrial DNA divergences in a tightly allied species complex of butterflies, but this conclusion would have been far stronger if the morphological assignments had been confirmed independently by several taxonomists (although these still might reflect an erroneous taxonomic hypothesis). It bears noting that blind tests of DNA barcoding *have* been carried out on several occasions. Indeed, DNA barcoding has passed double-blind tests, in which the taxonomist providing the specimens did not realize the full diversity of species present in a sample (that is, until further examination inspired by the barcoding results revealed key biological differences among them; e.g., Hebert et al., 2004a, 2004b). Providing empirical demonstrations that DNA barcodes are capable of consistent, accurate, and unambiguous identifications is a key aspect of barcoding research, and the same should be expected of alternative approaches.

7. *Should the completion of a DNA barcoding program ever occur, would this mark the beginning or end of taxonomic and biodiversity research, and what will be the role of systematists in a world where most identifications are done by "barcode"?*

DNA barcoding will increase the scale and success of biodiversity science by greatly increasing access to species identifications. An automated DNA-based system will free taxonomists from routine identifications, allowing them to direct their efforts to new collections, descriptions, and assessments of taxonomic relationships. Some DNA barcoding opponents have argued that

routine identifications are only a minor part of a taxonomist's work (Lipscomb et al., 2003; Wheeler, 2004; Will and Rubinoff, 2004), whereas others have lauded the potential utility of automated identification systems, but only if based on morphology (Gaston and O'Neill, 2004; Wheeler, 2004). We believe that species identifications are a rate-limiting step for many ecological and biodiversity investigations, as well as for taxonomic research, and that DNA barcoding will therefore both relieve a burden on taxonomists and fill a current need with important benefits to both taxonomy and biodiversity science.

In a barcoded world, taxonomists will retain their leadership role in the association, integration, and interpretation of knowledge about the character state variation that delineates species and what this implies for higher level taxonomy. As noted earlier, their work on new assemblages of life may often be expedited by using barcode results to enable an a posteriori approach to species recognition. Taxonomists will, of course, also continue to exploit other molecular and morphological approaches to explore deeper taxonomic relationships.

8. *Would the inevitable expansion of sequencing efforts that would come with a program of DNA barcoding be concomitant with a decline in the quality of taxonomic research?*

It has been suggested that sequencing is too expensive, difficult, or time-consuming for taxonomists to carry out (e.g., Dunn, 2003; Mallet and Willmott, 2003; Seberg et al., 2003). However, individual taxonomists are no more required to perform their own sequencing than individual photographers need to develop their own photographs. Barcoding has already moved to the "photomat" stage, with tens of thousands of specimens being analyzed at low cost in high-volume barcoding facilities (e.g., at the University of Guelph, Canada, and the Smithsonian Institution, USA). The direct "burden" on taxonomic collaborators involves fueling the analytical train by providing small tissue samples from identified, vouchered specimens to be barcoded. As microfluidic technologies mature over the next decade, one can expect the development of affordable, user-friendly, compact—if not handheld—devices that integrate all stages from DNA extraction through analysis of the barcode sequence to gain an identification (a movement to the "Polaroid" stage in the analogy with photography). Although we might expect such instruments to become standard equipment for both taxonomic research and for the broader community of organizations and individuals that need rapid access to species identifications, this certainly does not imply that barcoding will turn taxonomists into molecular biologists.

We believe the tools provided by DNA barcoding will add rigour to the generation and testing of taxonomic hypotheses. Taxonomy has generally been executed using discontinuities in *analog* (i.e., graded morphological) traits to infer species boundaries, an approach that has generated a total of 1.7 million taxonomic hypotheses over 250 years. DNA barcoding allows these

hypotheses to be tested using an independent *digital* (i.e., DNA nucleotide-based) data stream. Although there has been good correspondence between species recognized through morphological approaches with designations based on barcodes, there are discordances. These cases should be welcomed as they will strengthen both taxonomic hypotheses and methods for analyzing barcode differences, and may lead to new discoveries regarding evolution and ecology. All of these benefits have been evidenced in early barcoding efforts.

9. *Assuming the technical problems of DNA barcoding can be overcome, is it now, or will it ever be cost-effective relative to traditional methods to use DNA barcodes for bioinventory purposes?*

A major benefit of, and rationale for, DNA barcoding lies in its cost-effectiveness for species identification, especially in ambitious bioinventory and biomonitoring programs (Smith et al., 2005). As it stands, production-line systems for identifying even a small group of thoroughly known species through morphological approaches cost about \$2 per specimen (e.g., mosquito monitoring programs that deal with fewer than 60 species; F. C. Hunter, personal communication). When a team of taxonomic specialists targets a larger assemblage of species in a specific geographic area, costs rise substantially and the identification of single specimens can cost \$50 to \$100 if all costs are internalized. Today, a DNA barcode can be generated for about \$5 per specimen including labor and sequencing—and this cost is expected to plummet. In time, DNA barcoding programs have the potential to become self-sufficient by charging a small fee for identifications while still maintaining open access for academic researchers.

Cost is only one criterion in evaluating the utility of a taxonomic support system for biodiversity research. Speed, reliability, and accessibility are just as important, and we believe that DNA barcoding excels in these areas. By contrast, even smaller-scale biomonitoring programs based on morphology currently face a major challenge in delivering results in a rapid, cost-effective fashion. This can have dramatic economic consequences, as with the current difficulty in identifying invasive species early enough to suppress an outbreak. Yet, the economic benefits of excluding even a single noxious invader, such as the zebra mussel from North America, would have been sufficient after a decade to barcode most of the animal species on Earth.

10. *Hypothesis-driven research is the foundation upon which most research agencies assign funding priorities, yet taxonomy is discovery driven. How would your approach to taxonomy convince these agencies of the merits of taxonomic studies?*

We agree that hypothesis-driven science dominates small-scale funding competitions and that taxonomy fares poorly because of its discovery-driven nature. On the other hand, every “big science” initiative—from subatomic physics to the human genome to space

exploration—has been discovery-driven, and this will be the funding arena in which the global DNA barcoding program will operate if it rises to the challenge. As with most other big science projects, DNA barcoding has experienced claims that it is “not science,” and that it threatens the ability of smaller laboratories to carry out hypothesis-driven research. In past cases, such claims have always proved shortsighted. The repeated observation is that large-scale discovery science spins off hypotheses at a frenetic pace and reveals avenues of investigation that could never have been anticipated. In this sense, many of the criticisms levied at DNA barcoding are remarkably similar to those given a decade ago regarding the human genome project.

DNA barcoding has already been successful in attracting substantial funding from varied agencies and organizations that have not been traditional funders of taxonomy, but this has not been accomplished by selling “taxonomy” per se. Instead, the DNA barcoding initiative promotes the vision of a broadly accessible inventory of life’s diversity. It is only by emphasizing the benefits to society and by sparking interest among the taxpaying public that support for a global biodiversity initiative will be generated. Of course, that does not mean that taxonomy is set to become a “high tech service industry” for other biologists, as some have suggested (e.g., Lipscomb et al., 2003; Wheeler, 2004; Will and Rubinoff, 2004). A major goal of DNA barcoding is to enable the non-taxonomist majority of biologists—and indeed, *anyone*—to access taxonomic information directly while allowing professional taxonomists to focus on generating more such knowledge.

#### POSITION STATEMENT

Efforts to inventory eukaryotic diversity through morphological analyses have enjoyed much success. The generation of nearly two million taxonomic hypotheses over the past 250 years is an impressive feat that has provided a foundational understanding of biological diversity, but many details await clarification. DNA barcoding is positioned to aid the inventory of life by accelerating species discovery, by testing current taxonomic hypotheses, and by making species identifications more easily available. These contributions will not be made at the expense of core taxonomic values or funding. DNA barcoding does not seek to abandon “morphological studies in favor of a narrow and wholly molecular identification system” (Will and Rubinoff, 2004: 47). Rather, it strives to build alliances between molecular and morphological taxonomists (Hebert and Barrett, 2005). It seeks, as well, to preserve the Linnaean principles by which species are named and classified. DNA barcoding requires existing, morphologically derived species names for calibration, and it is these names that are recovered when barcoding is used for identification.

It is generally accepted that the study of biodiversity is seriously underfunded (Godfray, 2002). It is not easy to attribute this to the theme of investigation, as

biodiversity science is important and attracts much public interest. However, this area of research does suffer from a culture of conflict. Rather than mounting grand collaborations, the biodiversity community has a tradition of polarization and infighting. DNA barcoding is no stranger to invectives—it has been tagged as “theoretically vacuous technology” and as a “parlour trick” (Wheler, 2004; Will et al., 2005). The coupling of such comments with ad hominem attacks on DNA barcoding proponents brings little credit to the discipline.

Some critics charge that the DNA barcoding approach is fundamentally flawed, but the available data tell a very different story: the success of DNA barcoding has so far been startlingly impressive. As Smith (2005) notes, barcoding fared well in a test executed at the PEET conference. More importantly, a series of studies have now investigated the effectiveness of DNA barcoding in species assemblages from varied geographic settings, and from numerous taxonomic groups with divergent life history and evolutionary attributes. As a consequence of these sensitivity tests, barcode records are now available for more than 13,000 animal species (and accumulating rapidly) and they reveal resolution that is no illusion ([www.barcodinglife.org](http://www.barcodinglife.org)). In group after group, success in species identification exceeds 95% and the few cases of compromised resolution involve the inability to discriminate a small group of closely allied species (Hebert et al., 2004a, 2004b; Hebert et al., unpublished). Typical results resemble those in Figure 1, which shows the patterning of DNA barcode divergences for 31 species of *Acronicta*, one of the most diverse lepidopteran genera in North America. In this case, there is no evidence of the sequence sharing between taxa that would be expected if hybridization were occurring or if species were too young to be discriminated. Instead, there is clean separation of species with barcode cohesion for conspecifics even when they derive from disparate sites in eastern North America.

There is nothing exceptional about the barcode results for *Acronicta*—studies on soil invertebrates from the arctic and on lepidopterans from the tropics show similar success in species resolution (Hogg and Hebert, 2004; Janzen et al., 2005). This performance extends into marine settings: a barcode study which examined more than 200 morphologically defined species of Australian fishes generated 100% success in their discrimination (Ward et al., 2005). Sensitivity tests across 10-fold gradients in rates of mitochondrial evolution revealed high success in species identification from insect groups with both the lowest and highest rates of evolution (Ball et al., 2005; Smith et al., 2005). Shifts in nucleotide composition of the mitochondrial genome similarly fail to impact the resolution of DNA barcoding, as evidenced by success in groups, such as birds, with high G+C composition and others, such as spiders, with extreme A+T bias (Hebert et al., 2004b; Barrett and Hebert, 2005).

If these past studies are reflective of the general performance of barcodes across the animal kingdom, a comprehensive *cox1*-based system will deliver taxonomic

resolution in excess of 99.99% when viewed from a kingdom-level perspective. To grasp this, one need only imagine each of the 10,000 pits in Figure 2 as a repository for barcode data from a single species. Presuming that there are 10 million animal species, the barcode library for this kingdom could be represented by just 1000 of these pages. The global avifauna, which consists of about 10,000 bird species, will occupy only one of these pages. Barcode records for the fishes of the world will occupy three pages, while beetles will fill several hundred pages. Once all 10 million pits have been filled with barcode data, the analysis of any new barcode sequence will provide immediate transport to the correct page out of 1000, delivering 99.9% resolution. In fact, based on data for North American birds, the barcode sequence will provide perfect resolution by leading to an individual species pit on the single bird page in 96% of cases. In the remaining cases, the newly gathered barcode will match sequences in two or three adjacent pits. In summary, a short barcode will collapse the uncertainty in species identity from any one of 10 million species down to a single species in most cases, and to a small subset of closely allied species in other instances.

Funding is now in place to ensure that the DNA barcode library for animals will grow by at least 500,000 records over the next 5 years, providing coverage for some 50,000 species. Although this will be far from a complete registry of species, it will allow DNA barcodes to function as an effective identification tool for those taxonomic groups with comprehensive barcode records. For example, as barcode coverage for fishes, birds, and pest insects approaches completion, this will provide open access to the identification of these species regardless of life stage or condition. As this core of species records is joined by barcodes from other animals, a global identification system for this kingdom of life will rise.

Although we believe that the generality of barcoding has now been demonstrated for the animal kingdom, there remains a need to both establish and evaluate barcode protocols for the other kingdoms of life. The core principles of barcode analysis (minimalization and standardization of sequence targets) are surely applicable to these organisms, but the selection of gene regions and tests of their effectiveness remain in progress although early results on plants (Kress et al., 2005) and protists (G.W. Saunders, personal communication) provide cause for optimism. Aside from its success in separating known species, DNA barcoding will be a powerful aid in resolving other taxonomic issues. Overlooked species have been regularly revealed, even in well-studied groups such as North American birds (Hebert et al., 2004b), butterflies (Hebert et al., 2004a), and silk moths (Janzen et al., 2005). Its role in associating life stages (Beskansky et al., 2003) and genders (Janzen et al., 2005), and in clarifying synonymies, will also be of assistance in many other taxonomic investigations.

The activation of any major science program demands not just a strong scientific rationale, but a demonstration of societal relevance. DNA barcoding exhibits such

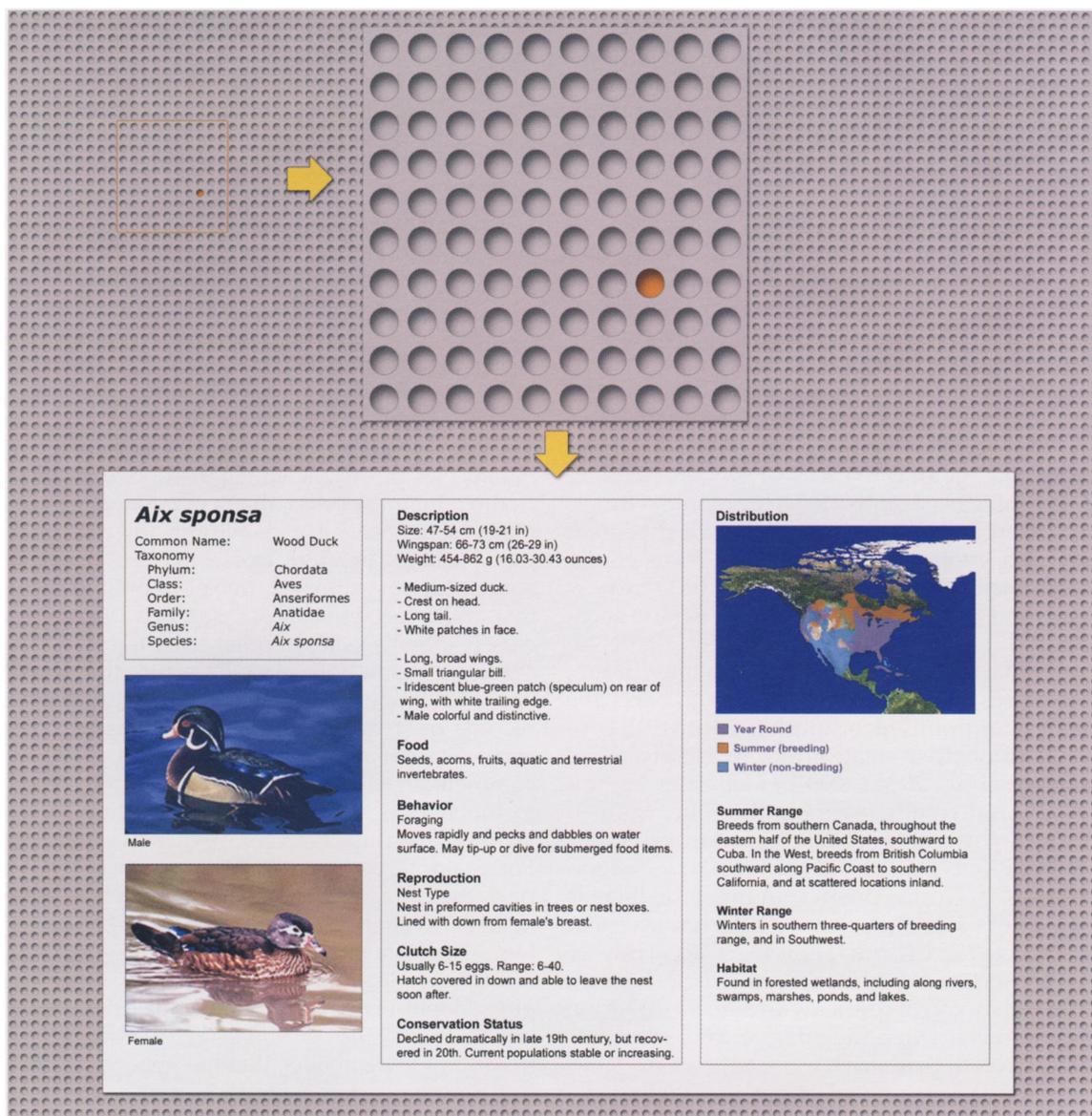


FIGURE 2. A graphic representation of a matrix that would store barcode data on 10,000 species (e.g., nearly all known bird species). One thousand such pages would house barcode records for 10 million species. In addition to providing a species epithet, such a system would act as a portal to all other collected information on a given species by linking into other comprehensive biological databases.

relevance by providing new access to identifications in varied contexts. Efforts to conserve life are currently constrained by the need for an identification system, and we believe that this need can only be met by DNA barcoding (see also Smith et al., 2005). The ability of barcodes to identify fragments of life has applications ranging from the resolution of cases of species substitution in the marketplace (Marko et al., 2004) to the protection of food security through, for example, screening animal feedstuffs for ruminant waste. More generally, the ability of DNA barcoding to deliver identifications quickly and cheaply has the potential to revolutionize humanity's relationship with biological diversity (Janzen, 2004).

If DNA barcoding proceeds on a large scale, it will generate important by-products for the scientific community. All DNA extracts produced during the barcode

analysis of vouchered specimens will be stored, allowing future efforts to examine patterns of sequence diversity in other gene regions, and the collection programs instigated by DNA barcoding will expand the specimens available for morphological analysis. The barcode initiative will also create a Web-based system delivering not just automated identifications, but also providing a portal to biological information for all species included in the registry. Although DNA barcoding will not create the "encyclopedia of life," it will generate its index and table of contents.

Because of both positive scientific results and its recognized societal benefits, there is growing enthusiasm for a large-scale DNA barcoding initiative. Two meetings at Cold Spring Harbor during 2003 clarified plans for action (Stoeckle, 2003), and more meetings have

followed. The most recent, which was hosted by the Natural History Museum in London, attracted more than 230 researchers (Marshall, 2005). The barcode movement also has a central organizing force: the Consortium for the Barcode of Life (CBOL), hosted by the Smithsonian Institution in Washington, which was launched in mid-2004. More than 80 organizations from 25 nations, including many prominent museums, have already joined CBOL ([www.barcoding.si.edu](http://www.barcoding.si.edu)). The first global barcode campaigns have been activated under its auspices; they include plans to barcode all 10,000 species of birds and all 15,000 marine fishes by 2010. Clearly, work needs to expand beyond individual laboratories to tackle major projects such as these, and national barcode networks are forming to establish specimen supply chains and to oversee core analytical facilities. The first of these, the Canadian Barcode of Life Network, which saw activation in May 2005 ([www.bolnet.ca](http://www.bolnet.ca)), intends to barcode at least 10,000 Canadian animal species over the next 5 years.

We view these signs of growing synergy among the various sectors of the biodiversity community as extremely hopeful. If developed to their full potential, history may view the DNA barcoding enterprise as one that not only enhanced access to taxonomic information, but also strengthened alliances among all those with interests in the documentation, understanding, and preservation of biodiversity—an exciting prospect indeed.

#### ACKNOWLEDGEMENTS

We thank Mark Stoeckle and Dan Janzen for valuable revisionary suggestions to earlier drafts of this article. We are also very grateful to the Gordon and Betty Moore Foundation, NSERC, CFI, and OIT for their support of DNA barcoding research at the University of Guelph.

#### REFERENCES

- Alroy, J. 2002. How many named species are valid? *Proc. Natl Acad Sci. USA* 99:3706–3711.
- Ball, S. L., P. D. N. Hebert, S. K. Burian, and J. M. Webb. 2005. Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *JNABS* 24:508–524.
- Barrett, R. D. H., and P. D. N. Hebert. 2005. Identifying spiders through DNA barcodes. *Can. J. Zool.* 83:481–491.
- Beskansky, N. J., D. W. Severson, and M. T. Ferdig. 2003. DNA barcoding of parasites and invertebrate disease vectors: What you don't know can hurt you. *Trends Parasit.* 19:545–546.
- Dunn, C. P. 2003. Keeping taxonomy based in morphology. *Trends Ecol. Evol.* 18:270–271.
- Ebach, M. C., and C. Holdrege. 2005. DNA barcoding is no substitute for taxonomy. *Nature* 434:697.
- Gaston, K. J., and M. A. O'Neill. 2004. Automated species identification: Why not? *Phil. Trans. R. Soc. Lond. B* 359:655–667.
- Godfray, H. C. J. 2002. Challenges for taxonomy. *Nature* 417:17–19.
- Hebert, P. D. N., and R. D. H. Barrett. 2005. Reply to the comment by L. Prendini on "Identifying spiders through DNA barcodes." *Can. J. Zool.* 83:505–506.
- Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. deWaard. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* 270:313–321.
- Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrapes fulgerator*. *Proc. Natl. Acad. Sci. USA* 101:14812–14817.
- Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis. 2004b. Identification of birds through DNA barcodes. *PLoS Biology* 2:1657–1663.
- Hogg, I. D., and P. D. N. Hebert. 2004. Biological identifications of springtails (Hexapoda: Collembola) from the Canadian arctic using mitochondrial DNA barcodes. *Can. J. Zool.* 82:749–754.
- Janzen, D. J. 2004. Now is the time. *Phil. Trans. R. Soc. Lond. B* 359:731–732.
- Janzen, D. H. M. Hajibabaei, J. M. Burns, W. Hallwachs, E. Remigio, and P. D. N. Hebert. 2005. Wedding biodiversity inventory of a large and complex lepidoptera fauna with DNA barcoding. *Phil. Trans. R. Soc. Lond. B* In press.
- Kress, W. J., K. J. Wurdack, E. A. Zimmer, L. A. Weigt, and D. H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA* 102:8369–8374.
- Lipscomb, D., N. Platnick, and Q. Wheeler. 2003. The intellectual content of taxonomy: A comment on DNA taxonomy. *Trends Ecol. Evol.* 18:65–66.
- Mallet, J., and K. Willmott. 2003. Taxonomy: Renaissance or Tower of Babel? *Trends Ecol. Evol.* 18:57–59.
- Marko, P. B., S. C. Lee, A. M. Rice, J. M. Gramling, T. M. Fitzhenry, J. S. McAlister, G. R. Horn, and A. L. Moran. 2004. Mislabelling of a depleted ref fish. *Nature* 430:309–310.
- Marshall, E. 2005. Will DNA bar codes breathe new life into classification. *Science* 307:1037.
- Seberg, O., C. J. Humphries, S. Knapp, D. W. Stevenson, G. Petersen, N. Scharff, and N. M. Andersen. 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends Ecol. Evol.* 18:63–65.
- Smith, M. A., B. L. Fisher, and P. D. N. Hebert. 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: The ants of Madagascar. *Phil. Trans. R. Soc. Lond. B* In press.
- Smith, V. S. 2005. DNA barcoding: Perspectives from a "Partnerships for Enhancing Expertise in Taxonomy" (PEET) debate. *Syst. Biol.* 54:841–844.
- Stoeckle, M. 2003. Taxonomy, DNA, and the bar code of life. *BioScience* 53:2–3.
- Wahlberg, N., R. Oliveira, and J. A. Scott. 2003. Phylogenetic relationships of *Phycoides* butterfly species (Lepidoptera: Nymphalidae): Complex mtDNA variation and species delimitations. *Syst. Entomol.* 28:1–17.
- Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. N. Hebert. 2005. A start to DNA barcoding Australia's fish species. *Phil. Trans. R. Soc. Lond. B* In press.
- Wheeler, Q. D. 2004. Taxonomic triage and the poverty of phylogeny. *Phil. Trans. R. Soc. Lond. B* 359:571–583.
- Will, K. W., B. D. Mishler, and Q. D. Wheeler. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54:844–851.
- Will, K. W., and D. Rubinoff. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20:47–55.

First submitted 25 April 2005; reviews returned 9 June 2005;

final acceptance 12 July 2005

Associate Editor: Vincent Savolainen