Research Focus

# DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you

**Nora J. Besansky, David W. Severson and Michael T. Ferdig**

Center for Tropical Disease Research and Training, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA

**Only a small fraction of extant species is known to science. Recently, there have been calls to census all biological diversity. Traditional morphology-based assessments are time-consuming and require specialists whose numbers are insufficient and dwindling. A DNA-based method called DNA barcoding has been proposed as a rapid means of cataloguing species. This initiative is especially attractive with respect to parasites for which morphometrics are difficult or impossible, and their mosquito hosts for which morphological expertise is rare. As an index of diversity, DNA barcoding is distinct from the science of circumscribing species and resolving their evolutionary relationships, but it can serve as a powerful scaffold both to motivate and guide these efforts.**

What if it was possible to catalogue all of animal diversity? Many taxonomists agree that this goal is beyond reach using traditional morphology-based taxonomic methods, considering that only ~10% of species on earth have been discovered in the first 250 years since Linnaeus [1]. Recently, several groups have argued that DNA-based identification should become the scaffold for taxonomic knowledge [2–4]. In practice, DNA-based identification is already well-established in the literature, and has become the standard for identification within many taxonomic groups. Hebert *et al*. [3] specifically suggest the employment of DNA sequences as taxon 'barcodes' and propose that the mitochondrial gene cytochrome oxidase I (COI) serves as the core of a global bioidentification system for animals. Response has been spirited and mixed. Some have argued that identifying species based only on a snippet of DNA risks impoverishing taxonomy by diminishing the hypothesis-driven aims of circumscribing and classifying species [5,6]. Others, pointing to the fact that DNA barcodes might not always correspond with species recognizable by standard criteria, question the utility of a system that could incorrectly identify two members of a species as separate species, or two separate species as the same. Indeed, why would anyone want to identify something that might not be a real species? Why would anyone want to identify a species not linked to information on higher categories? The answer is that the role of any molecular diagnostic is to aid research, not to serve as an end in itself. Barcoding as a method used to recognize species is independent of questions as to whether individual taxa are species, what species are (or should be), and where they fit in a unified tree of life. There is no doubt that the painstaking and hypothesis-driven aims of circumscribing and classifying species are vital and should not be ignored. However, we posit that DNA barcoding provides a powerful framework for organismal identification that will ultimately stimulate understanding of evolutionary relationships: organisms must be discovered and identified before they can be subjected to further study (G.G. Simpson as cited by Zavortink in Ref. [7]).

Nowhere is the gap in taxonomic knowledge more urgent than for medically important pathogens and their invertebrate vectors. In terms of the amount of morbidity and mortality inflicted on the human population, it can easily be claimed that parasites and their vectors are among the most dangerous groups of animals on earth. Not surprisingly, these groups have received more taxonomic attention than most other animal groups. Yet, even for groups such as mosquitoes, our taxonomic knowledge remains limited. Furthermore, barcodes offer distinct opportunities for the study of parasite taxa that are largely under-represented in the record of animal species.

## Convention or diagnosis?

The challenges of parasite identification can be extraordinary: in addition to being small, parasites develop through complex, multi-host life cycles, sometimes living deep in the host tissues. They also reside in and on hosts as assemblages of many parasite species and cryptic species complexes. Seberg *et al*. [5] have argued that 'naming' is a convention rather than a science. This perspective does not carefully consider parasitism. Because diversity and distributions of parasitic taxa are necessarily described in terms of their host organisms, naming is more than a cataloguing procedure but, instead, constitutes a diagnosis. The question 'what is it?' is central to understanding the host–parasite interactions that underlie parasitic diseases afflicting humans, domesticated animals and wildlife. Parameters unique to living on host organisms, such as host specificity, prevalence, intensity, virulence and transmission measures, are dependent, initially, on accurate diagnoses and, ultimately, on reconstruction of host and parasite ecological and evolutionary relationships, features that are significantly enhanced by molecular data. Primary diagnosis is crucial for understanding local epidemiological and ecological features of parasitic diseases [8], and the literature is rich with examples of application of sequence data in addressing both proximal

*Corresponding author:* Nora J. Besansky (nbesansk@nd.edu).

(identification) and historical questions in host–parasite systems. Proper parasite identification allows researchers to recognize important reservoir hosts, as well as distinguishing morphologically similar species that cause distinct pathologies [9,10]. In addition, host–parasite systems are excellent models for evolutionary studies, and molecular markers have been instrumental in establishing patterns of host–parasite co-evolution and host-switching [11,12], and in recognizing the evolutionary origins of transmission mechanisms, such as the acquisition of enhanced oral transmission in *Toxoplasma* [13].

## Still naming names

Of the >3500 known species of mosquitoes, only a handful are medically important as vectors of arboviruses, malaria parasites and/or nematodes. Except for a few of these medically important species, mosquito taxonomy has stagnated for >100 years at the stage of recognizing and naming species. Nevertheless, it is estimated that only a small fraction of extant species has been discovered [7,14], and there exists no comprehensive natural system of classification. The situation is made worse by the fact that expert mosquito taxonomists (generally sampled without replacement) are becoming the subjects of memorial lectures issued at the annual American Mosquito Control Association meeting (http://www.mosquito.org), begging the question: can initiatives like DNA barcoding improve the future of mosquito taxonomy?

'It is only through major projects like these that the diversity of mosquitoes becomes known. Taxonomic revisions produced by projects like these provide the data that will ultimately lead to the development of a natural classification. That classification will, in turn, suggest innumerable hypotheses that can be tested by studies at the gamma level of taxonomy, and may provide some basis for understanding the interrelationships between mosquitoes and the disease-causing organisms they vector' [7]. Taken from a memorial lecture delivered 13 years ago, these words did not refer to a DNA barcoding project, but to the monumental faunal studies of John Belkin, 'Mosquitoes of Middle America' and 'Southeast Asia Mosquito Project' [7]. Although DNA barcoding was unimaginable at that time, the message conveyed by Zavortink motivates our vision of DNA barcoding as a stimulus to the development of a natural classification.

## Moon shot of mosquito systematics?

Barcoding is not an end in itself, but will boost the rate of discovery. The unique contribution of DNA barcoding to mosquito taxonomy and systematics is a compressed timeline for the exploration and analysis of biodiversity. Yet, as Belkin did in his fieldwork, care should be taken to record the geographic source and habitat of each specimen, to obtain progeny rearings of all taxa wherever possible, and to retain voucher specimens, so that the different kinds of data, including morphometric, can be integrated. The discriminatory power of DNA barcoding would increase with the number of genes employed, and the additional sequence information would strengthen subsequent systematic research, but this benefit must be balanced against the added time and expense for a barcoding project. There are good grounds for believing that COI would be a good starting point [15]. In a minority of cases, regardless of which or how many genes are employed, barcoding will not work. DNA barcoding based on any mitochondrial DNA (mtDNA) (or nuclear) gene will fail when applied to species flocks, and where the members are very recently diverged and/or possibly hybridizing. There are known examples of closely related cryptic taxa that cannot be distinguished based on mtDNA, and suspected cases of interspecific hybridization that have resulted in gene flow between species [16,17]. These cases, although inherently interesting and important epidemiologically, are not typical of the vast majority of mosquito diversity as far as is known. Thus, despite its imperfections, the potential benefits of barcoding outweigh the costs. Inevitably, the flood of data will suggest 'innumerable hypotheses', the testing of which will enrich taxonomy. Already we bear witness to this phenomenon in the wake of recently completed genome sequencing projects that were not inherently hypothesis-driven and yet are profoundly impacting the study of biology.

## References


1 Pennisi, E. (2003) Modernizing the tree of life. *Science* 300, 1692–1697
2 Blaxter, M. (2003) Counting angels with DNA. *Nature* 421, 122–124
3 Hebert, P.D.N. *et al.* (2003) Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B. Biol. Sci.* 270, 313–321
4 Tautz, D. *et al.* (2003) A plea for DNA taxonomy. *Trends Ecol. Evol.* 18, 70–74
5 Seberg, O. *et al.* (2003) Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends Ecol. Evol.* 18, 63–65
6 Lipscomb, D. *et al.* (2003) The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends Ecol. Evol.* 18, 65–66
7 Zavortink, T.J. (1990) Classical taxonomy of mosquitoes – a memorial to John N. Belkin. *J. Am. Mosq. Control Assoc.* 6, 593–600
8 Monis, P.T. *et al.* (2002) Molecular biology techniques in parasite ecology. *Int. J. Parasitol.* 32, 551–562
9 Sterud, E. *et al.* (2002) The use of host specificity, pathogenicity, and molecular markers to differentiate between *Gyrodactylus salaris* Malmberg, 1957 and *G. thymalli* Zitnan, 1960 (Monogenea: Gyrodactylidae). *Parasitology* 124, 203–213
10 Hopkins, R.M. *et al.* (1997) Ribosomal RNA sequencing reveals differences between the genotypes of *Giardia* isolates recovered from humans and dogs living in the same locality. *J. Parasitol.* 83, 44–51
11 Ricklefs, R.E. *et al.* (2002) Diversification and host switching in avian malaria parasites. *Proc. R. Soc. Lond. B. Biol. Sci.* 269, 885–892
12 Johnson, K.P. *et al.* (2003) When do parasites fail to speciate in response to host speciation? *Syst. Biol.* 52, 37–47
13 Su, C. *et al.* (2003) Recent expansion of *Toxoplasma* through enhanced oral transmission. *Science* 299, 414–416
14 Krzywinski, J. and Besansky, N.J. (2003) Molecular systematics of *Anopheles*: from subgenera to subpopulations. *Annu. Rev. Entomol.* 48, 111–139
15 Gaunt, M.W. and Miles, M.A. (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol. Biol. Evol.* 19, 748–761
16 Besansky, N.J. *et al.* (1997) Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *An. arabiensis*, suggest extensive gene flow. *Genetics* 147, 1817–1828
17 Besansky, N.J. *et al.* (2003) Semipermeable species boundaries between *Anopheles gambiae* and *An. arabiensis*: evidence from multilocus DNA sequence variation. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10818–10823


1471-4922/$ - see front matter 
doi:10.1016/j.pt.2003.09.015