

Species Discovery versus Species Identification in DNA Barcoding Efforts: Response to Rubinoff

ROB DESALLE

Division of Invertebrates, American Museum of Natural History, 79th Street at Central Park West, New York, NY 10024, U.S.A., email desalle@amnh.org

Introduction

Rubinoff's (2006) essay in *Conservation Biology* raises some important issues relevant to the DNA barcoding initiative. Some of these issues are valid, others have been discussed recently in the DNA barcoding literature (Blaxter & Floyd 2003; Sperling 2003; Hebert et al. 2004; Moritz & Cicero 2004; Ebach & Holdrege 2005), and still others can be addressed by looking at DNA barcoding in a different light (DeSalle et al. 2005). Because papers in the DNA barcoding literature suggest that conservation biology is one of the major fields that will benefit from DNA barcoding, Rubinoff correctly argues that examining the validity of this claim is necessary.

Problems with DNA Barcodes and Species

It will be instructive for the conservation biology community to examine the problems raised by Rubinoff in detail and I hope this Comment generates interest in the discussion. I start with Rubinoff's second point that there are "problems with DNA barcodes and species." This "problem" has been at the heart of the discussion at the several meetings that have been held on DNA barcoding (Banbury, 1 March 2003; Banbury, 2 September 2003; Smithsonian Institution, Washington D.C., May 2004; CBoL London Meeting, February 2005; for reports on these meetings see <http://phe.rockefeller.edu/BarcodeConference/>). There were originally two purposes of DNA barcoding discussed and debated at these initial meetings. The first and most controversial aspect is species "discovery" (Matz & Neilsen, 2005). The second and more valid aspect, in my opinion, is species "identification." Species discovery is primarily the domain of taxo-

nomy. Species identification, which comes after established taxonomy, should be the domain of DNA barcoding.

In a taxonomic context DNA sequence information in the absence of other corroborating evidence can never be used by itself as an indicator of species delimitation. The same is true of morphological information: it is never used by itself without corroborating data. This is because any hypothesis about a species unit needs to be tested in a scientific context (Goldstein et al. 2000; Lipscomb et al. 2003; Sites & Marshall 2004; DeSalle et al. 2005). To look only at DNA sequences of organisms and then call these organisms species without a hypothesis test completely ignores the intellectual and scientific content of taxonomy (Dunn 2003; Lipscomb et al. 2003; Seberg et al. 2003). On the other hand DNA sequences can be used to augment taxonomy and species delimitation as either corroborating evidence for existing hypotheses or as starting points for further testing by other means (DeSalle et al. 2005).

Species "identification" based on DNA sequences does not violate the intellectual content of taxonomy or the scientific basis of taxonomy and species delimitation. This statement makes sense because species identification starts with known species designations on the basis of the work of taxonomists and biologists studying speciation. Species identification through DNA sequencing, then, uses the delimitations of taxonomy to establish a means for identifying unknown organisms by querying an existing database.

Rubinoff takes issue with the distance-based "barcode" approach as do several other authors (DeSalle et al. 2005; Meyer & Paulay 2005; Prendini 2005). When a character-based system in which DNA sequence information is used to generate discrete diagnostics (DeSalle et al. 2005) is implemented, the distance problem disappears. In fact, a DNA character-diagnostic approach then can become

an integral part of the usual way taxonomists examine morphological or behavioral attributes when describing species.

In addition, DNA barcoding can contribute to refining species discovery. Once a set of barcodes is established for a group of organisms, unidentified specimens (i.e., query specimens) can be examined. With an effective DNA identification or barcoding system, typically the query specimens will be identified to species based on the barcodes from a database. Nevertheless, some query specimens will not be identified to species. In these situations the DNA sequence information should raise a red flag about the existing taxonomy. The DNA barcodes, then, offer new hypotheses about the taxonomy of the group. These DNA sequence-based hypotheses are then open for testing and are the stimulus and starting point for further taxonomic revision of a particular group. Once the taxonomic community and researchers in DNA barcoding both get the point that DNA barcoding is an extension and augmentation of taxonomy and can be used for species identification (not discovery), the problems discussed by Rubinoff evaporate.

Problems with mtDNA as a Sole Source of Data

Hebert et al. (2003) suggest that there is enough power in short DNA sequences to provide unique signatures that could be used as character-based diagnostics. Davis and Nixon (1992) point out that species can be diagnosed by fixed and different attributes. These fixed and different characters are the gold standard of a character-based DNA identification system. DeSalle et al. (2005) also point out that DNA sequence positions in a short stretch have even more diagnostic character information than the positions that are fixed and different among species. Combinations of positions that are private and even polymorphic to particular species can be found and used in species identification (DeSalle et al. 2005). This leads me to suggest that a short, single nucleotide sequence the size of the Cytochrome Oxidase I (COI) fragment proposed by the DNA barcoding community is long enough to generate unique character-based DNA diagnostics for even very closely related cryptic species. Should one use the COI gene by itself to generate a phylogenetic hypothesis? The answer to that question based on the last decade of systematic research with multiple gene partitions is a resounding no (Gatesy et al. 2002; Rokas et al. 2003).

Problems with Application

The final problem of DNA barcoding application in the conservation community discussed by Rubinoff is also well taken. Conservation biologists have to work with the

public sector at many junctures. The public does need to be convinced that a conservation approach is scientifically valid. Any scientific approach needs to be viewed as reliable, consistent, honorable, and applicable in courts of law and in national and international governing agencies.

On the other hand, conservation biologists require methods that are rapid and informative because conservation biology is a crisis discipline (DeSalle & Amato 2004). Would a species identification system based on DNA sequences be reliable, consistent, and rapid? I think it would because once the problems of species discovery are removed from the equation, the reliability and consistency issues are moot. More important, a DNA identification system could be much more rapid than current morphological identification approaches. I am not advocating elimination of morphological diagnostics or approaches; rather, I argue that a barcode include DNA, morphology, behavior, geographic information, and any other attribute of the organism that facilitates its identification. If DNA methods are accessed with ease and rapidity in the field, then there is no reason they should not be used if a species identification system has been established. It is not unreasonable to assume that DNA technology will advance to the stage where field-based diagnostics can be accomplished. This development would greatly facilitate field conservation and biodiversity studies, and the utility of a field-based system of rapid species identification should be of obvious importance in conservation, biodiversity studies, and ecological assessments.

One of the hardest criticisms levied by Rubinoff concerns the skewed degree of funding to the DNA barcoding community when classical taxonomy could use the funding. This is a political as opposed to a scientific issue. Nevertheless, there are some underlying scientific points to be made that are relevant to this criticism. The DNA sequences in a species identification system can offer an excellent rationale for organizing and interacting with biodiversity databases. Workers are currently attempting to organize the information emanating from biodiversity studies. These initiatives include the Global Biodiversity Information Facility (GBIF), All Species Foundation, Morphobank, and Treebase, among many. The lack of uniformity of information being reported across these databases is a difficulty (Agosti 2003, 2006). Obtaining, archiving, and using a common DNA sequence or set of DNA sequences across a wide range of taxa with a uniform format for submission, accession, storage of tissues, and informatics would greatly enhance our understanding of biodiversity. DNA barcoding initiatives include such approaches to database construction as an integral component of the barcoding effort.

This is a plea not to throw the baby out with the bathwater, as Rubinoff suggests with DNA barcoding. His criticism is valid only if species discovery, exclusively through DNA, becomes the *raison d'etre* of DNA barcoding. If

DNA barcoding is defined and used as a species identification method, it has the potential to augment and draw funds to biodiversity assessment and conservation biology projects in general.

Acknowledgments

I thank R. Hanner for critical reading of an early draft of this paper. The writing of this Comment was supported by the Louis B and Dorothy Cullman Program in Molecular Systematics at the American Museum of Natural History and by a grant from the Lounsbury Foundation.

Literature Cited

- Agosti, D. 2003. Encyclopedia of life: should species description equal gene sequence? *Trends in Ecology & Evolution* **18**:273–274.
- Agosti, D. 2006. Biodiversity data are out of local taxonomists' reach. *Nature* **439**:392–393.
- Blaxter, M., and R. Floyd. 2003. Molecular taxonomics for biodiversity surveys: already a reality. *Trends in Ecology & Evolution* **18**:268.
- Davis, J. I., and K. C. Nixon. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Zoology* **41**:421–435.
- Desalle, R., M. G. Egan, and M. Siddall. 2005. The unholy trinity: taxonomy, species delimitation, and DNA barcoding. *Philosophical Transactions of the Royal Society, London B* **360**:1905–1916.
- Desalle, R., and G. A. Amato. 2004. The expansion of conservation genetics. *Nature Reviews Genetics* **5**:702–712.
- Dunn, C. P. 2003. Keeping taxonomy based in morphology. *Trends in Ecology & Evolution* **18**:270.
- Ebach, M. C., and C. Holdrege. 2005. DNA Barcoding is no substitute for taxonomy. *Nature* **434**:697.
- Gatesy, J., C. Matthee, R. DeSalle, and C. Hayashi. 2002. Resolution of a supertree/supermatrix paradox. *Systematic Biology* **51**:652–664.
- Goldstein, P. Z., R. DeSalle, G. Amato, and A. P. Vogler. 2000. Conservation genetics at the species boundary. *Conservation Biology* **14**:120–131.
- Hebert, P. D. N., S. Ratnasingham, and J. R. deWaard. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Philosophical Transactions of the Royal Society, London B. Series B* **270**(supplement):S96–S99.
- Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis. 2004. Identification of birds through DNA barcodes. *Public Library of Science, Biology* **2**:e312.
- Lipscomb, D., N. Platnick, and Q. Wheeler. 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends in Ecology & Evolution* **18**:64–66.
- Matz, M. V., and R. Nielsen. 2005. A likelihood ratio test for species membership based on DNA sequence data. *Philosophical Transactions of the Royal Society, London B* **360**:1969–1974.
- Meyer, C. P., and G. Paulay. 2005. DNA Barcoding: error rates based on comprehensive sampling. *Public Library of Science, Biology* **3**:e422.
- Moritz, C., and C. Cicero. 2004. DNA Barcoding: promises and pitfalls. *Public Library of Science, Biology* **2**:e354.
- Prendini, L. 2005. Comment on 'Identifying spiders through DNA barcoding'. *Canadian Journal of Zoology* **83**:498–504.
- Rubinoff, D. 2006. Utility of mitochondrial DNA barcodes in Species Conservation. *Conservation Biology* DOI:10.1111/j.1523-1739.2006.00372.x.
- Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**:798–804.
- 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 in Ecology & Evolution* **18**:63–65.
- Sites, J. W., Jr., and J. C. Marshall. 2004. Empirical criteria for delimiting species. *Annual Review of Ecology & Systematics* **35**:199–229.
- Sperling, F. 2003. DNA barcoding: deux et machina. University of Alberta, Edmonton, Alberta. Newsletter of the Biological Survey of Canada (terrestrial arthropods), opinion page. Available from http://www.biology.ualberta.ca/bsc/news22_2/contents.htm (accessed 1 February 2006).

