

NIH Public Access

Author Manuscript

Mol Ecol Resour. Author manuscript; available in PMC 2011 November 27.

Published in final edited form as:

Mol Ecol Resour. 2011 November ; 11(6): 937-942. doi:10.1111/j.1755-0998.2011.03054.x.

Comment on 'Birdstrikes and barcoding: can DNA methods help make the airways safer?'

SCOTT FEDERHEN

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, Bethesda, MD 20894, USA

Abstract

GenBank is the database of record for public sequence data. Results reported in the scientific literature that are based on sequence data cannot be evaluated if the underlying data is not in the public record.

Keywords

birdstrike; BoLD; COI; DNA barcode; GenBank; species identification

DNA barcoding is an extremely promising method for species identification that relies on the development of a reference library of barcode sequences from vouchered specimens that have been reliably identified by taxonomic experts. As with all sequence data, the database of record for public barcode sequences is the International Nucleotide Sequence Database (GenBank/EMBL/DDBJ). The Barcode of Life Data Systems (BoLD) is an online workbench for the collection and analysis of barcode data, much of which has been sequenced at the associated Biodiversity Institute of Ontario (BIO) at the University of Guelph. BoLD brokers the submission of barcode sequences to GenBank as a service to their user community.

Birdstrikes are an important and natural application of the barcoding method of species identification (Waugh *et al.* 2011). In this study, the authors obtained blood samples from 40 birdstrike incidents, of which 37 yielded enough DNA to sequence COI barcodes. These sequences were run against the BoLD identification engine, with 29 samples receiving unambiguous species identifications to 18 different species. Unfortunately, the data supporting these results are not available for public review, and the conclusions do not reach the standard of scientific evidence that is subject to independent examination and validation by others. There are two problems.

First, the sequences generated in this study were not submitted to GenBank and are not presented in the paper, so the analysis that is reported cannot be repeated. Most journals (including *Molecular Ecology Resources*) require GenBank accessions for all newly reported sequence data, for precisely this reason. We would still welcome the submission of these 'birdstrike environmental sample' sequences to GenBank, where they could at least be linked to the PubMed abstract in Entrez – perhaps the accessions could be published as a corrigendum.

The second problem is more subtle but also more serious. Most of the taxonomic identifications surfaced by the BoLD identification engine are from private entries – the

Correspondence: Scott Federhen, Fax: 301 402 9651; federhen@ncbi.nlm.nih.gov.

sequence and the specimen data that are associated with the species name are not accessible at BoLD and have never been submitted to GenBank. Of the 18 species reported as birdstrike identifications in Waugh *et al.*, three have absolutely no sequence data in GenBank (*Circus approximans, Chlidonias albostriatus* and *Anthus novaeseelandiae*), and six more have no COI barcode data (*Vanellus miles, Sterna striata, Gymnorhina tibicen, Charadrius bicinctus, Larus bulleri* and *Cygnus atratus*). The others do have public COI barcode data, but there is no way to tell whether the best hits reported in Waugh *et al.* are among them. The BoLD identification does provide a searchable 'Public Record Barcode Database', but it includes only 21% of the total number of barcode records at BoLD and would not have produced the results reported here.

In the present case, neither the study sequences nor the barcode reference sequences underpinning these identifications are in the scientific record. These results cannot be reproduced, and the evidence supporting the conclusions cannot be examined. Species identifications derived from the BoLD identification engine should be reported as 'unpublished results' with some discussion of the relevant caveats. The potential to identify birdstrikes with barcodes is clear, but it is not yet a scientific reality on the scale presented in the study of Waugh *et al.*

References

Waugh J, Evans MW, Millar CD, Lambert DM. Birdstrikes and barcoding: can DNA methods help make the airways safer? Molecular Ecology Resources. 2011; 11:38–45. [PubMed: 21429099]