

# OSAC RESEARCH NEEDS ASSESSMENT FORM



**Title of research need:**

Verifying the accuracy of sequences in public databases for forensic taxonomic identification of diverse biological materials

**Describe the need:**

In human forensic DNA analysis, curated reference databases have been developed for use in statistical analyses and comparisons. Databases meeting the standards expected by law enforcement and the criminal justice system for human DNA analyses do not exist for most non-human biological materials encountered in forensic casework, including but not limited to plants, fungi, and terrestrial and marine vertebrates and invertebrates. While some forensic laboratories have developed in-house reference databases for the taxa and associated loci commonly encountered in their casework, most practitioners also utilize DNA sequences contained in public sequence databases. Unfortunately, it has been well documented that public sequence databases such as GenBank contain erroneous data, therefore it is imperative that the accuracy of sequences in such databases be verified prior to use in taxonomic identification.

**Keyword(s):**

GenBank, wildlife, accuracy, nucleotide sequences, public databases

**Submitting subcommittee(s):**

Wildlife Forensic Biology

**Date Approved:**

10/7/2022

**Background Information:**

1. Does this research need address a gap(s) in a current or planned standard? (ex.: Field identification system for on scene opioid detection and confirmation)

This research need dovetails with an existing draft standard on the use of GenBank for taxonomic assignment. However, that standard does not discuss verification of the accuracy of the data included in such public sequence databases.

2. Are you aware of any ongoing research that may address this research need that has not yet been published (e.g., research presented in conference proceedings, studies that you or a colleague have participated in but have yet to be published)?

No. We are not aware of any extensive studies that focus on assessing the accuracy of public sequences for taxonomic identification of diverse non-human biological taxa encountered in forensic casework.

3. Key bibliographic references relating to this research need: (ex.: Toll, L., Standifer, K. M., Massotte, D., eds. (2019). Current Topics in Opioid Research. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-180-3)

[1] Meiklejohn KA, Damaso N, Robertson JM. (2019). Assessment of BOLD and GenBank – their accuracy and reliability for the identification of biological materials. PLoS ONE. 14(6): e0217084

[2] Nilsson RH, Ryberg M, Kristiansson E, et al. (2006). Taxonomic reliability in DNA sequences in public sequence databases: a fungal perspective. PLoS ONE. 1(1):e59.

[3] Ashelford KE, Chuzhanova NA, Fry JC et al. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl. Environ. Microbiol. 71:7724-7736.

[4] Bridge PD, Roberts PJ, Spooner BM et al. (2003). On the unreliability of published DNA sequences. New Phytol. 160:43-48.

[5] Vilgalys R. (2003). Taxonomic misidentification in public DNA databases. New Phytol. 160:4-5.

4. Review the annual operational/research needs published by the National Institute of Justice (NIJ) at <https://nij.ojp.gov/topics/articles/forensic-science-research-and-development-technology-working-group-operational#latest>? Is your research need identified by NIJ?

No.

5. In what ways would the research results improve current laboratory capabilities?

The ability to characterize the accuracy of sequences in public databases for a group of species and the informative locus sequenced for species identification (e.g., *rbcl* for plants, *COI* for insects, *Cytb* for vertebrates, etc.) would be highly valuable, and provide guidance to examiners as to whether such data should be used in casework.

6. In what ways would the research results improve understanding of the scientific basis for the subcommittee(s)?

The accuracy and associated potential error rates with using public sequence databases for taxonomic identification of species encountered in wildlife forensics is not known. Such information would provide valuable guidance for the broader wildlife forensics community when incorporating such data into the analysis.

7. In what ways would the research results improve services to the criminal justice system?

The ability to provide objective and scientifically sound taxonomic identifications for questioned biological materials encountered in casework could potentially a) allow determination of whether a violation has occurred, or b) based on the group of species present, predict the origin of a sample or provide discriminatory information for sample-to-sample comparisons. Such applications could be used to ensure violators are prosecuted, generate investigative leads, or provide valuable information for intelligence purposes.

8. Status assessment (I, II, III, or IV):

II

	Major gap in current knowledge	Minor gap in current knowledge
No or limited current research is being conducted	I	III
Existing current research is being conducted	II	IV

*This research need has been identified by one or more subcommittees of OSAC and is being provided as an informational resource to the community.*