OSAC 2022-S-0037 Standard for DNA-Based Taxonomic Identification in Forensic Entomology

Crime Scene Investigation & Reconstruction Subcommittee
Scene Examination Scientific Area Committee
Organization of Scientific Area Committees (OSAC) for Forensic Science

Draft OSAC Proposed Standard

OSAC 2022-S-0037 Standard for DNA-Based Taxonomic Identification in Forensic Entomology

Prepared by
Crime Scene Investigation & Reconstruction Subcommittee
Version 1.0
February 2023

Disclaimer:

This OSAC Proposed Standard was written by the Organization of Scientific Area Committees (OSAC) for Forensic Science following a process that includes an open comment period. This Proposed Standard will be submitted to a standards developing organization and is subject to change.

There may be references in an OSAC Proposed Standard to other publications under development by OSAC. The information in the Proposed Standard, and underlying concepts and methodologies, may be used by the forensic-science community before the completion of such companion publications.

Any identification of commercial equipment, instruments, or materials in the Proposed Standard is not a recommendation or endorsement by the U.S. Government and does not imply that the equipment, instruments, or materials are necessarily the best available for the purpose.

Foreword

This standard was developed to provide guidance to scene investigators and scene reconstructionist as a standard for DNA-based taxonomic identification in forensic entomology.

This document has been prepared by the Organization of Scientific Area Committees (OSAC) for Forensic Science Crime Scene Investigation and Reconstruction Subcommittee.

Keywords: *forensic entomology, death investigation, taxonomy, DNA-based identification, polymerase chain reaction, DNA sequencing*

Table of Contents

1. Scope

This document outlines the current best practices to be used when employing DNA-based techniques to identify species, or higher taxonomic categories, of entomological origin for forensic investigation purposes. Potential applications include the estimation of some portion of the postmortem interval (PMI); however, methods for estimating PMI after a specimen has been identified are beyond the scope of this document.

2. Normative Reference

There are no normative references.

See Annex A, (informative) Bibliography, for other references

3. Terms and Definitions

For purposes of this document, the following definitions and acronyms apply.

3.1
Allele
Any of the forms of the same gene that occur at the same locus on a homologous chromosome but differ in base sequence (OSAC Lexicon).

3.2
Amino acid
Any class of 20 molecules combined to form proteins in living things; the sequence of amino acids in a protein is determined by the exons of a gene (OSAC Lexicon).

3.3
Amplification
An increase in the number of copies of a specific DNA fragment; can be *in vivo* or *in vitro* (OSAC Lexicon).

3.4
Artifact
A non-allelic product of the DNA amplification or sequencing process (e.g., amplification of a pseudogene, non-template nucleotide addition, primer-dimer, or other non-specific product), an anomaly of the detection process (e.g., single or multichannel voltage spikes, or instrument noise), or a by-product of primer synthesis (e.g., “dye blob”) that may be observed on an electropherogram. Some artifacts may complicate the interpretation of DNA profiles when they cannot be distinguished from the actual alleles from a particular sample (OSAC Lexicon).

3.5
BLAST search
The Basic Local Alignment Search Tool (BLAST) is an algorithm for comparing and finding similarities between a query biological sequence and sequences deposited into the database and calculates a statistical significance of similarity. BLAST was developed and is currently run by the National Center for Biotechnology Information (NCBI) (Altschul et al, 1990).

3.6
Casework samples
Entomological material recovered from the scene or otherwise believed to be associated with a crime (Catts and Goff 1992).

3.7
Controls
Samples of known types run in parallel with experimental, reference, or evidence samples that are used to demonstrate that a procedure is working correctly (OSAC Lexicon). A negative control has no added template DNA; therefore, a result consistent with DNA present suggests contamination. A positive control has standard DNA of known genotype added; therefore, a result consistent with the expected genotype indicates that the protocol is reliable.

3.8
Deoxyribonucleic acid
DNA
A genetic material of organisms, usually double-stranded, is a biopolymer composed of nucleic acids, identified by the presence of deoxyribose, a sugar, and the four nucleobases. DNA is a stable molecule; variations in the DNA sequence between individuals and species permit DNA testing to distinguish individuals and species from each other (OSAC Lexicon).

3.9
Electropherogram
The graphic representation of the separation of molecules by electrophoresis in which the data appear as peaks along a line; the format in which DNA typing results are often presented, with the X-axis displaying the observed allele(s) and the Y-axis recording the relative amount of DNA detected based on the signal measured in relative fluorescent units (RFU) collected during analysis (OSAC Lexicon).

3.10
Haplotype
A set of linked DNA variations, or polymorphisms, tend to be inherited together (e.g., the insect mitochondrial DNA molecule) (OSAC Lexicon).

3.11
Locus (plural loci)
A unique physical location of a gene (or specific sequence of DNA) on a chromosome (OSAC Lexicon.)

3.12
Peak height
The maximum Y-axis value obtained for a data peak measured in relative fluorescence units (OSAC Lexicon).

3.13
Phred score
A Phred quality score is a numerical value associated with each sequenced nucleotide that corresponds to the quality of that nucleotide that is logarithmically linked to error probabilities. For example, a Phred score of 10 has a base call accuracy of 90%, and a Phred score of 20 has a base call accuracy of 99% (Ewing et al., 1998).

3.14
Polymerase chain reaction
PCR
A method whereby a specific sequence of nucleotides within a double-stranded DNA is amplified (OSAC Lexicon).

3.15
Reciprocal monophyly
A monophyletic group consists of a hypothetical common ancestor and all its descendants. Members of the group can be organisms or genotypes. Depicted on a phylogenetic tree, a monophyletic group is defined by a single branch from which all members, and no non-member, arise. Reciprocal monophyly exists for two species when the branches defining each, and no other species, arise from a single node on the phylogenetic tree. Figure 1 illustrates the terms applied to a taxonomic group according to the phylogenetic relationships of its member species (Schuh and Brower 2009).

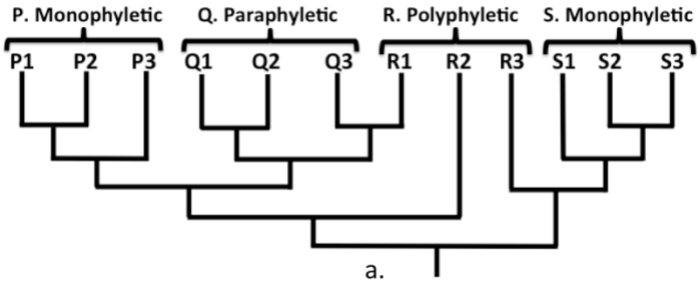

Figure 1: Phylogenetic Relationships

A hypothetical phylogeny illustrates the classification of a taxonomic group when each member organism (or genotype) is more related to any other member than it is to any outside species (monophyletic), one lineage of the group is more related to species outside the group (paraphyletic), or two or more member lineages are more related to species outside the group (polyphyletic) (Tarone et. al. 2015).

3.16
Paraphyly
A paraphyletic group contains a hypothetical common ancestor and some, but not all, of its descendants (Schuh and Brower 2009).

3.17
Primers
A short polynucleotide chain, usually 18-30 bases long, targets a specific region of the template DNA and allows a DNA polymerase to initiate the synthesis of a complementary strand; Two primers are required to amplify a particular section of double-stranded DNA. The primers are complementary to opposite strands and are designed to bracket the region of interest for amplification. Polymerization is initiated at the 3' end of the primer and extends in a 5´ to 3´ manner (OSAC Lexicon).

3.18
Protein-coding gene
The sequence of DNA which encodes for a protein. This usually includes a start and stop codon and is transcribed and translated into a protein (Watson et al., 2014).

3.19
Pseudogene
A DNA sequence that resembles a gene but has been mutated into an inactive form over the course of evolution. It often lacks introns and other essential DNA sequences necessary for function. Though genetically similar to the original functional gene, pseudogenes do not result in functional proteins (National Cancer Institute Dictionary).

3.20
Reference genotype
A genotype from a specimen is considered to have been accurately identified to a given taxonomic level (add to OSAC lexicon).

3.21
Sequencing
A method for determining the order of bases in a DNA molecule. In Sanger sequencing, this is based on the selective incorporation of chain-terminating dideoxynucleotides by a DNA polymerase during in vitro DNA replication (Hillis et al. 1996).

3.22
Sister group(s)
A pair of taxa is inferred to be more closely related to each other than to any other taxon (Schuh and Brower 2009).

3.23
Stop codon
A three-base DNA sequence that codes for or signals the end of the protein sequence. The exact sequence of nucleotides terminating the protein-coding region depends on the particular genetic locus (Watson et al., 2014).

3.24
Species
The level of taxonomic classification is denoted by a binomial (two-word) name in Latinate form (Notton 2001).

3.25
Taxonomic identification
Analyses to establish the classification of an organism to family, genus, species, etc. These analyses are based on class characters diagnostic for the taxonomic level in question (OSAC Lexicon).

3.26
Voucher
Biological specimen that is representative of its species in accordance with the relevant taxonomic authority and is therefore valid for comparative purposes. Voucher specimens are of known identity and are curated with available associated geographic, field collection, and life-history data (OSAC Lexicon).

4. Requirements

4.1 Laboratory Procedures

4.1.1 The laboratory shall have and follow procedures for cleaning and decontaminating facilities and equipment.

4.1.2 Pre-PCR and post-PCR activities of the laboratory shall be separated by space to avoid contamination of questioned samples.

4.1.3 Equipment and supplies shall not be transferred from post-PCR to pre-PCR areas unless decontaminated, e.g., using bleach, with associated documentation.

4.1.4 Casework and non-casework-related laboratory procedures shall be separated by space or time. Cleaning and decontamination of the area between procedures shall occur when using the same laboratory area.

4.2 Sample Processing

4.2.1 A detailed description of the specimen, including close-up images with a size scale, shall be recorded prior to any sample processing for DNA extraction.

4.2.2 When possible, insect morphological characteristics should be preserved and therefore not destroyed for DNA extraction purposes. For soft larvae, a flap may be cut in the cuticle to minimize the destruction of surface features when removing internal tissue. For adults with bilateral symmetry, dissection of body parts for DNA extraction should be attempted first on only one side of the body (for example, one to three legs). The remains of any specimen used for DNA extraction should be preserved for future examination.

4.2.3 Specimens potentially useful for future molecular genetic analysis should be preserved appropriately (e.g., by freezing or in >90% ethanol).

4.3 DNA Extraction

4.3.1 Written protocols shall exist and be available for inspection for all extraction methods used in the laboratory.

4.3.2 Each DNA extraction set shall include at least one reagent blank as a negative control, which is analyzed concurrently with casework samples.

4.3.3 The extraction of DNA from reference material shall be separated by time or space from the extraction of DNA from casework.

4.3.4 When extracted in the same space, samples suspected to have low quantities of DNA should be extracted before samples suspected to have high quantities of DNA.

4.4 Amplification

4.4.1 Written protocols shall exist for all amplification methods routinely used in the laboratory.

4.4.2 Primers used for PCR amplification shall be documented in the case file, and primer sequences shall be available in laboratory documentation.

4.4.3 Each PCR run shall include a positive control, a PCR reagent negative control, and the DNA extraction reagent blank.

4.4.4 All controls shall be amplified concurrently in the same instrument with the casework samples at all loci and with the same primers as the casework samples.

4.4.5 If the analysis includes genotyping reference specimens, these shall not be processed concurrently with casework specimens.

4.4.6 Results from casework samples shall not be accepted unless the positive control produces an expected genotype and unless the negative control(s) produces no genotype.

4.5 DNA Sequencing

4.5.1 Written protocols shall exist for all sequencing methods routinely used in the laboratory.

4.5.2 Redundant sequence data, for example, by sequencing both the forward and reverse strand of a PCR amplicon, should be obtained.

5 Analysis and Interpretation

5.1 General

5.1.1 Written protocols shall exist for all analysis and interpretation methods routinely used in the laboratory. These protocols shall include defined data quality indicators (e.g., Phred quality scores, signal intensities, peak heights).

5.1.2 If contamination in the negative control is present above laboratory-established acceptance parameters, then the results related to that negative control shall not be used for interpretation.

5.1.3 If the evidence genotype includes all or part of the DNA sequence of a protein-coding gene, that sequence shall be translated to the implied amino acid sequence. A misplaced stop codon or other result inconsistent with the biological characteristics of the target gene protein will suggest that the genotype is an artifact or an incorrect locus such as a pseudogene.

Protocols covering sequencing shall minimally include:

a) the process for nucleotide sequence editing and comparison to reference sequences

b) the process by which sequence contamination (e.g., detectable results in the negative controls) is evaluated and documented

c) the determination of minimum sequence quality (e.g., the Phred score of a peak on the electropherogram)

5.2 Taxonomic Identification

5.2.1 Identification of a given taxonomic level shall be made by comparison to reference genotypes considered to be reliable for this purpose. The reference genotype(s) used to identify each specimen shall be noted in the report for analysis.

5.2.2 To be reliable, the reference genotypes shall:

a) be associated with voucher specimens (Wheeler 2003)

b) have been reported in a scientific publication that describes the specimen characters considered to be diagnostic of that taxon, either through direct documentation or by reference to the published taxonomic key, including the species in question that was used to make the identification

5.2.3 A BLAST search of the entire genotype database of the National Center for Biotechnology Information (or comparable database) is acceptable for a preliminary identification, but it shall not be the sole basis for the final taxonomic determination.

5.2.4 The reference genotypes used for comparison to the evidence genotype should represent a selection of taxa appropriate for the setting and any tentative morphological identification of the casework specimen. Knowledge of the local carrion insect fauna is required to do this properly.

This knowledge can include being able to identify a higher taxonomic category (e.g., taxonomic family) of the evidence without yet knowing the genotype and the species in that higher taxonomic category that potentially occur in the circumstances of the scene (e.g., at that location and that time of year). Because intraspecific genotype variation is common, identification is not necessarily made by an exact match to a reference.

Failure to use an appropriate taxon selection could result in the evidence being apparently identified as the most closely related species in the database (Wells and Stevens 2008) when the true species is not included in the database. Morphological identification of a problematic species pair may narrow the need for a comprehensive database but still requires the inclusion of the species in question and their nearest genetic relatives.

5.2.5. Multiple protocols exist for making the comparison between evidence and reference genotypes.

For species identification, a common method is a phylogenetic analysis (Harvey et al. 2008), with the identification indicated by a sister-group relationship (Schuh and Brower 2009) between the evidence genotype and a reference species. When using this approach, the reference database should represent the genotype(s)/haplotype(s) of multiple unrelated individuals of each species.

Reliable species identification by phylogenetic analysis depends on a pattern of reciprocal monophyly for the gene(s) being analyzed. Therefore, particular care should be made to search the published literature and databases for evidence of non-monophyly (see Figure 1). Paraphyly of mitochondrial DNA is not unusual for closely related animal species.

The reporting of phylogenetic results shall include the phylogenetic data file and the software analysis log.

Annex A
(informative)

Bibliography

1] American National Standards Institute/American Academy of Forensic Sciences Standards Board. *ANSI/ASB Standard 048, First Edition*. Wildlife Forensic DNA Standard Procedures. 2019. www.asbstandardsboard.org/wp-content/uploads/2019/05/048_Std_e1.pdf.

2] Altschul, S.F. et al. "Basic local alignment search tool." *J Mol Biol.* Vol. 215, no. 3, 5 Oct 1990, pp. 403-10. DOI: 10.1016/S0022-2836(05)80360-2. PMID: 2231712.

3] Catts, E.P. and M.L. Goff. "*Forensic entomology in criminal investigations." Annual Review of Entomology*, vol 37, 1992, pp. 253-272.

4] Ewing, B. et al.. "Base-calling of automated sequencer traces using Phred. I. Accuracy assessment*." Genome research*, vol. 8, 1998, pp. 175-185.

5] Harvey, M.L. et al.. "A global study of forensically significant calliphorids: implications for identification." *Forensic Science International*, vol. 177, 2008, pp. 66-76.

6] Hillis, D.M. et al. "Nucleic acids IV: sequencing and cloning." *Molecular Systematics.* 2nd Ed, edited by Hillis, D.M., Moritz, C., Sinauer, M.B. 1996, pp. 321-381

7] Notton, D., et al.. "Best practice in the use of the scientific names of animals: Support for editors of technical journals." *Bulletin of Zoological Nomenclature*, vol. 68, no. 4, 2011, pp. 313-322.

8] Schuh, R.T., and A.V.Z Brower. *Biological Systematics.* Cornell University Press, 2009

9] Tarone, A.M., et al.. "Molecular Biology In Forensic Entomology." *Forensic Entomology: International Dimensions and Frontiers*. Edited by J.K. Tomberlin and M.E. Benbow. 2015. Figure reproduced by permission of Taylor and Francis Group, LLC, a division of Informa PLC.

10] Watson, J.D. et al.. *Molecular Biology of the Gene. 7th Ed*. Cold Spring Harbor Laboratory Press, 2014.

11] Wheeler, T.A. *The role of voucher specimens in validating faunistic and ecological research. Biological Survey of Canada (Terrestrial Arthropods) Document Series*, vol. 9, 2003, pp. 1-21.

12] Wells, J.D., and J.R. Stevens. "Applications of DNA-based methods in forensic entomology." *Annual Review of Entomology*, vol. 53, 2008, pp. 103-120.