Wildlife Forensics DNA Standards

DRAFT DOCUMENT

Wildlife Forensics DNA Standards

Keywords: wildlife forensics, DNA, sequencing analysis, STR analysis, taxonomic
identification, microsatellites

Abstract
This document provides minimum standards for forensic DNA analysis of wildlife.

Foreword

The wildlife forensic science community needs a wildlife specific DNA standards document to accommodate the diversity of species and substrates encountered during wildlife forensic analyses.

Differences in evolutionary histories for different taxonomic groups often necessitate the use of multiple markers for source attribution at the species, geographic, or individual level.

This standard was developed by SWGWILD and modified in the OSAC Wildlife Subcommittee DNA Task Group, reviewed by the OSAC Wildlife Subcommittee and presented to the Biology/DNA Scientific Area Committee for movement through the American Academy of Forensic Sciences' Academy Standards Board.

Acknowledgements

Table of Contents

Wildlife Forensics DNA Standards

1 Scope

This document provides minimum standards for forensic DNA analysis of wildlife evidence.

2 Normative References

The following referenced documents are indispensable for the application of this document.

Federal Bureau of Investigation. 2011. Quality assurance standards for forensic DNA testing laboratories. Available: http://www.fbi.gov/about-us/lab/codis/qas-standards-for-forensic-dna-testing-laboratories-effective-9-1-2011 [Accessed 9/24/12].

Federal Bureau of Investigation. 2010. SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Available: http://www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines [Accessed 9/24/12].

Scientific Working Group on Wildlife. 2012. SWGWILD Standards and Guidelines. Available: https://www.wildlifeforensicscience.org/wp-content/uploads/2016/07/swgwild-standards_and_guidelines_2-0_12192012.pdf [Accessed: 8/22/16]

Organization for Scientific Area Committees, Wildlife Forensics Subcommittee, Wildlife Forensics General Standards. Currently under review by the AAFS Standards Board (ASB).

3 Terms and Definitions

3.1 Allele is one of two or more versions of a genetic sequence at a particular location (a locus) in a genome.

3.2 Amplification is the process by which the number of copies of a specific DNA fragment is increased.

3.3 Analytical threshold is the minimum peak height (Relative Fluorescence Units, in RFUs*) requirement at and above which detected peaks on a DNA electropherogram can be reliably distinguished from instrument background noise.

3.4 Artifact is a non-allelic product of the DNA amplification process (e.g. stutter, non-template nucleotide addition, primer-dimer, or other non-specific product), and anomaly of the detection process (e.g. single or multichannel voltage spikes, instrument noise, or pull-up), 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.

3.5 Bin is an allele designation corresponding to the window of fragments sizes for each allele, determined by empirical testing.

3.6   Casework samples are biological material of unknown origin recovered from the crime scene or believed to have originated from an individual of interest that may be associated to a crime.

3.7   Casework reference samples are biological material obtained from a known individual and collected for purposes of comparison to samples.

3.8   Controls are samples of known types, run in parallel with experimental, reference, or evidence samples that are used to demonstrate that a procedure is working correctly. Controls include negative samples (no template DNA) and positive samples.

3.9   Deoxyribonucleic Acid (DNA) is a genetic material of organisms, usually double-stranded, 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 permits DNA testing to distinguish individuals from each other.

3.10   Extracted DNA is total cellular DNA isolated from a biological sample to include nuclear and mitochondrial DNA.

3.11   Genus is the level of taxonomic classification that defines a group of related species.

3.12   Haplotype is a set of linked DNA variations, or polymorphisms, that tend to be inherited together; a combination of alleles or a set of single nucleotide polymorphisms (SNPs) found on the same chromosome.

3.13   Heteroplasmy is the presence of more than one mitochondrial DNA (mtDNA) sequence of type within a single individual.

3.14   Database is a collection of reference material or sequences assembled for use in comparative analyses for purposes of taxonomic identification, individual source evaluation, population assignment, or mitotyping.

3.15   Inclusion is a conclusion statement that a biological sample may have originated from a particular source or individual.

3.16   Locus (plural loci) is a physical location of a gene or specific sequence of DNA on a chromosome.

3.17   Low copy number (LCN) is a term for DNA or a DNA analysis of small quantities of DNA that require modification to PCR protocols to enhance amplification success of the target molecule.

3.18   Negative Control is an analytical control that consists of the reagents used in various stages of testing without the introduction of sample; no results are expected from a negative control. For DNA testing, negative controls include extraction blanks, reagent blanks, and amplification blanks. A negative control in DNA testing is used to detect contamination introduced into the assay during the testing process via reagents, disposables or handling errors (which may impact the results observed from samples tested at the same time).

3.19   Peaks are the visual images of an allele or nucleotide in a DNA sequence as represented on an electropherogram by relative fluorescent units.

3.20   Peak height is the maximum Y-axis value obtained for a data peak, measured in relative fluorescence units.

3.21   Polymerase Chain Reaction (PCR) is the analytical process by which targeted segments of DNA are replicated during repetitive cycles of heating to denaturation, and cooling to anneal primer oligonucleotides and extend DNA sequences to enhance detection of DNA fragments.

3.22   Population is a group of organisms of the same species in a defined geographic area such that any pair of members can interbreed.

3.23   Positive Control is an analytical control sample that is used to determine if a test performed properly. This control consists of the test reagents and a known sample that will provide a positive response in the test.

3.24   Primers are short polynucleotide sequences which target a specific region of template DNA and allow a DNA polymerase to initiate synthesis of a complementary strand.

3.25   Pull-up is an artifact that may occur during analysis of fluorescently-labeled DNA fragments when signal from one dye color channel produces artificial peaks in another, usually adjacent color, at a similar position on the X axis in an electropherogram; sometimes referred to as bleed-through or matrix/ spectral calibration failure; one type of DNA testing artifact.

3.26   Reagent Blank is a sample that contains no analyte, but contains the buffer, test reagents, or other materials required to perform a particular test. This is a negative control where results are not expected. The observation where the result(s) in this control may indicate the presence of contaminated or compromised reagents, which may impact the DNA or the serological results observed from samples tested at the same time. For forensic DNA testing this control is treated the same as, and parallel to, the DNA samples being analyzed.

3.27   Reference materials are biological specimens, materials or substances of known identity and verified properties, or data derived from them.

3.28   Reference samples are samples of known origin collected for purposes of comparison to samples of unknown origin.

3.29   Sequencing (Sanger sequencing) is a method of DNA sequencing for determining the order of bases in a DNA molecule based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication.

3.30   Source attribution is a decision which identifies an individual as the source of the DNA that produced an evidentiary single-source or major contributor profile; This statement is often based on population frequency estimates that are more rare than some defined number, generally more than the Earth's population.

3.31 Species is a level of taxonomic classification that denotes interbreeding natural populations that are reproductively isolated from other interbreeding natural populations.

3.32 Taxonomic identifications are analyses used to establish the classification of an organism to family, genus, species, etc. These analyses are based on characters (e.g. morphological or genetics) diagnostic for the taxonomic level in question.

3.33 Theta correction is a method for calculating match probabilities, first described by Balding and Nichols (1994), to allow for population structure in the population for which a frequency database is constructed. It allows match probabilities for subpopulations to be calculated from whole population allele frequencies. It avoids the need to assume Hardy-Weinberg equilibrium at the whole-population level.

3.34 Uninterpretable is the inability to use results for comparisons due to poor or limited data quality.

4 General DNA Standards

4.1 Laboratory

4.1.1 Protocols covering laboratory facilities shall include the process by which facilities and equipment are cleaned and decontaminated.

4.1.2 Pre-PCR and post-PCR activities of the laboratory shall be separated by space.

4.1.3 Equipment and supplies shall not be transferred from post-PCR to pre-PCR areas unless decontaminated using generally accepted laboratory protocols.

4.1.4 Casework and non-casework-related research shall be separated by space or time.

4.2 DNA Extraction

4.2.1 Protocols shall exist for all extraction methods used in the laboratory.

4.2.2 Each DNA extraction set shall include at least one reagent blank that is analyzed concurrently with casework samples.

4.2.3 Extraction of DNA from reference material shall be separated by time or space from extraction of DNA from evidence.

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

4.3 Amplification

4.3.1 Protocols shall exist for all amplification methods routinely used in the laboratory.

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

4.3.3 Each PCR run shall include a positive control and a negative control for the PCR.

4.3.4 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.

4.3.5 Results from casework samples shall not be accepted unless the positive control produces an expected genotype, distinct from the casework samples.

4.4 Analysis and Interpretation

4.4.1 Protocols shall exist for all analysis and interpretation methods routinely used in the laboratory, and these protocols shall include defined data quality indicators.

EXAMPLE: peak quality scores, signal intensities, peak heights.

4.4.2 If contamination in the negative control is present above laboratory established acceptance parameters, then the data shall not be used for interpretation.

4.5   Statistics

NOTE: Statistical support is not necessary when determining exclusions or minimum number of individuals. Statistical support is not appropriate when conducting species identification.

4.5.1 Protocols shall exist for all statistical analyses routinely used in the laboratory.

4.5.2 The laboratory shall perform statistical analysis in support of any non-taxonomic source attribution that is determined to be relevant in the context of a case.

4.5.3 All databases used for statistical analysis shall be documented and identified in the laboratory documentation.

4.5.4 The laboratory shall not use inconclusive or uninterpretable data in statistical analysis.

4.5.5 Theta shall be incorporated in all appropriate statistical analyses.

   a)   For taxa with limited mobility or species with non-panmictic breeding, relevant estimates of population structure shall be acquired.

   b)   Theta shall be adjusted to reflect the acquired relevant estimates of population structure.

           NOTE: When Θ is not known for a particular species, a conservative adjustment shall be incorporated based on data available from taxa expected to have similar population structure.

5   Sequencing Analysis

5.1   General

5.1.1 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 (for example detectable results in the negative controls) is evaluated and documented,

   c)   the interpretation of sequence mixtures,

   d)   the determination of minimum sequence quality.

5.2   Taxonomic Identification

5.2.1 Taxonomic identification based on sequence data shall be supported by the following:

NOTE: If a particular area listed below is not applicable to the identification being completed, a clear narrative of the basis of this determination shall be present in case documentation.

a) comparisons to validated reference sequences that represent the alleged species and closely related species if available,

b) use of a locus for comparison that provides resolution for the taxonomic group of interest,

c) demonstration that genetic distances between closely related taxa support identification at the taxonomic level of interest,

d) information related to the biogeographic or ecological characteristics, life history, or taxonomic characters of the species of interest,

e) citations of published phylogenies for species not commonly analyzed in the laboratory.

NOTE: Well-described taxonomic groups do not require extensive support with published phylogenies. This requirement is particularly applicable to species not commonly analyzed in a laboratory that occasionally appear in wildlife forensic casework.

5.2.2 When a public or shared database is used to provide support for a taxonomic identification:

a) sequences shall be evaluated for suitability as forensic reference sequences prior to use in comparison to casework samples,

b) conclusions shall not be based on a single sequence from a public database.

5.2.3 If species-level identifications are not supported by comparison to laboratory reference materials or reference sequences obtained from public databases, results shall be reported as inconclusive or at a higher taxonomic level, and the limitations of the conclusion shall be clearly stated in the report.

5.3 Mitochondrial Haplotype Determination for Source Evaluation for Individuals

5.3.1 Both forward and reverse strands of DNA should be sequenced and compared for quality control purposes.

5.3.2 Only sequences where double coverage has been obtained shall be used for interpretations.

5.3.3 Procedures shall be defined for interpretation and documentation of heteroplasmy.

5.3.4 The method of defining haplotypes shall be documented.

5.3.5   All databases used for statistical analysis shall be:

    a)   thoroughly documented in case record,

    b)   identified in the report.

6 STR Analysis

NOTE: These standards apply to both autosomal and Y-chromosome short tandem repeat analysis.

6.1   General

6.1.1   Protocols for all STR analyses used in the laboratory shall include, at minimum:

    a)   a method for defining the analytical threshold for alleles used to assign genotypes,

        NOTE: These signal intensity criteria are determined by generally accepted values based on the collection platform or are determined empirically by internal validation.

    b)   a method for defining a set of criteria for allele designation,

    c)   a method for defining the number of loci required for interpretation,

    d)   a method for defining allele bins, if applicable,

    e)   a process to distinguish artifacts, stutter peaks and pull-up peaks, where applicable, from true allele peaks,

    f)   analysis and interpretation of degraded or low template DNA, if applicable,

    g)   a method for population assessment, specifically addressing reference databases used.

6.1.2   In laboratories that report on the composition of mixture samples, a method must be established that specifically addresses analysis and interpretation of such mixtures.

6.1.3   DNA samples shall be quantified prior to amplification unless validation demonstrates that quantification is not necessary.

6.2   Quality Controls

6.2.1   A size standard shall be run with samples.

6.2.2   The sample allele designation shall only be used if the largest and smallest alleles for that assay fall within the range covered by the size standard.

6.2.3   When data is shared between laboratories, allele calls shall be standardized by the use of samples of known genotype.

6.3 Data Interpretation

6.3.1 Relevant population databases shall be developed and maintained for:

a)   Source attribution for individuals,

b)   Population assessment.

6.3.2 Key values shall be defined and their determination based on validation data for:

a)   Analytical threshold required to assign alleles,

b)   Minimum number of loci required for comparison and statistical analysis for each species.

Annex B
(informative)
Bibliography

Butler, J. 2010. Fundamentals of Forensic DNA Typing, Academic Press, Burlington, MA.

Harris, D. J. 2003. Can you bank on GenBank? Trends in Ecology and Evolution 18:317-319.

Holland, M. M. and Parsons, T. J. 1999. Mitochondrial DNA sequence analysis—validation and use for forensic casework. Forensic Science Review 11:21-50.

Longo, M. S., O'Neill, M. J. and O'Neill, R. J. 2011. Abundant human DNA contamination identified in non-primate genome databases. PLoS One 6:e16410.

National Research Council. 1996. NRC II: The Evaluation of Forensic DNA Evidence, Washington, DC, National Academy Press.

Ross, H. A., Murugan, S. and Li, W. L. S. 2008. Testing the reliability of genetic methods of species identification via simulation. Systematic Biology 57:216-230.

Stephenson, J. J., Campbell, M. R., Hess, J. E., et al. 2009. A centralized model for creating shared, standardized, microsatellite data that simplifies inter-laboratory collaboration. Conservation Genetics 10:1145-1149.

Weir, B. S. and Cockerham, C. C. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

Weir, B. S. and Hill, W. G. 2002. Estimating F-statistics. Annual Review of Genetics 36:721-50.

Wilson, M. R., DiZinno, J. A., Polanskey, D., et al. 1995. Validation of mitochondrial DNA sequencing for forensic casework analysis. International Journal of Legal Medicine 108:68-74.