

# **OSAC 2022-S-0011**

# **Standards for Construction of Multilocus Databases**

Wildlife Forensic Biology Subcommittee  
Biology Scientific Area Committee (SAC)  
Organization of Scientific Area Committees (OSAC) for Forensic Science



# OSAC Proposed Standard

## OSAC 2022-S-0011

# Standard for Construction of Multilocus Databases

Prepared by  
Wildlife Forensic Biology Subcommittee  
Version: 2.0  
June 2024

---

### Disclaimer:

This OSAC Proposed Standard was written by the Wildlife Forensic Biology Subcommittee of 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 standard 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.

To be placed on the OSAC Registry, certain types of standards receive a Scientific and Technical Review (STR). The STR process is vital to OSAC's mission of generating and recognizing scientifically sound standards for producing and interpreting forensic science results. The STR shall provide critical and knowledgeable reviews of draft standards to ensure that the published methods that practitioners employ are scientifically valid, and the resulting claims are trustworthy.

The STR consists of an independent and diverse panel, which may include subject matter experts, human factors scientists, quality assurance personnel, and legal experts as applicable. The selected group is tasked with evaluating the proposed standard based on a defined list of scientific, administrative, and quality assurance based criteria.

For more information about this important process, please visit our website at: <https://www.nist.gov/organization-scientific-area-committees-forensic-science/scientific-technical-review-str-process>

## Foreword

This standard defines the minimum requirements that shall be met when developing multilocus population genetic databases for wildlife forensics.

The composition of a database intended for use in population genetic analyses is critical for accurate comparison among the individual subjects as well as statistically sound group assignment (e.g. individual, relatedness, population, geographic source, taxonomic grouping). Analysts must use their expert knowledge in assessing the scientific merit of results obtained from analysis of allele frequency and population genetic data, and in the subsequent reporting of these results.

**Keywords:** *wildlife forensics, population database, population genetics, multilocus, DNA*

## Table of Contents

1	Scope .....	6
2	Normative References.....	6
3	Terms and Definitions.....	6
4	Requirements .....	8
5	Conformance .....	10
	Annex A.....	11

## Standards for Construction of Multilocus Databases

### 1 Scope

This standard sets forth the minimum requirements that shall be met when developing multilocus population genetic databases for wildlife forensics. This document covers criteria for the identification of samples, inclusion of associated biological information, choice and evaluation of genetic markers, standard statistical evaluation of the reference database, and evaluation and quality assurance of databases. This document does not cover such specific applications, such as, individual and familial relationship evaluation, geographic assignment, or other scientific techniques performed on wildlife forensic casework. This document only applies to databases generated from reference samples and does not include samples derived from evidence items. This standard addresses what standards and technical procedures a laboratory should have but is not meant itself to constitute a standard for what specific validation studies are necessary (e.g., the representativeness of test samples, their choice of thresholds, and the like in order to meet scientific requirements of validity). This standard sets out only procedures for construction of multilocus databases and does not substitute for standards that will more specifically set forth minimum requirements or best practices for the other procedures mentioned here.

These minimum standards are not intended to replace standards in ISO 17025 or additional forensic laboratory standards, but shall guide laboratories that are working toward meeting those standards as well as laboratories constructing and modifying multilocus population genetic databases. Notes throughout this document offer clarifications and examples of how a lab may meet a specific standard.

### 2 Normative References

ANSI/ASB Standard 019, *Wildlife Forensics General Standards*

ANSI/ASB Standard 046, *Wildlife Validation Standards-STR Analysis*

ANSI/ASB Standard 047, *Wildlife Validation Standards-Validating New Primers for Sequencing*

ANSI/ASB Standard 048, *Wildlife Forensic DNA Standard Procedures*

ANSI/ASB Best Practices Recommendations 114, *Best Practice Recommendations for Internal Validation of Software used in Forensic DNA Laboratories*

### 3 Terms and Definitions

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

**3.1**  
**assignment**

A method for assigning individuals to predefined categories, based on a suite of characters (e.g., multilocus genotype) measured for samples from each category (e.g. potential source populations).

**3.2**  
**autocorrelation**

The degree of correlation between the values of the same variables across different observations in the data as a function of time or space.

**3.3**  
**coefficient of co-ancestry**

The probability that two alleles randomly sampled from different individuals are copies of the same ancestral allele without mutation, that is, the probability that they are identical by descent.

**3.4**  
**population**

A group of organisms of the same species in a defined geographic area, such that any pair of members can interbreed.

**3.5**  
**probability of identity**

The probability of rejecting a null hypothesis given that a specific alternative hypothesis is true.

**3.6**  
**probability of identity for siblings**

The probability that two sibling individuals drawn at random from a population have the same multilocus genotype.

**3.7**  
**sample**

A group of items, test results or portions of material, taken from a large collection of items, test results or portions of material, that serves to provide information that may be used as a basis for making a decision concerning the larger collection.

**3.8**  
**statistical power**

The power of a statistical test of a null hypothesis is the probability that it will lead to the rejection of the null hypothesis, i.e., the probability that it will result in the conclusion that the phenomenon exists.

## **4 Requirements**

The following requirements set forth criteria that shall be met for construction and evaluation of multilocus population genetic databases: including identification of database components, choice of genetic markers, procedures for statistical analysis, and evaluation and interpretation of results for general population genetic analyses. These criteria shall be addressed in laboratory validation studies according to established population genetic theory and practice. Species differ based on demographic, ecological, and evolutionary factors, so quantitative values for the minimum number of individuals and genetic markers needed for a reference database are expected to vary according to the species and populations of interest.

### **4.1 Inclusion Criteria for Genetic Database Samples**

#### **4.1.1 Technical procedures for constructing genetic databases shall include at least:**

- a) sample acquisition;
- b) establishment of parameters for inclusion samples;
- c) validation process for use of genetic markers;
- d) criteria for individual sample data quality;
- e) quality control/curation of sample information and genetic data.

#### **4.1.2 Quality control shall include adherence to standards in ANSI/ASB 019, ANSI/ASB 046, ANSI/ASB 047, and ANSI/ASB 048.**

#### **4.1.3 In determining database composition the laboratory shall assess, at minimum:**

- a) sample size needed to accurately represent source population genetic diversity;
- b) related taxonomic information, including but not limited to:
  - i. presence of subspecies
  - ii. evolutionary significant units (ESU);
  - iii. hybrids in the species group of interest;
  - iv. geographic range of the taxa in question.

#### **4.1.4 If known, the following metadata shall be documented for each sample:**

- a) geographic location of source samples (e.g. sampling location, breeding location, location of death);
- b) sex of individual;
- c) age class of individual;
- d) type of tissue sampled (e.g. fresh tissue, blood, bone, hair, antler, keratin, feces, etc);
- e) collection information (i.e. date, collector - both name and agency/institution, method of collection).

#### **4.1.5 At minimum, genetic markers shall be evaluated for the:**

- a) number of loci required, as determined by laboratory validation;



NOTE: The number of loci needed will vary by species/population and forensic application (e.g. individual evaluation, population assignment, paternity)

- b) genotyping
- c) genetic diversity measures, including but not limited to:
  - i. Hardy--Weinberg Equilibrium;
  - ii. linkage disequilibrium;
  - iii. allelic richness;
  - iv. allelic diversity;
  - v. heterozygosity measures within and among populations;
- d) presence of null alleles.

**4.1.6** Quality criteria shall be established for sample inclusion when adding genetic data to species/population databases. This shall include, at minimum:

- a) minimum acceptable completeness of genotype per sample;
- b) minimum genotype quality measures depending on genotyping platform [e.g. capillary electrophoresis - Relative Fluorescence Units (RFU); Next-generation sequencing - genotype quality score and read depth].

**4.2** Laboratories shall have technical procedures for data entry quality assurance and quality control.

**4.3** Once constructed, the database shall be evaluated for, if applicable:

- a) representative geographic coverage;
- b) power to discriminate species/population boundaries;
- c) power to identify natural groupings that are ecologically or biologically meaningful;
- d) population level allele frequencies;
- e) presence of spatial or temporal autocorrelation;
- f) sex-related bias;
- g) estimates for statistical power (i.e. probability of identity, probability of identity for siblings);
- h) presence of duplicated samples;
- i) coefficient of co-ancestry.

Note: Database composition will vary based on forensic application (e.g. individual and familial relationship evaluation, geographic assignment techniques).

**4.4** Laboratories shall have technical procedures for evaluation of software intended for use in statistical analysis, including commercial programs and programs developed in-house.

**4.5** Once initially constructed and validated, databases augmented with new samples or subsetted shall be re-evaluated as in 4.2.

- a) The most current version of the database should then be made available for public review, unless containing sensitive law enforcement information.

- b) If the multilocus database is made public then practitioners may anonymize the data beforehand. For example, geographic information metadata may be anonymized by:
- i. withholding geographic identifiers;
  - ii. relabeled using generic labels (e.g. "A" , "B" , etc);
  - iii. character masking identifiable information (e.g. 1011 vs 10xx).

**4.6** Laboratories shall have technical procedures for database archival and version control.

**4.7** Laboratories shall have documentation of required assessments/evaluations above for transparency purposes.

## **5 Conformance**

Conformance to the standards outlined in this document is measured by the availability of written documentation in the form of formal technical procedures and methods available for examination.

**Annex A**  
(informative)

**Bibliography**

The following information provides a list of the literature resources that may assist in defining the breadth and scope of this standard. This list is not meant to be all inclusive. The laboratory shall develop a list tailored to its specific needs. Updated references shall be added to the laboratory's list as new methods or technologies are incorporated into the laboratory's technical procedures.

- 1] Evett, IW and B.S. Weir. 1998. Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists. Sinaur Associates Inc, Sunderland, Massachusetts. 1998. Pp. 100
- 2] J.M. Butler. 2010. Fundamentals of Forensic DNA Typing. Elsevier Academic Press, San Diego, CA.
- 3] Cohen J.1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ.
- 4] Legendre, P and L. Legendre .1998) Numerical Ecology. 3rd ed. Elsevier, Amsterdam.
- 5] Ogden, R and A. Linacre. 2015, Wildlife forensic science: A review of genetic geographic origin assignment. Forensic Science International: Genetics, 18:152 - 159.  
<https://doi.org/10.1016/j.fsigen.2015.02.008>
- 6] Rousset F.. 2004. Genetic Structure and Selection in Subdivided Populations. Princeton University Press, Princeton, NJ.
- 7] Slatkin, M, and H. E. Arter. 1991. Spatial Autocorrelation Methods in Population Genetics. The American Naturalist, 138(2): 499–517. *JSTOR*, [www.jstor.org/stable/2462484](http://www.jstor.org/stable/2462484). <https://doi.org/10.1086/285228>. Accessed 11 Mar. 2021.
- 8] Smouse, P., R. Peakall.1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82:561–573.  
<https://doi.org/10.1038/sj.hdy.6885180>
- 9] Waits, L. P, G Luikart, P Taberlet. 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Molecular Ecology, 10(1): 249-256. <https://doi.org/10.1046/j.1365-294x.2001.01185.x>
- 10] Waples, R. S. 1995. Evolutionarily significant units and the conservation of biological diversity under the endangered species act: Evolution and the aquatic ecosystem: defining unique units in population conservation. American Fisheries Society Symposium, vol. 17: 8-27. <https://ci.nii.ac.jp/naid/10021852122/en/>