OSAC 2022-S-0011 Standards for Construction of Multilocus Databases

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

Draft OSAC Proposed Standard

OSAC 2022-S-0011 Standards for Construction of Multilocus Databases

Prepared by
Wildlife Forensics Biology
Version: 1.0
October 2021

Disclaimer:

This OSAC Proposed Standard was written by the (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 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.

To be placed on the OSAC Registry, certain types of standards first must be reviewed by a Scientific and Technical Review Panel (STRP). The STRP process is vital to OSAC's mission of generating and recognizing scientifically sound standards for producing and interpreting forensic science results. The STRP shall provide critical and knowledgeable reviews of draft standards or of proposed revisions of standards previously published by standards developing organizations (SDOs) to ensure that the published methods that practitioners employ are scientifically valid, and the resulting claims are trustworthy.

The STRP panel will consist of an independent and diverse panel, including subject matter experts, human factors scientists, quality assurance personnel, and legal experts, which will be tasked with evaluating the proposed standard based on a comprehensive list of science-based criteria.

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

1   # Standards for Construction of Multilocus Databases

2   ## Foreword

3   This standard defines the minimum requirements that shall be met when developing allele
4   frequency and population genetic databases for wildlife forensics.

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

11

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

13

14

15   **Table of Contents**

1 Scope

This document provides minimum standards to guide the construction of multilocus population genetic databases. This document covers criteria for the identification and collection of samples, inclusion of associated biological data, choice and evaluation of genetic markers, and standard statistical evaluation of the reference database. This document does not cover specific applications such as individual matching, familial matching, geographic assignment, or other wildlife forensic techniques to evidence in wildlife forensic casework. This document only applies to databases generated from reference samples and does not include evidence items.

These minimum standards are not intended to replace standards in ISO 17025 or additional forensic laboratory standards, but are designed to guide laboratories that are working toward meeting those standards. 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*[1]

ANSI/ASB Standard 046, *Wildlife Validation Standards-STR Analysis*[1]

ANSI/ASB Standard 047, *Wildlife Validation Standards-Validating New Primers for Sequencing*[1]

ANSI/ASB Standard 048, *Wildlife Forensic DNA Standard Procedures*[1]

ANSI/ASB Best Practices Recommendations 114, *Best Practice Recommendations for Internal Validation of Software used in Forensic DNA Laboratories*[1]

3 Terms and Definitions

For 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 the individual and for samples from each category (e.g. potential source populations).

3.2
autocorrelation
phenomenon where samples closer in geographic space or time tend to be more similar or dissimilar to each other than expected by chance alone, for a given variable such as allele frequencies.

56 **3.3**
57 **co-ancestry**
58 identical alleles that are copies of the same ancestral allele without mutation; this is a subset of
59 identical by state (IBS).
60
61 **3.4**
62 **familial matching**
63 a method for determining genetic family relationships such as paternity, siblingship and other
64 kinships.

65 **3.5**
66 **individual matching**
67 a method for comparing one DNA profile to another to determine if the DNA profiles are
68 consistent at the level of interest
69
70 **3.6**
71 **population**
72 a group of organisms of the same species in a defined geographic area, such that any pair of
73 members can interbreed.

74 **3.7**
75 **probability of identity**
76 the probability that two randomly drawn individuals within a given population will have
77 identical genotypes at multiple loci (reference Waits et al 2013 or Peakall and Sykes 1996)
78
79 **3.8**
80 **probability of siblings**
81 the probability that two randomly drawn individuals within a given population will have
82 identical genotypes at multiple loci when relatives are included in the sample.
83
84 **3.9**
85 **statistical power**
86 the probability of a hypothesis test finding an effect if there is an effect to be found.
87
88 # 4 Requirements

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

97 ## 4.1 Inclusion Criteria for Genetic Database Samples

98   **4.1.1** Protocols for constructing genetic databases shall include:

99   a. sample acquisition;
100   b. establishment of parameters for inclusion of samples;
101   c. validation process for use of genetic markers;
102   d. criteria for individual sample data quality;
103   e. quality control/curation of sample information and genetic data
104

105   **4.1.2** Quality control shall include adherence to standards in ANSI/ASB 019, ANSI/ASB 046,
106   ANSI/ASB 047, and ANSI/ASB 048

107   **4.1.3** In determining database composition the laboratory shall assess, at minimum:

108   a. sample size needed to accurately represent source population genetic diversity;
109   b. related taxonomic information, including
110   i.   presence of subspecies
111   ii. evolutionary significant units (ESU)
112   iii. hybrids in the species group of interest
113   iv. geographic range of the taxa in question

114   **4.1.4** Criteria shall be established for metadata associated with sample acquisition, to include at
115   minimum:

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

123   **4.1.5** At minimum, genetic markers shall be evaluated for the:

124   a. number of loci required, as determined by laboratory validation.

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

127   b. genotyping error rate;
128   c. genetic diversity measures, including but not limited to:
129   i.   Hardy-Weinberg Equilibrium,
130   ii. linkage disequilibrium,
131   iii. allelic richness,
132   iv. allelic diversity,
133   v.   heterozygosity measures within and among populations;
134   d. presence of null alleles.

135 **4.1.6** Quality criteria shall be established for sample inclusion when adding genetic data to
136 species/population databases. This shall include, at minimum:

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

141 **4.2** Once constructed, the database shall be evaluated for:

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

151 Note: Database composition will vary based on forensic application (e.g. individual matching,
152 familial matching, geographic assignment techniques).
153
154 **4.3** Laboratories shall have protocols for evaluation of software intended for use in statistical
155 analysis, including commercial programs and programs developed in-house.

156 **4.4** Once initially validated, databases augmented with new samples or subsetted shall be re-
157 evaluated as in 4.2.

158 **4.5** Laboratories shall have protocols for database archival and version control.

159 **5 Conformance**

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

162
163
164
165
166
167
168
169
170
171
172

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 protocols.

1] J.M. Butler. Fundamentals of Forensic DNA Typing. Elsevier Academic Press, San Diego, CA, 2010.

2] Legendre P and Legendre L (1998) Numerical Ecology. 3rd ed. Elsevier, Amsterdan.

3] 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

4] Rousset F., 2004. Genetic Structure and Selection in Subdivided Populations. Princeton University Press, Princeton, NJ.

5] Slatkin, M, and H. E. Arter. Spatial Autocorrelation Methods in Population Genetics. The American Naturalist, 138(2): 499–517. *JSTOR*, www.jstor.org/stable/2462484. Accessed 11 Mar. 2021.

6] Smouse, P., R. Peakall. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82:561–573 (1999). https://doi.org/10.1038/sj.hdy.6885180

7] 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

8] 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/