Standard for Construction of Multilocus Databases

Standard for Construction of Multilocus Databases

ASB Approved Xxxxxxx 2024

ANSI Approved Xxxxxxxx 2024

410 North 21st Street
Colorado Springs, CO 80904

This document may be downloaded from: *www.aafs.org/academy-standards-board*

Foreword

This standard addresses the requirements for 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 may 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.

The American Academy of Forensic Sciences established the Academy Standards Board (ASB) in 2015 with a vision of safeguarding Justice, Integrity, and Fairness through Consensus Based American National Standards. To that end, the ASB develops consensus based forensic standards within a framework accredited by the American National Standards Institute (ANSI), and provides training to support those standards. ASB values integrity, scientific rigor, openness, due process, collaboration, excellence, diversity and inclusion. ASB is dedicated to developing and making freely accessible the highest quality documentary forensic science consensus Standards, Guidelines, Best Practices, and Technical Reports in a wide range of forensic science disciplines as a service to forensic practitioners and the legal system.

This document was revised, prepared, and finalized as a standard by the Wildlife Forensics Consensus Body of the AAFS Standards Board. The draft of this standard was developed by the Wildlife Forensic Biology Subcommittee of the Biology Scientific Area Committee of the Organization of Scientific Area Committees (OSAC) for Forensic Science.

Questions, comments, and suggestions for the improvement of this document can be sent to AAFS-ASB Secretariat, asb@aafs.org or 410 N 21st Street, Colorado Springs, CO 80904.

All hyperlinks and web addresses shown in this document are current as of the publication date of this standard.

ASB procedures are publicly available, free of cost, at www.aafs.org/academy-standards-board.

Keywords: *TBD*

Table of Contents *(to be finalized prior to publication)*

Standard for Construction of Multilocus Databases

1 Scope

This standard sets forth the minimum requirements for developing multilocus population genetic databases for wildlife forensics, including criteria for the identification of samples, inclusion of associated biological information, selection and evaluation of genetic markers, standard statistical evaluation of the reference database, and evaluation and quality assurance of databases. This standard applies to databases generated from reference samples and excludes those derived from evidence items.

This document does not cover specific applications, such as individual and familial relationship evaluation, geographic assignment, or other scientific techniques performed on wildlife forensic casework. This standard addresses the technical procedures a laboratory requires, but does not specify what validation studies (e.g., representativeness of test samples, choice of thresholds) are necessary to meet scientific requirements of validity.

2 Normative References

For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ANSI/ASB Standard 019, *Wildlife Forensics General Standards*, 1st Ed., 2019[a].

ANSI/ASB Standard 046, *Wildlife Validation Standards-STR Analysis*, 1st Ed., 2019[a].

ANSI/ASB Standard 047, *Wildlife Validation Standards-Validating New Primers for Sequencing*, 1st Ed., 2019[a].

ANSI/ASB Standard 048, *Wildlife Forensic DNA Standard Procedures*, 1st Ed., 2019[a].

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 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.

---

[a] Available from: https://www.aafs.org/academy-standards-board

32 **3.3**
33 **coefficient of co-ancestry**
34 The probability that two alleles randomly sampled from different individuals are copies of the same
35 ancestral allele without mutation, that is, the probability that they are identical by descent.

36 **3.4**
37 **population**
38 A group of organisms of the same species in a defined geographic area, such that any pair of
39 members can interbreed.

40 **3.5**
41 **probability of identity**
42 The probability that two individual samples would have the same multilocus genotype.

43 **3.6**
44 **probability of identity for siblings**
45 The probability that two sibling individuals selected at random from a population have the same
46 multilocus genotype.

47 **3.7**
48 **sample (n)**
49 A group of items, test results, or portions of material, taken from a large collection of items, test
50 results or portions of material, that serves to characterize the larger population.

51 **3.8**
52 **statistical power**
53 The power of a statistical test of a null hypothesis is the probability that it will lead to the rejection
54 of the null hypothesis.

55 **4   Requirements**

56 **4.1   General**

57 **4.1.1**   This document sets forth criteria that shall be met for construction and evaluation of
58 multilocus population genetic databases, including: identification of database components,
59 selection of genetic markers, procedures for statistical analysis, and evaluation and interpretation
60 of results for general population genetic analyses.

61 **4.1.2**   Species and population of interest will differ based on demographic, ecological, and
62 evolutionary factors, so quantitative values for the minimum number of individuals and genetic
63 markers needed for a reference database are expected to vary.

64 **4.2   Inclusion Criteria for Genetic Database Samples**

65 **4.2.1**   Documented technical procedures for constructing multilocus genetic databases shall
66 include at minimum:

67 a)   sample acquisition;

68 b)   establishment of parameters for inclusion samples;

69   c)   validation process for use of genetic markers;

70   d)   criteria for individual sample data quality;

71   e)   quality control/curation of sample information and genetic data.

72   **4.2.2**   Quality control shall include adherence to requirements in ANSI/ASB Std 019, ANSI/ASB
73   Std 046, ANSI/ASB Std 047, and ANSI/ASB Std 048.

74   **4.2.3**   In determining database composition the laboratory analyst shall assess, at minimum:

75   a)   sample size needed to accurately represent source population genetic diversity;

76   b)   related taxonomic information, including but not limited to:

77   1)   presence of subspecies,

78   2)   evolutionary significant units (ESU),

79   3)   hybrids in the species group of interest,

80   4)   geographic range of the taxa in question.

81   **4.2.4**   If known, the following metadata shall be documented for each sample:

82   a)   geographic location of source samples (e.g., sampling location, breeding location, location of
83   death);

84   b)   sex of individual;

85   c)   age class of individual;

86   d)   type of tissue sampled (e.g., fresh tissue, blood, bone, hair, antler, keratin, feces);

87   e)   collector information (i.e., date of collection, collector - both name and agency/institution,
88   method of collection).

89   **4.2.5**   At minimum, genetic markers shall be evaluated for:

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

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

93   b)   suitability for genotyping;

94   c)   genetic diversity measures, including but not limited to:

95   1)   Hardy-Weinberg Equilibrium,

96   2)   linkage disequilibrium,

97   3)   allelic richness,

98   4)   allelic diversity,

99   5)   heterozygosity measures within and among populations;

100   d)   the presence of null alleles.

101   **4.2.6**   Quality criteria shall be established for sample inclusion when adding genetic data to
102   species/population databases. This shall include, at minimum:

103   a)   minimum acceptable completeness of genotype per sample;

104   b)   minimum genotype quality measures, depending on genotyping platform [e.g., capillary
105   electrophoresis - Relative Fluorescence Units (RFU); Next-generation sequencing -genotype
106   quality score and read depth].

107   **4.3**   **Inclusion Criteria for Data Entry, Quality Assurance, and Quality Control**

108   **4.3.1**   Laboratories shall have technical procedures for data entry quality assurance and quality
109   control.

110   **4.3.2**   Once constructed, the multilocus genetic database shall be evaluated for, if applicable:

111   a)   representative geographic coverage;

112   b)   power to discriminate species/population boundaries;

113   c)   power to identify natural groupings that are meaningful from the ecological, biological, or legal
114   perspective;

115   d)   population level allele frequencies;

116   e)   presence of spatial or temporal autocorrelation;

117   f)   sex-related bias;

118   g)   estimates for statistical power (i.e., probability of identity, probability of identity for siblings);

119   h)   presence of duplicated samples;

120   i)   coefficient of co-ancestry.

121   NOTE Multilocus Genetic Database composition will vary based on forensic application (e.g., individual and
122   familial relationship evaluation, geographic assignment techniques).

123   **4.3.3**   Laboratories shall have technical procedures for evaluation of software intended for use in
124   statistical analysis, including commercial programs and validated programs developed in-house.

125   **4.3.4**   Once initially constructed and validated, databases augmented with new samples or
126   subsetted shall be reevaluated as in 4.3.2.

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

129 b) If the multilocus genetic database is made public then practitioners may anonymize the data
130   beforehand. For example, geographic information metadata may be anonymized by:

131   1) withholding geographic identifiers;

132   2) relabeled using generic labels (e.g., "A", "B");

133   3) character masking identifiable information (e.g., 1011 vs 10xx).

134 **4.3.5** Laboratories shall have technical procedures for database archival and version control.

135 **4.3.6** Laboratories shall have documentation of required assessments/evaluations for
136 transparency purposes.

5 Conformance

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

141

Annex A

(informative)

Bibliography

The following bibliography is not intended to be an all-inclusive list, review, or endorsement of literature on this topic. The goal of the bibliography is to provide examples of publications addressed in the standard.

1] ANSI/ASB Best Practices Recommendations 114, *Best Practice Recommendations for Internal Validation of Software used in Forensic DNA Laboratories,* 1st Ed., 2022[b].

2] Evett, IW and B.S. Weir. 1998. Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists. Sinaur Associates Inc, Sunderland, Massachusetts. 1998. Pp. 100

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

4] Cohen J.1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ.

5] Legendre, P and L. Legendre .1998) Numerical Ecology. 3rd ed. Elsevier, Amsterdam.

6] Ogden, R and A. Linacre. 2015, Wildlife forensic science: A review of genetic geographic origin assignment. Forensic Science International: Genetics, 18:152 - 159[c].

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

8] 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[d].

9] Smouse, P., R. Peakall.1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82:561–573[e].

10] 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[f].

11] 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[g]..

---

[b] Available from: https://www.aafs.org/academy-standards-board
[c] Available from: https://doi.org/10.1016/j.fsigen.2015.02.008
[d] Available from: https://doi.org/10.1086/285228
[e] Available from: https://doi.org/10.1038/sj.hdy.6885180
[f] Available from: https://doi.org.10.1046/j.1365-294x.2001.01185.
[g] Available from: https://ci.nii.ac.jp/naid/10021852122/en

Academy Standards Board
410 North 21st Street
Colorado Springs, CO 80904

www.aafs.org/academy-standards-board