

MEMORANDUM

STATE OF ALASKA

Department of Public Safety

To: Ashley Lankford
Quality Assurance Manager



Date: 22 August 2017

From: Cheryl Duda
DNA Technical Manager



Subject: Acceptance of 2017 expanded
core loci population
frequency databases for
Native Alaska populations

The Alaska SCDL provided the FBI with reference samples from individuals self-declared as Athabaskan, Inupiat, and Yupik (~100 samples per population group). The FBI generated STR profiles from these samples using both PowerPlex Fusion and Global Filer, and used those profiles to create Athabaskan, Inupiat, and Yupik allele frequency databases for all the CODIS expanded core loci.

The reference profiles analyzed by the FBI included some samples from individuals who had been previously typed for the 2002 Alaska Native allele frequency databases, but many were from individuals not previously used. Because different sets of profiles were used for the 2002 and 2017 databases, it is expected that the calculated frequencies will not be precisely the same at the original core loci. The frequency of a given profile has been shown to vary up to 10-fold when calculated using different databases (NRCII).

To ensure that the 2017 Alaska Native allele frequency databases are concordant with the 2002 databases, a sample set of frequencies for the 13 original core loci were calculated for 10 single source profiles and 10 two-source mixtures, using the 2002 and 2017 Alaska Native population frequency databases. All calculated values from the 2017 databases were within the expected 10-fold range of the 2002 values. Based on the report provided by the FBI and the results of the concordance check, the 2017 Alaska Native allele frequency databases are deemed suitable for use in forensic casework.

The FBI has indicated that they ultimately plan to publish their results. In the meantime, they have given their permission for us to make available their report summary and population frequencies on the AK SCDL website.

Expanded Loci Genotyping and Statistical Evaluation of Three Alaskan Native Populations

The Alaska Department of Public Safety Scientific Crime Detection Laboratory provided the FBI Laboratory with reference exemplars corresponding to three Alaskan Native Populations – Inupiaq (N=104), Yupik (N=103) and Athabaskan (N=102). Samples were received as blood cuttings from either FTA or filter paper in 1.5 ml tubes. Tubes were stored refrigerated until processed.

Samples were extracted using Prepfiler Express chemistry and the associated Automate robotic platform (Life Technologies). Extracts were quantified using Quantifiler Duo (Life Technologies), diluted to target and amplified using both GlobalFiler and PowerPlex Fusion expanded loci kits. Data was separated in a 3500 Genetic Analyzer using the following injection parameters: 1.2 kV for 24 sec.

The generated electropherograms were analyzed by three independent individuals for accuracy and correctness. Genotype tables were generated from the verified electropherograms and data from each of the kits were compared for concordance at the shared loci.

Allelic frequencies were determined using Office Excel software (Microsoft Corp, Redmond, WA). PowerStats v1.2 (Promega Corp.) was used to calculate power of discrimination, matching probability, and polymorphism information content as well as power of exclusion and typical paternity indices for each marker. Data generated in PowerStats was also used to verify all calculated allelic frequencies.

To establish if the assessed populations were in Hardy-Weinberg Equilibrium (HWE), the observed and expected heterozygocities were determined using Arlequin v.3.5.2.2 (Excoffier, 2010). The p statistic obtained from these calculations is reported in the summary statistics table. Generally, loci with p values < 0.05 are considered to not be in equilibrium. Nevertheless, the Bonferroni correction can be applied to the data set to correct the significance level. Bonferroni allows the user to control the experiment error rate as opposed to the individual marker error rate; it is used to correct for multiple comparisons when independent statistical analyses are performed concurrently. The new alpha level, defined as $N/0.05$ (in this case $N = 21$ (# of markers)) is then used to determine if there is deviation from equilibrium. When applied to the Alaskan data set, all markers conformed to HWE.

The coefficient of inbreeding was also determined using Arlequin software and was determined to be **0.03** for the examined populations.

*Some samples within data set C, corresponding to the Athabaskan population, yielded poor results and not enough extract was available for re-processing. Therefore, some samples or markers corresponding to this data set were not used for analysis.

References:

Excoffier, L and H.E.L. Lischer (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour