Cadaveric blood cards: Assessing DNA quality and quantity and the utility of STRs for the individual estimation of trihybrid ancestry and admixture proportions

Frankie L. West, PhD [a, *], Bridget F.B. Algee-Hewitt, PhD [b]

[a] Forensic Science Program, Western Carolina University, USA
[b] Center for Comparative Studies in Race and Ethnicity, Stanford University, USA

ARTICLE INFO

ABSTRACT

As part a body donation program, blood samples were collected and stored on untreated (non-FTA) blood cards. The blood cards were evaluated in terms of DNA preservation and STR typing success with resulting profiles assessed with special consideration given to profile matching for positive identification and biogeographic ancestry estimation. While STR profiles were successfully generated for all samples, results indicate that the time interval between date of death and sample collection have an impact on DNA quantity and quality. There is a statistically significant decrease in relative fluorescent unit (RFU) values with increasing time interval between date of death and sample collection, indicating degradation in the blood card samples related to the post-mortem interval prior to sample collection. The STR profiles were used to estimate ancestry and admixture using the program STRUCTURE, demonstrating utility of these markers beyond individual identification purposes, with caveats for application based on population history.

1. Introduction

While next generation sequencing (NGS), or massively parallel sequencing (MPS), methods have dramatically altered the fields of medical genetics and paleogenomics, short tandem repeats (STRs) analyzed through traditional capillary electrophoresis remain the gold standard for forensic identification. Although new technologies integrating NGS/MPS approaches hold promise for ancestry estimation [1,2], phenotyping [3], and fluid identification [4], STRs remain the primary genotyping method due to extensive validation as a marker set and the availability of large databases of typed individuals. Since the 1990s, forensic analysts have focused most attention on a set of core STR loci, composing the Federal Bureau of Investigation's (FBI) Combined DNA Index System (CODIS), consisting of a set of 13 traditional markers, plus Amelogenin, and recently enlarged to include seven additional markers [5]. While new approaches apply NGS/MPS to STR typing as an alternative to traditional capillary electrophoresis (CE) and offer opportunities to expand beyond the core markers [6], the set of CODIS loci remains the primary means of genetic identification in forensic contexts [7]. Moreover, recent work demonstrates the utility of this marker set beyond identification for population inference, demonstrating a capacity for revealing biogeographic ancestry and patterns of admixture [8,9].

Here, we analyze DNA quantities and STR profile results from post-mortem blood drawn from 20 body donors and stored between 4 months and 4 years in ambient conditions on FITZCO untreated blood cards. While FTA cards have been validated for DNA preservation for a variety of sample materials, including blood [10,11], tissue [12], and saliva [13], the quality of DNA extracted from post-mortem blood samples stored on untreated cards is unknown. While FTA cards use proprietary technology to protect DNA from further degradation after samples are applied to the cards, the FITZCO FP705™ card is untreated and, so, does not lyse cells, denature proteins, or prevent microbial activity after sample deposition.

We focus on the quality of DNA from blood cards for long-term storage in forensic anthropology centers, with potential for other long-term storage applications. At the time of publication, there are eight forensic anthropology decomposition facilities in the United

* Corresponding author.
E-mail addresses: flwest@wcu.edu (F.L. West), bridgeta@stanford.edu (B.F.B. Algee-Hewitt).

States [14]. Collection of biological samples from donors, including blood and buccal swabs, is common practice for body donation programs. Long-term storage solutions are necessary in situations where DNA extraction and typing may not be conducted immediately after sample collection, including in medical examiners' offices [15], biobanking [16], and disaster victim identification [17,18]. Here, we test the applicability of the FITZCO FP705™ card for long-term storage of blood samples collected post-mortem and stored at room temperature ~4 °C.

STRs, or short-tandem repeats, have been the standard DNA profiling method for forensic identification since the 1990s [19]. STRs lend themselves to identification based on the large number of alleles at each locus, high discriminatory power provided by the combination of STR loci, suitability for multiplexing, and relatively small size (approximately 100–480 base pairs, or bp) which allows for use with degraded samples [19]. The core STR loci which make up the standardized CODIS set were primarily selected for their highly polymorphic qualities, enabling discrimination between unrelated individuals, with some overlap with the European Standard Set (ESS) [19].

In addition to autosomal STRs, STRs associated with sex chromosomes can also reveal important data for forensic identification, familial relationship determination, and deconvoluting mixtures. More broadly, pre-established panels of STRs, like those used in forensic profile matching, also serve as ideal markers for management of biological sample collections. Guidelines have been established by the American National Standards Institute for the authentication of human cell lines using STRs [20]. As a marker set, the CODIS set of STRs provides a cost effective and straightforward method for matching individual cell lines with their source individuals [21]. This approach could likewise be applied to skeletal collections and body donation programs, especially for elements which may become disassociated during decomposition and processing.

1.1. Limitations of STRs

While STRs are well-suited to identification based on their high heterogeneity between individuals and large number of alleles per locus, the large DNA fragments required for typing can present issues when dealing with degraded DNA samples. The set of original 13 CODIS core loci range in size between 100 and 400 bp [22]. Primers must be able to anneal on each side of the target amplicon during polymerase chain reaction (PCR) in order to amplify the target region. If the DNA sample is too degraded or if PCR inhibitors are present (including indigo dyes, humic acid from soil, heme from blood, to name a few), the reaction can fail to amplify the target loci, creating situations in which one allele at the target locus drops out or both alleles, resulting in locus drop out [19].

Allele and locus drop out are commonly seen with larger STR loci, resulting in electropherogram results that resemble a ski-slope pattern, in which smaller loci amplify in contrast to a reduction in amplification in larger loci, common in degraded samples [23]. Alternatives to traditional STR marker kits have been proposed to reduce the amplicon length of larger loci, suggesting "mini-STRs" to reduce chances of allelic/locus drop out [22]. While many mini-STRs have been included on expanded commercial kits, such as the Applied Biosystems AmpFℓSTR® MiniFiler™ PCR Amplification kit, they have not replaced traditional STRs as the typing method of choice [24].

The Amelogenin gene is present on both the X and Y chromosomes, with a distinguishing 6-bp deletion on the X chromosome not present on the Y chromosome. When typing the Amelogenin locus, a female profile will exhibit a single peak, whereas the male profile exhibits a separate peak for each chromosome. One of the issues complicating analysis of the Amelogenin locus is the phenomenon of Y-allele and X-allele dropout. In situations with degraded or inhibited DNA, Y-chromosomal specific DNA fragments can fail to amplify resulting in allelic drop-out, wherein the signal only amplifies the shorter fragment from the X-chromosome, or conversely there is dropout of the X-chromosomal Amelogenin marker. Dropout of the Y-chromosomal marker is much more common [25]. In cases of Y-allele dropout, an incorrect sex estimation can be made wherein the profile reads as female, X, X, rather than the true profile of X, Y.

Various biological sample types, including bone/tooth, blood, buccal cells/saliva, hair, and tissue present different challenges in DNA extraction and typing, resulting in differential yields and varying levels of potential PCR inhibitors. Expectations for DNA yields differ by sample type, with highest yields expected from blood [19]. Bones, teeth, blood, and hair all contain potential inhibitors that could interfere with PCR reactions, including calcium, heme, and melanin, respectively. Bone and tooth samples require extra demineralization steps to break down the hydroxyapatite matrix for DNA extraction [26–29]. Hair samples also require additional steps using DTT to lyse the keratin of the hair shaft [19,30]. As a substrate for sample storage, FTA (Flinders Technology Agreement) cards are a popular option for a variety of sample types, including blood and buccal cells. Extraction from FTA cards can be largely automated [31] and can be used for direct PCR when dealing with robust samples [13].

In contrast to robust samples, biological samples collected post-mortem may present difficulties in extraction and amplification, based on time since death and sample type, and are represented in far fewer studies regarding these sample types [10,11]. Rahikainen and colleagues determine that both post-mortem femoral blood and buccal cell samples transferred to FTA paper produce successful profiles at 16 STR loci with the caveat that each produced low-quality DNA when evaluated by UV absorbance [11]. Assess DNA quality and quantity from blood collected post-mortem from autopsy samples and stored on FTA cards. The authors show that post-mortem interval and storage time both have a significant impact on DNA quantity and quality as assessed by relative fluorescence units (RFUs). FTA cards are a commonly used substrate for long-term sample storage. Another study using multiple replicates from a single bloodstain sample stored at room temperature for 20 years found significant degradation and locus dropout in STR typing [32].

Despite the importance of understanding the constraints placed upon DNA results, given the potential for degradation issues, no studies have assessed quantity and quality of DNA from blood samples stored on untreated blood cards for analyses of interest to forensic and anthropological geneticists, especially in the context of pursuing research using bio-banked blood samples collected from deceased individuals. The inability to produce complete CODIS profiles places limitations on individual identification and impacts random match probabilities. When conducting ancestry estimation using STR marker sets, a reduced number of markers limits the resolution of ancestry inference as shown by Algee-Hewitt and colleagues [8].

In this study, we test 20 untreated blood cards collected post-mortem to assess the quality and quantity of DNA extracted. DNA quantity and the presence of inhibitors are assessed through qPCR. The relationship between time intervals between date of death and sample collection (IDDC) and sample collection and STR analysis (CST) and DNA quantity are evaluated. DNA quality is measured through a variety of methods, including peak height ratios and RFUs. Microvariants and off-ladder alleles are identified for each individual. We also assess the utility of these typed loci for generating ancestry and admixture proportions using the unsupervised clustering methods implemented via the program STRUCTURE [33].

2. Materials and methods

The FITZCO FP705™ blood card used here was originally designed for use by the U.S. military, beginning in 1991 (FITZCO). The collection area of the card is made of biological grade cotton linter paper which prevents sample diffusion off of the substrate surface. The card design consists of four circles with a "fold-over" flap to reduce contamination risk following collection. Unlike FTA blood cards which are treated to lyse cells, deactivate nucleases, and deter microbial activity [34], the FITZCO FP705™ card is untreated.

Blood cards (FITZCO FP705™) were collected postmortem from donors of the Body Donation Program at the Forensic Anthropology Center (FAC) at The University of Tennessee, Knoxville (UTK). Blood was drawn from the aorta or subclavian artery of each cadaver using a syringe and placed on blood cards as part of the standard intake process, which involves documentation of the individual donor and sample (blood, hair, nails) collection for future research [35]. Anonymized sample IDs and time interval between date of death (DoD) and sample collection (IDDC) as well as interval between collection and DNA analysis (CST) are shown in Table 1. Information on individual donor demographics, including geographic ancestry, or identity, was collected prior to or during the donation process. The blood card donors included individuals designated as pre-donors, individuals donated by family members, and one individual donated by a medical examiner's office. A total of nine individuals were pre-donors, individuals who planned donation and provided self-identified demographic data, including sex and ancestry/race/ethnicity. Ten individuals were donated by family members and their identities offered by next-of-kin. One individual was donated by the office of a medical examiner, thus the record of self/group-identifiers was based on the post-mortem assessment of the medical examiner rather than self or familial description.

The blood cards were stored in a desiccator until sealed in plastic FoodSaver bags with a silica-based desiccant. One half-inch circle (outlined by the manufacturer) of the blood card was removed using sterilized scissors and placed in a DNA-free 50-mililiter (ml) conical tube. All samples were sent to Bode Cellmark Forensics for DNA extraction, quantification, and fragment analysis.

2.1. Sample treatment

All samples were extracted at Bode Cellmark Forensics laboratories using the automated Qiagen EZ-1 Investigator Kit with an initial incubation, storage at 4 °C overnight and extraction on the following day. Samples were quantified using the proprietary BodeQuant quantitative PCR (qPCR) for low-copy number samples. This qPCR method includes a nuclear DNA target to assess quantity of nuclear DNA as well as an Internal Positive Control (IPC) to assess presence of inhibitors within the sample extract. Following quantification, samples were amplified using the Applied Biosystems® Identifiler kit. This multiplex PCR kit included the thirteen original CODIS loci plus the D2S1338 locus, the D19S433 locus, and Amelogenin. STR typing through kit-based approaches, including the Applied Biosystems® Identifiler kit, uses fluorescent dyes attached to primers for each of the multiplexed loci. Samples were run on the Applied Biosystems® 3130 Genetic Analyzer for capillary electrophoresis. The instrument detects fluorescence of the labeled fragments and reports this output as relative fluorescence units (RFUs) which are used to interpret quality thresholds and fragment sizes when compared against an internal size standard and allelic ladder. Positive and negative controls were used throughout the entire process.

2.2. STR quantity and quality assessment

Sample quantities were first compared to the IDDC and CST using a linear regression model to test for linear relationships. Next, the non-parametric Spearman's ρ (rho) test was executed to evaluate the strength and direction of any association that exists between the two variables, time intervals and DNA quantity. We suggest that an increased time interval between donor death/sample collection as well as increased time between collection and STR typing will result in lower average DNA yield.

To assess STR quality and impact of time intervals between DoD, collection, and extraction, RFUs were averaged across sample and locus size class and compared to IDDC and CST using a linear regression model, followed by the Spearman's ρ test. Locus size classes were grouped on the basis of size as per [11] with Class 1

Table 1
Demographic data and time interval information for post-mortem blood donors.

Sample ID	Sex	Age (in years)	Interval DoD/Collection (in days) - IDDC	Interval Collection/Storage Time (in days) - CST
1	M	70	1	1568
2	F	75	1	1551
3	M	64	1	1548
4	F	65	0	1537
5	M	79	0	1519
6	M	64	3	1513
7	F	29	17	1492
8	M	50	74	1448
9	F	75	3	1443
10	F	58	1	1436
11	F	71	12	911
12	M	60	2	1206
13	F	94	1	1184
14	M	74	16	1181
15	F	79	3	1180
16	F	62	3	782
17	M	58	4	782
18	M	51	1	154
19	F	66	0	490
20	M	81	3	133

Table 2
Locus size classes with size range and loci included; average RFUs per locus size class across all individuals included.

Class	Locus Size Range	Loci Included	Average RFU/Class
1	<130 bp	D3S1358, D19S433, D10S1248	1342
2	130-200 bp	vWA, TH01, D5S818	1011
3	200-300 bp	D21S11, D13S317, D7S820, D16S539	805
4	>300 bp	CSF1PO, TPOX, D18S51, FGA, D2S1338	698

(<130bp), Class 2 (130—200 bp), Class 3 (200—300 bp), and Class 4 (>300 bp) as seen in Table 2. As with DNA quantity, we suggest that an increased time interval between donor death/sample collection as well as increased time since collection/STR typing will result in a reduction in DNA quality. Through assessing degradation from a decrease in RFUs across locus size, we determined whether patterns of differential amplification are present in the profiles generated using the blood cards.

Peak height ratios (PHR) were calculated for each individual and locus by dividing the lower peak (Peak A) RFU by the higher (Peak B) RFU as outlined by the Scientific Working Group on DNA Analysis Methods (SWGDAM) guidelines. Peak height ratios of below 70% were designated as severe imbalance, a threshold indicative of multiple contributors or other issues [36]. From single source samples (i.e. not mixtures), peak height ratio imbalance can be attributed to several issues, including low starting DNA template, preferential amplification, DNA degradation, the presence of inhibitors, or a combination of these factors [37]. Profiles were also checked for the presence of stutter and off-ladder alleles.

2.3. STR analysis for sex determination & microvariants

The Amelogenin marker was typed and compared to self-reported biological sex and skeletal estimations. As a smaller marker, we expected that the Amelogenin markers would successfully amplify and match recorded biological sex. We also identified off ladder alleles, i.e., those not found within the allelic ladder for each STR kit. Off-ladder alleles can include full repeats, which are uncommon within known typed populations. Microvariants, a form of off-ladder allele with incomplete repeat units, were also noted for each individual profile. An example of a microvariant would include a simple tetranucleotide (4 bp) locus with 14 repeats, but with the addition of a partial repeat of 2 bases, making the allele call 14.2. For each microvariant and other off-ladder alleles, the frequency relative to the U.S. population was also assessed.

2.4. Population inference from amplified STRs

Trihybrid ancestry estimation was conducted using the unsupervised clustering program STRUCTURE, v. 2.3.4 [33,54]. Thirteen CODIS loci were compiled for 332 individuals from the Human Genome Diversity Panel (HGDP-CEPH) H1048 subset [8,38], including 94 individuals from Sub-Saharan Africa, 83 individuals from the Americas, and 155 individuals from Europe, who together served as the parental reference sample. An initial STRUCTURE run was used to determine the optimal range of K, or number of clusters. Here, and for the other reported runs, parameters were set at 10,000 for burn-in and 10,000 Markov Chain Monte Carlo (MCMC) repetitions (reps). For the next run, we pre-set the number of K clusters to 1, 2, or 3, thereby limiting the analysis to the maximum number of populations under a trihybrid ancestry model. We used the No Admixture model which assumes origin of individuals from only one population and is appropriate for discrete populations [52]. We assumed that allele frequencies were independent among populations with parameters of alpha (α) and lambda (λ) set at 1.

Post-processing was performed using Structure Selector [39] which integrates several approaches for data interpretation, including the Puechmaille [40] method and Clumpak [41].

To evaluate admixture, we performed a second analysis with STRUCTURE, for which we used a subset of the National Institute of Standards and Technology (NIST) dataset [6,7] in place of the HGDP-CEPH parental populations. These reference data included 149 self-identified African Americans, 151 European Americans, and 101 Hispanics. We used the Admixture model, operating under the assumption that each of the individuals shares genetic ancestry with one or more of the clusters included (Pritchard et al., 2010), and that allele frequencies were independent between populations, with the α and λ set at 1. K was set between 1 and 3.

3. Results

Of the 20 blood card samples, five were re-extracted twice and one sample three times to obtain sufficient quantities of DNA to produce a complete STR profile. Final sample nuclear DNA quantities ranged from 15.72 ng/uL to 153.81 ng/uL (Table 3). Five samples exceeded the average internal positive control (IPC) threshold of 20.64 by more than 2 cycles for the standards, indicating the presence of inhibitors in those samples.

Correlations between DNA quantity in ng/uL and the time interval (in days) between a) date of death and collection (IDDC) and b) sample collection and STR testing (CST) were assessed using a linear regression model and Spearman's correlation coefficient, ρ. A linear regression model was used to assess whether time intervals were significant predictors of DNA quantity. Modeled with IDDC, there was no significance detected, with a p-value of 0.2933, F-statistic of 1.176, $R^2 = 0.065$, and 17° of freedom. With CST, no significance was detected with a p-value of 0.8505, F-statistic of 0.0366, $R^2 = 0.00215$, and 17° of freedom. Non-normal distribution of the variables representing IDDC was confirmed by a Shapiro-Wilk normality test, yielding, respectively, significant p-values of 1.101e-07 and 0.001574 when $\alpha = 0.01$ thus the Spearman's ρ statistic. We find a small positive association with DNA quantity for IDDC with DNA quantity for both IDDC (Spearman's $\rho = 0.0823$, p-value = 0.7301) and a small negative association for CST ($\rho = -0.0519$, p-value = 0.8279), both associations being statistically insignificant. These results suggest that time, when measured as IDDC and CST intervals, does not have a significant relationship to DNA quantity — a finding that is contrary to our original expectations.

3.1. STR quality assessment results

Peak height ratios were averaged across each sample (reported in Table 3) and across each locus (reported in Table 4). Several samples did not meet the 70% peak height ratio threshold, indicating that those samples were imbalanced, likely due to degradation rather than possibility of a mixture due to lack of more than 2 alleles at multiple loci. Peak height imbalance can be attributed to sample degradation as well as potential mixed profiles, which include more than one contributor. In the profiles generated, only 1—2 alleles were present at each locus across the profile as a whole,

Table 3
DNA quantities, internal positive control cycle threshold (IPC CT), and average peak height ratios and RFUs across all loci. Those that exceed the IPC CT and indicate the presence of inhibitors are in bold.

Sample ID	DNA Quant. (ng/uL)	IPC CT	Average Peak Height Ratios	Average RFUs Across all Loci
1	33.65	20.77	81%	837
2	87.22	**22.52**	86%	1369
3	106.59	**22.98**	86%	1455
4	42.56	20.26	90%	1074
5	131.02	**23.24**	91%	1861
6	82.66	20.38	82%	262
7	90.91	20.33	79%	409
8	69.19	19.91	85%	1896
9	84.18	20.70	90%	1087
10	96.70	20.22	84%	645
11	75.25	19.43	82%	434
12	15.72	19.60	85%	1033
13	64.56	19.68	82%	359
14	153.81	**22.77**	84%	416
15	68.48	19.65	88%	1194
16	146.88	**22.88**	85%	683
17	118.96	19.96	83%	597
18	62.29	19.47	87%	618
19	131.18	20.79	87%	1296
20	19.30	19.70	87%	1039

indicating no sign of a potential second contributor (a major/minor mixture). Rather, the imbalance in peak heights can be attributed to increased degradation which is responsible for differential amplification of damaged DNA fragments, wherein one allele is replicated at a higher number than the other, producing differences in fluorescent units within the same locus.

Average RFUs by locus class are reported in Table 2. Results assessing the impact of IDDC and CST on DNA quality as shown through RFUs were conducted using a linear regression and the Spearman's ρ test statistic. A single donor was stored frozen after date of death for a total of 74 days, creating an outlier in terms of statistical analysis. This outlier was removed prior to statistical testing. Using a linear model to assess whether IDDC was a significant predictor, significant results were found in each RFU class. For Class 1, the IDDC was a significant predictor of RFU values, with a p-value of 0.02043, F-statistic of 6.536, $R^2 = 0.2777$, and 17° of freedom. For Class 2, the linear regression results indicate a p-value of 0.01407, F-statistic of 7.488, $R^2 = 0.3058$, and 17° of freedom. For Class 3, results indicate a p-value of 0.02822, F-statistic of 5.752, $R^2 = 0.2528$, and 17° of freedom. For Class 4, results indicate a p-value of 0.0418, F-statistic of 4.849, $R^2 = 0.2219$, and 17° of freedom.

Using Spearman's ρ, the association between IDDC and RFUs,

associations for Class 1 ($\rho = -0.6519$, p-value $= 0.0025$), Class 2 ($\rho = -0.5278$, p-value $= 0.0201$) and Class 3 ($\rho = -0.6089$, p-value $= 0.0056$) were all significant, while associations for Class 4 ($\rho = -0.4432$, p-value $= 0.0573$) were not. All associations between IDDC and RFUs for all class sizes indicate a negative correlation between time and fluorescence, demonstrating that as number of days post-mortem before sample collection increase, fluorescence values decrease across all class sizes.

In assessing the association between CST and RFUs from each size class (1–4) using a linear model, no significant relationships were identified (Class 1 - p-value $= 0.1701$, F-statistic of 1.053, $R^2 = 0.1077$, and 17 df, Class 2 - p-value $= 0.6523$, F-statistic of 0.2104, $R^2 = 0.0122$, and 17 df, Class 3 - p-value $= 0.1701$, F-statistic of 1.053, $R^2 = 0.1077$, and 17 df, and Class 4 - p-value $= 0.8766$, F-statistic of 0.0249, $R^2 = 0.0015$, and 17 df. No significant relationships were identified using Spearman's ρ (Class 1: $\rho = 0.4242$, p-value $= 0.0623$; Class 2: $\rho = 0.2399$, p-value $= 0.3082$, Class 3: $\rho = 0.2595$, p-value $= 0.2692$ Class 4: $\rho = 0.0684$, p-value $= 0.7743$).

3.2. STR analysis results

Comparisons of the Amelogenin marker returned total agreement between the genetic sex markers, self-reported biological sex, and skeletal sex estimations. One off-ladder allele was recorded at locus D21S11 as microvariant 29.3 in individual 15 and confirmed by a second fragment analysis run. This allele is found at a frequency of 0.0005 in the combined U.S. population. Other microvariants not considered off-ladder alleles were typed at 3 loci

Table 4
STR loci, average peak height ratios per locus, and number of samples below peak height ratio of 70%.

STR Locus	Average Peak Height Ratio	Number of Samples Below 70% PHR
Amelogenin	86%	–
D3S1358	88%	–
D19S433	88%	–
D8S1179	86%	–
D5S818	88%	–
TH01	87%	–
vWA	87%	1
D21S11	81%	2
D13S317	85%	3
TPOX	87%	–
FGA	85%	1
D7S820	86%	1
D16S539	89%	–
D18S51	86%	–
CSF1PO	85%	1
D2S1338	73%	8

Table 5
Frequencies of microvariants found in surveyed STR profiles. * Denotes a lack of reported frequencies for a particular allele in the NIST database.

Locus	Allele Variant	Number of Ind.	Frequency in U.S. Population
D19S433	13.2	1	*
D19S433	15.2	2	0.0569
TH01	9.3	3	0.2056
D21S11	24.2	1	0.0005
D21S11	29.3	1	0.0005
D21S11	30.2	1	0.0217
D21S11	31.2	2	0.0772
D21S11	32.2	4	0.0912
D21S11	33.2	4	0.0328

(D19S433, TH01, and D21S11), in three, four, and three individuals, respectively. These frequencies are reported in Table 5.

STRUCTURE analysis for ancestry estimation was conducted using the HDGP-CEPH populations and the 20 blood card samples; the number of populations, *K*, was set at 3. This was also the optimal number of ancestry clusters identified by computational methods. This optimal value of *K* was determined using STRUCTURE Selector [39], implementing the MedMeaK, MaxMeaK, MedMedK, MaxMedK methods [40] for choosing the best *K* among a range of *k*-clusters. The MedMeaK, MaxMeaK, MedMedK, MaxMedK approaches all outperformed traditional deltaK methods for determining the optimal number of clusters in situations with uneven sample sizes [40]. Each of the 20 individuals was assigned to one of three population clusters, with membership coefficients representing the posterior probability that the individual is from selected population (shown in Table 6). Results are visualized in the structure plot shown in Fig. 1 generated by Clumpak [41].

When adopting a hard classification or single cluster approach to ancestry inference [8,42], the documented group identifier for 17 of the 20 individuals matched the continental population cluster to which the individual was assigned, assuming that there is positive correlation between the social "quasi-ancestry," ethnic and race-based identifiers and continental ancestry, e.g., Black ~ African, White ~ European, Native American ~ Indigenous American. In one case, the individual self-identified as White but was assigned a membership coefficient of 0.844 for the African cluster and 0.156 for the European cluster. Two other individuals were documented, one self-identified and one familial identification, as Black but were grouped into the European cluster with membership coefficients of 0.906 and 0.950 respectively. The only non-self-identified or non-familial identified individual was labeled by the medical examiner's office as Black: here, African ancestry is estimated with high probability, with the STRUCTURE analysis producing membership coefficients of 0.965 for the African cluster and 0.014 for the European cluster.

Results from the second STRUCTURE analysis using the Admixture model present *K* = 2 clusters when analyzed using STRUCTURE Selector [39]. Using the Puechmaille [40] method, the MedMeaK, MaxMeaK, MedMedK, MaxMedK preferred two distinct clusters using the NIST sub-dataset. Results are visualized in the barplot shown in Fig. 2 generated by Clumpak [41].

The best fit number of clusters was 2, with the inferred cluster assignments between two groups (Table 7). Those identifying as Black had higher correlation coefficients with Cluster 1 (Table 8). Those identifying as White had higher correlation coefficients on average with Cluster 2. The individual who identified as White/Native American and the individual who identified as Hispanic were split between each cluster.

4. Discussion and conclusion

In this study, we have evaluated the effectiveness of the FITZCO FP705™ untreated blood card as a reliable substrate for long-term storage. Our samples were extracted after room temperature storage at time intervals between 4 months and 4 years. We address practical laboratory concerns for the successful recovery of nuclear DNA after longer periods of time. We find that sufficient amounts of nuclear DNA can be recovered from the sampled blood cards to amplify the original 13 CODIS core loci, although several samples had to be re-extracted due to insufficient DNA recovery during the initial extraction. The 13 original CODIS loci were adequate for ancestry estimation using the program STRUCTURE, for which 17 of 20 samples were hard classified into the ancestry group that most likely corresponded with their self-reported group identity. The individuals who self-identified as Hispanic and White/Native American were classified, with higher membership coefficients greater than (>0.98) into the European cluster under the trihybrid ancestry model. These individuals display opposite trends, however, when subjected to the admixture analysis using the NIST population samples as the reference dataset. Their admixture proportions were distributed similarly to the NIST samples across the two inferred clusters. This appears to capture White and non-White variation, as the Hispanic individual represents high (about 60%) European admixture and the White/Native American individual, for whom dual-identity was explicitly recorded, represents low (about 42%) European admixture. These findings recall what has been argued elsewhere for forensic ancestry and what cannot be overstated: that Hispanics represents some mixture of ancestries and the terminology, "Hispanic," itself is uninformative given the known diversity in population history of the many geographic groups who fall under this category and the expression of this history in terms of the calculated ancestry proportions [9,43].

Table 6
Correspondence between the membership coefficients obtained from the trihybrid ancestry analysis using STRUCTURE and documented population identifier. Individuals were assigned to the population cluster with the highest degree of membership. Reported identity is included, with the source of the identity assignment. Abbreviations are S: self-identified, F: familial identification; ME: identity assigned by the medical examiner. *Denotes a potential disagreement between genetically-inferred ancestry and reported population identity.

| Sample ID	European	African	Indigenous American	Reported Identity & Source
1	0.903	0.097	–	White – S
2	1.000	–	–	White – S
3	0.998	0.002	–	White – F
4	1.000	–	–	White – F
5	1.000	–	–	White – S
6	0.931	0.069	–	White – S
7	0.156	0.844	–	White * – S
8	0.999	0.001	–	White – F
9	0.691	0.011	0.299	White – F
10	0.961	0.001	0.035	White – F
11	0.950	0.050	–	Black * – S
12	0.906	0.094	–	Black * – F
13	0.856	0.143	0.001	White – F
14	0.014	0.965	.021	Black – ME
15	0.978	0.022	–	White/American Indian – F
16	0.900	0.092	0.008	White – F
17	1.000	–	–	White – F
18	0.988	0.010	0.002	Hispanic – S
19	0.161	0.839	–	Black – S
20	0.999	0.001	0.001	White – S
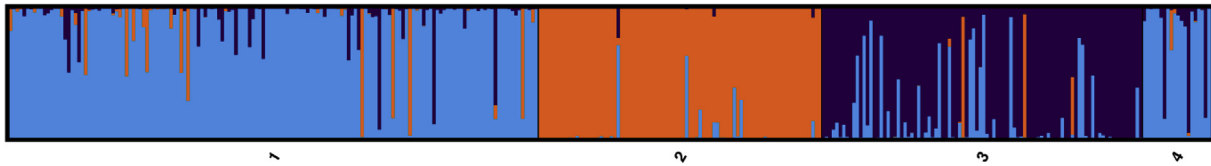
Fig. 1. STRUCTURE plot depicting $K = 3$ ancestry clusters by population, generated in Clumpak [41]. Each individual is represented by a single bar partitioned into 3 colored segments, which gives the individual's proportion of membership across the 3 "parental ancestry" clusters. Blue = European, Orange = Indigenous American, and Purple = African. Numbers correspond to the 3 identity categories self-reported by or ascribed to the sampled individuals: 1) White, 2) American Indian or Hispanic 3) Black. The unknown samples shown are labeled by 4. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
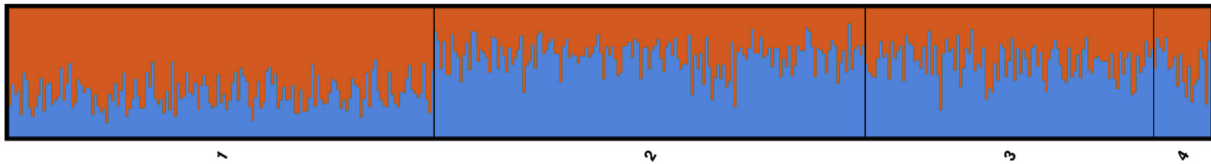

Fig. 2. STRUCTURE plot depicting admixture results from NIST dataset for $K = 2$ cluster solution, generated using Clumpak [41]. Each individual is represented by a single bar partitioned into 2 colored segments, which gives the individual's proportion of membership across the 2 clusters. Groups are 1) African Americans, 2) European Americans, 3) Hispanics, and 4) unknowns from blood cards.

Table 7

Correspondence between the membership coefficients obtained from the admixture analysis, using STRUCTURE and the NIST reference dataset, and the documented population identifier. The optimal $K = 2$ model was identified computationally. Individuals were assigned to one of two population clusters with the highest degree of membership.

Sample ID	Cluster 1	Cluster 2	Reported Identity
1	0.471	0.529	White
2	0.251	0.749	White
3	0.361	0.639	White
4	0.361	0.681	White
5	0.249	0.751	White
6	0.477	0.523	White
7	0.462	0.538	White
8	0.391	0.609	White
9	0.603	0.397	White
10	0.380	0.620	White
11	0.567	0.433	Black
12	0.682	0.318	Black
13	0.460	0.540	White
14	0.714	0.286	Black
15	0.578	0.422	White/American Indian
16	0.556	0.444	White
17	0.323	0.677	White
18	0.401	0.599	Hispanic
19	0.732	0.268	Black
20	0.282	0.718	White

Table 8

Proportion of membership of each pre-defined NIST population in each of $K = 2$ clusters.

NIST Population	Cluster 1	Cluster 2	Number of Individuals
African Americans	0.651	0.349	149
European Americans	0.382	0.618	151
Hispanics	0.431	0.569	101

quality of STRs typed. While Rahikainen and colleagues also report a decrease in DNA quantities over time in FTA cards, part of the reduction seen in RFUs may be indicative of DNA degradation exacerbated by nuclease activity, which was not halted in the untreated blood cards Rahikainen and colleges [11] were able to recover DNA from FTA cards stored up to 16 years, however, this longer time interval may result in increased degradation and reduced yields in non-treated cards.

An additional aspect of the untreated blood cards of value to consider is the potential for pathogen exposure. Since FTA cards lyse the cells upon contact, pathogens are inactivated [44]; however, pathogens can persist in the analysed cells on the untreated substrate. While viruses such as HIV are typically undetectable within a week to a month, Hepatitis C has been identified in dried blood spots after 4 weeks and on blood in needle syringes for up to 8 months [45]. All potentially biohazardous material should be treated with universal precautions, yet this precautionary aspect of blood sample storage may be a concern for forensic body donation programs, providing an additional reason to consider FTA cards over untreated cards.

Body donation programs often collect sample material for subsequent genotyping; however, DNA typing is often not the main focus of attention for decomposition facilities and budgets are limited. Based on our results, we suggest that FTA-based cards may provide a more dependable method for long-term storage, despite the lower cost of untreated cards. If typing of large-scale marker sets may be desirable for future applications, untreated cards may not produce the high quantities and quality of DNA required for expansive SNP panels or typing combinations of multiple marker types (STRs, Y-STRs, SNPs). We recommend long-range planning for future genotyping needs when selecting sample storage substrates.

4.1. Implications for long-term storage

While full profiles were typed from each of the 20 sampled untreated blood cards, several issues emerged during the analysis. The presence of inhibitors in five of the 20 samples may present a concern for downstream amplification of STRs and other markers. One potential source may include heme from red blood cells, a known inhibiting substance [19]. In contrast, Rahikainen [11] reported no inhibition in DNA extracts from FTA cards. We also noted a reduction in RFUs from smaller to larger loci, as shown through the decrease in RFUs from Class 1 through Class 4 in the untreated blood cards indicating degradation. We also show that a statistically significant reduction in RFUs is associated with increased time intervals between donor death and sample collection. Increased time intervals between the date of death and collection lower the

For short-term preservation, untreated cards may be adequate for STR typing but for extended storage duration, FTA cards provide an option that lyses cells, limits nuclease activity, and demonstrates DNA recovery from post-mortem collected blood samples after a more than a decade of storage.

4.2. CODIS markers for ancestry/admixture estimation

Trihybrid ancestry analysis in STRUCTURE produced membership coefficients for three ancestral groups. Out of 20 samples, 17 individuals were classified into the population which best matched their reported group identity. Two of the individuals, one self-identified and one familially-identified as Black, had membership coefficients that more closely aligned them with the European cluster, whereas one individual who self-identified as White had membership coefficient of 0.844 in the African ancestry cluster. Population history in the U.S. reflects admixture between groups of different continental ancestries and it has been noted that African Americans carry a wide range of proportions of European ancestry [46,47]. In a large-scale health study on Genetic Epidemiology Research on Adult Health and Aging (GERA) of over 100,000 individuals, researchers found that, of those identifying as African American, 91% had European ancestry [48]. From the same study, only 0.4% of self-reported Europeans displayed some quantity of African ancestry. These discordant results between the higher population membership coefficients and reported identities may be a result of directional admixture, reflecting the unique conditions of the post-contact Americas, capturing especially the vestiges of the trans-Atlantic slave trade, practices of assortative mating, and the past and present history of racism that shapes population interactions.

Our STRUCTURE analyses of ancestry and admixture of the unknown individuals typed from blood cards and known-source database samples produced different results, owing to the different number of population clusters, or values of *K,* identified computationally. One cause for the discrepancy between the model-based clusters in STRUCTURE is surely the difference in population datasets used for each analysis and the fundamental conceptual difference between a parental source and a contemporary reference population. While the initial No Admixture model for ancestry estimation used the HGDP-CEPH populations, the admixture analysis used a subset of the NIST population dataset. The HGDP-CEPH populations were sampled from individuals world-wide and are routinely taken to represent parental populations — in this particular case from each of three continental regions of Africa, Europe, and the Americas. In contrast, the NIST population subset is composed of individuals from the U.S., specifically those self-identifying as African American, European American, and Hispanic, collected from the Interstate Blood Bank in Memphis, Tennessee or the DNA Diagnostics Center in Fairfield, Ohio.

It has been previously noted that populations in the U.S. reflect varying levels of continental admixture based on the complex population history of the country [47,49]. Considering that the admixture analysis used U.S. populations, all of which are known to carry on average some quantities of ancestry from each of the three major U.S. source populations [9,42,47], the best number of clusters was estimated at *K* = 2. STRUCTURE analysis of African Americans by Lawson and colleagues (2018) demonstrated similar clustering of each into two "ancestral" population clusters based on recent admixture [53]. Algee-Hewitt has also shown, for both genetic and proxy quantitative skeletal traits similar, 2 cluster patterns for Latinos, largely of Mexican descent, and African Americans [9]. The authors further reported only trivial levels of admixture for European Americans, as also noted by Banda and colleagues [48].

While the CODIS STRs meet the recommended qualities of markers for STRUCTURE analysis in that they reflect low mutation rates, are selectively neutral, and are in linkage equilibrium [33,50], alternative sets markers provide more ancestry information. Explorations of sets of forensic STRs, with different characteristics or comprising more markers, have demonstrated increased recovery of ancestry information and greater differentiation between individuals using STRUCTURE [8]. While these particular 13 CODIS loci provide valuable insights into ancestral origin on the continental scale, the limitations must be considered when extending this panel of markers beyond its intended scope for individual identification to admixture estimation, especially for populations with complex population histories and peoples with potentially high levels of admixture.

Declaration of competing interest

The authors, FL West and BFB Algee-Hewitt, declare no conflicts of interest for this publication.

CRediT authorship contribution statement

Frankie L. West: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing - original draft, Writing - review & editing. **Bridget F.B. Algee-Hewitt:** Conceptualization, Data curation, Funding acquisition, Formal analysis, Investigation, Methodology, Project administration, Writing - review & editing.

Acknowledgements

References

[1] O. Bulbul, A.J. Pakstis, U. Soundararajan, C. Gurkan, J.E. Brissenden, J.M. Roscoe, B. Evsanaa, A. Togtokh, P. Paschou, E.L. Grigorenko, D. Gurwitz, S. Wootton, R. Lagace, J. Chang, W.C. Speed, K.K. Kidd, Ancestry inference of 96 population samples using microhaplotypes, Int. J. Leg. Med. 132 (2018) 703–711.
[2] J.L. King, J.D. Churchill, N.M.M. Novroski, X. Zeng, D.H. Warshauer, L.- H Seah, B. Budowle, Increasing the discrimination power of ancestry- and identity-informative SNP loci within the ForenSeq™ DNA Signature Prep Kit, Forensic. Sci. Int.: Genetics 36 (2018) 60–76.
[3] S. Walsh, L. Chaitanya, K. Breslin, C. Muralidharan, A. Bronikowska, E. Pospiech, J. Koller, L. Kovatsi, A. Wollstein, W. Branicki, F. Liu, M. Kayser, Global skin colour prediction from DNA, Hum. Genet. 136 (2017) 847–863.
[4] C. Bartling, M. Hester, J. Bartz, E. Heizer, S. Faith, Next-generation sequencing approach to epigenetic-based tissue source attribution, Electrophoresis 35 (2014) 3096–3101.
[5] D. Hares, Selection and implementation of expanded CODIS core loci in the United States, Forensic Sci. Int.: Genetics 17 (2015) 33–34.
[6] C.R. Steffen, M. Coble, K.B. Gettings, P.M. Vallone, Corrigendum to 'U.S. population data for 29 autosomal STR loci' [Forensic Sci, 2013, Int. Genet. 7 (2017) e82–e83. Forensic Sci. Int.: Genetics..
[7] K.B. Gettings, L. Borsuk, C. Steffen, K. Kiesler, P. Vallone, Sequence-based U.S. population data for 27 autosomal STR loci, Forensic Sci. Int.: Genetics 37 (2018) 106–115.
[8] B. Algee-Hewitt, M. Edge, J. Kim, J. Li, N. Rosenberg, Individual identifiability predicts population identifiability in forensic microsatellite markers, Curr. Biol. 26 (2016) 935–942.
[9] B.F.B. Algee-Hewitt, C.E. Hughes, B.E. Anderson, Temporal, geographic and identification trends in craniometric estimates of ancestry for persons of Latin American origin, Forensic Anthropology 1 (2018) 4–17.
[10] S. Tredoux, S. Mfolozi, K. Shires, Efficiency of buccal DNA sampling device in the mortuary, J. Forensic Invest. 3 (2) (2015) 6.
[11] A.-L. Rahikainen, Palo Ju, W. de Leeuw, B. Budowle, A. Sajantila, DNA quality and quantity from up to 16 years old post-mortem blood stored on FTA cards,

Forensic Sci. Int. 261 (2016) 148–153.

[12] G.D.C. Santos, FTA cards for preservation of nucleic acids for molecular assays: a review on the use of cytologic/tissue samples, Arch. Pathol. Lab Med. 142 (2018) 308–312.

[13] D. Hall, R. Roy, An evaluation of direct PCR amplification, Croat. Med. J. 44 (2014) 655–661.

[14] K. Zejdlik, N. Passalacqua, J. Williams, A force for good: the what, where, and why of human decomposition facilities, Pathologist 6 (2018) 22. https://thepathologist.com/outside-the-lab/a-force-for-good.

[15] K. Pincknard, Memorial eckert paper for 2007 forensic DNA analysis for the medical examiner, The Am. Journ. of Forensic Med. And Path 29 (4) (2008) 375–381.

[16] P. Elliott, T. Peakman, The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine, Int. J. Epidemiol. 37 (2) (2008) 234–244.

[17] J. Gill, 9/11 and the New York city office of chief medical examiner, Forensic Sci. Med. Pathol. 2 (2006) 29–32.

[18] M. Prinz, A. Carracedo, W. Mayr, N. Morling, T. Parsons, A. Sajantila, et al., DNA Commission of the International Society for Forensic Genetics (ISFG): recommendations regarding the role of forensic genetics for disaster victim identification (DVI), Forensic Sci. Int.: Genetics 1 (1) (2007) 3–12.

[19] J. Butler, Advanced Topics in Forensic DNA Typing: Methodology, Elsevier, New York, 2012.

[20] R. Barallon, S.R. Bauer, J. Butler, A. Capes-Davis, W.G. Dirks, et al., Recommendation of Recommendation of short tandem repeat profiling for authenticating human cell lines, stem cells, and tissues. In Vitro Cell, Dev. Biol-Animal 46 (2010) 727–732.

[21] R.W. Nims, G. Sykes, K. Cottrill, P. Ikonomi, E. Elmore, Short tandem repeat profiling: part of an overall strategy for reducing the frequency of cell misidentification. In Vitro Cell, Dev. Biol-Animal 46 (2010) 811–819.

[22] J.M. Butler, Y. Shen, B.R. McCord, The development of reduced size STR amplicons as tools for analysis of degraded DNA, J. Forensic Sci. 48 (2003) 2003043.

[23] B. McCord, K. Opel, M. Funes, S. Zoppis, L.M. Jantz, An Investigation of the Effect of DNA Degradation and Inhibition on PCR Amplification of Single Source and Mixed Forensic Samples, NIJ Final Report, 2011, p. 236692.

[24] F. Nieuwerburgh, D. Van Hoofstat, C. Van Neste, D. Deforce, Retrospective study of the impact of miniSTRs on forensic DNA profiling of touch DNA samples, Sci. Justice 54 (2014) 369–372.

[25] K.Y. Kim, Y. Kwon, M. Bazarragchaa, A.J. Park, H. Bang, W.B. Lee, J. Lee, K.H. Lee, B.J. Kim, K. Kim, A real-time PCR-based Amelogenin Y allele dropout assessment model in gender typing of degraded DNA samples, Int. J. Leg. Med. 127 (1) (2013) 55–61, https://doi.org/10.1007/s00414-011-0663-5.

[26] O. Loreille, T. Diegoli, J. Irwin, M. Coble, T. Parsons, High efficiency DNA extraction from bone by total demineralization, Forensic Sci. Int.: Genetics 1 (2) (2007) 191–195.

[27] S. Amory, R. Huel, A. Bilic, O. Loreille, T. Parsons, Automatable full demineralization DNA extraction procedure from degraded skeletal remains, Forensic Sci. Int.: Genetics 6 (2012) 398–406.

[28] S. Lee, J. Shewale, DNA extraction methods in forensic analysis, in: R.A. Meyers (Ed.), Chapter in the Encyclopedia of Analytical Chemistry, John Wiley, Chichester, 2017, https://doi.org/10.1002/9780470027318.a1104m pub2. Published June 19 2017. Available at: http://onlinelibrary.wiley.com/doi/10.1002/9780470027318.a1104m.pub2/otherversions.

[29] H. Correa, F. Carvalho, M. Brito, E. O'Dantas, D. Junior, F. Silva, Forensic DNA typing from demineralized human bone sections, SM J. Forensic Res. Criminology 2 (2018) 1016.

[30] K. Grisedale, G. Murphy, H. Brown, M. Wilson, S. Sinha, Successful nuclear DNA profiling of rootless hair shafts: a novel approach, Int. J. Leg. Med. 132 (2018) 107–115.

[31] M. Stangegaard, B. Hjort, T. Hansen, A. Hoflund, H. Mogensen, A. Hansen, N. Morling, Automated extraction of DNA from biological stains on fabric from crime scenes. A comparison of a manual and three automated methods, Forensic Sci. Int.: Genetics 7 (3) (2013) 384–388.

[32] M. Hara, H. Nakanishi, K. Yoneyama, K. Saito, A. Takada, Effects of storage conditions on forensic examinations of blood samples and bloodstains stored

[33] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959.

[34] H.A. Ahmed, E.T. MacLeod, G. Hide, S.C. Welburn, K. Picozzi, The best practice for preparation of samples from FTA®cards for diagnosis of blood borne infections using African trypanosomes as a model system, Parasites Vectors 4 (2011) 68.

[35] G. Vidoli, D.W. Steadman, J. Devlin, L.M. Jantz, History and development of the first anthropology research facility, Knoxville, Tennessee, Taphonomy of Human Remains (2017), https://doi.org/10.1002/9781118953358.ch35.

[36] J. Glider, C. Inman, W. Shields, D. Krane, Magnitude-dependent variation in peak height balance at heterozygous STR loci, Int. J. Leg. Med. (2009), https://doi.org/10.1007/s00414-009-0411-2.

[37] C. Word, DNA Mixture Interpretation Webcast, National Institute of Standards and Technology, 2013. April 12, https://strbase.nist.gov/training/MixtureWebcast/6_ComplexMixtures-Word.pdf.

[38] N. Rosenberg, Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives, Ann. Hum. Genet. 70 (6) (2006) 841–847.

[39] Y.L. Li, J.X. Liu, Structure Selector: a web based software to select and visualize the optimal number of clusters using multiple methods, Mol. Ecol. Res. 18 (2018) 176–177.

[40] S.J. Puechmaille, The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem, Mol. Ecol. Resour. 16 (2016) 608–627.

[41] N.M. Kopelman, J. Mayzel, M. Jakobsson, N.A. Rosenberg, I. Mayrose, Clumpak: a program for identifying clustering modes and packaging population structure inferences across K, Mol. Ecol. Resour. 15 (5) (2015) 1179–1191.

[42] B.F.B. Algee-Hewitt, Population inference from contemporary American craniometrics, Am. J. Phys. Anthropol. 160 (2016) 604–624.

[43] K. Hunley, H. Edgar, M. Healey, C. Mosley, G. Cabana, F. West, Social identity in New Mexicans of Spanish-speaking descent highlights limitations of using standardized ethnic terminology in research, Hum. Biol. 89 (3) (2017) 217–228.

[44] O. Serra, R. Frazzi, A. Perotti, L. Barusi, A. Buschini, Use of FTA classic cards for epigenetic analysis of sperm DNA, Biotechniques 64 (2018). https://www.future-science.com/doi/full/10.2144/btn-2017-0101.

[45] S. Thompson, C. Boughton, G. Dore, Blood-borne viruses and their survival in the environment: is publich concern about needlestick exposures justified? Aust. N. Z. J. Publ. Health 27 (2002) 602–607. https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1467-842X.2003.tb00606.x.

[46] H. Tang, T. Quertermous, B. Rodriguez, S. Kardia, X. Zhu, A. Brown, et al., Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies, Am. J. Hum. Genet. 76 (2) (2005) 268–275.

[47] K. Bryc, E.Y. Durand, J.M. Macpherson, D. Reich, J.L. Mountain, The genetic ancestry of African Americans, Latinos, and European Americans across the United States, Am. J. Hum. Genet. 96 (2015) 37–53.

[48] Y. Banda, M. Kvale, T. Hoffmann, S. Hesselson, D. Ranatunga, et al., Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic Epidemiology research on Adult health and aging (GERA) cohort, Genetics 4 (2015) 1285–1295.

[49] F. Montinaro, G.B.J. Busby, V.L. Pascali, S. Myers, G. Hellenthal, C. Capelli, Unravelling the hidden ancestry of American admixed populations, Nat. Commun. (2015) [Internet] 6. Available from: http://www.nature.com/articles/ncomms7596.

[50] L. Porras-Hurtado, Y. Ruiz, C. Santos, C. Phillips, A. Carracedo, M. Lareu, An overview of STRUCTURE: applications, parameter settings, and supporting software, Front. Genet. 4 (2013) 98.

[52] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: dominant markers and null alleles, Mol. Ecol. Notes 7 (2007) 574–578.

[53] D.J. Lawson, L van Dorp, D. Falush, A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots, Nat. Commun. 9 (2018) 3258.

[54] Pritchard JK, Wen X, Falush D. Documentation for structure software: Version 2.3.:39. http://www.ccg.unam.mx/~vinuesa/tlem09/docs/structure_doc.pdf.