Correspondence

Corrigendum to 'U.S. Population Data for 29 Autosomal STR Loci' [Forensic Sci. Int. Genet. 7 (2013) e82–e83]

Carolyn R. Steffen, Michael D. Coble, Katherine B. Gettings, Peter M. Vallone*

National Institute of Standards and Technology, Material Measurement Laboratory, Gaithersburg, MD 20899-8314, United States

Dear Editor,

In 2013, we reported the genotypes and allele frequencies for 1036 unrelated samples in the U.S. population using capillary electrophoresis (CE) [1]. Since then, multiplex STR assays designed for sequencing technologies have become available, and we have re-analyzed our set of 1036 samples to determine sequence-based allele frequencies (manuscript in preparation). As a quality control for this sequence data and to evaluate back-compatibility, the calculated length-based genotypes from the sequence data were compared to the 2013 published CE genotypes. This comparison resulted in a list of differences which were further evaluated via sequence- and CE-data review. Instances in which the difference was not attributable to the sequencing assay were further evaluated with additional CE-based genotyping. This evaluation has resulted in revisions to the 2013 publication [1], detailed below.

We have categorized the reasons for revisions as: (1) polymerase chain reaction (PCR) primer design differences, (2) change in the reporting of tri-alleles, (3) laboratory error, and (4) data analysis error. In summary, revisions have been made for a total of 13 STR loci, four of which are U.S. core loci (D5S818, D7S820, D13S317, and TPOX). The remaining nine loci are D6S1043, F13A01, F13B, FESFPS, LPL, Penta C, Penta D, Penta E, and SE33. The revisions affect 12 separate samples in the 1036 data set (12/1036 = 1.16%) and are summarized in Table 1. The distribution of revisions among the four populations is as follows: four African American samples (4/342 = 1.12%), three Caucasian samples (3/361 = 0.83%), four Hispanic samples (4/236 = 1.69%), and one Asian sample (1/97 = 1.03%). The revisions affect 37 genotypes out of 30,044 total genotypes (37/30,044 = 0.123%), not including the change in reporting of tri-alleles. The tri-allelic genotypes detected at TPOX (9, 10, 11) and Penta D (11, 14, 15) were reported as bi-allelic in 2013 (TPOX reported as 9, 11 and Penta D reported as 11, 14). In the revised data set, these genotypes have been removed. This change not only impacts the frequencies of the removed alleles, but also results in a sample number change at these loci: TPOX revised global n = 1035 and revised African American n = 341; Penta D revised global n = 1035 and revised Hispanic n = 235. Any change in sample number results in a change in all allele frequencies at the affected

locus/population. A detailed presentation illustrating each of the revisions can be found at http://strbase.nist.gov/NISTpop.htm. Tables 2a–2d provides a summary of the revisions by population, locus, and specific allele(s) affected: original, revised, and the difference of revised − original. The maximum change in allele frequency by population was as follows 0.15% (African American), 0.28% (Caucasian), 0.71% (Hispanic), 1.0% (Asian). The greatest overall single change of 1.0% was observed for the 11 allele at D7S820 in the Asian population (n = 97).

Similar to the 2015 Federal Bureau of Investigation (FBI) allele frequency revisions [2,3], empirical comparisons of random match probabilities (RMP) calculated from the original allele frequencies and the revised allele frequencies were performed on 100 randomly generated profiles for the two populations where U.S. core loci have been affected (African American and Asian). Comparisons were based on the original 13 U.S. core loci, as the expanded loci were not affected by the allele frequency changes. The random profiles were generated using *DNA Profile Builder* software (http://www.nucfs.ac.uk/dna-profile-builder/) using the allele frequencies from the NIST original. RMP calculations for the 100 random profiles were generated with the LSAM (Laboratory Statistical Analysis Module) software (Future Technologies Inc., Fairfax, VA) using a theta correction of 0.01 for homozygous loci. Since the corrections only affected markers in the original 13 U.S. core loci, we only calculated statistics on these markers. The differences in the African American population RMP calculations were within 1.0004-fold and the differences in the Asian population RMP calculations were within 1.3262-fold. This falls within a 2-fold change in RMP (comparable to the FBI's analysis [3]) and well within the 10-fold difference expected by using a different set of allele frequencies for that population as suggested by previous studies and the National Research Council [4–7], as shown in Fig. 1 and Table 3.

RMP scenarios were calculated for each population assuming homozygosity at the affected loci and using a theta correction of 0.01. The analysis was performed to understand the scope of the "worst case" effect of the revisions. The bounds of less rare and more rare RMPs as a function of commonly used STR kits are tabulated in Supplemental Table 1. Using the Asian population as an example, RMPs of 1.22 fold less rare and 1.61 fold more rare were calculated for the loci contained

Table 1
Detailed summary of revised allele calls from the NIST 1036 U.S. population data set. Revised alleles are bolded. Reasons for revisions are numbered as follows: (1) PCR primer design differences (2) change in the reporting of tri-alleles (3) laboratory error and (4) data analysis error.

Sample	Population	Sample Name	Locus	Original	Revised
1	African American	C28B	D5S818	12,12	**7**,12
2	African American	OT05588	TPOX	9,11	removed
3	Hispanic	C88H	Penta D	11,14	removed
4	African American	C37B	D6S1043	20,20	**18**,20
5	African American	C63B	D6S1043	15,15	**13**,15
6	Asian	C66A	TPOX	8,11	**9,10**
	Asian	C66A	D5S818	12,13	**11**,12
	Asian	C66A	D7S820	11,11	**8,10**
	Asian	C66A	D13S317	8,11	11,**12**
7	Hispanic	C82H	D6S1043	11,14	11,**19**
	Hispanic	C82H	Penta C	5,11	5,**12**
	Hispanic	C82H	Penta D	11,12	**10**,11
	Hispanic	C82H	Penta E	7,12	7,**15**
	Hispanic	C82H	F13A01	6,7	**7**,7
	Hispanic	C82H	FESFPS	11,13	11,**11**
8	Hispanic	C84H	D6S1043	12,20.3	12,**15**
	Hispanic	C84H	Penta C	13,13	**12**,13
	Hispanic	C84H	Penta D	10,12	10,**10**
	Hispanic	C84H	Penta E	12,17	**15**,17
	Hispanic	C84H	F13A01	3.2,6	3.2,**5**
	Hispanic	C84H	F13B	9,9	**6**,9
	Hispanic	C84H	FESFPS	11,13	**12**,13
	Hispanic	C84H	LPL	10,10	10,**12**
9	Hispanic	C86H	D6S1043	13,14	13,**21.3**
	Hispanic	C86H	Penta C	13,13	**12**,13
	Hispanic	C86H	Penta D	10,12	10,**10**
	Hispanic	C86H	Penta E	7,9	**9,11**
	Hispanic	C86H	F13A01	5,7	**7**,7
	Hispanic	C86H	F13B	10,10	**9**,10
	Hispanic	C86H	FESFPS	12,13	12,**12**
10	Caucasian	OT07767	D6S1043	11,12	11,**13**
11	Caucasian	MT97180	SE33	18,20.2	**18.3**,20.2
12	Caucasian	C67C	F13A01	6,6	**5**,6
	Caucasian	C67C	F13B	10,10	**6,8**
	Caucasian	C67C	FESFPS	11,12	**10**,11
	Caucasian	C67C	LPL	10,10	**11,11**
	Caucasian	C67C	Penta C	11,11	**9**,11

For more details see http://strbase.nist.gov/Vallone-Error-Management-July-25-2017.pdf.
Tri-allele: 9, 10, 11.
Tri-allele: 11, 14, 15.

Table 2a
African American population allele frequency revisions to the 1036 data set. U.S. core loci are italicized. Negative and positive values reflect a decrease or an increase, respectively, in allele frequency. * denotes the removal of the tri-allele that decreased the denominator count by one.

African American			
D5S818	Original	Revised	delta
7	0.0015	0.0029	0.0015
12	0.3699	0.3684	−0.0015
*TPOX**	Original	Revised	delta
6	0.08918	0.08944	0.00026
7	0.01754	0.01760	0.00006
8	0.36696	0.36804	0.00108
9	0.19591	0.19501	−0.00090
10	0.08626	0.08651	0.00025
11	0.21637	0.21554	−0.00083
12	0.02632	0.02639	0.00007
13	0.00146	0.00147	0.00001
D6S1043	Original	Revised	delta
13	0.0965	0.0980	0.0015
15	0.0541	0.0526	−0.0015
18	0.1067	0.1082	0.0015
20	0.0731	0.0716	−0.0015

Table 2b
Asian population allele frequency revisions to the 1036 data set. U.S. core loci are italicized. Negative and positive values reflect a decrease or an increase, respectively, in allele frequency.

Asian			
D5S818	Original	Revised	delta
11	0.2680	0.2732	0.0052
13	0.1650	0.1598	−0.0052
D7S820	Original	Revised	delta
8	0.1289	0.1340	0.0051
10	0.2577	0.2629	0.0052
11	0.3608	0.3505	−0.0103
D13S317	Original	Revised	delta
8	0.2217	0.2165	−0.0052
12	0.2062	0.2113	0.0051
TPOX	Original	Revised	delta
8	0.5516	0.5464	−0.0052
9	0.0773	0.0825	0.0052
10	0.0258	0.0309	0.0052
11	0.2990	0.2938	−0.0052

in Identifiler, Globalfiler (Thermo Fisher), PowerPlex 16, PowerPlex Fusion/6C (Promega), and Investigator 24plex QS (Qiagen) STR kits.

The revised genotypes for the 1036 U.S. population data set are provided in Supplemental Table 2 and the revised allele frequencies for the full data set and each population group are provided in Supplemental Table 3. The revised data have been provided to the FBI CODIS unit for review and dissemination and are also available on STRBase at http://strbase.nist.gov/NISTpop.htm. We encourage the forensic community to further evaluate the effects of these changes.

Table 2c
Caucasian population allele frequency revisions to the 1036 data set. U.S. core loci are italicized. Negative and positive values reflect a decrease or an increase, respectively, in allele frequency.

Caucasian			
D6S1043	Original	Revised	delta
12	0.2368	0.2355	−0.0014
13	0.0859	0.0873	0.0014
SE33	Original	Revised	delta
18	0.0734	0.0720	−0.0014
18.3	0.0000	0.0014	0.0014
F13A01	Original	Revised	delta
5	0.1925	0.1939	0.0014
6	0.3504	0.3490	−0.0014
F13B	Original	Revised	delta
6	0.0942	0.0956	0.0014
8	0.2452	0.2465	0.0014
10	0.3920	0.3892	−0.0028
FESFPS	Original	Revised	delta
10	0.2812	0.2826	0.0014
12	0.2368	0.2355	−0.0014
LPL	Original	Revised	delta
10	0.4252	0.4224	−0.0028
11	0.2618	0.2645	0.0028
PentaC	Original	Revised	delta
9	0.1482	0.1496	0.0014
11	0.3961	0.3947	−0.0014

Table 2d

Hispanic population allele frequency revisions to the 1036 data set. U.S. core loci are italicized. Negative and positive values reflect a decrease or an increase, respectively, in allele frequency. * denotes the removal of the tri-allele that decreased the denominator count by one.
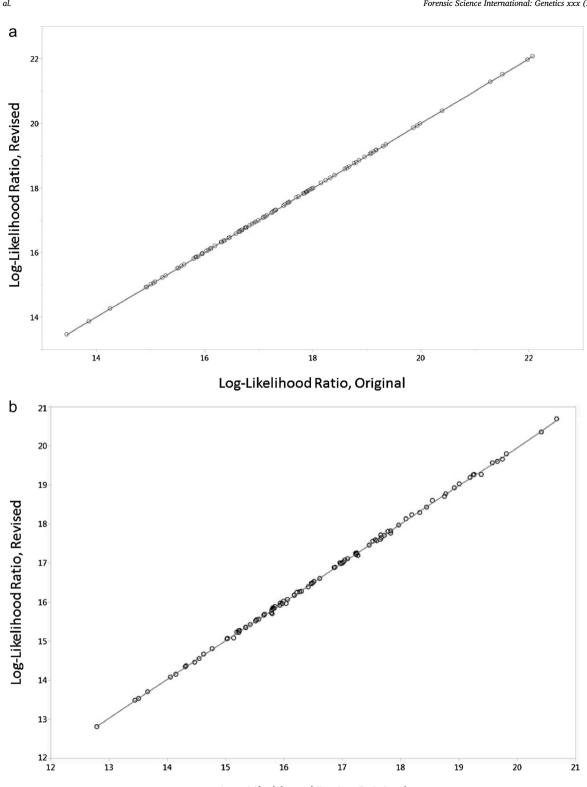
	Hispanic		
D6S1043	Original	Revised	delta
14	0.1356	0.1314	−0.0042
15	0.0297	0.0318	0.0021
19	0.0763	0.0784	0.0021
20.3	0.0127	0.0106	−0.0021
21.3	0.0403	0.0424	0.0021
PentaC	Original	Revised	delta
11	0.3326	0.3305	−0.0021
12	0.2034	0.2098	0.0064
13	0.1081	0.1038	−0.0042
PentaD*	Original	Revised	delta
2.2	0.01695	0.01702	0.00007
3.2	0.00212	0.00213	0.00001
5	0.00636	0.00638	0.00002
6	0.00212	0.00213	0.00001
7	0.00212	0.00213	0.00001
8	0.01907	0.01915	0.00008
9	0.24153	0.24255	0.00102
10	0.15678	0.16383	0.00705
11	0.15678	0.15532	−0.00146
12	0.16314	0.15745	−0.00569
13	0.14407	0.14468	0.00061
14	0.07203	0.07021	−0.00182
15	0.01059	0.01064	0.00005
16	0.00424	0.00426	0.00002
17	0.00212	0.00213	0.00001
PentaE	Original	Revised	delta
7	0.1186	0.1165	−0.0021
11	0.0742	0.0763	0.0021
12	0.1737	0.1695	−0.0042
15	0.0911	0.0953	0.0042
F13A01	Original	Revised	delta
6	0.1716	0.1674	−0.0042
7	0.3030	0.3072	0.0042
F13B	1036 (2013)	1036 (2017)	delta
6	0.1186	0.1208	0.0021
10	0.4407	0.4386	−0.0021
FESFPS	Original	Revised	delta
12	0.2140	0.2182	0.0042
13	0.0784	0.0742	−0.0042
LPL	Original	Revised	delta
10	0.4852	0.4831	−0.0021
12	0.2119	0.2140	0.0021

Fig. 1. (a) Plot of Log-Likelihood Ratios calculated from the original versus the revised allele frequencies for the African American population. The data represents a comparison of 100 randomly generated profiles for the 13 U.S. core loci. The data points are within a two-fold difference with a maximum of a 1.0004-fold difference in RMP. (b) Plot of Log-Likelihood Ratios calculated from the original versus the revised allele frequencies for the Asian population. The data represents a comparison of 100 randomly generated profiles for the 13 U.S. core loci. The data points are within a two-fold difference with a maximum of a 1.3262-fold difference in.

Table 3

The maximum and minimum observed ratios of the profile probabilities calculated using the original allele frequencies and the corrected allele frequencies data for the 100 randomly generated profiles.

Original RMP versus Revised RMP	African American	Asian
Max (new RMP is more rare)	1.0002	1.1501
Min (new RMP is less rare)	1.0004	1.3262

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2017.08.011.

References

[1] C.R. Hill, D.L. Duewer, M.C. Kline, M.D. Coble, J.M. Butler, U.S. population data for 29 autosomal STR loci, Forensic Sci. Int. Genet. 7 (2013) e82–e83.

[2] B. Budowle, T.R. Moretti, A.L. Baumstark, D.A. Defenbough, K.M. Keys, Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians Hispanics, Bahamians, Jamaicans and Trinidadians, J. Forensic Sci. 44 (1999) 1277–1286.

[3] T.R. Moretti, B. Budowle, J.S. Buckleton, Erratum: population data on the thirteen CODIS core short tandem repeat loci in african americans, US, caucasians, hispanics, bahamians, jamaicans and trinidadians, J. Forensic Sci. 60 (4) (2015) 1114–1116.

[4] National Research Council, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington D.C, 1996.

[5] K.L. Monson, B. Budowle, Effect of reference database on frequency estimates of polymerase chain reaction (PCR) − based DNA profiles, J. Forensic Sci. 43 (1998) 483–488.

[6] B. Budowle, K.L. Monson, A.M. Giusti, B. Brown, The assessment of frequency estimates of Hae III-generated VNTR profiles in various reference databases, J. Forensic Sci. 39 (1994) 319–352.

[7] B. Budowle, K.L. Monson, A.M. Giusti, B. Brown, Evaluation of Hinf I-generated VNTR profile frequencies determined using various ethnic databases, J. Forensic Sci. 39 (1994) 988–1008.