ARC Journal of Forensic Science
Volume 3, Issue 1, 2018, PP 22-26
ISSN 2456-0049
http://dx.doi.org/10.20431/2456-0049.0301005
www.arcjournals.org

Evaluation of Forensic STR Population Data from Samples Analysed in Leipzig

Michael Kohl[1*], Frank Götz[2], Jan Dressler[1], Jeanett Edelmann[1]
[1]*Institute of Forensic Medicine, University of Leipzig, Leipzig, Germany*
[2]*Qualitype GmbH Dresden, Germany*

Corresponding Author:** ***Michael Kohl, *Institute of Forensic Medicine, University of Leipzig, Leipzig, Germany,* ***Email:*** *michael.kohl@medizin.uni-leipzig.de*

Abstract: *Forensic DNA analysis makes it possible to provide biostatistically assessable evidence in criminal proceedings. The basics of these calculations are the allele frequencies or population data used in the respective calculation models. Here we present a new comprehensive population data collection, which comprises profiles from our forensic case studies of the last seven years. The data include typing results using the complete German standard marker set (European standard set plus additional loci D2S1358, D16S539, D19S433 and SE33) for 10.011 independent genotypes representing the entire forensic case work of the Leipzig Institute of Legal Medicine from the data collection period, without consideration of the geographical origin of each sample. Therefore the data collection can be assumed as typical mixed in our forensic case work. The fact that the exact proportions of individual ethnic groups are unknown should be compensated by the large number of samples. For statistical evaluation, the allele frequency data of all loci showed no deviation from the Hardy-Weinberg equilibrium. Compared with the ENFSI data, nearly similar values were found for the combined genotype frequencies.*

Keywords: *STR multiplexes; Population data; European standard set of markers*

1. Introduction

Forensic DNA analysis is performed using common sets of short tandem repeat (STR) loci which are defined to be typed routinely for entry of DNA genotype data into national or international databases [1]. The extension of core sets of genetic markers used for routine forensic DNA testing and human identification will aid international data-sharing capabilities and reduce the potential for random matches as DNA databases grow in size [2]. The selection of STR loci was based on collaborative exercises from the European DNA Profiling (EDNAP) Group and the recommendations of the European Network of Forensic Science Institutes (ENFSI) [3-5].Thus, the currently recommended European Standard Set (ESS) contains the 12 STR lociFGA, TH01, VWA, D1S1656, D2S441, D3S1358, D8S1179, D10S1248, D12S391, D18S51, D21S11 and D22S1045 as a minimum to enable the international comparison of DNA profiles [3].

There are commercial test kits available providing these markers as well as the additional loci D2S133, D16S539, D19S433 and SE33, which are part of several national databases [6-8].

For statistical analysis of DNA results each laboratory initially used its own allele frequency data specific to the corresponding region [9-13]. At first, only a few frequency data were available for the new European standard set of loci [14] until the ENFSI population data were published for calculations [15]. Since that time uniform European allele frequency data are recommended to use and the statements of reports became more comparable.

In the course of a present increasing global mixing of populations the impact of allele frequencies used for statistical calculations has changed. In Europe, more and more African, Asian or other foreign people are integrated so that a clear differentiation is often not possible. The use of adequate frequency data is thus becoming increasingly problematic. The best way to take this situation into account in statistical evaluations is a much larger frequency database.

2. Materials And Methods

2.1. Sample Collection

In total, 10.011 independent complete genotype data were evaluated. They were collected in our forensic case studies over the last years (2011-

2017) and comprise results of paternity investigations, reference samples, profiles of unknown single stain causers and reference samples from victims or suspects of the whole forensic case work of the Institute of Legal Medicine, Leipzig (Germany). Thus, in this work, DNA profiles of 7046unrelated persons and 2965 trace samples have been incorporated. The data were evaluated without reference to the geographical origin of the samples, representing a typical mixed population. All incorporated data were part of already executed investigation.

Subsequent data analysis proceeded anonymous. The responsible ethics committee was informed and agreed to analyze the already collected data.

2.2. Data Analysis

All data for the evaluation were independently confirmed in duplicate with the PCR amplification kits Power Plex ESI-17 and ESX-17 (Promega) or the ESS plex SE kit (Qiagen). It has been verified that each genotype appears only once in the database. Allele variants within the allelic ladder were allocated to an allele number corresponding to their length.

2.3. Allele Frequencies

The Allele frequencies were determined by counting the alleles in a Microsoft Excel® spreadsheet. The data were tested for Hardy-Weinberg Equilibrium (HWE) using the Guo and Thompson Fisher Exact test algorithm [16] using GenoProof software (Qualitype GmbH, Germany).

Based on the allelic frequencies, the following statistical parameters of genetic and forensic efficiency have been estimated.

2.4. Heterozygosity (H)

Observed heterozygosity (H_{obs}) is calculated by dividing the number of heterozygote individuals at a locus by the total number of all individuals at that locus. Expected heterozygosity (H_{exp}) is calculated according Edwards et al. 1992 [17] using GenoProof software package.

2.5. Polymorphism Information Content (PIC)

This parameter indicates the polymorphic level of a locus and was calculated according [18].

2.6. Power of Discrimination (PD)

It is defined as the probability that two individuals selected at random from the population will not have an identical genotype at the locus [19].

2.7. Power of Exclusion (PE)

It means the probability of excluding a random person as the contributor of an allele at the locus. [$PE = H^2 (1 - (1 - H) H^2)$]

2.8. Data Comparison

The combined genotype frequencies were calculated for each profile using the current population data and the ENFSI data [15] with a minimum frequency of 0.001 per allele. For the calculation of the ratio between the genotype frequencies, the ENFSI values were divided by the Leipzig values.

3. Results And Discussion

A total of 364 different alleles were determined in this study by evaluating 10.011 independent 16 STR marker profiles. Only results of confirmed duplicate determinations and complete profiles have been included. In comparison to the ENFSI study, the data show 79 additional alleles in the 16 STR systems considered. These alleles were observed 148 times in a total of 320.352 counted alleles. Conversely, 23 alleles of the ENFSI study (observed 29 times in 161.352 counted alleles) were not found in our dataset.

As in the ENFSI study, the largest number of different alleles could be observed for SE33, D21S11 and FGA. The greater the number of individual alleles per locus, the higher is their heterozygosity. Accordingly, the highest heterozygosity was found in SE33 with 0.949.The mean value of heterozygosity for the 16 STR loci is 0.826 with the lowest value observed in D22S1045 (0.730).Observed and expected allele frequencies for each locus found to be in Hardy-Weinberg equilibrium (table 1).

The data prove that there is a high level of polymorphism of the selected microsatellites. Basically, the number of different detected alleles depends on the size of the corresponding database and the geographical spread of the samples collected in the database. With rising population size and extended geographical origin, the number of observed alleles tends to increase, due to the presence of unique alleles and alleles which occur in lower frequencies. Since there are some alleles which were sampled only once or a few times in a dataset, a minimal allele frequency of 0.001 is recommended for forensic calculations [20].

The most common genotype profiles, reflecting the utility of the assay, were calculated for our and the ENSFI data using the product rule for the two most common alleles per locus in each

dataset [21]. The resulting genotypes were highly similar (table 2). The only differences were observed in the loci D12S391 and SE33.

The overall genotype frequencies were calculated as 1.9 x 10^{-17} (current study) and 2.9 x 10^{-17} (ENFSI data). This shows the close resemblance between the allele frequencies of the two datasets.

Furthermore, the combined genotype frequency of all samples was calculated using both sets of data. The computed values were only slightly different from each other. The observedratios (combined genotype frequency$_{ENFSI}$/ combined genotype frequency$_{Leipzig}$) were calculated to be 0.1275to 5.5296 (table 3), with 96.6 % of the samples ranging from 0.5 to 2.0.

Table1. *Statistic parameters of the STR markers*

Marker	N	Alleles	HET	PIC	PD	PE	MEC	HWE
D3S1358	10011	12	0.7937	0.7615	0.9045	0.5874	0.5892	confirmed
D19S433	10011	25	0.7897	0.7626	0.8945	0.5801	0.6026	confirmed
D2S1338	10011	16	0.8770	0.8650	0.9667	0.7487	0.7531	confirmed
D22S1045	10011	13	0.7304	0.6869	0.8254	0.4768	0.5003	confirmed
D16S539	10011	11	0.7685	0.7326	0.8752	0.5419	0.5551	confirmed
D18S51	10011	25	0.8776	0.8649	0.9677	0.7499	0.7514	confirmed
D1S1656	10011	28	0.9011	0.8926	0.9791	0.7977	0.7990	confirmed
D10S1248	10011	13	0.7651	0.7272	0.8724	0.5360	0.5455	confirmed
D2S441	10011	16	0.7569	0.7192	0.8604	0.5216	0.5401	confirmed
TH01	10011	11	0.7873	0.7544	0.8971	0.5757	0.5807	confirmed
VWA	10011	12	0.8087	0.7819	0.9170	0.6153	0.6224	confirmed
D21S11	10011	32	0.8468	0.8293	0.9471	0.6886	0.6971	confirmed
D12S391	10011	27	0.8896	0.8796	0.9735	0.7742	0.7776	confirmed
D8S1179	10011	13	0.8128	0.7891	0.9185	0.6230	0.6371	confirmed
FGA	10011	28	0.8650	0.8501	0.9602	0.7246	0.7274	confirmed
SE33	10011	81	0.9490	0.9466	0.9946	0.8962	0.8967	confirmed

N = population size, HET = heterozygosity, PIC = polymorphism information content, PD = power of discrimination, PE = Power of exclusion, MEC = mean exclusion chance

Table2. *Frequencies of the two most common alleles per locus (most frequent genotype profiles). Alleles in bold indicate variation between the data sets*

	Currentdata			**ENFSI data**		
	most common alleles		genotype frequency	most common alleles		genotype frequency
D3S1358	15	16	0.1223	15	16	0.1284
D19S433	13	14	0.1576	13	14	0.1646
D2S1338	17	20	0.0579	17	20	0.0638
D22S1045	15	16	0.2388	15	16	0.2438
D16S539	11	12	0.1813	11	12	0.1833
D18S51	14	15	0.0478	14	15	0.0504
D1S1656	15	17.3	0.0355	15	17.3	0.0377
D10S1248	13	14	0.1685	13	14	0.1702
D2S441	11	14	0.1971	11	14	0.1979
TH01	6	9.3	0.1366	6	9.3	0.1389
VWA	17	18	0.1178	17	18	0.1167
D21S11	29	30	0.1011	29	30	0.1009
D12S391	**18**	**21**	0.0447	**18**	**20**	0.0444
D8S1179	13	14	0.1311	13	14	0.1403
FGA	21	22	0.0657	21	22	0.0645
SE33	**19**	**28.2**	0.0112	**27.2**	**28.2**	0.0116
Combined (P)			1.905E-17			2.954E-17

Table3. *Ratio of the combined genotype frequencies of both datasets (genotype frequency$_{ENFSI}$/ genotype frequency $_{Leipzig}$).Shown are 5 samples with the lowest and highest combined values ranging from 0.1275 to 5.5296.*

sample	P07953	P07952	P05118	P04321	P06349	P00774	P04461	P09904	P00459	P01103
D3S1358	1.0110	0.9008	1.0344	1.0110	0.9057	1.0344	0.9990	1.0110	1.0496	1.0165
D19S433	0.8685	0.9867	1.0501	0.8145	0.8638	1.0330	1.0444	1.0387	1.0919	1.0501
D2S1338	1.0093	0.8937	1.0145	0.9305	0.9449	0.9839	0.9665	0.8983	1.0093	1.0489

D22S1045	1.0151	1.1119	1.0151	1.0622	1.0209	1.1119	1.0146	1.0209	1.0209	1.0209
D16S539	0.9058	1.0252	0.8321	1.0574	1.0110	0.9709	1.0545	1.0258	0.9709	1.0258
D18S51	0.7544	0.8235	0.7936	1.1024	0.9256	0.8988	1.1193	1.0233	1.0409	1.1830
D1S1656	0.9646	0.8960	0.8885	0.9977	0.9778	1.0964	0.9953	1.0073	0.9893	1.0960
D10S1248	1.0097	0.8946	1.0369	1.0097	1.0603	1.0325	1.0249	0.9760	1.0097	1.0603
D2S441	1.0041	1.0041	0.6114	1.0041	0.9525	0.9525	0.9844	1.1795	0.9994	0.9994
TH01	0.5475	0.5475	0.5426	0.9988	0.7781	0.7852	1.0054	1.0083	1.0054	1.0079
vWA	1.0339	0.9978	0.9769	0.9697	0.9906	0.9697	1.0262	0.9978	1.0861	1.0264
D21S11	1.0479	1.0904	1.0177	0.2506	0.4604	0.9990	0.9616	1.0098	1.0274	0.9186
D12S391	0.9078	0.9279	0.9926	0.9886	1.0176	1.0048	0.9886	0.8995	0.9721	1.2267
D8S1179	0.8919	0.9380	0.8502	0.8820	1.0071	1.1525	1.0701	1.0777	1.1607	1.1607
FGA	0.5471	0.5379	1.0304	1.0533	0.8905	1.1379	1.0867	1.0508	0.8648	0.9406
SE33	0.8074	0.9355	0.8074	0.8963	1.0114	3.2142	2.6752	3.0713	3.0135	2.6973
combined values	0.1275	0.1565	0.1576	0.1917	0.2284	3.5436	3.7000	3.7259	3.9299	5.5296

The frequency data and the resulting statistical parameters make our database suitable for use in forensic case work. This allows us to provide the most comprehensive population data set ever evaluated for this application. The data can be used primarily for the European population, but also for mixtures with persons of unknown origin. There is a high degree of consistency between our frequency data and the currently used ENFSI data, so a summary of both data sets would be possible. We believe that, due to the large sample size, the data collected are particularly well-suited for robust statistical assessments. If endorsed by the Stain Commission, these data could be uniformly recommended for statistical calculations and could provide the basis for appropriate computer programs.

REFERENCES

[1] Martin P. D., Schmitter H., Schneider P. M., A brief history of the formation of DNA databases in forensic science within Europe, Forensic Sci. Int. 119, 225-31(2001).

[2] Butler J. M., Hill C. R. and Coble M. D., Applied Genetics Group, National Institute of Standards and Technology, Gaithersburg, Maryland, USA. http://strbase.nist.gov/pub_pres/Profiles-in-DNA_Variability-of-New STR-Loci.pdf (2012).

[3] ENFSI DNA Working Group: DNA Database Management. Review and Recommendations http://enfsi.eu/wpcontent/uploads/2017/09/DNA-databasemanagement-review-and recommendatations-april-2017.pdf (2017).

[4] Gill P., Fereday L., Morling N., Schneider P. M., The evolution of DNA databases - recommendations for new European STR loci, Forensic Sci. Int. 156, 242–4 (2006).

[5] Gill P., Fereday L., Morling N., Schneider P. M., New multiplexes for Europe-amendments and clarification of strategic development, Forensic Sci. Int. 163, 155–7(2006)

[6] Schneider P. M., Expansion of the European Standard Set of DNA Database Loci – the current situation, Profiles in DNA 12, 6–7(2009).

[7] Budowle B., Ge J., Chakraborty R., Eisenberg A. J., Green R., Mulero J. et al. Population genetic analyses of the NGM STR loci, Int. J. Legal Med. 125, 101–9(2011).

[8] Tucker V. C., Hopwood A. J., Sprecher C. J., McLaren R. S., Rabbach D. R., Ensenberger M. G., Thompson J. M., Storts D. R., Developmental validation of the PowerPlex ESI 16 and PowerPlex ESI 17 Systems: STR multiplexes for the new European standard, Forensic Sci. Int. Genet. 5, 436–448(2011).

[9] Ćurić G., Gašić V., Plužarić V., Smiljčić D., Genetic parameters of five new European Standard Set STR loci (D10S1248, D22S1045, D2S441, D1S1656, D12S391) in the population of eastern Croatia, Croat. Med. J. 53(5), 409–415(2012).

[10] Vanek D., Silerova M., Urbanova V., Dubska J., Saskova L., Ehler E., Allele frequencies of the new European Standard Set (ESS) loci in the population of Czech Republic, Forensic Sci. Int. Genet.3, e395-e396, (2011).

[11] Seider T., Fimmers R., Betz P., Lederer T., Allele frequencies of the five miniSTR loci D1S1656, D2S441, D10S1248, D12S391 and D22S1045 in a German population sample, Forensic Sci. Int. Genet. 4, e159–60 (2010).

[12] Berti A., Brisighelli F., Bosetti A., Pilli E., Trapani C., Tullio V., Franchi C., Lago G., Capelli C., Allele frequencies of the new European Standard Set (ESS) loci in the Italian population, Forensic Sci. Int. Genet. 5 (5) 548–549(2011).

[13] Previderé C., Grignani P., Alessandrini F. et al., The 2011 GeFI collaborative exercise. Concordance study, proficiency testing and Italian population data on the new ENFSI/ EDNAP loci D1S1656, D2S441, D10S1248, D12S391, D22S1045, Forensic Sci. Int. Genet. 7 e15–e18 (2013).

[14] Barbora E., Ehlera E., Vanek D., Missing population studies on the new European Standard set loci (ESS) and its impact on population statistics, Forensic Sci. Int. Genet. Suppl. Series 5, e671-e673 (2015).

[15] Welch L. A., Gill P., Phillips C., Ansell R., Morling N., Parson W., Palo J.U., Bastisch I., European Network of Forensic Science Institutes (ENFSI): Evaluation of new commercial STR multiplexes that include the European Standard Set (ESS) of markers, Forensic Sci. Int. Genet. 6, 819-826(2012).

[16] Guo S. W., Thompson E. A., Performing the exact test of Hardy-Weinberg proportion for multiple alleles Biometrics 48, 361-372(1992).

[17] Edwards A., Hammond H. A., Jin L., Caskey C. T., Chakraborty R., Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups, Genomics, 12, pp. 241-253(1992).

[18] Botstein D., White R. L., Skolnick M., Davis R. W., Construction of a genetic linkage map in man using restriction fragment length polymorphism, Am. J. Hum. Genet., 32, pp. 324-331(1980).

[19] Brenner C., Morris J., Paternity index calculations in single locus hyper variable DNA probes. Proceedings of the international symposium on human identification, 1989, Promega Corporation, Long Beach, California, pp. 21-53(1990).

[20] Ulbrich W., Anslinger K., Bäßler G., Eckert M., Fimmers R., Hohoff C., Kraft M., Leuker C., Molsberger G., Pich U., Razbin S., Schneider H., Templin M., Wächter A., Weirich V., Zierdt H., Schneider P. M., Gemeinsame Empfehlungen der Projektgruppe „Biostatistische DNA-Berechnungen“ und der Spurenkommission zur biostatistischen Bewertung von DNA-analytischen Befunden. Rechtsmedizin 26, 291-298(2016).

[21] Edwards A., Hammond H. A., Jin L., Caskey C. T., Chakraborty R., Genetic variation atfive trimeric and tetrameric tandem repeat loci in four human population groups, Genomics 12, 241–253(1992).

Citation: *Michael Kohl, Frank Götz, Jan Dressler, Jeanett Edelmann. Evaluation of Forensic STR Population Data from Samples Analysed in Leipzig. ARC Journal of Forensic Science. 2018 3(1):22-26. http://dx.doi.org/10.20431/2456-0049.0301005*