

Between country comparison of a large Caucasian STR database collected as part of the Standardisation in DNA Profiling project

D. Syndercombe Court^{a,*}
European DNA Profiling group (EDNAP)^b

^a*Haematology Department, Barts and The London, Queen Mary's School of Medicine and Dentistry,
Turner Street, London E1 2AD, UK*

^b*EDNAP, UK*

Abstract. *Background:* Standardisation in DNA Profiling (STADNAP) was a project funded by the European Union between 1998 and 2000. One aim was to collect a large amount of frequency data for all the loci being used at the time. The purpose was to provide frequency information and to examine if there was any between country variability. *Methods:* A request for short tandem repeat (STR) genotype information was sent to 165 laboratories known to be working in the field, either as criminal or paternity practitioners. A detailed questionnaire was included so that information about the laboratory's quality practices and the typing methodology they used could be examined with the data. An exact test was used to examine the frequency data set each locus and anomalous patterns examined. *Results:* Thirty-four laboratories provided genotype data from over 20,000 Caucasian individuals, with 95% of the information originating from within Europe. Twenty-nine different STRs were reported, but larger amounts of data were generally seen amongst those loci available in available commercial multiplexes. Rare intermediates were reported in all systems except VWA, TPOX and D16S539. The exact test revealed significant differences in those loci related to the use of older technologies, where difficulties can be experienced in distinguishing alleles that differ by one base pair. © 2004 Elsevier B.V. All rights reserved.

Keywords: Short tandem repeat (STR); Caucasian frequencies; STADNAP

1. Introduction

Standardisation in DNA Profiling (STADNAP) was a project funded by the European Union between 1998 and 2000. A group of forensic scientists from European countries, currently also members of EDNAP, initiated the collection of short tandem repeat (STR)

* Corresponding author. Tel.: +44-2073-777076; fax: +44-2073-777629.

E-mail address: y.d.syndercombe-court@qmul.ac.uk (D. Syndercombe Court).

URL: <http://www.isfg.org/ednap/ednap.htm>.

frequency data from scientists working in the field, in order to provide a large amount of frequency data for all the loci being used at the time. The purpose was to provide frequency information and to examine if there was any between country variability.

2. Materials and methods

Letters were sent to 165 laboratories, throughout Europe and outside, seeking collaboration. In addition, a questionnaire was sent, requesting information on the laboratory's methodologies at the time. We wanted, in particular, to know what technology had been used to produce the database and how the alleles had been defined, including information on how the allelic ladders were produced. We also asked whether the laboratory participated in proficiency testing.

Submitted genotype data was screened for typographical errors and other discrepancies and individual STR genotypes that were in doubt deleted. Formatting differences were standardised before analysis of the data. The exact test was used to look at allele distribution on sets of STR data where more than 3000 genotypes had been collected.

3. Results

Caucasian genotype data was received from 34 laboratories, representing 12 European and 4 non-European countries. Laboratories provided a range of genotype information from between 100 and 6000 individuals resulting in a total of 21,544 profiles, of which 95% were from a European source. Data from 29 different STRs was provided, 16 of which were in common usage because of their availability in commercially available multiplexes, plus SE33, a locus commonly employed in German laboratories. These common loci each provided between 3000 and 15,000 genotypes.

Most samples were typed using automated technology. Applied Biosystems 310 and 377 instruments were the most popular instruments at the time, but a large amount of data was typed using older instrumentation, including ABI 373, Alf and AlfExpress. In addition, a substantial minority of laboratories used silver staining to provide manual types.

Virtually, all the laboratories participated in proficiency testing and all used allelic ladders run concurrently to type the samples. Many allelic ladders were of commercial origin, but homemade ladders were also extensively used. Virtually, all ladders had also been sequenced.

Rare intermediates were reported at around a 0.2% frequency, although none were seen in the VWA (24,000 alleles), TPOX (11,000 alleles) and D16S539 (6000 alleles) loci. The exact test revealed nothing of significance on examination of the D2S1338, D3S1358, D5S818, D13S317, D16S539, D21S11 and TPOX loci.

Significant differences were seen on the exact test in relation to the D18S51 ($p=0.035$), CSF1PO ($p=0.007$), D19S433, FGA, TH01, VWA and SE33 (all $p<0.00001$). Multiple testing could have resulted in the D18S51 locus significance and this was not considered further. The observation of a slight excess of 9 and 14 homozygotes in the CSF1PO locus could also have been due to chance, but other loci differences were likely to be real effects. Examination of the D19S433 locus revealed one laboratory typing data two repeats less than the others. Removal of this data resulted in the exact test becoming nonsignificant. In the FGA locus, there was a tendency to underreport the 0.2 intermediates and, at the TH01 locus,

an excess of 10/10 homozygotes reported. Examination of the VWA data revealed nothing obvious other than a slight excess of homozygotes, and more variability comparing observed and expected where genotypes differed by only one repeat. In SE33, there appeared to be less 24 alleles than expected.

4. Conclusions

The only significant difference observed between allele distributions was associated with a difference in nomenclature.

It was notable that the only other differences observed in the exact test were within those loci popular in the early days of STR testing (TH01, FGA, VWA and SE33 all fall into that category). The reduced ability of the early genotyping instruments or manual methods to reliably detect the 0.2 intermediates seen in FGA, and distinguish 9.3 and 10 alleles in TH01, is likely to be the explanation for the observed differences. SE33 is also a complex locus that may have resulted in typing difficulties, although the reason for a particular allele to be underrepresented is not clear. VWA has been the subject of several reports of significant differences over time, but this too is an 'old' locus and examination of data typed only with newer technology may clarify this and the other issues.

Collection of a large amount of data from across Europe has provided a set of reliable STR frequencies for the majority of loci commonly used across the region. It has also revealed the importance of technology that can reliably distinguish single base pair differences.

Acknowledgements

The group would like to acknowledge the European Union for their funding of the STADNAP project, and also to thank the following organisations for their generous contribution of data for this project: University of Buenos Aires, Argentina; Universities of Innsbruck and Vienna, Austria; Universities of Antwerp and Leuven, Belgium; DNA Reference Laboratory, Brazil; University of Copenhagen, Denmark; Scientific Inter-Regional Police Laboratory, Toulouse, France; Institutes of Blood Grouping, Legal Medicine, Forensic Medicine and Human Genetics of Berlin, Cologne, Dresden, Geestacht, Heidelberg, Homburg, Leipzig, Mainz, Marburg, Rostock, LKA Berlin and LKA Nordrhein Westfalen, Germany; Ireland Forensic Science Laboratory; Rome Catholic University, Italy; University of Okinawa, Japan; University of Oslo, Norway; Institutes of Legal Medicine, Coimbra, Lisbon and Porto, Portugal; University of Cape Town, South Africa; Centre for Genetic Analysis (CAGT) and Institutes of Legal Medicine, Madrid and Seville, Spain; Northern Ireland Forensic Science Laboratory and Haematology at Barts and The London, UK.