

Dynamics of Molecular Genetic Diversity in the East Midlands (UK): Some Forensic and Paternity Implications

S. S. Mastana¹, D. R. Lee¹, P. P. Singh² and M. Singh²

1. *Human Genetics Lab., Department of Human Sciences, Loughborough University Loughborough. U.K.*
2. *Department of Human Biology, Punjabi University, Patiala, India*

KEY WORDS VNTRs; STRs; East Midlands; Derbyshire; Leicestershire and Nottinghamshire

ABSTRACT The main objectives of this investigation were to establish a database of minisatellite (VNTRs) and microsatellite (STRs) allele frequencies for the regionally sub-divided populations of the East Midlands, which is suitable for forensic and population genetic investigations. The secondary objective was to determine if Caucasian sub-population heterogeneity exists within the United Kingdom and how this affects the forensic and paternity inferences. Blood samples were collected from random blood donors belonging to five regions of the East Midlands. Using standard molecular genetic techniques, we analysed MS1, MS31, YNH24, MS43a VNTRs and HumTHO1, HumF13A, HumFESFPS, HumTPOX, HumCSF1PO and HumvWA STRs. There was appreciable genetic variation in regionally subdivided populations. Overall efficiency of these loci for forensic and paternity work in the East Midlands is at par with other Caucasian populations though estimates show considerable variation within the region.

INTRODUCTION

The study of genetic diversity is the cornerstone of all population and forensic genetics investigations. Firstly, the newly isolated molecular genetic markers have been shown to exhibit extensive variation amongst the human species as compared to conventional/classical (blood serum) markers. Secondly, only very small quantities of material are required for analysis, due to recent refinements in amplification procedures. Finally, and most importantly, by analysing the variation within various populations and sub-populations of the world, a picture of the genetic structure of these populations will evolve. The applications of such knowledge are far-reaching and multidisciplinary.

Address for Correspondence: Dr. Sarabjit Mastana, Human Genetics Laboratory, Department of Human Sciences, Loughborough University, Loughborough. LE11 3TU. UK
Tel: +44-1509-223041, *Fax:* +44-1509-223941
E-mail: S.S.Mastana@LBORO.AC.UK

The new genetic techniques have been shown to be a highly effective and informative tool in the areas of genomic mapping, evolution, and forensic and paternity investigations (Budowle et al. 1991a, b, 1995; Deka et al. 1994, Brinkmann et al. 1998; Mastana 1999; Mastana and Papiha 2001)

The aim of the present study was to determine the level and extent of the variation of molecular genetic markers, Variable Number of Tandem Repeats (VNTRs) and Short Tandem Repeats (STR) amongst five sub-populations of the East Midlands region (UK) as a part of our ongoing genetic investigations (Mastana and Sokol 1998). A Comparative study of the five populations (Northwest Derbyshire, Northeast Derbyshire, and South Derbyshire, Nottinghamshire and Leicestershire) was performed by variety of multivariate statistical methods. They will be referred to in tables, figures and the following text as NWD, NED, SD, NOT and LEI respectively. Standard forensic and paternity indices were calculated for these markers and compared in order to elucidate the validity of these markers in forensic investigations and paternity disputes.

Previous studies have observed that NWD has shown some isolation from the other four populations due to the geographical nature of the land (the Peak District), thus restricting migration into, and out of, this region (Mastana and Sokol 1998). NED has shown some isolation from SD probably caused by the geographical barrier of the river Trent, or as a reflection of the settlement pattern of the continental European populations (Danish) in this region (Mastana and Sokol 1998). So, it was deemed important to analyse molecular markers to evaluate if there are any potential issues of population substructure in forensic usage of these genetic markers.

The VNTR loci analysed in this investigation were MS1 (D1S7), MS31 (D7S21), MS43a

(D12S11) and YNH24 (D2S44). These loci display very high levels of heterozygosity (between 96% to 99%) and low mutation rates (0% to 5.2%). The STRs analysed in this investigation were HumTH01 (D18S51), HumTPOX, HumCSF1PO, HumvWA, HumFESFPS and HumF13A.

SUBJECTS AND METHODS

Blood samples were collected from random blood donors with the individual's consent. These samples were analysed using standard molecular genetic techniques for a battery of genetic markers (Papiha et al. 1996; Mastana 1999). The base pair size data was subsequently utilized to determine allele frequencies, heterozygosity levels and Hardy-Weinberg Expectations. The sample sizes (number of individuals) for the VNTR loci varied from 50 in SD to 57 NED and for STR loci 43 (NOT) to 49 (NWD). VNTR data was binned into 31 FBI fixed bins (Budowle et al. 1991b) and further analyses were performed on the fixed bins. The gene diversity coefficient, (G_{st}), and Nei's genetic distance, D_A , were also computed. The forensic and paternity indices, Probability of Match (PM) and Power of Exclusion (PE) were calculated after Jones (1972)

RESULTS AND DISCUSSION

Although the sample sizes for each population studied is small, it was clear that the distributions of the bin/allele frequencies showed interesting variation at the analysed loci. Selected examples of this variation are shown in figure 1. General pattern of bin/allele frequency variation amongst these populations is similar to many European Caucasian populations (Budowle et al. 1991b). All populations were in Hardy Weinberg Equilibrium (HWE) which was tested by two independent tests- Maximum

Likelihood Ratio and Exact Test (Weir 1996; Guo and Thompson 1992).

The VNTR loci have higher heterozygosity level compared to STR loci. Overall a narrow range of average heterozygosity was observed across populations (82 to 87%). The coefficient of gene differentiation (G_{st}) is also notably greater with the molecular genetic markers (0.009 for VNTRs and 0.028 for STRs), than the conventional blood- serum markers (0.003) (Mastana and Sokol 1998).

The D_A genetic distance calculated showed small but significant inter-population differences, a UPGMA dendrogram derived from it is given in figure 2. Overall, Leicestershire stands out from other populations of the region (Fig. 2). Also, it can be seen that North East Derbyshire shows some degree of increased genetic distance/isolation from other Derbyshire populations. This finding is in line with our previous work (Mastana and Sokal 1998) and has been attributed to the settlement patterns of the continental European populations in this region and so isolated reproductive groups may have evolved. Comprehensive comparisons with European populations were not possible due to lack of consistent data sets.

There are interesting forensic implications of this work. PM and PE values of different populations are given in table 1. VNTR markers have greater level of sensitivity than STR markers for forensic and paternity calculations. However, due to the discrete nature and small size, the STR loci are much easier and faster to analyse in the laboratory. They subsequently offer less scope for interpretation error and are easier to incorporate into a multiplexing gel substrate analysis system. There are also some interesting differences in the populations studied here. Probability of match for STR loci varied from 1 in 170412 (SD) to 1 in 344650, which is of middle order, while for the whole region this value is relatively

Table 1: Probability of match and paternity exclusion in five populations of the East Midlands

	<i>STR-PM</i>	<i>1 in VNTR-PM</i>	<i>1 in</i>	<i>Combined PM</i>	<i>1 in</i>	<i>STR-PE</i>	<i>VNTR-PE</i>	<i>Overall PE</i>	
NWD	3.8E-06	265579	3.2E-06	308985	1.2E-11	8.21E+10	98.04%	99.46%	100.00%
NED	2.9E-06	344650	4.6E-06	218400	1.3E-11	7.53E+10	99.53%	99.49%	100.00%
SD	5.9E-06	170412	5.6E-06	177436	3.3E-11	3.02E+10	99.15%	98.82%	100.00%
LEICS	3.0E-06	333938	2.7E-06	364394	8.2E-12	1.22E+11	98.64%	99.78%	100.00%
NOT	4.2E-06	240285	2.1E-06	475443	8.8E-12	1.14E+11	97.37%	99.59%	100.00%
Whole Region	1.0E-06	988127	5.2E-07	1927228	5.3E-13	1.9E+12	98.40%	99.43%	100.00%

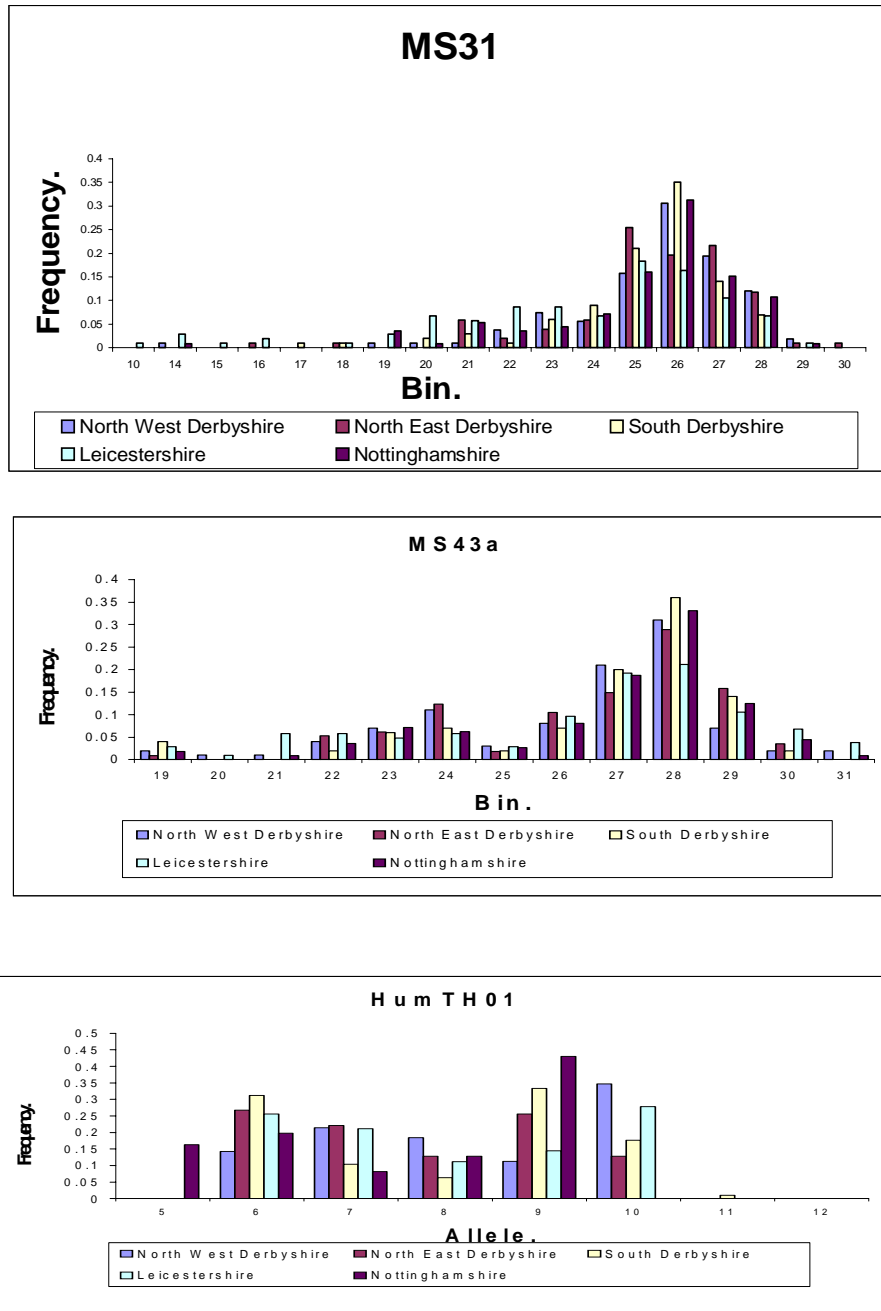


Fig. 1. Distribution of selected VNTR and STR loci in the East Midlands

better (1 in 988127). In comparison 4 VNTR loci also provide similar or higher PM estimates, where for the whole region it is 1 in 1927227.

Overall using both VNTRs and STRs one can correctly discriminate between forensic samples with very low match probabilities (1 in 1.9×10^{12})

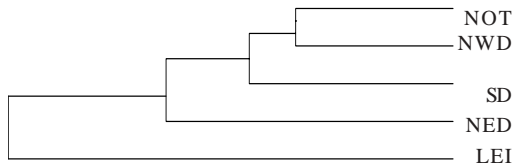


Fig. 2. UPGMA dendrogram derived from Nei's genetic distance (D_A) for the combined molecular data.

and high paternity exclusion values. These molecular genetic marker systems have been shown to demonstrate a much greater level of heterozygosity than conventional marker systems (Mastana and Sokal 1998) and can therefore be considered as a useful tool in anthropogenetic, forensic and paternity investigations.

ACKNOWLEDGEMENT

SSM would like to thank Loughborough University for research support and all blood donors for samples.

REFERENCES

- Brinkmann B, Junge A, Meyer E, Wiegand P 1998. Population genetic diversity in relation to microsatellite heterogeneity. *Human Mutation*, **11**: 135-44.
- Budowle B, Chakraborty R, Guisti AM, Eisenberg AJ, Allen RC 1991a. Analysis of the VNTR locus D1S80 by the PCR following high-resolution PAGE. *American Journal of Human Genetics*, **48**: 137-144.
- Budowle B, Guisti AM, Waye JS, Baechtel FS, Fournay RM, Adams DE, Presley LA, Deadman HA, Monson KL 1991b. Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. *American Journal of Human Genetics*, **48**: 841-55.
- Deka R, Shriver MD, Yu LM, Ferrell RE, Chakraborty R 1995. Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis*, **16**: 1659-64.
- Guo SW, Thompson EA 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**: 361-372.
- Jones DA 1972. Blood samples: Probability of discrimination. *Journal of Forensic Science Society*, **12**: 355-359.
- Mastana SS 1999. Genetic analysis of the D1S80 locus in five North Indian populations. *Annals of Human Biology*, **26**: 405-411.
- Mastana SS, Papiha SS 2001. D1S80 distribution in world populations with new data from the UK and the Indian sub-continent. *Annals of Human Biology*, **28**: 308-18.
- Mastana SS, Reddy PH, Das MK, Reddy P, Das K 2000. Molecular genetic diversity in 5 populations of Madhya Pradesh, India. *Human Biology*, **72**: 499-510.
- Mastana SS, Sokol RJ. 1998. Genetic variation in the east Midlands. *Annals of Human Biology*, **25**: 43-68.
- Papiha SS, Mastana SS, Purandare CA, Jayasekara R, Chakraborty R 1996. A population genetic study of three VNTR loci (D2S44, D7S22, D12 S11) in five ethnically defined populations of Indian Subcontinent. *Human Biology*, **68**: 819-835.
- Weir BS 1996. *Genetic Data Analysis II*. Massachusetts: Sinaur.