

Wei Wang · Huiling Jia · Qin Wang · Zhiming Cai
Lihui Wei · Debing Wang · Alan H. Bittles

STR polymorphisms of “forensic loci” in the northern Han Chinese population

Received: 25 February 2003 / Accepted: 7 April 2003 / Published online: 17 June 2003
© The Japan Society of Human Genetics and Springer-Verlag 2003

Abstract A preliminary Chinese DNA database has been constructed by the analysis of samples from 2,211 Han Chinese in Liaoyang City, northeast China. Thirteen autosomal tetranucleotide short tandem repeats (STRs) widely used in forensic identification were selected for the DNA profiling, together with the X-Y homologous gene *Amelogenin* for sex determination. Only one of the 13 autosomal loci showed significant deviation from Hardy-Weinberg equilibrium in the individuals genotyped. The cumulative discrimination power and power of exclusion of the 13 loci were greater than 0.999999999 and 0.9999888, respectively, giving an average match probability of 5.5×10^{-15} for the population. Allelic distributions at the vWA, TH01, D13S317, and D16S539 loci differed from African-Americans and US Caucasians, and more detailed population data at these four loci may be needed to ensure their applicability for

forensic purposes in Chinese populations. Previously unreported alleles were detected at several loci (some at relatively high frequencies), suggesting the need for their inclusion in the reference allelic ladder to meet the practical standard of forensic profiling in certain Chinese ethnic sub-populations. The preliminary DNA database provides base-line information applicable to the construction of a National Index System for criminal DNA profiling in PR China.

Keywords Forensic DNA typing · DNA database · STR polymorphisms · Han Chinese · PR China

W. Wang (✉) · Q. Wang · Z. Cai · L. Wei
D. Wang
Central Laboratory, Peking University Shenzhen Hospital,
Shenzhen Medical Centre,
Peking University-Hong Kong University of Sciences
and Technology, Shenzhen, PR China
E-mail: w.wang@ecu.edu.au
Tel.: +86-755-83066402
Fax: +86-755-83910721

W. Wang
School of Public Health,
Peking University Health Science Centre,
Beijing, PR China

H. Jia
Bureau of Public Hygiene and Health,
Liaoyang City Government,
Liaoyang, PR China

Z. Cai
Reproductive Centre,
Peking University Shenzhen Hospital,
Shenzhen, PR China

W. Wang · A. H. Bittles
Centre for Human Genetics,
Edith Cowan University,
Perth, Australia

Introduction

DNA profiling has provided key information in forensic criminal identifications of sexual assault and murder, parentage and sibling relationship testing, and war casualty investigations. There is also an increasing demand for DNA typing in clinical medicine, including cell line monitoring, laboratory contamination control, prenatal diagnosis, and donor and recipient matching in organ and bone marrow transplantations, and in compensation medicine involving victim identification in car, train, and aircraft accidents (Jack 1997).

The first national DNA database was introduced in the UK in 1995 and has since assisted in solving a large number of crimes by linking offenders with crime scenes, even in cases in which no suspects had been previously identified (Wrettt 1997). The subsequent creation of the US Federal Bureau of Investigation (FBI) Combined DNA Index System (CODIS) in 1998 has allowed national searches for criminals in all 50 US states (Hoyle 1998). By July 2001, there were 707,867 DNA profiles of offenders in the USA, and 28,711 stored profiles from unsolved crime scenes. Some 1,211 offenders have been identified in 30 states, and 667 unidentified crime scenes have been linked with other crimes. The remains of over 100 missing persons have also been identified following

the attack on the World Trade Centre in 2001 (Wertz 2002).

In addition to the UK and USA, other developed countries including Australia, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, Norway, Portugal, Russia, Sweden, and Switzerland, are also in the process of compiling or have created national DNA databases (Lincoln 1997; Schneider 1997; Wrettt 1997). The European DNA Profiling Group was established in 1989 and has grown rapidly from the eight initial members to 20 members representing all states of the European Community and associated western European countries (Schneider 1997; Martin et al. 2001; Schneider and Martin 2001).

The Peoples Republic China (PR China), with a current population of 1,260 million, is facing the major challenge of creating a national DNA database to meet the rapidly growing demand for DNA forensic profiling. In 1999, we started the construction of a preliminary Chinese DNA Database by genotyping 13 short tandem repeats (STRs) by using the AmpFL STR Profiler Plus Kit and AmpFL STR Cofiler Kit (Applied Biosystems), plus the sex-specific locus *Amelogenin*. Here, we present a summary of the resultant genotyping data that will provide base-line information for the ongoing task of constructing a national DNA index system in PR China.

Subjects and methods

Subjects

Blood samples were collected from 2,211 Han Chinese (1,111 males and 1,100 females) in northeast PR China, all of whom voluntarily took part in the project. Ethical clearance for the sampling was obtained from the Bureau of Public Hygiene and Health, Liaoyang City Government, P.R. China.

Finger-prick blood samples from each individual were collected on two Whatman 3MM cards. One card was bar-coded and despatched for DNA analysis, and the other was stored at -80°C .

Genotyping

DNA was extracted by the Chelex extraction method (Walsh et al. 1991). The amplification of 13 autosomal STR loci, viz., TPOX (2p13), D3S1358 (3p), FGA (4q28), D5S818 (5qter), CSF1PO (5q33.3), D7S820 (7pter), D8S1179 (8pter), TH01 (11p15.5), vWA (12p13.3), D13S317 (13q22), D16S539 (16q23.1), D18S51 (18q21.33), and D21S11 (21q21), together with the *Amelogenin* locus (Xp22.31–22.1 and Yp12.1) for sex testing, was performed on a GeneAmp PCR System 9600 (Applied Biosystems) by using the AmpFL STR Profiler Plus Kit and AmpFL STR Cofiler Kit (Applied Biosystems) according to manufacturer's recommendations (Perkin-Elmer 1998). Separation of the subsequent PCR products was undertaken by capillary electrophoresis on an ABI Prism 310 DNA Analyzer (Applied Biosystems). The Genotyper program (Genotyper 2.1) was used to detect and analyze PCR products by reference to allelic ladders (Applied Biosystems). The nomenclature system for allele designation was based on the number of repeat units contained in each allele according to the DNA Commission of ISFH (1994) regarding PCR-based polymorphism in STR systems.

Statistical analysis

Basic statistical computations including allele frequency, observed heterozygosity, pairwise independence of genotypic frequencies for each combination of loci, and Hardy-Weinberg equilibrium (HWE) tests were performed by using the GDA program (Lewis and Zaykin 2000). An exact test was used to assess the significance of deviation from HWE (Guo and Thompson 1992). The power of discrimination (DP), average power of paternity exclusion (PE) per locus and for the combined loci, and polymorphism information content (PIC), were computed according to established methods (Botstein et al. 1980; Odelberg and White 1990; Edwards et al. 1992; Weir 1996).

Results

Allelic distributions

As shown in Table 1, a total of 161 alleles were detected at the 13 autosomal loci, ranging from seven alleles at TH01 to 26 alleles at FGA. An average of 12.4 alleles per locus was recorded among the 4,422 chromosomes genotyped. The shortest allele was allele-11 (110 bp) at D3S1358, and the longest allele was allele-15 (317 bp) at the CSF1PO locus. Rare alleles were found at FGA (alleles 15 and 16), D5S818 (allele 17), D8S1179 (allele 18), D18S51 (alleles 6, 24, 25, 26, 27), and D21S11 (alleles 23.2, 27.2, 28.2). With the exception of loci vWA, TH01, D13S317, and D16S539 ($P < 0.05$; Fig. 1), the majority of loci showed comparable allelic distributions to African-American and US Caucasian data ($P > 0.05$; Perkin-Elmer 1998).

Tests of HWE and genotypic disequilibria

Of the 13 loci investigated, only locus D8S1179 deviated significantly from HWE ($0.01 < P < 0.05$) in the Han Chinese ($n = 2,211$). The results of the genotypic equilibrium tests are shown in Table 2. Significant dependence in pairwise genotypic frequencies were found at nine pairs of loci in the Han Chinese: TPOX-D7S820, D8S1179-D3S1358, D8S1179-FGA, D8S1179-D5S818, vWA-FGA, D16S539-D7S820, D18S51-FGA, D21S11-FGA, D21S11-D8S1179.

Calculation of forensic statistics

The forensic statistics are summarized in Table 3. Of the 13 autosomal loci, locus D18S51 showed the highest values of DP (0.96), PIC (0.84), and PE (0.72). The accumulated DP and PE were 0.999999999 and 0.9999888, respectively (Fig. 2), giving a match probability for the population of 5.5×10^{-15} .

The X-Y homologous gene *Amelogenin* for sex determination was not included in the all calculations, because of its different inheritance pattern from autosomes. *Amelogenin* itself gives 50% exclusion power regarding questions concerning the gender of the suspected criminal.

Fig. 1 Comparative distribution of TH01, vWA, D13S317, and D16S539 by ethnic studies

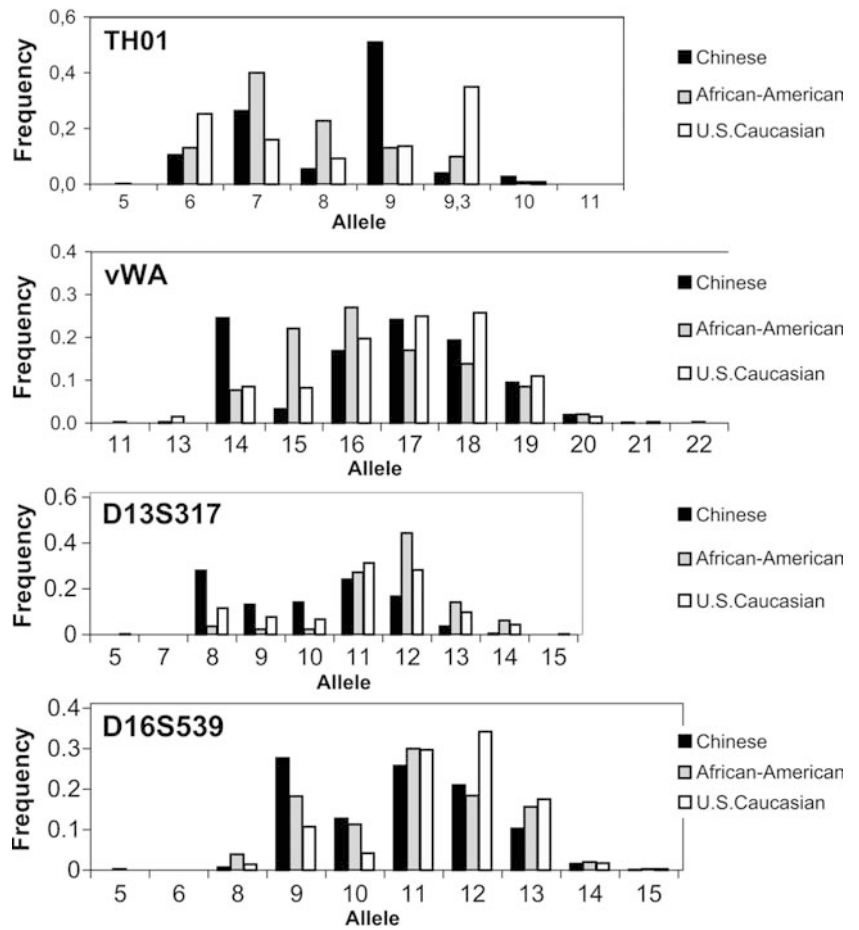


Table 2 Loci displaying genotypic disequilibrium in Han Chinese

Loci	D7S820	D8S1179	vWA	D16S539	D18S51	D21S11
TPOX	$P < 0.03$					
D3S1358	$P < 0.01$					
FGA	$P < 0.04$		$P < 0.02$		$P < 0.04$	
D5S818	$P < 0.05$					
D7S820	$P < 0.01$					
D8S1179	$P < 0.02$					

Discussion

As the frequency of specific STR alleles can vary according to the ethnic origin of the individuals sampled, separate DNA databases may need to be constructed from the results obtained from different major populations (Fig. 1). Studies have also indicated the potentially confounding effects of intra-community and consanguineous marriage on STR profiling in various ethnic groups (Weir 1994; Balding and Nichols 1995; Wang et al. 2000, 2003; Black et al. 2001; Zhivotovsky et al. 2001; <http://www.consang.net/>). To meet such a challenge in PR China, frequency profiling of the

Table 3 Forensic statistics for the Chinese study population across the 13 STR loci (*DP* power of discrimination, *PE* average power of paternity exclusion per locus, *PIC* polymorphism information content)

Locus	DP	PE	PIC
TPOX	0.8023	0.3702	0.5687
D3S1358	0.8772	0.4816	0.6764
FGA	0.9621	0.7090	0.8369
D5S818	0.9122	0.5629	0.7389
CSF1PO	0.8446	0.4988	0.6899
D7S820	0.9139	0.5626	0.7376
D8S1179	0.9570	0.6862	0.8246
TH01	0.8332	0.4197	0.6091
vWA	0.9333	0.6125	0.7762
D13S317	0.9297	0.6025	0.7691
D16S539	0.9203	0.5781	0.7520
D18S51	0.9648	0.7211	0.8443
D21S11	0.9461	0.6537	0.7991

forensic loci from each of the 56 officially recognized ethnic populations (Chinese Family Planning Commission 1998) is required. In addition, the possible effect of geographic differentiation needs to be investigated, particularly in the majority Han Chinese community, which accounts for over 92% of the total

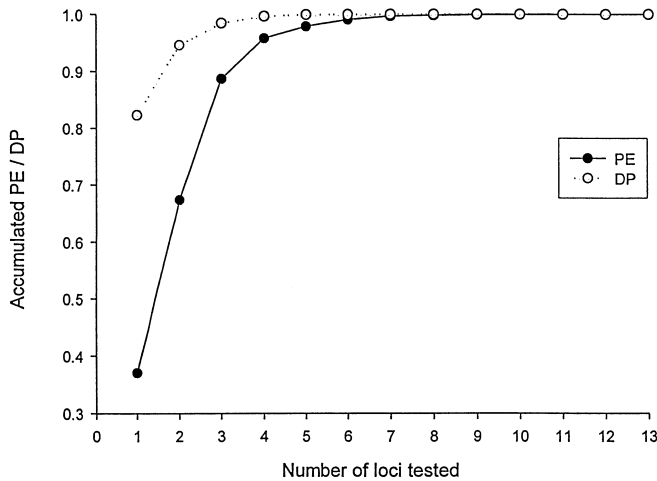


Fig. 2 Plot of the cumulative power of exclusion (*PE*) and discrimination power (*DP*) for the Han Chinese population, by number of loci tested

population of 1,260 million, and which is dispersed across an area of 9.6 million square kilometres. In the present context, it is possible that the genotypic disequilibrium observed at some of the loci on various chromosomes (Table 2) may have resulted from admixture of genetically substructured Han (Black et al. 2001; Wang et al. 2003).

The allele distributions at each autosomal locus were compared with data from African-Americans and US Caucasians and differed only at TH01, vWA, D13S317, and D16S539 (Fig. 1). A series of previously unreported alleles and rare alleles, which serve to increase the range of genetic variation, were identified in the Han Chinese population (Table 1, Fig. 1). The inclusion of a number of these rare alleles into the ladder marker system and the establishment of a specific Chinese DNA database at such loci could significantly improve the match probabilities in forensic examinations in PR China.

In practice, genetic markers used in paternity testing are required to provide a cumulative power of exclusion of >99%. Figure 2 summarizes the cumulative powers of discrimination and exclusion for the 13 forensic loci. After genotyping approximately six markers, the forensic kits reached a power of exclusion greater than the required 99% minimum limit. The 13 loci thus provide a powerful battery of DNA markers that are appropriate for use in paternity testing and individual identification in Chinese populations.

Acknowledgements The assistance of the Liaoyang Municipal Committee in the organization of the survey is acknowledged with gratitude. The authors especially thank the Han people of Liaoyang City for their participation. Financial support was provided by the Edith Cowan University through a University Infrastructure Grant.

References

- Balding DJ, Nichols RA (1995) A method for quantifying differentiation between populations at multi-allelic loci and its implication for investigating identity and paternity. *Genetica* 96:3–12
- Black M, Wang W, Bittles AH (2001) A genome-based study of the Muslim Hui community and the Han population of Liaoning Province, PR China. *Hum Biol* 73:801–813
- Botstein P, White RL, Skolnick M (1980) Construction of a genetic linkage map in Man using restriction length polymorphism. *Am J Hum Genet* 32:314
- Chinese Family Planning Commission (1998) Chinese Family Planning Yearbook 1998. Family Planning Commission, Beijing, pp 512–513
- DNA Commission of ISFH (1994) DNA recommendations—1994 report concerning further recommendations of the DNA commission of ISFH regarding PCR-based polymorphism in STR (short tandem repeat) systems. *Int J Legal Med* 107:159–160
- Edwards A, Hammond HA, Jin J (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12:241–253
- Guo S, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372
- Hoyle R (1998) The FBI's national DNA database. *Nat Biotech* 16:987
- Jack B (1997) Mass disaster genetics. *Nat Genet* 15:329–331
- Lewis PO, Zaykin D (2000) Genetic data analysis. Version 1.0, <http://chee.unm.edu/gda/>
- Lincoln PJ (1997) Criticisms and concerns regarding DNA profiling. *Forensic Sci Int* 88:23–31
- Martin PD, Schmitter H, Schmitter PM (2001) A brief history of the formation of DNA database in forensic science within Europe. *Forensic Sci Int* 119:225–231
- Odelberg SJ, White R (1990) Repetitive DNA: molecular structure, polymorphism and forensic application. In: Year Book Medical Publishers (eds) DNA and other polymorphisms in forensic science. Year Book Medical Publishers, Chicago, pp 26
- Perkin-Elmer (1998) AmpFl STR Profiler Plus PCR amplification kit: user's manual. AmpFl STR Cofiler PCR amplification kit: user bulletin. Perkin-Elmer Corporation, USA
- Schneider PM (1997) Basic issues in forensic DNA typing. *Forensic Sci Int* 88:17–22
- Schneider PM, Martin PD (2001) Criminal DNA database: the European situation. *Forensic Sci Int* 119:232–238
- Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10:506–518
- Wang W, Sullivan SG, Ahmed S, Chandler D, Zhivotovsky LA, Bittles AH (2000) A genome-based study of consanguinity in three endogamous Pakistan communities. *Ann Hum Genet* 64:41–49
- Wang W, Wise C, Baric T, Black ML, Bittles AH (2003) The origins and genetic structure of three co-resident Chinese Muslim populations: the Salar, Bo'an and Dongxiang. *Hum Genet* (in press)
- Weir BS (1994) The effects of inbreeding on forensic calculations. *Annu Rev Genet* 28:597–621
- Weir BS (1996) Genetic data analysis II. Sinauer Associates, Sunderland, Mass., pp 161–201
- Wertz D (2002) DNA forensics: professional and patient attitudes internationally. HGM 2002 Abstracts, Shanghai, pp 114
- Wrettt DJ (1997) The National DNA database. *Forensic Sci Int* 88:33–42
- Zhivotovsky LA, Ahmed S, Wang W, Bittles AH (2001) Deep genetic differentiation among co-resident endogamous communities and its forensic DNA implications. *Forensic Sci Int* 119:269–272