TECHNICAL NOTE

Li Jin,[1] *Ph.D.; Peter A. Underhill,*[2] *Ph.D.; Martin Buoncristiani,*[2] *Ph.D.; and James M. Robertson,*[2] *Ph.D.*

Defining Microsatellite Alleles by Genotyping Global Indigenous Human Populations and Non-Human Primates

REFERENCE: Jin L, Underhill PA, Buoncristiani M, Robertson JM. Defining microsatellite alleles by genotyping global indigenous human populations and non-human primates. J Forensic Sci 1997;42(3):496–499.

ABSTRACT: Polymorphisms at variable number of tandem repeat (VNTR) loci have been used in forensic science for almost decade. Microsatellite loci, especially tri-, tetra-, and pentanucleotide repeat loci, have shown great potential in personal identification and paternity testing. In this report, we describe results of genotyping three tetranucleotide repeat loci (D5S818, D7S820, D13S317) in 16 worldwide indigenous human populations and one chimpanzee population which were being developed for forensic applications. We demonstrate the utility of typing globally diverse populations in defining microsatellite alleles: Specifically (i) investigating the measurement errors of each allele using semi-automatic genotyping instrumentation and software, (ii) assessing the range of alleles, (iii) understanding the extent of allele frequency differences across worldwide populations, and (iv) identifying possible anomalous alleles with complex structures.

KEYWORDS: forensic science, DNA typing, short tandem repeats, microsatellite, tetranucleotide loci, genetic markers

Polymorphisms at variable number of tandem repeat (VNTR) loci have been used for personal identification in forensic science for almost decade (1–3). Currently in the United States, DNA profiles presented as evidence in the judicial system involve genotyping minisatellite loci (i.e., the loci with the size of repeat unit range from nine to several hundreds base pairs) by Southern hybridization-based detection schemes. Genotyping VNTR loci is, however, not without shortcomings. Inability to make precise allele determinations have aroused controversies in the applications of VNTRs to personal identification (4–12).

Microsatellite loci (or short tandem repeat loci, STR), especially those tri-, tetra-, and pentanucleotide repeat loci, however, have shown great potential in personal identification and paternity testing (13–18). The appeal of microsatellite genotyping is due to the fact that not only can allele sizes at those loci be determined with great precision on polyacrylamide gels, the use of PCR, robotics, and non-radioactive size quantification technology allows automation of the genotyping with a minimal quantity of DNA. Microsatellite loci typing permits state-of-art measurement of fragment sizes and therefore, many problems related to the resolving power of the detecting system are considerably reduced.

A fragment may be assigned to a wrong allele in microsatellite loci typing if there is an anomalous allele which only slightly differs in size due to the insertion or the deletion of even one single nucleotide (16–17). The allele frequencies of such complex anomalous alleles can be high in human populations (16–17). Genotyping worldwide diverse indigenous human populations and non-human primates provides better opportunity to define all alleles for each locus including any anomalous ones. The typing of worldwide indigenous populations will also help to identify alleles which are extremely rare in cosmopolitan populations while being quite abundant in isolated populations due to various genetic and/or demographic reasons.

The present report describes an effort we made to type three tetranucleotide repeat loci (D5S818, D7S820, D13S317) developed for forensic applications in 16 indigenous populations and one Chimpanzee population. The objectives of this project are (i) to investigate the measurement errors of each allele using ABI 373A DNA Sequencer, GeneScan™ 672, and Genotyper™ software; (ii) to assess the range of alleles in world populations; and (iii) to identify possible anomalous alleles with complex structures.

Materials and Methods

Three microsatellite loci (D5S818, D7S820, D13S317, see Table 1 for detailed description) were co-amplified in a single 5-μL volume PCR reaction described in Table 2. Sufficient cocktail (based upon the 1 × recipe) was prepared for the desired number

TABLE 1—*Three microsatellite loci studied.*

Loci	Dye	Primer Sequence (5′ to 3′)
D5S818	JOE	GGG TGA TTT TCC TCT TTG GT TGA TTC CAA TCA TAG CCA CA
D7S820	JOE	TGT CAT AGT TTA GAA CGA ACT AAC G CTG AGG TAT CAA AAA CTC AGA GG
D13S317	JOE	ACA GAA GTC TGG GAT GTG GA GCC CAA AAA GAC AGA CAG AA

[1] Department of Genetics, Stanford University School of Medicine, Stanford, CA.
[2] Applied Biosystem Division, Perkin Elmer, Inc., Foster City, CA.
Received 27 July 1995; and in revised form 19 July 1996 and 19 Sept. 1996; accepted 23 Sept. 1996.

TABLE 2—*Recipe for 5-μL PCR reaction.*

Amount	Reagent	Final Concentration
0.5 μL	10 × Buffer	10 mM Tris-HCL, pH 8.3 1.5 mM $MgCl_2$, 50 mM KCl 0.1 mg/mL gelatin
0.1 μL	25 mM $MgCl_2$	0.5 mM
0.8 μL	1.25 mM dNTP mix	200 μM each dNTP
0.05 μL	Taq polymerase, 5 U/μL	0.25 units
0.1 μL	Each primer, 10 mM	0.2 mM each
1.95 μL	Water	
1.0 μL	Genomic DNA (50 ng/μL)	10 ng/μL

of reactions, 4 μL dispensed per tube and then 1 μL genomic DNA added. A "touchdown" PCR regime (19,20) was run on a Perkin Elmer 9600 thermal cycler to enhance specificity. All cycles had a denaturation at 94°C for 20 s and an extension at 72° for 60 s. During each of the initial 14 cycles, the annealing temperature (60-s duration) was decreased from 63°C by 0.5°C increments per cycle. For the last 20 cycles, the annealing temperature was held constant at 56°C for 60 s. Following amplication, a 1-μL aliquot was combined with 0.5μL ABI GS2500 ROX size standard (p.n. 401100) and 2-μL formamide based loading buffer, incubated at 95–100°C for 2 min and loaded on a standard 6% denaturing sequencing gel. Electrophoresis was conducted at 30 W for 5 h using 24 cm well to read size gels and an ABI 373A sequencer. GeneScan 672 software was used to collect data, track lanes, measure fragment sizes, and to verify the internal size standard assignments. Genotypes were called by Genotyper (v1.0).

Sixteen human populations and one chimpanzee population were typed. The human populations typed were Zaire and the Central African Republic Pygmies (ZAI and CAR), Lisongo Africans (LIS), Chinese (CHI), Japanese (JAP), Cambodian (CAM), Surui and Karitiana Amazonians (SUR and KAR), Maya from the Yucatan peninsula (MAN), North Italian (ITA), North European (NEU), Native Australian (AUS), Melanesian (MEL), New Guinean (NGN) and two linguistically diverse Pakistani groups (PK1 and PK2). The number of individuals typed for each population is presented in Table 4.

Results

Initially, all three loci were amplified and analyzed individually. Subsequently, co-amplification of all three loci in a single PCR reaction tube was achieved using a 'touchdown' PCR regime. We evaluated each locus independently and did not see any difference from those in the multiplexing experiments. All peak topologies were reproducible and readily interpretable. Each of the three loci display distinctive and simple tetranucleotide repeat signatures. However, the non-templated nucleotide addition to the PCR products (21) should be taken into account in the interpretation of the D7S820 locus. Nonetheless, because we were able to confidently interpret all D7S820 alleles, no attempt to alter cycling conditions was made to obtain a better signature for this locus.

The distributions of fragment size for each allele are showed in Tables 3a–c. Both average and standard deviation (in parenthesis) are presented for each allele, along with the count and the range of fragment sizes. The measurement error for these three loci are generally small (<0.55 base pair).

Genotypes were determined in 100%, 97%, and 97% of 180 individuals across 16 human populations at D5S818, D7S820, D13S317, respectively. Chimpanzee individuals cannot be amplified at D7S820. This is either due to nucleotide substitution(s) at priming regions or an absence of the locus in the Chimp genome. Each individual was amplified at most twice or was dropped if the amplification failed in both trials.

TABLE 3a—*Distribution of fragment sizes of D5S818 alleles.*

Allele Designated	Average Size	Count	Smallest Fragment	Largest Fragment
123	122.95 (.05)	2	122.93	122.97
125	124.82 (.00)	1	124.82	124.82
127	126.83 (.10)	17	126.69	126.96
133	132.63 (.08)	9	132.53	132.76
137	136.62 (.02)	7	136.57	136.69
141	140.63 (.12)	15	140.50	140.87
145	144.63 (.05)	46	144.51	144.81
149	148.69 (.00)	108	148.54	148.95
153	152.75 (.16)	100	152.57	152.93
157	156.85 (.00)	47	156.69	156.98
161	160.96 (.05)	8	160.88	161.05

TABLE 3b—*Distribution of fragment sizes of D7S820 alleles.*

Allele Designated	Average Size	Count	Smallest Fragment	Largest Fragment
201	201.40 (.10)	2	201.32	201.47
205	205.40 (.16)	53	205.28	205.65
209	209.41 (.14)	20	209.25	209.54
213	213.38 (.15)	73	213.08	213.63
217	217.45 (.19)	96	217.21	217.63
221	221.48 (.04)	72	221.26	221.64
225	225.97 (.15)	8	225.70	226.12
229	230.22 (.08)	4	230.14	230.29

TABLE 3c—*Distribution of fragment sizes of D13S317 alleles.*

Allele Designated	Average Size	Count	Smallest Fragment	Largest Fragment
174	173.78 (.07)	41	173.71	173.89
178	177.85 (.07)	48	177.77	177.92
182	181.91 (.09)	39	181.80	182.00
186	185.97 (.00)	78	185.90	186.09
190	189.85 (.12)	93	189.70	190.00
194	193.78 (.08)	32	193.69	193.87
198	197.66 (.17)	11	197.40	197.88

The number of alleles, heterozygosity, and the number of chromosomes typed for each population at all three loci are listed in Table 4. High level of heterozygosities are observed in all three loci across 16 human populations included in this study. The allele frequencies are not presented due to the small sample sizes, but are available upon request. Generally, populations from the same continent tend to have similar allele frequency distributions. D5S818 alleles for Chimpanzees differ by increments of two base pair instead of four suggesting that there may be a dinucleotide repeat associated with the locus. No anomalous alleles due to simple insertion or deletion events were detected within the human samples examined.

Discussion

Achieving a balanced fluorescent signal for accurate size analysis is desirable when numerous loci are analyzed in a single gel

TABLE 4—*Observed number of alleles and heterozygosity in worldwide populations at D5S818, D7S820, D13S317.*

Population	D5S818			D7S820			D13S317		
	n_A	H	n	n_A	H	n	n_A	H	n
CAR	6	74.5	20	5	74.0	20	4	67.0	20
LIS	5	75.5	20	5	69.0	20	5	73.5	20
ZAI	3	56.5	20	5	74.7	18	6	78.0	20
CAM	6	76.0	20	4	71.5	20	6	72.0	20
CHI	6	79.0	20	4	71.5	20	6	79.0	20
JAP	5	72.5	20	5	75.5	20	6	77.5	20
KAR	4	62.5	20	3	46.0	20	3	53.5	20
SUR	5	74.0	20	4	41.0	20	5	58.5	20
MYN	6	60.0	20	6	74.1	18	6	79.5	20
ITA	6	69.2	32	6	77.1	28	7	76.9	30
NEU	5	66.4	28	7	79.1	28	7	74.0	28
AUS	6	78.0	20	4	70.5	20	5	70.4	18
MEL	5	66.0	20	4	69.1	18	4	66.1	18
NGN	7	77.5	20	5	66.5	20	6	76.0	20
PK1	4	70.0	20	6	69.0	20	7	77.0	20
PK2	5	64.0	20	5	71.0	18	5	77.3	16
APE*	3	26.5	20	0	0.00	0	5	76.4	12

NOTE:—n_A, H, n are the number of alleles, heterozygosity, and number of chromosomes typed, respectively.

*There are three alleles (123, 125, and 127 bp) in Chimpanzee at D5S818. Chimpanzee cannot be amplified at D7S820.

lane. Experimental protocols should be adjusted such that one locus is not significantly overloaded with respect to any others. We chose to exploit the genotyping efficiency inherent in the instrumentation design by both co-amplifying the three loci in a single PCR reaction and then analyzing all alleles in a single lane. To minimize fluorescent signal difference between loci, only one fluorophore (JOE) was used to label all three loci, with discrimination being based on size alone. Alternative fluorescent labeling and gel loading schemes are possible. Although the most stringent analysis conditions would involve amplifying one locus per PCR reaction and analyzing only one locus per lane, genotyping efficiency would be sacrificed. The techniques described here have sufficient flexibility to accommodate different priorities and objectives.

The presence of anomalous alleles due to the insertion or the deletion of one or two single nucleotides may introduce misidentification and reduce the exclusionary power of microsatellites in forensic applications (16–17). In this study of tetranucleotide repeats, we noticed that D5S818 may have such anomalous alleles because 2-bp differences were detected for some alleles in chimpanzees all of which have never been observed in humans. The use of instrumentation and procedures that can resolve one base pair difference between DNA fragments (e.g., ABI 373A and ABI 377 sequencers) is essential to detect any anomalous alleles. By genotyping worldwide indigenous human populations and non-human primates, one can define the existence of such alleles and characterize their frequencies in human populations.

The local Southern (22) algorithm was used by Genescan 672 software in determining the sizes of DNA fragments. However, the size standard used in this study (ABI GS2500 ROX) may introduce systematical measurement errors to alleles with certain sizes because of an occurrence of split peaks. This is most evident with 225 and 229 base pairs alleles at D7S820. Appropriate precautions must be taken if the fragments with unexpected sizes are observed. The range of fragment sizes for each allele presented in Tables 3a–c can be used in Genotyper to create category groups for three markers studied. The fragment size can be measured more accurately by using an improved fluorescent size standard (GS350) which do not display any split peaks.

One controversial problem, unrelated to the issue of measuring error, in application of DNA evidence in personal identification is the choice of reference populations in the computation of match probability (7,23). It was claimed that the reference populations used may not be suitable or they may be substructured (4,23) which could introduce deviation from Hardy-Weinberg expectations and consequently invalidate the computation of match probability. The National Research Council (7) suggested the concept of 'ceiling principal' as a conservative guide line for match probability estimation. It has been shown that Hardy-Weinberg equilibrium generally holds in cosmopolitan populations as well as indigenous populations at almost all microsatellite loci examined (13–15,24) and the selection of reference populations is not crucial for minisatellite markers (25–26). One rigorous approach to study this issue is to type numerous indigenous populations in addition to cosmopolitan populations for each microsatellite locus proposed for use in forensic application, and to compare allele frequency distributions of those indigenous populations with those of cosmopolitan populations. Unfortunately, our current sample size is insufficient to address this issue.

We would like to indicate that microsatellite loci are also extremely useful in studies of human genome diversity. It has been shown that such loci can provide a general picture of human evolution and migration with high resolution (24,27).

Acknowledgment

Support was provided by a NIH grant GM28428 to Dr. Luca Cavalli-Sforza and by a NIH training grant T32-GS08404 to L. Jin.

References

1. Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. Nature 1985;314:67–73.
2. Jeffreys AJ, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. Nature 1985;316:76–9.
3. Jeffreys AJ, Brookfield JFY, Semeonoff R. Positive identification of an immigration test-case using human DNA fingerprints. Nature 1985;317:818–9.
4. Lander ES. DNA fingerprinting on trial. Nature 1989;339:501–5.
5. Chakraborty R, Kidd KK. The utility of DNA typing in forensic work. Science 1991;254:1735–9.
6. Budowle B, Giusti AM, Waye JS, Baechtel FS, Fourney RM, Adams DE, et al. Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR.
7. National Research Council. DNA technology in forensic science. National Academy Press, Washington, DC, 1992.
8. Devlin B, Risch N, Roeker K. No excess of homozygosity at loci used for DNA fingerprinting. Science 1990;249:1416–20.
9. Chakraborty R, de Andrade M, Daiger SP, Budowle B. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Ann Hum Genet 1992; 56:45–57.
10. Chakraborty R, Zhong Y, Jin L, Budowle B. Non-detectability of restriction fragments explain the observed departures from independence of DNA-fragment sizes within loci in the current forensic databases. Am J Hum Genet 1994;55:391–401.
11. Tully G, Sullivan KM, Gill P. Analysis of 6 VNTR loci by 'multiplex' PCR and automated fluorescent detection. Hum Genet 1993;92:554–62.
12. Chakraborty R, Jin L. Heterozygote deficiency, population substructure and their implications in DNA fingerprinting. Hum Genet 1992;88:267–72.
13. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 1991;49:746–56.
14. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 1992;12:241–53.

15. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 STR loci for use in personal identification applications. Am J Hum Genet 1994;55:175–89.
16. Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW. Identification of microsatellite heterogeneity at the polymorphic STR locus. HUMTH01 [AATC] n and re-assignment of alleles in population analysis using a locus-specific allelic ladder. Am J Hum Genet 1993;53:953–8.
17. Puers C, Hammond HA, Caskey CT, Lins AM, Sprecher CJ, Brinkmann B, et al. Allelic ladder characterization of the short tandem repeat polymorphism located in the 5′ flanking region to the human coagulation factor XIII A subunit gene. Genomics 1994;23:260–4.
18. Alford RL, Hammond HA, Coto I, Caskey CT. Rapid and efficient resolution of parentage by amplification of short tandem repeats. Am J Hum Genet 1994;55:190–5.
19. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. "Touchdown" PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 1991;19:4008.
20. Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, et al. Construction of human Y-chromosome haplotypes using a new polymorphic A to G transition. Hum Mol Genet 1994;12:2159–61.
21. Clark JM. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymoerases. Nucleic Acids Res 1988;16:9677–86.
22. Elder JK, Southern EM. Computer-aided analysis of one-dimensional restriction fragment gels. In: Bishop MJ, Rawlins CJ, editors. Nucleic acid and protein sequence analysis. Oxford: IRL Press, 1987;165–72.
23. Lewontin RC, Hartl DL. Population genetics in forensic DNA typing. Science 1991;254:1745–50.
24. Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser J, et al. Population genetics of dinucleotide (dC-dA)n(dG-dT)n polymorphisms in world populations. Am J Hum Genet 56:1995;461–74.
25. Chakraborty R, Jin L, Zhong Y, Srinivasan MR, Budowle B. On allele frequency computation from DNA typing data. Int J Legal Med 1993;106:103–6.
26. Chakraborty R, Jin L, Deka R. Intra- and inter-population variations at short tandem repeats, polymarkers, VNTR loci and their implications in forensic and parentage analysis. Proceedings from The Fifth International Symposium on Human Identification 1995. In press.
27. Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, CavalliSforza LL. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 1994;368:455–7.

Additional information and reprint requests:
Dr. Li Jin
Human Genetics Center
University of Texas—Houston
School of Public Health
P.O. Box 20334
Houston, TX 77225