TECHNICAL NOTE

Danielle Desmarais,[1] Ph.D.; Yixi Zhong,[2] M.D.; Ranajit Chakraborty,[2] Ph.D.; Claude Perreault,[3] M.D.; and Lambert Busque,[1,3] M.D.

Development of a Highly Polymorphic STR Marker for Identity Testing Purposes at the Human Androgen Receptor Gene (HUMARA)

REFERENCE: Desmarais D, Zhong Y, Chakraborty R, Perreault C, Busque L. Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). J Forensic Sci 1998;43(5):1046–1049.

ABSTRACT: We developed a non-isotopic method which improves the technical quality of the X-linked HUMARA locus typing process. The use of formamide and a low concentration of acrylamide increased resolution and sharpness of HUMARA alleles in silver-stained polyacrylamide gels. In addition, the construction of an allelic ladder containing amplified sequence of 9 alleles (even-numbered alleles) of the HUMARA locus, allows confident, rapid and precise assigment of discretely defined alleles. Allele and genotype frequencies for the HUMARA locus were determined in a French Canadian population sample. Observed genotype frequencies in females conformed to Hardy-Weinberg expectations. Furthermore, the HUMARA locus is highly polymorphic with 18 observed alleles and an heterozygosity value of 89.3%. Also, this locus has average powers of discrimination of 97.8% and 88.7% for testing samples of female and male origin, respectively. In the French Canadian population, the average probability of excluding a random man as the father in paternity analysis when both mother and daughter are tested for this locus is 88.0%. Together, the results indicate that the HUMARA locus provides a highly discriminatory system that is appropriate for the purposes of forensic identification and paternity testing involving a female child.

KEYWORDS: forensic science, paternity testing, HUMARA, short tandem repeat, electrophoretic resolution, allele frequencies, French Canadian

Recently a large number of short tandem repeat (STR) systems have been made available for identity testing as performed in the fields of paternity determinations, forensic sciences and medical diagnosis (1–3). Nevertheless, most of them are less polymorphic than VNTR loci (4–6). Although this disadvantage is consequently reduced by the analysis of more loci, there is a need to look for highly polymorphic STR loci that will allow an increase in the precision of genotyping done by STR analysis.

In this study, we report the development of a simple non-isotopic method which allows confident, rapid and precise assigment of discretely defined alleles of a highly polymorphic STR locus located within the coding region of exon 1 of the human androgen receptor gene (HUMARA). The HUMARA locus consists of CAG repeat unit and is located at chromosome Xcen-q13 (2,7). We also demonstrate, via a French Canadian population study, that the HUMARA locus is very informative for forensic identification of individuals and for paternity determinations involving female children.

Materials and Methods

Sample Preparation—Genomic DNA was isolated from whole blood from 358 unrelated healthy individuals as previously described (8).

HUMARA Typing—For population study, DNA amplification was achieved using the primers described by Edwards et al. (7) and performed in 25 μL reaction volumes containing 25 to 50 ng template DNA; 10 mM Tris-HCL, pH 8.3; 50 mM KCl; 1.5 mM $MgCl_2$; 4% DMSO; 200 μM of each of the four dNTP; 12.5 pmoles of each primer, and 0.5 unit of Amplitaq DNA polymerase (Perkin Elmer-Cetus). For sensitivity studies, K562 cell line DNA (Promega Corporation, Madison, WI) was serially diluted down to 3 pg. The PCR was carried out in a Perkin Elmer 9600 Thermal Cycler and subject to 30 cycles of denaturation at 94°C for 45 s, primer annealing at 60°C for 30 s, and primer extension at 72°C for 30 s with an initial denaturation step at 94°C for 3 min. PCR products were mixed with 15 μL of formamide loading solution (95% formamide; 10 mM EDTA; 0.05% bromophenol blue; 0.05% xylene cyanole), denatured at 94°C for 3 min and chilled on ice prior to loading. Six microliters of each sample were electrophoresed through a 0.4 mm-thick 4% denaturing polyacrylamide gel (19:1 ratio of acrylamide to bis-acrylamide) which contained 7 M urea, 32% formamide, and 1X TBE (0.089 M Tris Base; 0.089 M Boric acid; 2 mM EDTA) onto a vertical PAGE system (SA 32

[1] PRO-ADN Diagnostic, Polyservices Maisonneuve-Rosemont, Montreal, Quebec, Canada.
[2] Human Genetics Center, University of Texas, Houston, TX.
[3] Centre de Recherche Guy Bernier, Hopital Maisonneuve-Rosemont, Montreal, Quebec, Canada.

Received 22 Sept. 1997; and in revised form 27 Jan. 1998; accepted 29 Jan. 1998.

apparatus, GIBCO BRL, Gaithersburg, MD). All gels were chemically cross-linked to one glass plate (Bind Silane, Pharmacia, Uppsala, Sweden). Electrophoresis was performed at 70 watts in 1X TBE for approximately 2.5 h. Following electrophoresis, the gels were silver stained. Briefly, the gel was successively fixed in 10% glacial acetic acid for 20 min, washed in deionized water for 6 min and silver stained in 0.01% $AgNO_3$ and 0.056% formaldehyde solution for 40 min. Then, the gel was rapidly rinsed in deionized water and reduced in a solution containing 0.28 M sodium carbonate, 0.056% formaldehyde, and 2 μg/mL sodium thiosulfate. When the desired image intensity had developed, reduction was stopped with 10% glacial acetic acid. Allele designations were determined by comparison of the base pair sizes of the sample fragments with those of the allelic ladder. The alleles were numbered according to the number of tandem repeats present in the amplicons.

Allelic Ladder—Several samples of French Canadian Caucasian males were typed at the HUMARA locus. Four relevant HUMARA alleles from the haploid male DNA were sequenced to determine the number of repeated CAG motifs in each allele. Briefly, the PCR products were subcloned in pUC18 and sequenced by the dideoxy-chain-termination method using the Sequenase kit (Sequenase 2.0, United States Biochemicals, Cleveland, OH) according to the manufacturer's specifications. Then, these four alleles containing 15, 18, 26 and 31 repeats were used as reference for assessing allele size of several DNA samples. To construct the allelic ladder, consisting of nine even-numbered HUMARA alleles (14, 16, 18, 20, 22, 24, 26, 28 and 30), relevant chromosomal DNA were chosen and amplifyed in 100 μL PCR reaction volumes. Subsequently, all the PCR products were mixed, purified by phenol/chloroform extraction, precipitated with 95% ethanol and resuspended in TE (10 mM Tris-HCL pH 8.0; 1 mM EDTA).

Statistical Analysis—Gene count estimates of allele frequencies, unbiased estimates of heterozygosity, and their standard errors were calculated as described by Edwards et al. (7). Comparison of allele frequencies in males and females were performed by the exact test of a R × C contingency table analysis (9). Tests (applicable to data on females) for Hardy-Weinberg expectations (HWE) are based on four test criteria: (*a*) The chi-square test based on total heterozygosity; (*b*) the likelihood ratio test criterion (10); (*c*) the exact test (11); and (*d*) the intraclass correlation (12). For each test, the levels of significance were determined empirically by permutating all alleles across the individuals samples with the algorithm used by Chakraborty et al. (12). Pairwise independence of genotype frequencies for seven STR loci was performed by computing interclass correlation of allele sizes (13) as well as by performing the test of independence based on the 2 × 2 chi-square test as described by Hammond et al. (2). An approximate test of random association of alleles at different loci (gametic phase equilibrium) was performed considering the distribution of the number of heterozygous loci (14,15) as described by Edwards et al. (7). Locus-specific fixation indice within individuals (i.e., kinship coefficient) was computed by the maximum-likelihood method (16). The potential usefulness of the HUMARA locus in identity testing and paternity analysis was assessed by calculating the average power of discrimination (PD) and the expected probability of exclusion (PE).

The PD for the HUMARA locus is given by:

1. For a sample of female origin, $PD = 1 - 2(\sum p_i^2)^2 + \sum p_i^4$
2. For a sample of male origin, $PD = 1 - \sum p_i^2$

The PE (for parentage testing of a girl) was calculated using the formula:

1. When both mother and child are tested, $PE = 1 - \sum p_i^2 + \sum p_i^4 - (\sum p_i^2)^2$
2. When only the child is tested, $PE = 1 - 2(\sum p_i^2) + \sum p_i^3$

Results and Discussion

Improvement of HUMARA Allele Resolution

As described by Edwards et al. (7) HUMARA alleles appear as widely spaced doublets in 6% acrylamide denaturing (7 M urea) gel, such that adjacent alleles overlap, making allele designation ambiguous (Fig. 1*A*). This was due to the complementary DNA strands of each allele displaying different electrophoretic mobilities. To minimize confusion caused by DNA strand separation in silver-stained gels, we modified the electrophoretic conditions using both non-denaturing and denaturing (5.6 M, 7 M and 7.5 M urea) polyacrylamide gels where total acrylamide concentrations ranged from 4 to 10%. Also, we tested four oligonucleotide primers giving four possible combinations in PCR. The smaller amplification products ranged between 148 and 208 bp whereas the longer amplification products ranged between 255 and 315 bp. Although in general the primer set described by Edwards et al. (7) gave the much closer spaced doublet bands, the results were globally unsatisfactory. However, adding 32% formamide to a 4% denaturing gel (7 M urea) had advantageous effects on the resolution and sharpness of the DNA fragments and improved the technical quality of the HUMARA locus typing process (Fig. 1*B*). Under these conditions, alleles appear as closely spaced doublet bands that allow precise allele designation (compare Fig. 1*A* and 1*B*). Also, the intensity of the stutter bands (principally due to slipped strand mispairing during PCR) is relatively low compared with the true allele product and did not compromise typing.

Allelic Ladder

In order to determine the polymorphism of the X-linked HUMARA locus and investigate parentage and forensic efficiency of this locus, we developed an allelic ladder as size markers for

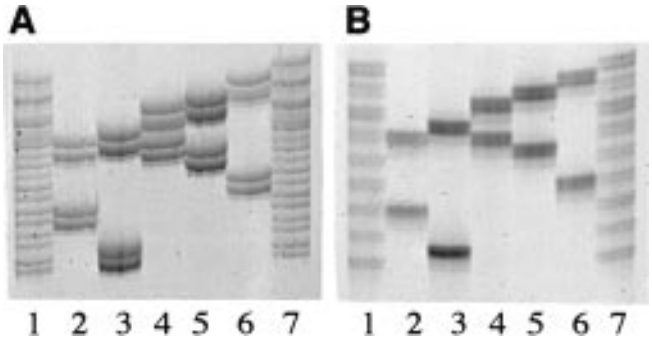

FIG. 1—*Comparative study of the HUMARA typing process. Examples of HUMARA genotypes obtained with silver stained polyacrylamide gels. (A) 6% (19:1) acrylamide: bis-acrylamide, 7 M urea. (B) 4% (19:1) acrylamide: bis-acrylamide, 7 M urea, 32% formamide. Lanes 1 and 7, allelic ladder containing even-numbered alleles (14, 16, 18, 20, 22, 24, 26, 28 and 30). Lanes 2 to 6, amplification of DNA from five different women. The HUMARA types from left-to-right are: 18–24; 15–25; 24–27; 23–28; 20–29.*

genotyping. The HUMARA allelic ladder contains nine alleles of 14, 16, 18, 20, 22, 24, 26, 28 and 30 repeat units, respectively. Since individual alleles that differ by only one repeat unit are juxtaposed, the odd-numbered alleles were not included in the allelic ladder, creating a gap that simplifies identification of the alleles (Fig. 1*B*). This allelic ladder allows confident, rapid and precise assignment of discretely defined alleles.

HUMARA Locus Genotype

The allelic ladder was used to evaluate 358 DNA samples (570 chromosomes) from unrelated French Canadian Caucasians of Quebec. A representative example of the data is shown in Fig. 1*B*. Table 1 presents gene count estimates of allele frequencies from genotype data on 212 females and simple counts from 146 males, along with their standard errors. Comparison of allele frequencies in males and females by the exact test of a R $\times$ C contingency table analysis showed that the HUMARA allele frequencies are not significantly different in the female and male gene pool ($\chi^2 = 11.97$, $p = 0.846$ in 10,000 replications of permutations). Therefore, Table 1 also presents the allele frequencies in the pooled French Canadian sample. Analysis of allele and genotype frequency data at this locus were also presented by Edwards et al. (7) for four major human populations from a sample in Houston, Texas. A quantitative comparison of allele frequencies in the gene pool of French Canadians with the American Caucasians from Houston indicates that allele frequency distributions in these two samples are very similar (data not shown).

Tests of HWE and Independence Across Loci

Genotype data from females of the present sample were subjected to four different test procedures to examine whether the genotype frequencies at this X-linked locus conform to Hardy-Weinberg expectations. The results of these analyses are summarized in Table 2. Clearly, there is no evidence of departure of independence of alleles at the HUMARA locus in French Canadians. Even though a priori, an X-linked locus should segregate independently from all autosomal loci, it is worthwhile to examine if there is any evidence of dependence of genotype frequencies at this locus with any of the six autosomal loci already examined for this population sample (8). The tests of independence of alleles at each of these six loci with the HUMARA locus were performed by computing interclass correlation of allele sizes as well as by performing the test of independence based on the 2 $\times$ 2 chi-square test (Table 3). With the single exception of the ARA-F13A01 locus interclass correlation value ($r = -0.095$, $p = 0.044$) showing marginally significant result, there is no evidence of allelic dependence of this X-linked locus with any of the six autosomal ones in this population (Table 3). We ascribe this single exception to chance occurrence for three reasons. First, the major reason for an X-linked locus to be dependent with an autosomal locus is that the population could be of heterogenous origin. In this case, positive interclass correlation would have been expected (17,18). Second, the chi-square test results do not show any evidence of dependence of alleles at this locus with any of the autosomal ones. Third, based on seven locus genotype data on 113 females, the test for global disequilibrium shows that the observed variance of the number of heterozygous loci in individuals is 1.478 with 95% confidence interval of 0.910–1.534 (based on the assumption of gametic phase equilibria across loci). This implies that there is no evidence of gametic phase disequilibrium among pairs of these loci. These informative validations indicated that the HUMARA locus and the six autosomal STR loci (8) can be advantageously combined according to the current practice as suggested by the National Research Council (19).

Forensic Identification and Parentage Testing Evaluations

The X-linked HUMARA locus exhibited large number of alleles ($k = 18$), high heterozygosity value of 89.3% $\pm$ 4.5% (in females) and insignificant kinship value ($\alpha = 0.016 \pm 0.017$) which make this locus very informative for identity testing purposes. Also, sensitivity studies demonstrated that standard PCR yielded typeable results with as little as 100 pg of DNA. In addition, from the

TABLE 1—*HUMARA alleles frequencies and their standard errors (S.E.) in the French Canadian Caucasian population.*

Allele Sizes	Allele Frequencies ($\pm$ S.E.) %		
	Females	Males	Pooled
14	0.47 $\pm$ 0.33	0.69 $\pm$ 0.68	0.53 $\pm$ 0.30
15	0.24 $\pm$ 0.24	0.69 $\pm$ 0.68	0.35 $\pm$ 0.25
16	0.47 $\pm$ 0.33	0.69 $\pm$ 0.68	0.53 $\pm$ 0.30
17	1.42 $\pm$ 0.57	2.06 $\pm$ 1.17	1.58 $\pm$ 0.52
18	5.43 $\pm$ 1.10	8.22 $\pm$ 2.27	6.14 $\pm$ 1.01
19	9.67 $\pm$ 1.44	11.64 $\pm$ 2.65	10.18 $\pm$ 1.27
20	14.15 $\pm$ 1.69	15.07 $\pm$ 2.96	14.39 $\pm$ 1.47
21	17.45 $\pm$ 1.84	17.12 $\pm$ 3.12	17.37 $\pm$ 1.59
22	12.97 $\pm$ 1.63	10.96 $\pm$ 2.59	12.46 $\pm$ 1.38
23	12.03 $\pm$ 1.58	11.64 $\pm$ 2.65	11.93 $\pm$ 1.36
24	8.96 $\pm$ 1.39	8.90 $\pm$ 2.36	8.95 $\pm$ 1.20
25	7.31 $\pm$ 1.26	6.85 $\pm$ 2.09	7.19 $\pm$ 1.08
26	3.07 $\pm$ 0.84	4.11 $\pm$ 1.64	3.33 $\pm$ 0.75
27	2.12 $\pm$ 0.70	0.69 $\pm$ 0.68	1.75 $\pm$ 0.55
28	1.65 $\pm$ 0.62	0.00 $\pm$ 0.00	1.23 $\pm$ 0.46
29	1.18 $\pm$ 0.52	0.00 $\pm$ 0.00	0.88 $\pm$ 0.39
30	1.18 $\pm$ 0.52	0.00 $\pm$ 0.00	0.88 $\pm$ 0.39
31	0.24 $\pm$ 0.24	0.69 $\pm$ 0.68	0.35 $\pm$ 0.25
Total*	424	146	570

* Number of chromosomes typed.

TABLE 2—*Test of Hardy-Weinberg expectations of genotype frequencies at the X-linked HUMARA locus in 212 French Canadian females of Quebec.*

Test Procedure	Test Statistic	Level of Significance
Heterozygosity	$\chi^2 = 1.33$	0.243
Likelihood Ratio Test	G2 = 114.35	0.362
Exact Test	...	0.412
Intraclass Correlation Test	$r = 0.043$	0.568

TABLE 3—*Test of independence of alleles at HUMARA locus versus six autosomal STR loci in French Canadians.*

Locus-pairs	Interclass Correlation			Chi-square Test	
	n	*r*	*p*	χ^2	*p*
ARA-CSF1PO	147	−0.037	0.441	0.154	0.735
ARA-TPOX	147	−0.044	0.340	0.487	0.572
ARA-TH01	147	−0.012	0.786	2.553	0.116
ARA-F13A01	147	−0.095	0.044*	1.534	0.225
ARA-FESFPS	113	−0.036	0.531	0.174	0.721
ARA-vWA	147	0.057	0.220	0.181	0.751

n = Number of individuals.
* Significant with $p < 0.05$.

observed and expected genotype distributions, the most common genotype at this locus (20,21) is expected to occur at a reasonably low frequency (below 5%) in French Canadians. Moreover, the average discriminatory powers at this locus are 97.8% (0.022) and 88.7% (0.113) for testing samples of female and male origin, respectively. Then, the HUMARA locus could be a more powerful tool than the highly polymorphic VNTR D1S80 locus which is widely used in forensic and paternity investigations (20–23). Even though the D1S80 locus has a high discrimination potential of about 90%, there will be times when the significance of a match will not be as informative as desired because two alleles were found to be relatively common in Caucasians having frequencies of about 0.25 and 0.36, respectively (21–23). In contrast, the most frequent HUMARA alleles have frequencies of 0.17 and 0.15, respectively. Also, the X-linked HUMARA locus appears especially useful in parentage testing of female child and will be most valuable in situations were autosomal polymorphism have limited power. Since the father's genotype at a X-linked locus specifies her daughter's genotype at that locus, exclusion power is increased over autosomal loci. The average exclusion power when both mother and daughter are tested for the HUMARA locus is 88.0% and the expected chance of excluding a random man from the French Canadian population when testing on a female child alone is 79.5%. Addition of the HUMARA locus to a battery of six autosomal STR loci—CSF1PO, TPOX, TH01, F13A01, FESFPS, and vWA (8)—yields a combined exclusion efficiency of about 99.85% when the mother and the female child are typed. Thus, the HUMARA locus will be a valuable addition to the arsenal of tests available for personal identification issues in forensic science and parentage testing.

Conclusions

We provide evidence for the applicability of the X-linked trimeric HUMARA STR loci for routine analysis used in forensic and paternity testing. First, we developed a high-resolution approach which improves the technical quality of the HUMARA locus typing process. Second, the construction, characterization, and application of the allelic ladder for the HUMARA locus allowed confident, rapid and precise assigment of discretely defined alleles. Third, as calculated using the French Canadian Caucasian database, the HUMARA locus has high discriminatory powers of 97.8% and 88.7% for testing samples of female and male origin, respectively.

Acknowledgment

The authors are grateful to Mr. Robert Mio for revision of the manuscript.

References

1. Alford R, Hammond H, Coto I, Caskey C. Rapid and efficient resolution of parentage by amplification of short tandem repeats. Am J Hum Genet 1994;55:190–5.
2. Hammond HA, Jin L, Zhong Y, Caskey T, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 1994;55:175–89.
3. Frankel W, Chan A, Corringham RET, Shepherd S, Rearden A, Wang-Rodriguez J. Detection of chimerism and early engraftment after allogeneic peripheral blood stem cell or bone marrow transplantation by short tandem repeats. Am J Hematol 1996;52:281–7.
4. Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 1987;235:1616–22.
5. Balazs I, Baird M, Clyne M, Meade E. Human population genetic studies of five hypervariable DNA loci. Am J Hum Genet 1989; 44:182–90.
6. Smith JC, Anwar R, Riley J, Jenner D, Markham AF, Jeffrey AJ. Highly polymorphic minisatellite sequences: alleles frequencies and mutation rates for five locus-specific probes in a Caucasian population. J Forensic Sci 1990;30:19–32.
7. Edwards A, Hammond HA, Jin L, Casey T, Chakraborty R. Genetic variation at five trimeric and tetrameric repeat loci in four human population groups. Genomics 1992;12:241–53.
8. Busque L, Desmarais D, Provost S, Schumm JW, Zhong Y, Chakraborty R. Analysis of allele distribution for six short tandem repeat loci in the French Canadian population of Quebec. J Forensic Sci 1997;42:1145–51.
9. Roff DA, Bentzen P. The statistical analysis of mitochondrial DNA polymorphisms: chi 2 and the problem of small samples. Mol Biol Evolution 1989;6:539–45.
10. Weir BS, Genetic data analysis-II, Sunderland, MA: Sinauer, 1996.
11. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 1992;48:361–72.
12. Chakraborty R, Srinivasan MR, DeAndrade M. Intraclass and interclass correlations of allele sizes within and between loci in DNA typing data. Genetics 1993;133:411–9.
13. Karlin S, Cameron EC, Williams PT. Sibling and parent offspring correlation estimation with variable family size. Proc Natl Acad Sci 1981;78:2664–8.
14. Brown AHD, Feldman MW, Nevo E. Multilocus structure of natural populations of Hordeum spontaneum. Genetics 1980;96: 523–36.
15. Chakraborty R. Detection of nonrandom association of alleles from the distribution of the number of heterozygous loci in a sample. Genetics 1984;108:719–31.
16. Yasuda N. Estimation of the inbreeding coefficient from phenotype frequencies by a method of maximum likelihood scoring. Biometrics 1968;24:915–35.
17. Weir BS. Independence tests for VNTR alleles defined by quantile bins. Am J Hum Genet 1993;53:1107–13.
18. Chakraborty R, Zhong Y, Budowle B. Nondetectability of restriction fragments and independence of DNA fragment sizes within and between loci in RFLP typing of DNA. Am J Hum Genet 1994; 55:391–401.
19. National Research Council. The evaluation of forensic DNA evidence. National Research Council, National Academy Press, Washington, DC, 1996.
20. Hochmeister MN, Budowle B, Jung J, Borer UV, Comey CT, Dirnhofer R. PCR-based typing of DNA extracted from cigarette butts. Int J Legal Med 1991;104:229–33.
21. Sajantila A, Budowle B, Strom M, Johnsson V, Lukka M, Peltonen L, et al. PCR amplification of alleles at the D1S80 locus: Comparison of a Finnish and a North American Caucasian population sample, and forensic casework evaluation. Am J Hum Genet 1992;50: 816–25.
22. Deka R, DeCroo S, Jin L, McGravey ST, Rothhammer F, Ferrell RE, et al. Population genetic characteristic of the D1S80 locus in seven human populations. Hum Genet 1994;94:252–8.
23. Budowle B, Baechtel FS, Smerick JB, Presly KW, Giusti AM, Parsons G, et al. D1S80 population data in African Americans, Caucasians, Southeastern Hispanics, Southwestern Hispanics, and Oriental. J Forensic Sci 1995;40:38–44.

Additional information and reprint requests:
Lambert Busque MD, FRCPC
Centre de Recherche Guy Bernier, Hôpital Maisonneuve-Rosemont
5415 boul. de l'Assomption, Montréal, Québec, Canada, H1T 2M4