TECHNICAL NOTE

Lambert Busque,[1] M.D.; Danielle Desmarais,[1] Ph.D., Sylvie Provost,[1] M.Sc.; James W. Schumm,[2] Ph.D.; Yixi Zhong,[3] M.D.; and Ranajit Chakraborty,[3] Ph.D.

Analysis of Allele Distribution for Six Short Tandem Repeat Loci in the French Canadian Population of Québec

REFERENCE: 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 Québec. J Forensic Sci 1997;42(6):1147–1153.

ABSTRACT: Short tandem repeat (STR) loci represent a rich source of highly polymorphic markers in the human genome which are useful for the purposes of forensic identification and determination of biological relatedness of individuals. Here, as a part of an ongoing extensive study, we report the analysis of a multilocus genotype survey of 642 to 870 chromosomes in the French Canadian Caucasian population of Québec at six STR loci. The loci HUMCSF1PO, HUMTPOX, HUMTH01, HUMF13A01, HUMFESFPS, and HUMvWA were typed using two multiplex polymerase chain reactions (PCR). Amplified DNA samples were subsequently analyzed by polyacrylamide gel electrophoresis followed by silver staining. The heterozygote frequencies of the loci range from 0.614 to 0.820 (0.661 to 0.818 expected) and the number of alleles from 7 to 12 per locus. Although statistically significant deviation from Hardy-Weinberg expectations of genotype frequencies was noted at some loci by one or more tests, in general, the genotype frequencies are well estimated from the product of allele frequencies at all loci. The most frequent six-locus genotype is expected to occur in the French Canadian population with a frequency of 3.50 by 10^{-5} and together, these six loci have an average probability of discrimination of 0.9999985. The study presented here indicates that these six STR loci are informative genetic markers for identity testing purposes in the French Canadian Caucasian population of Québec.

KEYWORDS: forensic science, STR, CSF1PO, TPOX, TH01, F13A01, FESFPS, vWA, PCR, allele frequency, population database, French Canadian, paternity testing

Tandemly repeated DNA sequences in the human genome are a source of genetic markers useful for genetic studies, medical diagnosis, paternity testing, and personal identification in forensic science. A subgroup of the highly polymorphic variable number of tandem repeat (VNTR) loci is the short tandem repeat (STR) loci (1–4). STR alleles which are generally less than 350 bp in length, can be successfully amplified by polymerase chain reaction (PCR; (5)) whereas VNTR loci are generally typed through restriction fragment length polymorphism (RFLP) analysis (6). Analysis of STR polymorphisms by PCR-based method offers certain advantages over RFLP typing. First, STR loci can be typed with a high degree of specificity and sensitivity in a relative short time period. Second, typing of multiple loci can be accomplished in a single multiplex reaction (7). Third, because STR loci are smaller in size, discrete allele resolution can be obtained. Four, these loci can also be successfully amplified from limited amount of DNA even when it is partially degraded (8–10). Furthermore, STR loci are more easily amenable to automation and high throughput analysis than is possible for the VNTR typed by RFLP (11).

Nevertheless, the use of genetic markers, such as STR, in identity testing must accompany allele/genotype frequency data from relevant populations for estimating the frequency of any particular genetic profile. As a part of ongoing surveys in several ethnically-defined populations, this paper reports data on discrete allelic variation at six STR loci in the French Canadian Caucasian population of Québec. The French Canadian population of Québec originates mainly from some 8500 founders who settled in new France between 1608 and 1759 (12). After the British conquest of 1759, this population remained quite isolated and grew very quickly by natural increase (13). This population demonstrated some genetic founder effects, manifested by an increased prevalence of several genetic disorders (14–16). Thus, it was of interest to evaluate if founder effects could affect genetic equilibrium at any of the STR loci currently being used for identity testing and paternity analysis.

We have analyzed the genotype and allele frequency distribution and investigated the validity of assumptions of allelic independence within and among the loci HUMSCF1PO, HUMTPOX, HUMTH01, HUMF13A01, HUMFESFPS, and HUMvWA in the French Canadian Caucasian population of Québec. The analyses presented here indicate that the six STR loci are highly polymorphic, their genotype frequencies can be adequately estimated by the Hardy-Weinberg predictions without major systematic bias, and there is little evidence of gametic phase disequilibrium among loci. These, in turn demonstrate that these six loci provide valuable potential for personal identification and parentage testing in populations with the unique history of the French Canadian Caucasians of Québec.

Materials and Methods

Sample Preparation

DNA was extracted from whole blood samples collected from unrelated and healthy French Canadian Caucasian individuals.

[1]Centre de Recherche Guy Bernier, Hôpital Maisonneuve-Rosemont, Montréal, Québec, Canada.
[2]Promega Corporation, Madison, WI.
[3]Human Genetics Center, University of Texas, Houston, TX.
Received 26 Nov. 1996; and in revised form 12 Feb. 1997; accepted 14 Feb. 1997.

Briefly, 5 mL of peripheral blood was lysed with 40 mL of lysis solution (0.32 M sucrose; 10 mM Tris pH 7.5; 5 mM $MgCl_2$; 1% Triton X-100). The cells were then resuspended in 2 mL $NaCl_2$ buffer (75 mM $NaCl_2$; 25 mM EDTA) and digested with 150 μL of proteinase K solution (2 mg/mL proteinase K; 5% SDS) for 24 h at 37°C. The DNA was subsequently extracted with phenol, phenol/chloroform and chloroform, and precipitated with $^1/_{10}$ volume of 3M sodium acetate and 2 volumes of absolute ethanol. DNA was resuspended in TE buffer (10 mM Tris pH 7.6; 1 mM EDTA). A total of 642 to 870 chromosomes were analyzed for each locus.

Typing

The loci studied, chromosomal assignments, and range of product sizes are given in Table 1. Genotyping was achieved using the GenePrint STR systems CTT triplex (CSF1PO, TPOX, TH01) and FFv triplex (F13A01, FESFPS, vWA) with 50–100 ng of genomic DNA and Amplitaq enzyme (Perkin-Elmer) in 25 μL reaction volumes according to the manufacturer's recommendations (Promega Corporation, Madison, WI). PCR was carried out in a Perkin-Elmer Cetus GeneAmp PCR system 9600 thermocycler following the recommended amplification parameters. Control reactions without added DNA were included in every set of amplification to monitor contamination artifacts. Amplified DNA samples were diluted one fold in formamide loading buffer (10 mM EDTA; 95% formamide; 0.05% bromophenol bleu; 0.05% xylene cyanole), denatured at 94°C for 5 min and then, 6 μL were loaded onto a 5% acrylamide:bis-acrylamide (38:2) denaturing (7M urea) gel. Allelic ladder (supplied by Promega Corporation) was run at frequent intervals across each gel. The gels were subsequently silver stained with the Promega's DNA Silver Staining System. Allele designations were determined by comparison of the base pair sizes of the sample fragments with those of the allelic ladders. The alleles are numbered according to the number of tandem repeats present in the amplification products. Microvariants of alleles (produced by imperfect repeats) were also detectable at two loci (TH01 and F13A01) by this allele scoring method.

Statistical Approaches

Since the above genotyping protocol detects codominant alleles at each of the autosomal loci, gene count estimates of allele frequencies with their respective standard errors were obtained by the method given in Li (17), as used for other STR loci (2,3,18). Unbiased estimate of expected heterozygosity was calculated according to Nei (19). Tests for Hardy-Weinberg expectations (HWE) are based on four test criteria: (1) The chi-square test based on total heterozygosity; (2) the likelihood ratio test criterion (20); (3) the exact test (21); and (4) the intraclass correlation test (22). For each test, the levels of significance were determined empirically by permuting all alleles across the individuals sampled with the algorithm used in Chakraborty et al. (22). For the loci that showed statistically significant departure from HWE, frequencies of observed genotypes were plotted against the expected genotypes to examine if any of the genotype frequencies are grossly underestimated when they are estimated by using the product rule. Pairwise independence of genotype frequencies was tested for detecting potential disequilibrium between loci (23,24). Independence across all six loci was also performed using the observed variance (s_k^2) of the number of heterozygous loci (25). An alternative test of mutual independence of six loci was performed by evaluating the distribution of allele sharing in multilocus genotypes of all possible pairs of individuals (in all 50,086 = 317 by 316/2 comparisons) in the sample and comparing it with the one expected under independence (26). In addition, locus-specific fixation indices within individuals (i.e., kinship coefficient) were computed by the maximum likelihood method (27). The minimum allele frequency estimates are given following the method described in Budowle et al. (28) and Weir (23). The variance of allele sizes in units of repeat lengths was determined as described by Kimmel et al. (29). The potential usefulness of the six studied loci in identity testing and paternity analysis was examined by computing the average probability of discrimination (30) and probability of paternity exclusion (31). In addition, the most common genotype frequencies were also computed for each locus, and the probability of paternity exclusion was determined for the most likely mother-child combination at each locus. These computations were done with the assumption of HWE at each locus and linkage equilibrium among all loci.

Results

Allele Frequency Estimates

Table 2 presents the gene count estimates of allele frequencies along with their standard errors at each locus. While DNA was extracted from 435 French Canadian Caucasians of Québec, the number of chromosome typed varied from locus to locus (Table 2). All alleles differed in size by one repeat unit for all loci, except for the HUMTH01 allele 9.3 which is one base pair smaller than allele 10 (32) and for the HUMF13A01 allele 3.2 which is a length microvariant missing two bases distal to the repeat region (33). The HUMTH01 allele 9.3 is relatively common (28.9%) in the French Canadian Caucasians of Québec, while the HUMF13A01 allele 3.2 has an estimated frequency of 7.7% in this population.

TABLE 1—*Polymorphic short tandem repeat loci studied.*

GenePrint™ STR Systems	Repeat Sequence	Chromosome	Product Length (bp)	Gene
HUMCSF1PO	AGAT	5q33.3-34	299–323	c-fms proto-oncogene for CSF-1 receptor
HUMTPOX	AATG	2p23-2pter	224–252	Thyroid peroxidase
HUMTH01	AATG	11p15.5	179–203	Tyrosine hydroxylase
HUMF13A01	AAAG	6p24-25	283–331	Coagulation factor XIII a subunit
HUMFESFPS	AAAT	15q25-qter	222–250	c-fes/fps proto-oncogene
HUMvWA	AGAT	12p12-pter	139–167	von Willebrand factor

TABLE 2—*Allele frequencies and their standard errors (SE) at six STR loci in the French Canadian Caucasian population.*

Loci	Allele Sizes	Allele Frequencies (± SE) (%)	Loci	Allele Sizes	Allele Frequencies (± SE) (%)
HUMCSF1PO	7	0.1 ± 0.1	HUMF13A01	3.2	7.7 ± 0.9
(2n = 864)	8	0.5 ± 0.2	(2n = 866)	4	2.8 ± 0.6
	9	2.7 ± 0.5		5	22.1 ± 1.4
	10	27.0 ± 1.5		6	26.2 ± 1.5
	11	29.1 ± 1.5		7	37.5 ± 1.6
	12	32.2 ± 1.6		8	0.6 ± 0.3
	13	7.2 ± 0.9		9	—
	14	1.2 ± 0.4		10	—
	15	0.2 ± 0.2		11	0.2 ± 0.2
				12	0.1 ± 0.1
				13	0.5 ± 0.2
				14	0.1 ± 0.1
HUMTPOX	6	0.2 ± 0.2		15	1.4 ± 0.4
(2n = 870)	7	—		16	0.8 ± 0.3
	8	48.2 ± 1.7			
	9	10.2 ± 1.0	HUMFESFPS	7	0.3 ± 0.2
	10	6.9 ± 0.9	(2n = 642)	8	1.1 ± 0.4
	11	30.1 ± 1.6		9	0.3 ± 0.2
	12	4.1 ± 0.7		10	24.9 ± 1.7
	13	0.2 ± 0.2		11	40.8 ± 1.9
				12	26.6 ± 1.7
HUMTH01	5	—		13	5.8 ± 0.9
(2n = 870)	6	24.1 ± 1.5		14	0.2 ± 0.2
	7	17.6 ± 1.3			
	8	12.0 ± 1.1	HUMvWA	13	1.1 ± 0.3
	9	15.1 ± 1.2	(2n = 856)	14	10.6 ± 1.1
	9.3	28.9 ± 1.5		15	10.0 ± 1.0
	10	2.3 ± 0.5		16	20.7 ± 1.4
	11	0.1 ± 0.1		17	25.4 ± 1.5
				18	21.4 ± 1.4
				19	9.5 ± 1.0
				20	1.3 ± 0.4
				21	0.1 ± 0.1

2n = number of chromosomes typed for the locus.
A dash indicates the absence of that allele in the sample.

Allele frequency distributions are unimodal for the loci HUMCSF1PO, HUMFESFPS, and HUMvWA, clearly bimodal for both HUMTPOX and HUMTH01 and more complex for HUMF13A01.

Tests of Hardy-Weinberg Equilibrium

To determine if genotype frequencies at each locus deviated from the Hardy-Weinberg expectation (HWE), four tests were carried out: (1) The chi-square test based on total heterozygosity at the locus; (2) the likelihood ratio test criterion which contrasts the observed and expected frequencies of all possible genotypes for each locus; (3) the exact test that checks the conformity of the observed genotype frequencies with their respective HWE, and (4) the intraclass correlation test based on correlation of allele sizes within each individual's genotype. None of the tests detected any deviation from HWE at the loci HUMCSF1PO, HUMF13A01, and HUMvWA (Table 3). HUMTH01 showed significant departure from the HWE by the likelihood ratio and exact tests, whereas no deviation from HWE is detected by the intraclass correlation test, as well as by total heterozygosity at this locus (Table 3). Likewise, HUMFESFPS showed departure from HWE by the likelihood ratio and intraclass correlation tests (Table 3). The HUMTPOX locus showed statistically significant departure by all but the likelihood ratio test (Table 3). The observed heterozygosity at the TPOX locus (61.4%) was also significantly below its prediction from HWE (66.1%; see Table 5). Non detectable null-alleles can produce an excess of apparent homozygotes and thus a deficiency of heterozygotes. Although, our genotyping protocol was not specifically designed to look for allelic non-detectability, nor differential amplification of alleles of certain sizes, when we considered null-allele occurrence with the algorithm used in Chakraborty et al. (34), all significant departures from HWE by each of the four tests disappeared. Null-allele frequencies required to explain the departure from HWE were not substantial. For example, for the HUMTPOX locus, the maximum likelihood estimate of null allele frequency required was 1.26%, and for the HUMTH01 and HUMFESFPS loci, they were 0.02 and 0.07% respectively. Nevertheless, in spite of the seven significant departures noted for the loci HUMTPOX, HUMTH01, and HUMFESFPS, there is no data for which their genotype frequencies estimated by using HWE is significantly (at $p = 0.05$) underestimated, compared with the observed proportions of genotypes in the sample (Fig. 1).

TABLE 3—*Test of Hardy-Weinberg expectations of genotype frequencies at six STR loci in the French Canadian Caucasian population.*

Tests	Loci					
	CSF1PO	TPOX	TH01	F13A01	FESFPS	vWA
n	432	435	435	433	321	428
Heterozygosity Test: χ^2	1.147	4.360*	0.214	0.075	0.139	0.018
P	0.301	0.021	0.647	0.776	0.706	0.899
Likelihood Ratio Test: L	19.565	25.280	29.916*	34.238	29.344*	24.603
P	0.851	0.082	0.029	0.767	0.028	0.839
Exact Test: *P*	0.691	0.043*	0.018*	0.566	0.051	0.813
Intraclass Corr. Test: r	0.008	0.116*	−0.055	0.020	0.159†	0.058
P	0.890	0.012	0.228	0.660	0.004	0.228

n = number of individuals scored.
*Significant with $p < 0.05$.
†Significant with $p < 0.01$.

Tests for Population Subheterogeneity

Nonrandom association of alleles from genetically unlinked loci has often been interpreted as evidence for either population subheterogeneity or selection. Thus, allelic association between any pair of loci were analyzed by two methods (Table 4). Karlin's nonparametric interclass correlation of allele sizes depict two significant ($p < 0.05$) deviations from zero for the pairs TH01-vWA and F13A01-FESFPS (Table 4). However, the negative interclass correlation for F13A01-FESFPS cannot be explained by population substructure (22). By contrast, the contingency table χ^2 analyses of independence of pairwise two-locus genotype frequencies did not detect any significant dependence between loci (Table 4). Furthermore, as shown by Brown et al. (35), data on multi-locus genotype of individuals, summarized in the form of a distribution of the number of heterozygous loci in their multi-locus genotypes, provide a test of mutual independence of loci. Using this theory in the present survey, from data on 317 individuals for whom genotypes at all six loci were available, the observed variance of the number of heterozygous loci (s_k^2) statistic was used to detect linkage disequilibrium (25). The s_k^2 value was 1.17 with 95% confidence interval of 0.97–1.31 (under assumption of mutual independence of all six loci). This result indicates that there is no

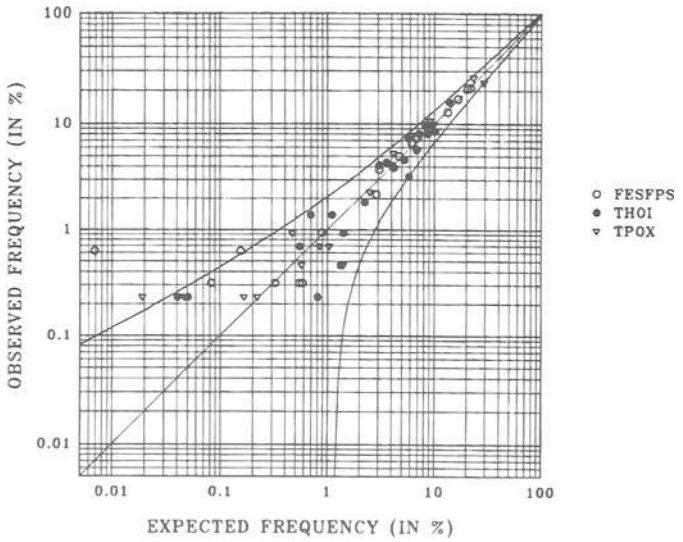

FIG. 1—*Observed versus expected (under HWE) genotype frequencies (in percent) at the three tetranucleotide repeat loci (HUMTPOX, HUMTH01, and HUMFESFPS) in French Canadian Caucasians of Québec. The diagonal line represents equality of observed and expected frequencies, while the two curved lines are the upper and lower 95% confidence intervals of observed frequencies (in percent).*

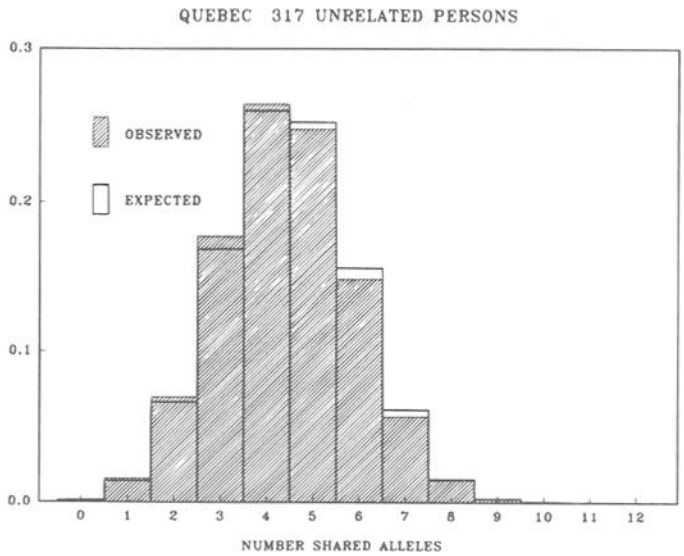

FIG. 2—*Distribution of the number of shared alleles at six STR loci among all pairs of 317 unrelated French Canadian Caucasian individuals from Québec. Individuals having complete six-locus genotype data are used in these computations (50,086 = 317 by 316/2 pairwise comparisons). The shaded bars represent the observed distribution, and the blank bars are the expected distribution of allele sharing under the assumptions of HWE and linkage equilibrium, assuming that the individuals are all biologically unrelated.*

gametic phase linkage disequilibrium among the six loci studied. Data on these 317 individuals were also used to test mutual independence of loci through an alternative test of allele sharing (26). The number of shared alleles can range from 0 (complete non-identity) to 12 (complete match at six-loci) for six typed loci. As shown in Fig. 2, the observed distribution and its expectation (based on HWE and linkage equilibrium) are statistically indistinguishable. Finally, examination of the maximum likelihood estimates of kinship coefficient allows evaluation of whether the significant departure from HWE is due to population substructure (27). These values along with their standard errors are shown in Table 5. Except the HUMF13A01 locus, all kinship coefficients are non-significant ($p < 0.05$). However, the HUMF13A01 locus did not show any evidence of departure from the HWE by any of the four tests (Table 3). Thus, we ascribe the significant kinship coefficient at this locus to chance alone, perhaps contributed by the highest number of rare alleles found at this locus (see Table 2). Together, these results suggest that the product rule is a reasonable good approximation for computing multilocus genotype frequencies for this population using these six loci.

TABLE 4—*Tests of independence of genotype frequencies between all pairs of loci in French Canadian Caucasians.*

Pair of Loci	Number of Individuals	Interclass Correlation		Chi-Square Test	
		Karlin's p	p	χ^2	p
CSF1PO—TPOX	432	−0.046	0.087	704.31	0.101
CSF1PO—TH01	432	0.013	0.604	487.19	0.410
CSF1PO—F13A01	430	0.002	0.947	693.71	0.698
CSF1PO—FESFPS	320	0.037	0.229	247.75	0.687
CSF1PO—vWA	425	−0.007	0.793	823.05	0.193
TPOX—TH01	435	−0.008	0.755	361.53	0.428
TPOX—F13A01	433	−0.024	0.339	865.22	0.165
TPOX—FESFPS	321	0.015	0.650	200.17	0.674
TPOX—vWA	428	0.038	0.149	452.26	0.778
TH01—F13A01	433	−0.0001	0.996	943.58	0.139
TH01—FESFPS	321	−0.015	0.608	289.43	0.504
TH01—vWA	428	0.057*	0.017	799.68	0.079
F13A01—FESFPS	319	−0.076*	0.014	541.96	0.213
F13A01—vWA	426	0.025	0.319	1034.38	0.451
FESFPS—vWA	320	−0.012	0.707	331.67	0.889

*Significant with $p < 0.05$.

Rare Alleles

The minimum allele frequencies were determined to ascertain that there can be confidence that DNA profile frequency estimates are meaningful even with small size databases. The minimum allele frequencies were computed by using two levels of confidence, 99 and 95% (α = 0.01 and 0.05 respectively), to designate the probabilities that alleles with a frequency of 1% or more are all represented in the database (Table 6). The minimum allele frequencies prescribed by Method 1 gives conservative results because it takes into account the extent of polymorphism at the locus (i.e., the number of alleles observed and heterozygosity at the locus) in addition to the size of the database. In contrast, the

TABLE 5—*Measures of variability at six STR loci in French Canadian Caucasians.*

Loci	2n	Number of Alleles	Heterozygosity		Kinship co-eff.	Variance of Allele Sizes
			Observed	Expected	Estimate ± S.E.	
CSF1PO	864	9	0.757	0.734	−0.016 ± 0.017	1.159
TPOX	870	7	0.614	0.661	0.025 ± 0.020	2.088
TH01	870	7	0.800	0.791	0.0004 ± 0.020	1.877
F13A01	866	12	0.730	0.736	0.041* ± 0.014	3.645
FESFPS	642	8	0.689	0.698	0.001 ± 0.021	0.915
vWA	856	9	0.820	0.818	0.001 ± 0.017	2.359

2n = number of chromosomes typed for the locus.
*Significant with $p < 0.05$.

TABLE 6—*Minimum allele frequency estimates (%) at six STR loci for the French Canadian Caucasian database.*

Loci	Sample Size 2n	α = 0.01* Method 1†	α = 0.01* Method 2‡	α = 0.05* Method 1†	α = 0.05* Method 2‡
CSF1PO	864	0.77	0.53	0.58	0.35
TPOX	870	0.74	0.53	0.55	0.34
TH01	870	0.79	0.53	0.60	0.34
F13A01	866	0.77	0.53	0.58	0.35
FESFPS	642	1.01	0.71	0.76	0.47
vWA	856	0.82	0.54	0.63	0.35

2n = number of chromosomes typed for the locus.
*(1 − α) represents the probability which with all alleles with a frequency of 1% or more are represented in the sample of 2n chromosomes.
†Method 1: using the algorithm of Budowle et al. (28).
‡Method 2: using the theory of Weir (23).

minimum allele frequencies derived from Method 2 depend only on the size of database (the number of chromosome sampled, 2n). These computations indicate that even though 16 of the 52 segregating alleles found at the six loci (see Table 2) have an estimated allele frequency below 1% in the database, the use of the threshold minimum allele frequencies is recommended to guard against providing very small multilocus genotype frequencies. In this context it may be noted that more recently a simpler approach of estimating minimum allele frequencies has been suggested (36), which is base on the concept of rebinning of DNA fragment sizes in RFLP typing of VNTR loci (37). Using this approach (where the minimum allele frequency would have been 5/2n, n being the number of individuals typed at the locus), the estimates of the minimum allele frequencies would have ranged from 0.57% (for HUMTPOX and HUMTH01) to 0.78% (for HUMFESFPS). We advocate the use of Method 1 for estimating the minimum allele frequency (as shown in Table 6) over the others, because it provides more conservative estimates and it exploits the extent of genetic variation at a locus as well as the size of the data base, while Method 2 (23) as well as the suggestion of NRC (36) are simply dependent on data size alone.

Parentage and Forensic Evaluations

The extent of polymorphism at each locus in the French Canadian population of Québec is reflected by the number of segregating alleles, the heterozygote frequencies and the variance of allele sizes in units of repeat lengths (Table 5). The number of alleles range from 7 to 12 per locus and the heterozygote frequencies of the loci from 61.4% to 82.0%. To examine the utility of these six loci for the purposes of personal identification and paternity testing, we computed: (1) the average exclusion probability for each locus (for an average mother-child pair), (2) the probability of discrimination, (3) the most common genotype and its frequency, and (4) the respective estimates for all six loci combined (Table 7). The most common genotype frequency is an indicator of the utility of a locus for both purposes, since the more common a genotype is, it becomes less efficient for discrimination as well as parentage exclusionary purposes. At none of these six loci, there is any genotype that would occur above 30% frequency (Table 7). In combination, the six loci would discriminate two unrelated individuals with a probability of 0.9999985, and in 98.74% cases a random male will be excluded in paternity analysis when the mother and the child are typed for these six loci (Table 7). The combined six locus genotype frequency is always below 3.69 by 10^{-5} (i.e., less frequent than 1 in 27,000 individuals). The agreement of expected (HWE) and observed frequencies of most common genotypes (locus-specific as well as combined, Table 7) also supports the results of test of allelic independence within and across loci.

Discussion and Conclusions

We have performed a genotype survey at six tetrametric STR loci (HUMCSF1PO, HUMTPOX, HUMTH01, HUMF13A01, HUMFESFPS, and HUMvWA) in the French Canadian Caucasian population of Québec. Analysis of the STR loci in this population revealed that their allele frequencies are distributed in a similar fashion to the one observed in other Caucasian populations (38) and are consistent with a general forward-backward model of evolution of new repeat-alleles (39). In terms of number of alleles and heterozygosity, substantial genetic variation exists at these six loci. In addition, we have shown that even if some loci (HUMTPOX, HUMTH01, and HUMFESFPS) are not in conformity with the predictions of HWE, there is little evidence of linkage disequilibrium between loci. These findings indicate that the observed allele frequencies may be used to estimate genotype frequencies and the product rule used to calculate combined genotype frequencies across loci in this population although the French Canadian Caucasian population of Québec demonstrated some founder effects of its recent settlement and growth (14–16). Additionally, Monson et al. (40) have also shown that there are no genetic disequilibrium between four VNTR loci and/or VNTR frequencies differing from other Caucasians despite the well-documented genetic founder effect in one of the subgroups of the French Canadian population studied.

The seven significant departures from HWE, which occur at three of the six loci as well as comparisons of levels of polymorphism by the different measures of genetic variation are suggestive indication of founder effect. While we cannot exclude the possibility that some of these departures could have caused by the presence of null alleles (of frequency not exceeding 1.3% at any locus), we do not believe that null-alleles are responsible for all of the noted departures. In fact, in a population experiencing recent growth, slight departure from HWE is expected and the different measures of genetic variation will not be perfectly correlated (29,41). For example, the variance of allele sizes is not perfectly correlated with either the locus heterozygosity nor with the number of alleles (Table 5), whereas a perfect correlation of these measures of variation is expected if the population was at mutation-drift equilibrium at these loci. A more detailed theory for studying such effects of small founding size and recent expansion of populations is under investigation.

Nevertheless, our results indicate that, founder effect, even if present, does not compromize the utility of these loci for forensic identification and parentage analyses. In fact, if the assumption of allelic independence is somewhat violated, Hardy-Weinberg deviations are generally towards being conservative (Fig. 1) and do not produce systematic bias of predicting genotype frequencies. Thus, we conclude that the use of allele frequency data for both purposes does not cause any problem particularly when conservative estimate of genotype frequencies is recommended (36). However, to achieve an average power of paternity exclusion above 99%, perhaps these six loci should be used in conjunction with other genetic markers for parentage determination. For identity testing, the combined power of discrimination (0.9999985) for

TABLE 7—*Probability of exclusion, probability of discrimination, and the most frequent genotype frequencies at the six STR loci in the French Canadian Caucasian population.*

Loci	Probability of Paternity Exclusion	Probability of Discrimination	Most Common Genotype/Frequency		
			Genotype	Frequency	
				Observed	Expected
CSF1PO	0.4900	0.8809	(11, 12)	0.201	0.187
TPOX	0.4129	0.8315	(8, 11)	0.237	0.290
TH01	0.5860	0.9239	(6, 9.3)	0.156	0.139
F13A01	0.5024	0.8862	(6, 7)	0.196	0.197
FESFPS	0.4383	0.8529	(11, 12)	0.212	0.218
vWA	0.6359	0.9411	(17, 18)	0.119	0.109
Combined	0.9874	0.9999985		3.69×10^{-5}	3.50×10^{-5}

these six loci is sufficiently high. Even, the most common genotype should not be found with frequency above 1 in 27,000 individuals.

Together, the loci HUMCSF1PO, HUMTPOX, HUMTH01, HUMF13A01, HUMFESFPS, and HUMvWA appear to be informative genetic markers for identity testing purposes in the French Canadian Caucasian population of Québec.

Acknowledgment

The statistical analysis was supported by US Public Health Service Research grants GM 41399, GM 45861, and GM 58545 (to RC).

References

1. 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.
2. 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.
3. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 1994;55:175–89.
4. Sprecher CJ, Puers C, Lins AM, Schumm JW. A general approach to analysis of polymorphic short tandem repeat loci. BioTechniques 1996;20:266–76.
5. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 1988;239:487–91.
6. Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culyer M, et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 1987;235:1616–22.
7. Lins AM, Sprecher CJ, Puers C, Schumm JW. Multiplex sets for amplification of polymorphic short tandem repeat loci—silver stain and fluorescent detection. BioTechniques 1996;20:882–9.
8. Hochmeister MN, Budowle B, Jung J, Borer UV, Comey CT, Dirnhofer R. PCR-based typing of DNA extracted from cigarette butts. Int J Leg Med 1991;104:229–33.
9. Walsh DJ, Corey AC, Cotton RW, Forman L, Herrin GLJr, Word CJ, et al. Isolation of deoxyribonucleic acid (DNA) from saliva and forensic science samples containing saliva. J Forensic Sci 1992;37:387–95.
10. Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, et al. Identification of the remains of the Romanov Family by DNA analysis. Nature Genet 1994;6:130–5.
11. Kimpton C, Fisher D, Watson S, Adams M, Urquqhart A, Lygo J, et al. Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. Int J Leg Med 1994;106:302–11.
12. Charbonneau H, Robert H. The French origin of the Canadian population 1608–1795. In: Harris RC, Matthews G, editors. Historical Atlas of Canada, Vol. 1, University of Toronto Press, Toronto, 1987;118–9.
13. Charbonneau H, Desjardins B, Guillemette A, Landry Y, Légaré J, Nault F. Naissance d'une population: les Français établis au Canada au XVIIe Siècle. Presses Universitaires de France (INED), Paris; Presses de l'Université de Montréal, Montréal, 1987.
14. De Braekeleer M, Dionne C, Gagné C, Julien P, Brun D, Murthy MRV, et al. Founder effect in familial hyperchylomicronemia in French Canadians of Québec. Hum Heredity 1991;41:168–73.
15. De Braekeleer M, Hechtman P, Andermann E, Kaplan F. The French Canadian Tay-Sachs disease deletion mutation: identification of probable founders. Hum Genet 1992;89:83–7.
16. Normand T, Bergeron J, Fernandez-Margallo T, Bharucha A, Murthy MRV, Julien P, et al. Geographic distribution and genealogy of mutation 207 of the Lipoprotein Lipase Gene in the French Canadian Population of Québec. Hum Genet 1992;89:671–5.
17. Li CC. First course in population genetics. Boxwood Press, Pacific Grove, CA. 1976.
18. Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser H, et al. Population genetics of dinucleotide $(dC\text{-}dA)_n$-$(dG\text{-}dT)_n$ polymorphisms in world populations. Am J Hum Genet 1995; 56:461–74.
19. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 1978;89:583–90.
20. Weir BS. Genetic data analysis II. Sinauer, Sunderland, MA. 1996.
21. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 1992;48:361–72.
22. 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.
23. Weir BS. Independence of VNTR alleles defined by fixed bins. Genetics 1992;130:873–87.
24. Karlin S, Cameron EC, Williams PT. Sibling and parent offspring correlation estimation with variable family size. Proc Natl Acad Sci USA. 1981;78:2664–8.
25. 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.
26. Chakraborty R, Jin L. Determination of relatedness between individuals by DNA fingerprinting. Hum Biol 1992;65:875–95.
27. Yasuda N. Estimation of the inbreeding coefficient from phenotype frequencies by a method of maximum likelihood scoring. Biometrics 1968;24:915–35.
28. Budowle B, Monson KL, Chakraborty R. Estimating minimum allele frequencies for DNA profile frequency estimates for PCR-based loci. Int J Leg Med 1996;108:173–6.
29. Kimmel M, Chakraborty R, Stivers DN, Deka R. Dynamics of repeat polymorphisms under a foward-backward mutation model: within- and between-population variability at microsatellite loci. Genetics 1996;143:549–55.
30. Fisher R. Standard calculations for evaluating a blood group system. Heredity 1951;5:95–102.
31. Chakravarti A, Li CC. The effect of linkage on paternity calculation. In: Walker RH, editor. Inclusion probabilities in parentage testing. American Association of Blood Banks, Arlington, VA, 1983; 411–22.
32. Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW. Identification of repeat sequence heterogeneity at the polymorphic short

tandem repeat locus HUMTH01 $[AATG]_n$ and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 1993;53:953–8.
33. 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.
34. Chakraborty R, Zhong Y, Jin L, Budowle L. Nondetectability of restriction fragments and independence of DNA fragment sizes within and between loci in RFLP typing DNA. Am J Hum Genet 1994;55:391–401.
35. Brown AHD, Feldman MW, Nero E. Multilocus structure of natural populations of Hordeum spontaneum. Genetics 1980;96:523–36.
36. National Research Council. The evaluation of forensic DNA evidence. National Research Council, National Academy Press, Washington, DC, 1996.
37. 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 loci, for use in forensic comparisons. Am J Hum Genet 1991;48:841–55.
38. Gill P, Evett I. Population genetics of short tandem repeat (STR) loci. Genetica 1995;96:69–87.
39. Shriver MD, Jin L, Chakraborty R, Boerwinkle E. VNTR allele frequency distribution under the stepwise mutation model: a computer simulation approach. Genetics 1993;134:983–93.
40. Monson KL, Moisan J-P, Pascal O, McSween M, Aubert D, Giusti A, et al. Description and analysis of allele distribution for four VNTR markers in French and French Canadian populations. Hum Hered 1995;45:135–43.
41. Kimmel M, Chakraborty R. Measures of variation at DNA repeat loci under a general stepwise mutation model. Theor Pop Biol 1996;45:345–67.

Additional information and reprint requests:
Lambert Busque, M.D., FRCPC
Centre de Recherche Guy Bernier
Hôpital Maisonneuve-Rosemont
5415 boul. de l'Assomption, Montréal
Québec, Canada, H1T 2M4
TEL: (514) 252-3495; FAX: (514) 254-5094.