

Rapid communication

Joint match probabilities for Y chromosomal and autosomal markers

Bruce Walsh^a, Alan J. Redd^b, Michael F. Hammer^{a,b,*}

^aDepartment of Ecology and Evolutionary Biology, University of Arizona Tucson, AZ 85721, USA

^bArizona Research Laboratories, Division of Biotechnology, University of Arizona, Tucson, AZ 85721, USA

Received 12 July 2006; received in revised form 13 February 2007; accepted 18 March 2007

Available online 20 April 2007

Abstract

Empirical tests of association between Y chromosome and autosomal markers are presented and a theoretical framework for determining a joint match probability is recommended. Statistical analyses of association were performed in 16 US populations between the autosomal genotypes from loci CSF1PO, FGA, THO1, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S512, D21S11 and Y chromosome haplotypes from loci DYS19, DYS385ab, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, and DYS439. The sample populations include individuals of European-, African-, Hispanic-, Native-, and Asian-American ancestry. The results are consistent with independence of Y and autosomal markers, although small amounts of dependence would likely have escaped our tests. Given the data in hand, we suggest it is appropriate to compute joint match probabilities by multiplying the Y haplotype frequency with the appropriately corrected autosomal frequency. In addition to correcting for autosomal frequency differences between groups, a further correction may be required. Since two individuals sharing the same Y haplotype are likely to be more recently related than two randomly chosen individuals, the autosomal frequencies have to be adjusted to account for this, akin to the θ correction used to account for population substructure. The structure imposed on the autosomal frequencies conditioned in a Y match is a function of the number of markers scored and their mutation rate. However, in most settings $\theta < 0.01$. When population structure is already present in the autosomes, the additional effect due to conditioning on the Y is small. For example, if the amount of structure in the population is $\theta = 0.01$ or 0.03 (the NRCII range), then the effect of conditioning on the Y results in only a trivial increase in θ to 0.02 – 0.04 , respectively.

© 2007 Elsevier Ireland Ltd. All rights reserved.

Keywords: Y chromosome; Autosomal loci; Y-STR; STR; Match probability

1. Introduction

In several forensic settings, one may need to obtain a joint match probability for a set of autosomal and Y chromosome markers. The ability to combine data from the autosomes and the Y chromosome is particularly important in cases where the signal from the male autosomal markers is largely obscured by the overwhelming amount of a female victim's DNA, such that only a subset of the tested autosomal markers can be detected in the resulting mixture [1,2]. In these cases Y-linked markers might be easily amplified, resulting in an n -locus haplotype for the perpetrator's Y chromosome. How

can we estimate the joint match probabilities for autosomal and Y chromosome markers with a random draw from the reference population?

Following the 1996 (US) National Research Council recommendations [3] (henceforth NRC II), the product rule is typically used to compute the match probability for autosomal markers, multiplying the single locus genotype frequencies (suitably corrected for population structure and differences in allele frequencies among distinct subpopulations) of the m scored markers to obtain the m -locus genotype frequency. However, the product rule does not apply for obtaining the frequency of a particular Y haplotype from estimates of individual marker allele frequencies. Y-linked markers (i.e., from the non-recombining region) are typically in strong linkage disequilibrium. Thus, the frequency of the particular haplotype must be estimated from a reference database. Because the Y chromosome and autosomes are unlinked, this might suggest using the product of the autosomal

* Corresponding author at: Division of Biotechnology, University of Arizona, Tucson, AZ 85721, USA. Tel.: +1 520 621 9828; fax: +1 520 621 9247.

E-mail addresses: mfh@u.arizona.edu, mfh@email.arizona.edu (M.F. Hammer).

and Y match probabilities for the joint matching probability. There are some potential concerns with this assumption. First, disequilibrium between the mitochondrial DNA and the autosomes is known to exist in many hybrid zones in natural populations of various organisms [4–6]. The level of Y-autosomal disequilibria (or lack thereof) should be empirically quantified in population samples rather than assumed. Sinha et al. [7] have tested for associations between Y-STR alleles and 13 autosomal STRs alleles in an African-American and a European-American sample from Louisiana. Our study differs in that we compared Y chromosome haplotypes and autosomal genotypes (as opposed to comparing individual alleles), which we believe is the forensically more relevant comparison.

Even if there is no disequilibrium among autosomal markers given a particular Y haplotype, there are several additional complications in computing Y-autosomal joint match probabilities. The first is that Y chromosome haplotypes may be highly informative as to which subpopulation an individual belongs, and this in turn may change the autosomal allele frequencies used to compute the autosomal match probabilities. The second, and potentially more important, complication is that two individuals sharing the same Y haplotype are likely to be more closely related than two random individuals from the population. To account for this potentially increased resemblance, it is necessary to correct the probability of the two individuals sharing an autosomal genotype given that they both share the same Y haplotype. Expressed another way, NRC II recommends dealing with small amounts of population substructure within any particular ethnic group through a substructure correction fraction θ . Here we obtain the expected value of this θ correction by considering autosomal frequencies among the set of individuals that share a common Y haplotype. We will show that while conditioning on a Y match does indeed impose a small amount of structure, the effect is usually very small.

2. Materials and methods

2.1. Samples

Blood stains from New York City European-Americans, African-Americans, Hispanic-Americans, and Asian-Americans were provided by M. Prinz. Blood stains from North Carolina European-Americans, African-Americans, and Hispanic-Americans were provided by R. Pendergraft and M. Nelson. DNA samples from Vermont European-Americans, African-Americans, Hispanic-Americans, Asian-Americans, and Native-Americans were provided by E. Buel and J. Nicklas. Blood samples from European-American and Hispanic-American from Mesa Arizona were provided by V. Smart. Blood samples of Apache and Navajo were provided by D. Duplissa, R. Vossbrink, and S. Narveson. Additional Navajo samples and Tucson-Asian samples were collected by M. Hammer [8]. DNA extraction and quantification methods were previously published [9].

2.2. Y chromosome STRs

The US core Y-STRs (DYS19, DYS385ab, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, and DYS439) were typed using previously published methods [9] with the exception that DYS438 was incorporated into multiplex II using primers from Butler et al. [10].

2.3. Autosomal STRs

The genotypes of sampled individuals at 13 CODIS loci (CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S512, D21S11) were kindly provided by M. Prinz, R. Pendergraft, M. Nelson, E. Buel, Janice Nicklas, V. Smart, D. Duplissa, Russell Vossbrink, and Sue Narveson. In addition, we used the Identifier kit from Applied Biosystems to type some of the Vermont African-American, Apache and Navajo samples.

2.4. Statistical analyses

An exact test of linkage disequilibrium between Y chromosome haplotypes and autosomal genotypes was carried out using a Markov chain methodology in the Arlequin program (V2.0) [11–13]. This method is an extension of the Fisher exact probability test on contingency tables [12], with a Markov chain used to search the space of all contingency tables [14,15]. The P -value of the test indicates the proportion of tables that have a probability smaller or equal to the observed contingency table. A total of 100,000,000 steps were performed in the Markov chain following 10,000 demorization steps (the initial burn-in for the Markov sampler).

3. Results and discussion

3.1. The product rule applies

As shown in Table 1, the results are largely consistent with independence of Y haplotypes and autosomal genotypes. Results are reported for each sampled subpopulation within the sampled ethnic groups and also for the composite populations for each ethnic group formed by pooling the samples. Within populations there were only 2 associations out of 208 comparisons (less than 1%). One of these associations, the Vermont Hispanic-Americans at D18S512, was just significant at the $P = 0.05$ level, while that of the Arizona (Mesa) European-Americans at vWA was significant at the $P < 0.01$ level. When populations were pooled by ethnic group there were three additional associations between the Y and the autosomal genotypes (3 of 65, less than 5% of comparisons) including: two in African-Americans at TPOX ($P < 0.01$) and D5S818 ($P < 0.01$), and one in Asian-Americans at CSF1PO ($P < 0.05$).

Table 1 also shows the power of the various comparisons, which is a function of sample size. Over half of the sample sizes have power to detect associations larger than $r^2 = 0.05$, while a few of the smaller samples have reduced power (Table 1). However, there are caveats as to why we may have missed potential associations, especially if they are weak. There are also concerns in the literature that the exact test we employ may be under-powered in some settings [16], in particular when there is considerable admixture within the sample. There is evidence of this in that roughly half of our P -values are close to one, while the remainder of the P -values follow the more expected uniform distribution.

Thus, small amounts of association may not be detectable under these sample sizes with the methods we employed. However, given that both our study and that of Sinha et al. [7], who used a very different approach, fail to find consistently strong associations, we are very comfortable with using the product rule, suitably modified as discussed below.

Table 1
Exact test results of associations between Y chromosome haplotypes and autosomal genotypes

Y haplotype vs. population	<i>N</i>	$r^2(50)$	CSF1PO	FGA	THO1	TPOX	vWA	D3S1358	D5S818	D7S820	D8S1179	D13S317	D16S539	D18S512	D21S11
European-American	380	0.02	0.56	0.86	0.30	0.52	0.31	0.28	0.49	0.56	0.87	0.61	0.37	0.91	0.77
New York City	43	0.09	1.00	1.00	0.17	0.19	1.00	0.20	1.00	1.00	1.00	1.00	0.15	1.00	1.00
North Carolina	94	0.04	0.58	0.74	1.00	0.67	1.00	1.00	1.00	0.51	1.00	0.32	0.49	0.75	0.68
Vermont	200	0.02	0.76	0.82	0.49	0.70	0.27	0.14	0.64	0.58	0.91	0.70	0.69	0.82	1.00
Arizona (Mesa)	43	0.09	0.13	1.00	1.00	1.00	0.00	0.52	0.41	0.33	1.00	0.42	0.34	0.38	0.14
African-American	187	0.02	1.00	1.00	0.35	0.00	0.35	0.21	0.00	0.42	0.68	0.49	0.48	1.00	0.75
New York City	41	0.10	0.64	1.00	1.00	0.72	0.46	0.06	0.48	0.60	0.34	1.00	0.10	1.00	1.00
North Carolina	96	0.04	1.00	1.00	0.48	0.36	0.59	0.25	0.59	0.56	1.00	0.63	0.54	1.00	1.00
Vermont	50	0.09	1.00	1.00	1.00	0.12	1.00	1.00	0.15	1.00	1.00	1.00	1.00	1.00	0.32
Hispanic-American	117	0.03	0.53	1.00	0.73	0.46	1.00	0.24	0.77	1.00	1.00	0.49	1.00	0.77	0.84
New York City	38	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
North Carolina	15	0.26	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Vermont	17	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.05	1.00
Arizona (Mesa)	47	0.09	1.00	1.00	1.00	0.40	1.00	0.29	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Native-American	187	0.02	0.31	0.87	0.35	0.43	0.45	0.10	0.51	0.36	0.69	0.53	0.67	1.00	0.62
Apache	85	0.04	0.25	1.00	0.49	0.59	1.00	0.57	0.63	0.55	1.00	1.00	1.00	1.00	1.00
Navajo	83	0.04	0.09	0.59	0.08	0.46	0.11	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29
Vermont	19	0.21	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Asian-American	46	0.09	0.04	1.00	1.00	1.00	0.33	1.00	1.00	0.28	1.00	1.00	1.00	1.00	0.27
New York City	38	0.10	0.25	1.00	1.00	1.00	1.00	1.00	1.00	0.21	1.00	1.00	1.00	1.00	0.18
Vermont	8	0.49	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Power to detect association given the sample size is indicated in the column labeled $r^2(50)$, which, for the given sample size, is the smallest value of r^2 that would give a significant test for an association (using a 95% two-sided test of $r^2 = 0$) at least 50% of the time (i.e., 50% power). Significant *P*-values are shown in bold. Values are reported for both specific populations of each ethnic group and for composite populations by combining all subpopulations within each sampled ethnic group.

Formally, we can express this joint probability by conditioning first on a match at the Y:

$$\Pr(Y, \text{autosomal genotype}) = \Pr(\text{autosomal genotype} | Y) \cdot \Pr(Y) \quad (1)$$

We, of course, need to account for the possibility that the autosomal genotype frequencies vary over groups. A more subtle issue is that two individuals sharing the same Y haplotype are potentially more recently related than two randomly chosen individuals, and the autosomal frequencies have to be corrected to account for this increased resemblance. This is akin to the θ correction for autosomal frequencies under population structure suggested in NRC II [3].

3.2. Accounting for autosomal structure within identical Y haplotypes

Two individuals with identical Y haplotypes are likely to be more closely related to each other than are two random individuals. Hence, the probability of an autosomal match, conditioned on the two individuals sharing the same Y haplotype, needs to be computed.

To do so, we first note that in a population genetics framework, we expect the time back to the common (Y-contributing) ancestor of two individuals with identical Y haplotypes to be more recent than the time back to a common ancestor for two randomly chosen individuals. In essence, the common Y haplotype imparts a subgroup structure and we

need to account for this. Balding and Nichols [17] show that the correct single-locus (autosomal) expressions for match probabilities when individuals come from the same subpopulation is a function of the average coefficient of coancestry θ , with

$$\Pr(A_i A_i | A_i A_i) = \frac{[2\theta + (1 - \theta)p_i][3\theta + (1 - \theta)p_i]}{(1 + \theta)(1 + 2\theta)} \quad (2a)$$

$$\Pr(A_i A_j | A_i A_j) = \frac{2[\theta + (1 - \theta)p_i][\theta + (1 - \theta)p_j]}{(1 + \theta)(1 + 2\theta)} \quad (2b)$$

If one draws a random allele from each of these individuals, recall that the coefficient of coancestry θ for a pair of individuals is the probability that these alleles are identical by descent [18,19].

How do we compute the expected value of θ given two individuals are identical at all n tested Y chromosome markers? We do so by conditioning on the distribution of time back to a common Y ancestor given the number of identical markers scored on the Y. This uses the results of Walsh [20], who developed Bayesian estimators for the distribution of the time back to this most recent common ancestor (MRCA) given the distribution of marker differences between two Y haplotypes. Given this distribution of times back to the MRCA, we can then compute an average value θ .

If the mutation rate per marker is μ , then the chance that two haplotypes are identical at n markers given they last shared a

common Y ancestor t generations ago is

$$\Pr(n \text{ marker match} | t \text{ generations to MRCA}) = (1 - \mu)^{2nt} \tag{3}$$

This is the likelihood function for t given the mutation rate and number of markers. A more detailed analysis under the stepwise mutation model in Walsh [20] shows that, when all n markers match, that this simple expression based on the infinite alleles model is essentially equivalent to the more exact analysis under a stepwise model that allows for back mutations. To obtain a formal posterior distribution for the time to the MRCA given the molecular match data for the Y, we first need a prior on the time to the MRCA.

Population genetics theory provides such a prior distribution for the time to MRCA for two randomly drawn Y chromosomes in the absence of any marker information [20]. The time back to MRCA follows a geometric distribution with parameter $\lambda = 1/N_e$, the reciprocal of the effective population size (since the Y chromosome is haploid). Hence, the prior becomes

$$\Pr(t \text{ generations to MRCA}) = \lambda(1 - \lambda)^{t-1} \tag{4}$$

The resulting posterior distribution (from Eqs. (3) and (4)) becomes

$$\Pr(t \text{ generations to MRCA} | n\text{-marker match}) = C(1 - \mu)^{2nt} \lambda(1 - \lambda)^{t-1} \tag{5}$$

where the constant C assures that the probabilities sum to one. (Walsh [20] used a continuous time approximation, while Eqs. (3)–(5) are exact for discrete generations.)

In particular, the posterior distribution for the time t to MRCA given two individuals exactly match at n markers is

$$p(t) = C(\lambda, \mu)(1 - \mu)^{2nt} (1 - \lambda)^{t-1} \tag{6a}$$

where

$$C(\lambda, \mu) = \left(\sum_{\tau=1}^{\infty} (1 - \mu)^{2n\tau} (1 - \lambda)^{\tau-1} \right)^{-1} \tag{6b}$$

Since the coefficient of coancestry for two individuals that shared a common ancestor t generations ago is just $(1/2)^{2t+1}$, e.g. [19], the expected coefficient of coancestry θ thus becomes

$$\begin{aligned} E[\theta] &= \sum_{t=1}^{\infty} \theta(\text{tMRCA} = t) \times P(\text{tMRCA} = t) \\ &= \sum_{t=1}^{\infty} \left(\frac{1}{2^{2t+1}} \right) p(t) \\ &= C(\lambda, \mu) \sum_{t=1}^{\infty} \left(\frac{1}{2} \right)^{2t+1} (1 - \mu)^{2nt} (1 - \lambda)^{t-1} \\ &= \frac{1 - (1 - \lambda)(1 - \mu)^{2n}}{8 - 2(1 - \lambda)(1 - \mu)^{2n}} \end{aligned} \tag{7}$$

For a Y match with $n = 11$ STR markers, Eq. (7) gives the following values for $E[\theta]$ for different mutation rates μ

N_e	$\mu = 0.001$	$\mu = 0.002$	$\mu = 0.004$
∞	0.004	0.007	0.014
500	0.004	0.008	0.014
100	0.005	0.009	0.015
50	0.007	0.010	0.017

The reason for this small amount of structure is that while two individuals with identical Y haplotypes do indeed have a more recent common ancestor than two random individuals for the population, the mean time back to that ancestor is 46, 23, and 12 generations (respectively) for $\mu = 0.001$, 0.002, and 0.004 in a large population. Thus, while related, they are on average very distant relatives and hence the increased sharing of autosomal alleles is very small.

Given that most estimates of STR mutation rates are around 0.002–0.004, taking $E(\theta) = 0.01$ –0.02 is expected to be conservative (i.e., most favorable to a defendant by generating higher match probabilities). Thus for an unstructured population, conditioning on a complete match at the Y imposes a small amount of structure when computing autosomal match probabilities conditional on this match.

Given that conditioning on a Y match can impose some structure on autosomal frequencies even when the population is unstructured, what effect does conditioning on a Y match have when the population does show structure? One simple way to approach this is as follows. In the US, the NRCII recommendations suggest considering a θ value of between 0.01 and 0.03 for standard autosomal match probabilities when there is some concern about structure. We can generate this amount of structure among two randomly drawn individuals from the population by assuming a prior with hyperparameter $\lambda = 0.061$ to give a base structure of $\theta = 0.01$ or $\lambda = 0.190$ to give a $\theta = 0.03$. Assuming these values for the hyperparameter in the prior gives values of θ when conditioning on a Y match as follows:

λ	$\mu = 0.001$	$\mu = 0.002$	$\mu = 0.004$
0.061	0.013	0.017	0.022
0.190	0.032	0.035	0.040

Thus, under this simple model if we start with an intrinsic θ value of 0.01, the additional amount of structure imposed by conditioning on a Y match increases θ to between 0.013 and 0.022, while if the base θ is 0.03, conditioning on a Y increases θ to between 0.032 and 0.040. Hence, when structure is present, conditioning on a Y match only results in a very slight increase in θ over the unconditional θ value for the population.

3.3. Recommendation

A conservative approach to computing the joint Y-autosomal matching probability is thus as follows. For each of the major ethnic groups under consideration, compute the autosomal match probability using the product of single-locus genotype frequencies obtained from using Eqs. (2a) and (2b) with $\theta = 0.04$ and the allele frequencies appropriate for each group (note that this is just Recommendation 4.2 of NRC II). A

conservative estimate of the joint probability is obtained by multiplying the largest value of these group autosomal match probabilities by the estimated matching probability for the Y.

As to the issue of estimating the Y match probability from a database, we favor a conservative modification [17] of the simple counting method:

$$\widehat{\Pr}(y) = \frac{i + 2}{n + 2} \quad (9)$$

where there are i matches in an initial data base of n haplotypes. Making the conservative (favoring a suspect) assumption that a crime sample came from a different individual than the suspect, we have two more samples (hence $n + 2$ total samples) and matches for the suspect, the crime sample and i matches from the databases (hence $i + 2$ matches). While one can compute a match probability using Eq. (9) for each ethnic group in the database, this tends to overweight those groups with the fewest samples. For example, if there are 4000 European samples but only 200 Native- (North) Americans, then if there are no matches, the probability is $2/4002 = 0.0005$ for Europeans but $2/202 = 0.0099$ for Native-Americans.

To avoid this concern we suggest the following weighting method given a specified reference population.

3.4. Reporting a single match probability for a specified reference population

Our final comment is that the NRC II recommendations of reporting match probabilities for each of the major ethnic subpopulations can be improved upon [21,22]. Specifically, if $\Pr(i)$ is the match probability for ethnic group i , then an overall match probability for a given reference population is simply:

$$\sum_i \Pr(\text{match for group } i) \cdot \Pr(\text{random individual is from group } i) \quad (10)$$

This returns a single match probability given a specified reference population, which we suggest provides a more meaningful value to tryers-of-fact than simply reporting a series of match probabilities for different groups. Further, one can easily compute match values for different reference populations. Consider a crime occurring in Orange County in Southern California. Depending on the circumstance, the reference population might be a particular neighborhood in this county, or it might be all of southern California. Providing the tryers-of-fact with match probabilities assuming the perpetrator was from the neighborhood, as opposed to a random individual from Southern California (or other scenarios), provides them with additional information to weight against other appropriate facts for their particular case.

Acknowledgements

We thank our colleagues from crime laboratories who provided the US population samples and associated autosomal profiles (as detailed in the text). Two reviewers provided very

useful comments which significantly improved this manuscript. This research was supported by the National Institute of Justice Grant 2000-IJ-CX-K006 to MFH.

References

- [1] W. Parson, H. Niederstätter, S. Köchl, M. Steinlechner, B. Berger, When autosomal short tandem repeats fail: optimized primer and reaction design for Y-chromosomal short tandem repeat analysis in forensic case work, *Croat. Med. J.* 42 (2001) 285–287.
- [2] N. Cerri, U. Ricci, I. Sani, A. Verzeletti, F. De Ferrari, Mixed stains from sexual assault cases: autosomal or Y-chromosome short tandem repeats? *Croat. Med. J.* 44 (2003) 289–292.
- [3] National Research Council, *The Evaluation of Forensic DNA Evidence*, National Academic Press, Washington, DC, 1996.
- [4] R.G. Latta, Y.B. Linhart, J.B. Mitton, Cytonuclear disequilibrium and genetic drift in a natural population of ponderosa pine, *Genetics* 158 (2001) 843–850.
- [5] Z. Redenbach, E.B. Taylor, Evidence for bimodal hybrid zones between two species of char (*Pisces: Salvelinus*) in northwestern North America, *J. Evol. Biol.* 16 (2003) 1135–1148.
- [6] A.L. Roca, N. Georgiadis, S.J. O'Brien, Cytonuclear genomic dissociation in African elephant species, *Nat. Genet.* 37 (2005) 96–100.
- [7] S.K. Sinha, B. Budowle, R. Chakraborty, A. Paunovic, R.D. Guidry, C. Larsen, A. Lal, M. Shaffer, G. Pineda, S.K. Sinha, E. Schneida, H. Nasir, J.G. Shewale, Utility of the Y-STR typing systems Y-PLEX 6 and Y-PLEX 5 in forensic casework and 11 Y-STR haplotype database for three major population groups in the United States, *J. Forensic Sci.* 49 (2004) 691–700.
- [8] A.J. Redd, V.F. Chamberlain, V.F. Kearney, D. Stover, T. Karafet, K. Calderon, B. Walsh, M.F. Hammer, Genetic structure among 38 populations from the United States based on 11 core Y chromosome STRs, *J. Forensic Sci.* 51 (2005) 580–585.
- [9] A.J. Redd, A.B. Agellon, V.A. Kearney, V.A. Contreras, T. Karafet, H. Park, P. de Knijff, J.M. Butler, M.F. Hammer, Forensic value of 14 novel STRs on the human Y chromosome, *Forensic Sci. Int.* 130 (2002) 97–111.
- [10] J.M. Butler, R. Schoske, P.M. Vallone, M.C. Kline, A.J. Redd, M.F. Hammer, A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers, *Forensic Sci. Int.* 129 (2002) 10–24.
- [11] S. Schneider, J.-M. Kueffer, D. Roessli, L. Excoffier, *ARLEQUIN ver 2.0. A Software for Population Genetic Data Analysis*, University of Geneva, Geneva, 1997.
- [12] M. Slatkin, Linkage disequilibrium in growing and stable populations, *Genetics* 137 (1994) 331–336.
- [13] M. Slatkin, L. Excoffier, Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm, *Heredity* 76 (1996) 377–383.
- [14] S. Guo, E.A. Thompson, Performing the exact test of Hardy–Weinberg proportion for multiple alleles, *Biometrics* 48 (1992) 361–372.
- [15] M. Raymond, F. Rousset, An exact test for population differentiation, *Evolution* 49 (1995) 1280–1283.
- [16] B. Law, J.S. Buckleton, C.M. Triggs, B.S. Weir., Effects of population structure and admixture on exact tests for association between loci, *Genetics* 164 (2003) 381–387.
- [17] D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, *Forensic Sci. Int.* 64 (1994) 125–140.
- [18] B. Weir, *Genetic Data Analysis*, Sinauer Associates, Sunderland, MA, 1990.
- [19] M. Lynch, B. Walsh, *Genetics and Analysis of Quantitative Traits*, Sinauer Associates, Sunderland, MA, 1997.
- [20] B. Walsh, Estimating the time to the MRCA for the Y chromosome or mtDNA for a pair of individuals, *Genetics* 158 (2001) 897–912.
- [21] J.S. Buckleton, K.A.J. Walsh, G.A.F. Seber, D.G. Woodfield, A stratified approach to the compilation of blood group frequency surveys, *J. Forensic Sci. Soc.* 27 (1987) 103–122.
- [22] K.A.J. Walsh, J.S. Buckleton, Calculating the frequency of occurrence of a blood type for a “random man”, *J. Forensic Sci. Soc.* 31 (1991) 49–58.