Population-specific F_{ST} values for forensic STR markers:
A worldwide survey.

John Buckleton¹, James Curran², Jérôme Goudet³, Duncan Taylor⁴, Alexandre Thiery⁵ and B.S. Weir⁶

¹ Environmental and Scientific Research Ltd, Auckland, New Zealand, and
National Institute of Standards and Technology, Gaithersburg MD, USA

² Department of Statistics, University of Auckland, Auckland, New Zealand

³ Department of Ecology and Evolution, UNIL-Sorge, Lausanne, Switzerland

⁴ Forensic Science South Australia, Adelaide SA, Australia, and
School of Biological Sciences, Flinders University, Adelaide SA, Australia

⁵ Bioinformatics Core Facility, Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

⁶ Department of Biostatistics, University of Washington, Seattle WA, USA

March 9, 2016

Abstract

The interpretation of matching between DNA profiles of a person of interest and an item of evidence is undertaken using population genetic models to predict the probability of matching by chance. Calculation of matching probabilities is straightforward if allelic probabilities are known, or can be estimated, in the relevant population. It is more often the case, however, that the relevant population has not been sampled and allele frequencies are available only from a broader collection of populations as might be represented in a national or regional database. Variation of allele probabilities among the relevant populations is quantified by the population structure quantity F_{ST} and this quantity affects matching proportions. Matching within a population can be interpreted only with respect to matching between populations and we show here that F_{ST} , can be estimated from sample allelic matching proportions within and between populations. We report such estimates from data we extracted from 250 papers in the forensic literature, representing STR profiles at up to 24 loci from nearly 500,000 people in 446 different populations. The results suggest that theta values in current forensic use do not have the buffer of conservatism often thought.

Key words: Forensic DNA, STR marker, theta, coancestry

INTRODUCTION

The use of microsatellite, or short tandem repeat (STR), loci for forensic identification is now well established and analyses of large STR frequency databases have been published. Silva et al. [1] made use of STR allele frequencies accessed from a then-online database strdna-db (Pamplona et al., [2]), while Steele et al. [3] used data from people living in the United Kingdom (UK), or wishing to migrate to the UK on the basis of relatedness to a UK resident, as well as “reference” data collected by the UK Forensic Science Service. Both these papers provide useful reviews of the forensic uses of STR profiles and both describe the population genetic structure revealed by STR data. Silva et al. [1] commented on the overall similarity between conclusions that can be drawn from forensic and other population databases about the structure of human populations and its relationship to the movement of people from East Africa, the most likely place of origin of anatomically modern humans [4]. This is in spite of forensic markers being chosen to maximize the diversity among individuals. Steele et al. [3] gave detailed analyses of the population structure quantity F_{ST} that can be regarded as measuring the evolutionary relatedness between two individuals in the same population.

Here we continue the discussion of F_{ST} by describing a new approach to estimating this quantity and then by applying this approach to allele frequencies at 24 STR loci in forensic use published for 446 populations. Central to our approach is the recognition that F_{ST} values are statements about pairs of alleles within a population *relative to* some other population or collection of populations, and this reference frame needs to be specified. Our results suggest that a somewhat higher value of F_{ST} than previously thought should be used in forensic match probability calculations, especially when the continental ancestry of the relevant population is unknown.

Allelic Matching Theory

Population genetic theory (e.g. [5]) distinguishes between three types of probability of A_u , the u th allele at locus **A**. First, there is the sample frequency $\tilde{p}_{iu}$ in a sample of individuals from population i . Second, there is the actual frequency $\check{p}_{iu}$ of A_u in population i from which this sample has been taken. Note that $\check{p}_{iu}$ is the expected value of $\tilde{p}_{iu}$ over samples from that population, irrespective of sample size, so $\tilde{p}_{iu}$ is an unbiased and consistent estimate of $\check{p}_{iu}$. With random sampling of individuals, and if the population in question is in Hardy-Weinberg equilibrium, then the number of A_u alleles in a sample of n_i genotypes has a binomial distribution with parameters $2n_i$ and $\check{p}_{iu}$

so the within-population variance of $\tilde{p}_{iu}$ is $\text{Var}_W(\tilde{p}_{iu}) = \check{p}_{iu}(1 - \check{p}_{iu})/(2n_i)$.

Any population genetic theory aimed at calculating match probabilities must address another level of sampling, namely that inherent to the underlying evolutionary process. An actual allele frequency $\tilde{p}_{iu}$ is just one of many values that are possible for a given evolutionary history, and the expected value of $\tilde{p}_{iu}$ taken over many different realizations of this same history, written as p_u , is usually referred to as the allele frequency of A_u . It may be better to say “allele probability” here but we will defer to common usage and use “frequency” both for sample and population proportions and the corresponding expected values or probabilities. The same value p_u applies here to all populations and it is the probability that an allele drawn at random from any population is of type A_u . Note that the total expectation of the sample frequency $\tilde{p}_{iu}$, taken over samples and over populations is also equal to p_u . There is an implicit assumption that natural selection has not been acting differentially among populations. For a wide class of evolutionary models, the variance of the actual allele frequency $\tilde{p}_{iu}$ among replicates of that population has the form $\text{Var}_A(\tilde{p}_{iu}) = p_u(1 - p_u)\theta_i$, where θ_i is the probability that two alleles drawn randomly from population i are identical by descent (ibd). Alleles are ibd if they have a common origin – this is necessarily a statement about the past and there is variation among all the possible populations that might descend from an ancestral population. The total variance of the sample frequency $\tilde{p}_{iu}$, over samples from a population and over realizations of a population, is $\text{Var}_T(\tilde{p}_{iu}) = p_u(1 - p_u)[\theta_i + (1 - \theta_i)/(2n_i)]$.

The evolutionary perspective given by this introduction of θ_i allows account to be taken of some alleles in the population having a shared ancestry and being ibd. Shared ancestry of some alleles increases the chance of allele matching, or identity in state (ibs), for alleles drawn randomly from the population.

If there is Hardy-Weinberg equilibrium within a (large) population i for which the frequency of allele A_u is $\check{p}_{iu}$, then the probability that two alleles in that population are of type A_u is $\text{Pr}(A_u A_u)_i = \check{p}_{iu}^2$, and this has an expected value over replicates of the population of $p_u^2 + p_u(1 - p_u)\theta_i$. If, instead of a population frequency $\check{p}_{iu}$ for population i , or its expected value p_u , we have only a sample value $\tilde{p}_u$ from a larger population, or collection of populations, then we must consider the sampling variance of $\tilde{p}_u$. Some of the alleles contributing to that sample value are from the target population i for which θ_i is appropriate and some are from different populations j , $j \neq i$ for which an analogous quantity θ_{ij} is appropriate. This new measure is the probability that an allele from population i is ibd to an allele from population j . The estimated two-allele probability for population i is $\tilde{p}_u^2 + \tilde{p}_u(1 - \tilde{p}_u)\beta_i$, where $\beta_i = (\theta_i - \theta_B)/(1 - \theta_B)$ and θ_B is the average over all

pairs of populations within the sampled collection of population-pair θ_{ij} 's. We are assuming a large (unknown) number of populations within the sampled collection. The corresponding allelic match probability is $\Pr(A_u|A_u) = p_u + \theta_i(1 - p_u)$ and we can estimate this as $\tilde{p}_u + \beta_i(1 - \tilde{p}_u)$.

Balding and Nichols [7] extended the discussion of matching to genotypes and gave the match probabilities for a population as:

$$\begin{aligned}\Pr(A_u A_u | A_u A_u)_i &= \frac{[2\theta_i + (1 - \theta_i)p_u][3\theta_i + (1 - \theta_i)p_u]}{(1 + \theta_i)(1 + 2\theta_i)} \\ \Pr(A_u A_v | A_u A_v)_i &= \frac{2[\theta_i + (1 - \theta_i)p_u][\theta_i + (1 - \theta_i)p_v]}{(1 + \theta_i)(1 + 2\theta_i)}, \quad u \neq v\end{aligned}\tag{1}$$

If the particular population within a collection of populations is not specified, and we wish to express a match probability that would apply to any of a large number of populations within the collection, then we replace θ_i by the average value θ_W over populations. To estimate the matching probabilities with the total sample allele frequencies $\tilde{p}_u$ in place of the true values p_u , we use $\beta_W = (\theta_W - \theta_B)/(1 - \theta_B)$, the average of the β_i 's, in place of θ_W .

We note that most discussions in the literature, e.g. [1], [2], do not make explicit mention of between-population values θ_{ij} or θ_B , and we also note that β_W is the quantity usually referred to as F_{ST} . Our formulation stresses that the concept of identity by descent of alleles within a population requires a point of reference, which we take as identity by descent between pairs of populations. We see below that there is an immediate translation to allelic matching within and between populations.

There is widespread use in forensic science of the profile probabilities $\Pr(A_u A_u)$ or $\Pr(A_u A_v)$, $u \neq v$, and their product rule estimates $\tilde{p}_u^2$ or $2\tilde{p}_u \tilde{p}_v$, rather than the match probabilities shown in (1). To allow for departures from Hardy-Weinberg equilibrium in the population collection that results from variation in allele frequencies among populations, the profile probabilities in a subpopulation can be estimated as

$$\Pr(A_u A_u)_i \hat{=} \tilde{p}_u[F_i + (1 - F_i)\tilde{p}_u], \Pr(A_u A_v)_i \hat{=} 2\tilde{p}_u \tilde{p}_v(1 - F_i)\tag{2}$$

although, following the National Research Council [8], the $(1 - F_i)$ is often omitted in the expression for heterozygotes. In these last equations, the symbol $\hat{=}$ means “is estimated by” and the equations provide estimates for population genotype frequencies in terms of population-collection sample allele frequencies. The quantity F_i is the total inbreeding coefficient for population i . It is the average of Equations 2 that would be appropriate if the particular population was not identified and here we

write the average over populations of F_i as F_W , although it is generally written as F_{IT} and is often written as F_{IT} . If there is Hardy-Weinberg equilibrium within populations, then $F_i = \theta_i$, $F_W = \theta_W$.

Regardless of whether the average $(1 - F_W)$ term is included for heterozygotes, it is not clear that estimated profile probabilities are of great forensic interest since the relevant quantities are match probabilities: the chance that an untyped person will have a certain genotype given that a typed person has that genotype. When account is taken of the evolutionary history of a population, $F_i, \theta_i > 0$, the match probabilities exceed the profile probabilities. Match probability estimates require the quantities β_i , or their average β_W .

Since their introduction by Balding and Nichols [7], the match probabilities in Equations 1 have been extended to allow for inbreeding [9], mixtures [10], and relatedness [11]. They have been of substantial benefit for the interpretation of matching autosomal profiles and they were endorsed by the US National Research Council [8].

Estimation of F_{ST}

There is a logical difficulty in estimating $F_{ST} = \beta_W$ or the population-specific values β_i . We have shown how it arises in match probabilities to account for variation in allele frequencies over evolutionary replicate populations. We could estimate β_i from standard methods, such as those of Weir and Cockerham [12] (who assumed $\theta_B = 0$ and $\theta_i = \theta$ for all values of i) or Weir and Hill [13] (who relaxed those assumptions), *if* we had data from more than one actual population to provide an indication of evolutionary variation, but if we had those frequencies we could use them directly in a within-population analysis and not need the θ formulation. We adopt a pragmatic work-around by exploiting the many published sets of allele frequencies, grouping them into broad ancestry or geographic sets: an estimate from a set of European-ancestry samples, for example, can be used in situations where a European-ancestry θ is required.

We take this opportunity to simplify the expressions for our earlier estimation procedures [12],[13] while allowing for different values θ_i for different populations i and for analogous quantities θ_{ij} for pairs i, j of populations. Bhatia et al. [14] found it convenient to work with sample heterozygosities within and among populations, and use the complementary matching proportions (i.e. sample homozygosities). The within-population sample matching proportions can be found from the counts n_{ilu} of allele A_{lu} at locus l sampled from population i as $\tilde{M}_{il} = \sum_u n_{ilu}(n_{ilu} - 1)/[n_{il}(n_{il} - 1)]$, where $n_{il} = \sum_u n_{ilu}$ is the total number of alleles sampled at that locus. The sample allele frequencies are $\tilde{p}_{ilu} = n_{ilu}/n_{il}$, and we can write $\tilde{M}_{il}$ as $n_{il}[\sum_u \tilde{p}_{ilu}^2 - 1/n_{il}]/(n_{il} - 1)$ or, approximately,

as $\tilde{M}_{il} = \sum_u \tilde{p}_{ilu}^2$. For populations i and j , the between population-pair sample matching proportions are $\tilde{M}_{ijl} = \sum_u n_{ilu}n_{jlu}/(n_{il}n_{jl})$ or $\tilde{M}_{ijl} = \sum_u \tilde{p}_{ilu}\tilde{p}_{jlu}$. For sets of r populations, we take averages over populations and over pairs of populations of the matching proportions for locus l : $\tilde{H}_{Wl} = \sum_{i=1}^r \tilde{H}_{il}/r$ and $\tilde{H}_{Bl} = \sum_{i \neq j} \tilde{H}_{ijl}/[r(r-1)]$.

We estimate the β 's by comparing within-population matching proportions to between population-pair matching proportions with the ratios $\hat{\beta}_{il} = (\tilde{M}_{il} - \tilde{M}_{Bl})/(1 - \tilde{M}_{Bl})$ and $\hat{\beta}_{Wl} = (\tilde{M}_{Wl} - \tilde{M}_{Bl})/(1 - \tilde{M}_{Bl})$. Approximating the expectation of these ratios as the ratio of expectations, the Weir and Hill model [13] leads to $\hat{\beta}_{il}, \hat{\beta}_{Wl}$ being unbiased estimates of β_{il}, β_{Wl} .

We will see below that there is variation among loci of the locus-specific values $\hat{\beta}_{il}$ and these differences may reflect both different locus-specific mutation rates as well as large sampling variances. We provide estimates of locus-average values, evaluated as explained in the Supplementary Material, that will be unbiased if each locus has the same value. The average matching proportions over loci are $\tilde{M}_i = \sum_{l=1}^L \tilde{M}_{il}/L$, $\tilde{M}_W = \sum_{i=1}^r \tilde{M}_i/r$, and $\tilde{M}_B = \sum_{l=1}^L \tilde{M}_{Bl}/L$ for a set of L loci. The β estimates are $\hat{\beta}_i = (\tilde{M}_i - \tilde{M}_B)/(1 - \tilde{M}_B)$ and $\hat{\beta}_W = (\tilde{M}_W - \tilde{M}_B)/(1 - \tilde{M}_B)$: these ‘‘ratio of averages’’ estimates have smaller variances than the averages of the $\hat{\beta}_{il}$ or $\hat{\beta}_{Wl}$ ratios.

It is not possible to estimate the within-population θ 's other than as being ‘‘relative to’’ the between-population-pair values. The compound quantities β measure how much more matching there is within populations than there is between pairs of populations. Weir and Hill [13] showed that it is these compound parameters that, for a pure drift model of evolution, are functions of the time since the set of populations diverged from an ancestral population. For two populations i and j in the pure drift case, the parameter $(\beta_i + \beta_j)/2$ is proportional to the time since those two populations diverged from an ancestral population and so it can be used as a distance measure in constructing evolutionary dendrograms (see Supplementary Material).

To illustrate how populations i differ in match probabilities it would be preferable to examine values of θ_i . Although only estimates of θ_i relative to θ_B are available, comparisons are still possible. If $\beta_i > \beta_j$, for example, then it can be inferred that $\theta_i > \theta_j$ and it will be assumed reasonable that $\hat{\beta}_i > \hat{\beta}_j$ also implies $\theta_i > \theta_j$ when the same between-population-pair matching proportion $\tilde{M}_B$ is used for each. In the empirical values shown below, the estimates $\hat{\beta}_i$ for population i , in a group of populations such as those with common continental ancestry, use data from all the populations in that group in order to calculate $\tilde{M}_B$. Within-population match probabilities are appropriately estimated with population-collection sample allele frequencies $\tilde{p}_{lu}$ in place of population frequencies p_{lu} and estimates $\hat{\beta}_W$ values in place of θ_W . The formulation given here of β_W or F_{ST} shows

this quantity describes allelic matching within populations relative to matching between pairs of populations within a collection of populations.

DATA AND DATA HANDLING

Data were obtained from 250 published population reports in the following journals: Forensic Science International, Forensic Science International: Genetics, International Journal of Legal Medicine, Journal of Forensic Sciences, and Legal Medicine. The data were gathered from electronic sources, either from the PDF text or from published electronic files, or from hard copy using OCR and post-processing. We are aware that we have not been comprehensive and that even more data are available. The populations used and the associated references are given in the Supplementary Material. This process yielded 446 distinct populations. The populations and the papers from which we extracted data are also shown in the Supplementary Material. We do not have access to the individual genotypes, so many of the data-cleaning steps described by Pemberton et al. [15] are not available.

The data required a significant amount of post-data handling from cleaning to storage. In particular, not all of the data collected in this study were originally collected for forensic purposes. A proportion was collected more probably for anthropological reasons than validation of forensic multiplexes for casework. The net effect of this fact is that we had some issues regarding non-standard loci and non-standard allele designations, as well as small samples.

Errors in the data were also apparent in a significant number of cases. Most of these errors were typographical and in many cases the correct data could be deduced. In some cases we determined that an error was present because the allele frequencies did not add sufficiently close to one or because allele frequencies multiplied by twice the number of individuals in the sample were not sufficiently close to integers. In one case the allele frequencies had not been correctly determined from the genotype frequencies and in another case D3S1358 and D13S317 had been swapped. In approximately half of the cases of perceived error we received a helpful answer from the corresponding author to our enquiries. In the other cases we have either omitted the data or used our own deductions as to the correct values. We have also struggled with different nomenclatures for rare alleles, however this should have only a minor effect on the analysis. We assigned the populations to geographic groups using our best judgment, and we show below that these groups of populations cluster together with principal component analysis. We therefore have some confidence that the

grouping has both genetic and geographic coherence although we concede that further adjustments could be made. The motivation for grouping at all, apart from simplifying the presentation of results, is that we wish to provide θ values (now interpreted as β_W) that would result in appropriate match probability estimates for any population within a broadly-defined group of populations. The group names, numbers of populations within groups, and sample sizes within groups, are shown in Table 1. The data we discuss represent observations on 494,473 individuals. The loci for which we present results are displayed in Table 2.

RESULTS

Principal Coordinates Analysis

As a first examination of the data, and as a check on our assignment of populations to geographic groups, we performed a principal component analysis of the allele frequencies in the whole data set. We used only those populations for which at least 25 individuals were scored. In Figure 1 we show a parallel coordinates plot of the first 10 principal coordinates. There is clear separation of populations by geographic group: principal coordinate (PC) 1 separates out Asian, Caucasian and Indo-Pakistan groups, PC2 separates out the Hispanic and Native American groups, PC3 separates out the African group, PC4 separates out the Australian Aborigine group, PC5 separates out the Inuits and Polynesians. There does not appear to be any population assigned to the wrong group and the overall impression is that the forensic STR markers can distinguish geographic groups, as has been noted previously [1]. The results depicted in Figure 1 suggest that further F_{ST} -based analyses are appropriate.

Population-specific F_{ST}

A complete list of the population- and locus-specific estimates $\hat{\beta}_{il}$ is given in the Supplementary Material. In Figure 2 we display estimates of the population-specific parameters β_i for populations i . The estimates use all the loci scored in each population, as explained in the Appendix, and we bootstrap over loci to provide the 95% confidence intervals indicated by vertical lines in the figure. The between-population-pair matching values $\tilde{M}_B$ used to estimate the β 's were for all pairs of populations in the survey. The estimates have been ordered by size and colored by geographic group. The smallest values are for the African populations, reflecting the greater diversity within those older populations, whereas the largest values are for the smaller and less diverse Native Amer-

ican and Inuit populations. The Asian values are generally higher than the Caucasian values, with both lying between the African and Native American values. This pattern is not unexpected but is based on larger sets of data than have been used previously. Even when up to 24 loci are used to estimate β_i , there are large sampling variances. We now turn to a closer look among loci and among geographic groups.

Locus-specific F_{ST}

In Figure 3 we display box-plots of the estimates of β_{il} for loci l and populations i . The between-population pair matching values $\tilde{M}_B$ used to estimate these β 's were for all pairs of populations in the survey. The colored dots show the mean values over populations for each of the nine geographic groups.

There is relatively little variation over loci in the averages over populations of locus-specific values, but there is substantial variation in the range of values over populations at each locus. The variation appears to reflect the numbers of geographic groups and/or numbers of populations typed at each locus. Locus D3S1358, for example, was typed in over 400 populations from all geographic groups, whereas locus D6S1043 was typed in only eight populations in four geographic groups. This sampling range seems to mask any effect of mutation rate in the figure. The loci have been ordered on the X axis in decreasing order of published mutation rates (Table 14.5 in [16]): SE33 on the left has the highest reported rate, of 0.0016, whereas the D1S1656, D2S441, D10S1248, D22S1045 loci on the right have no reported mutations.

Geographic-group-specific F_{ST}

In Figure 4 we display population-specific estimates of β_i for each continental group of populations. Two sets of estimates are shown: those on the right, with a shaded box plot, use only the populations in the group to calculate the between-population-pair matching proportion $\tilde{M}_B$. The resulting β_W estimate is therefore the average over populations of θ relative to that group or region, and we could write the estimates as $\hat{\beta}_{PR}$. These are the estimates to be used if the relevant group of populations is known. The estimates on the left for each group, with an open box, use all pairs of populations across all groups to calculate $\tilde{M}_B$. Now the estimates are relative to the total collection of populations and we could write them as $\tilde{\beta}_{PT}$. These are appropriate if there was no information about the group to which the relevant population belongs.

As the population versus total estimates are to be used when there is no information about the

group to which a population belongs, it is reasonable that those estimates $\tilde{\beta}_{PT}$ are greater than the population versus region estimates $\hat{\beta}_{PR}$: this increases the matching probability and reduces the evidentiary strength of a match. A substantial difference in the two estimates $\hat{\beta}_{PT}$ and $\hat{\beta}_{PR}$, is seen for the Inuit group of populations: the average within-population matching proportion, averaged over loci, is $\tilde{M}_W = 0.4379$ whereas the average between-population-pair matching proportions are 0.1726 for pairs within the group and 0.0090 for all pairs in the study. The Inuit populations are more similar to each other than are any pair of populations in the study. We find that $\hat{\beta}_{PR} = 0.0205$ and $\hat{\beta}_{PT} = 0.1057$. The theta correction has much less effect if attention can be confined to the Inuit group.

The values for the African group show the opposite relationship from the Inuit group. Now the average within-population matching proportion is $\tilde{M}_W = 0.1884$ and the average between-population-pair averages are 0.1691 within the African region and 0.1726 for all pairs of populations. There is more divergence between pairs of African populations than there is among all pairs of populations and $\hat{\beta}_{PR} = 0.0082$ and $\hat{\beta}_{PT} = 0.0020$. The theta corrections for African populations have little effect regardless of the reference group.

The locus- and group-specific values of $\hat{\beta}_{WI}$ are shown in Tables 3 and 4, where the group or the world serve as reference collections of populations. We advocate the use of the locus-average values. Except for Africa and India-Pakistan, the values are greater when the world rather than the group serves as a reference, and we note the different rankings of the groups in the two tables: Africa clearly has the lowest average value in Table 4 whereas it is comparable to the Australian Aboriginal and Hispanic values in Table 3. The Inuit group has the highest value in Table 4, but is second behind Native Americans in Table 3.

From the perspective of a forensic scientist wanting to assign an F_{ST} value for a broad geographic group, if an upper limit (the upper end of each whisker in Figure 4) was considered to be the most appropriate [3], then we would suggest values around 0.05 except for Native Americans and Inuits when the value is above 0.10. Rather than the upper limit, however, we have some preference for the median values (the center of each box in Figure 4), and these suggest values around 0.01–0.03 except for Native Americans and Inuits when the value are higher and more susceptible to the group or the world being the reference set. The table in the Supplementary Material shows much higher variation for locus- and population-specific values. As we discuss below, our recommendation is to use the multi-locus values in the last lines of Tables 3 and 4.

DISCUSSION

We have presented an analysis of an extensive set of autosomal forensic STR allele frequencies with the aim of assisting forensic scientists assign F_{ST} or theta values for use in match probability calculations.

If allele frequencies $\check{p}_{iu}$ are known for the population thought to be relevant, meaning that it is the population from which an unknown contributor to an evidence profile is supposed to be drawn randomly, then those frequencies may be used directly. Assuming independence of alleles within and between loci, and assuming the population is large, the probability of the two profiles (for the person of interest and the unknown person with a matching profile) is just the product of these frequencies and the match probability is the same as the profile probability.

Allele frequencies in a population are not known, however, and sampling variation must be considered. For reasonably large sample sizes, the evolutionary variation is larger than the sampling of individuals from the population and the “theta corrections” of Equations 1 are appropriate. Values of θ_i need to be assigned: they cannot be estimated with data from a single population. Taking evolutionary variation into account, in effect, allows the evolutionary relatedness of the person of interest and the unknown person to be accommodated. If the relevant population is not sampled, but allele frequencies are available from some collection of populations, maybe of some specific geographic region or ancestry, then θ is replaced by β , with an average value β_W seeming most appropriate.

Silva et al. [1] showed that measures of population structure based on STR data provide good separation of geographic groups whether the loci were selected for evolutionary or forensic studies, with attenuated separation for forensic markers because they have been chosen to have the high diversity needed for individual identification. These authors made most use of the statistic R_{ST} , in which alleles are compared by the squared difference in the number of repeat units, rather than simply matching or not. As did these authors we also examined the survey of non-forensic STR data reported by Pemberton et al. [15] and, like them, found similar but attenuated separation for forensic markers with β instead of R_{ST} (results not shown). It is interesting that “global” F_{ST} value in [1] of 0.027 from allele frequency data is comparable to our 0.024 value shown in Table 4. Silva et al. did not focus on estimation of match probabilities.

Steele et al. [3] reported F_{ST} values for a collection of populations self-identified by UK residents or potential immigrants. Their “direct” method of estimation made use of data from a reference

population with similar continental ancestry for each population under consideration, so their approach is closest to our $\hat{\beta}_{PR}$ values shown in Table 3. Their “indirect” method uses allele frequencies from the whole set of populations and so corresponds to the approach we used to generate the $\hat{\beta}_{PT}$ values shown in Table 4. Steele et al. use a Bayesian method and we would expect the medians of their posterior distributions to be comparable to the moment estimates we display. They concluded with a general recommendation of an F_{ST} of no more than 0.03 being generally applicable. For populations within a group, we are in general agreement with Steele et al, provided data from the appropriate group are used. If a world-wide dataset is to be used then higher values are necessary for Native American and Inuit populations. For individual Asian populations, however, we often find higher values than those shown in [3].

Steele and Balding [17] repeated the recommendation in [3] that an F_{ST} of 0.03 is sufficiently large to be almost always conservative. They advocate the use of a database “most appropriate” for the unknown donor of an evidentiary sample.

As the forensic task is to predict profile matching probabilities from an available set of data, we find intuitive appeal in phrasing our approach in terms of allelic matching proportions, especially by emphasizing the need to consider matching within and between populations. A matching proportion within a population is high in the present context only if it is substantially higher than matching between pairs of populations. The importance of the appropriate terms of reference is illustrated by the plots in Figure 4.

The situation most likely to confront a forensic scientist is having to decide on a value of F_{ST} for match probability predictions for a population when the allele frequencies are going to be based on a sample from a collection of populations of similar ancestry or geographic region. We show the locus-specific values for each region in Tables 3 and 4. These $\hat{\beta}_{Wl}$ values are highly variable across loci and values are ranked differently within each region. We note a few negative locus-specific values, indicating less allelic matching at those loci within populations than between pairs of populations in a region. Because of the large variances in locus-specific values, we place little confidence in any one of them. We suggest, instead, the values we show that are based on all the loci typed for a population. In Figure 2 we showed the 95% confidence intervals based on bootstrapping over loci and we note that Steele et al. [3] advocated using quantities analogous to the upper ends of these intervals. Our recommendation, however, is to use the all-loci values in the last rows of Tables 3 and 4.

We have approached allelic match probabilities from the expression $\Pr(A_u A_u) = \check{p}_{iu}^2$ for pairs

of alleles in a population, i , for which each allele has probability $\check{p}_{iu}$ of being of type A_u . The actual population frequencies $\check{p}_{iu}$ are generally unknown, so we took expectations over replicates of the evolutionary process to obtain $\Pr(A_u A_u) = p_u^2 + p_u(1 - p_u)\theta_i$. If only an estimate $\tilde{p}_u$ is available for p_u , we had the estimate $\tilde{p}_u^2 + \tilde{p}_u(1 - \tilde{p}_u)\beta_W$. Steele et al. [2] also considered the case of matching probabilities for individuals in two populations i, j . Our development leads to $\Pr(A_u A_u) = p_u^2 + p_u(1 - p_u)\theta_{ij}$ in that case, with sample values requiring $\beta_{ij} = (\theta_{ij} - \theta_B)/(1 - \theta_B)$. If populations i, j are in the same region R then we have an estimation problem: we cannot estimate β_{ij} other than by comparison to the average θ_B over all pairs of populations in that region and the only possibility we have is to compare each $\hat{\beta}_{ij} = (\tilde{M}_{ij} - \tilde{M}_B)/(1 - \tilde{M}_B)$ to their average. This average is zero by construction. The reference point for a single population is (the average over) pairs of populations: for a single region there is no meaningful reference point for pairs of populations. We can make progress in estimating β_{ij} for two populations in the same region by taking θ_B to refer to pairs of populations, one from each of two regions. If we have a hierarchical structure limited to populations and regions, however, we cannot address estimation of θ_{ij} for populations in different regions. Steele et al. [2] do not address the decomposition of F_{ST} into components within and between populations and so do not fully address the issue of allelic matching between populations.

The concept we stress in this paper is that forensic match probabilities, designed to recognize the shared ancestry between a person of interest and an unknown donor of an evidentiary sample, can be expressed in terms of identity by descent of pairs of alleles drawn from the relevant population. If sample allele frequencies, drawn from a larger set of populations, are to be used in calculating match probabilities then it is necessary to consider the identity between alleles drawn from pairs of these populations. It is the compound parameter F_{ST} , comparing allelic identity within populations to that between populations, that is of forensic importance. The value of this parameter clearly depends on the set of populations furnishing allele frequencies. The same point was made by Steele et al. [1] and we endorse their suggestion for the sampled set of populations to be generally of similar ancestry to the population of interest. A possible exception is for populations of African ancestry that tend to have retained greater diversity than those in the rest of the world.

ACKNOWLEDGMENTS

This work was supported in part by grants 2011-DN-BX-K541 and 2014-DN-BX-K028 from the US National Institute of Justice, and grant IZK0Z3_157867 from the Swiss National Science Foundation. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Departments of Justice or Commerce or of the Swiss National Science Foundation.

LITERATURE CITED

- [1] Silva NM, Pereira L, Poloni ES, Currat M. 2012. Human neutral genetic variation and forensic STR data. PLoS One 7:e49666.
- [2] Pamplona JP, Freitas F, Pereira L. 2008. A worldwide database of autosomal markers used by the forensic community. Forensic Science International: Genetics Supplement Series 1:656-657.
- [3] Steele CD, Syndercombe Court D, Balding DJ. 2014. Worldwide F_{ST} estimates relative to five continental-scale populations. Annals of Human Genetics 78:468-477.
- [4] Stringer CB, Andrews P. 1988. Genetic and fossil evidence for the origin of modern humans. Science 239:1263-1268.
- [5] Cockerham CC. 1969. Variance of gene frequencies. Evolution 23:72-84.
- [6] Wright S. 1931. Evolution in Mendelian populations. Genetics 16:97-159.
- [7] Balding DJ, Nichols RA. 1994. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Science International 64:125-140.
- [8] National Research Council. 1996. National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence. National Academy Press, Washington, D.C.
- [9] Ayres KL, Overall ADJ. 1999. Allowing for within-subpopulation inbreeding in forensic match probabilities. Forensic Science International 103:207-216.
- [10] Curran JM, Triggs CM, Buckleton JS, Weir BS. 1999. Interpreting DNA mixtures in structured populations. Journal of Forensic Sciences 44:987-995.
- [11] Weir BS. 2007. The rarity of DNA profiles. Annals of Applied Statistics 1:358-370.
- [12] Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.
- [13] Weir BS, Hill WG. 2002. Estimating F -statistics. Annual Reviews of Genetics 36:721-750.

- [14] Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting F_{ST} : the impact of rare variants. *Genome Research* 23:1514-1521.
- [15] Pemberton TJ, DeGiorgio M, Rosenberg NA. 2013. Population structure in a comprehensive genomic data set on human microsatellite variation. *G3 Genes, Genomes, Genetics* 3:891-907.
- [16] Butler JM. 2014. *Advanced Topics in Forensic DNA Typing: Interpretation*. Academic Press, New York.
- [17] Steele DC, balding DJ. 2014. Choice of population database for forensic DNA profile analysis. *Science and Justice* 54:487-493.

Table 1 Sample characteristics for geographic groups.

Group	Name	Number of		Sample size [†]	
		Populations	Loci*	Minimum	Maximum
Africa	African	37	24	342	7998
Andaman Islands	Andam	3	9	61	97
Asian	Asian	85	24	1153	23557
Australian Aborigine	AusAb	17	15	1686	18441
Caucasian	Caucn	173	24	861	233753
Hispanic	Hisp	41	24	236	167872
India Pakistan	IndPk	30	23	488	4525
Inuit	Inuit	3	21	194	403
Native American	NatAm	34	17	618	4234
Polynesian	Polyn	4	15	2614	33593
Unknown	Unknn	15	24	205	9072

* Number of loci scored in at least one population in the group.

[†] Number of individuals scored at a single locus over all populations in the group.

Table 2 Numbers of alleles and numbers of populations for each locus.

Locus	No. of	Number of populations per geographic group									
	Alleles	African	Andam	Asian	AusAb	Caucn	Hisp	IndPk	Inuit	NatAm	Polyn
CSF1PO	27	33	—	76	2	109	39	25	2	33	2
D1S1656	26	3	—	7	—	34	2	2	1	—	—
D2S441	33	3	—	9	—	36	2	2	1	—	—
D2S1338	33	20	—	63	2	97	21	14	1	25	4
D3S1358	30	36	3	82	17	161	41	27	3	31	4
D5S818	25	34	2	81	17	131	40	27	2	34	2
D6S1043	32	1	—	4	—	2	1	—	—	—	—
D7S820	37	34	3	81	17	131	40	27	2	34	2
D8S1179	30	37	3	77	17	161	41	27	3	34	4
D10S1248	16	3	—	9	—	36	2	2	1	—	—
D12S391	38	3	—	7	—	31	3	2	1	—	—
D13S317	28	34	3	81	17	131	40	27	2	34	2
D16S539	26	35	—	74	2	135	39	28	3	30	4
D18S51	56	36	2	80	17	161	41	28	3	34	4
D19S433	36	20	—	65	2	99	21	14	1	25	4
D21S11	70	37	3	80	17	161	41	28	3	34	4
D22S1045	13	3	—	8	—	36	2	2	1	—	—
FGA	86	37	3	81	17	162	41	27	3	34	4
PENTAD	40	6	—	22	—	37	10	7	—	4	—
PENTAE	46	6	—	23	—	37	10	7	—	4	—
SE33	85	2	—	3	—	11	2	2	1	—	—
TH01	30	36	—	77	2	142	41	28	3	34	4
TPOX	22	33	—	75	2	109	39	25	2	34	2
vWA	31	36	3	81	17	162	41	30	2	34	4

Table 3 Estimated values $\hat{\beta}_{Wl}$ of locus-specific β 's for each geographic region, and the value for all loci within a region. Each estimate is for θ_W relative to pairs of populations within that region. There are no estimates if a locus is scored in less than two populations for a region, or if less than 25 individuals were typed in a region. Values in the body of the table are for each population, averaged over populations in a region. The bottom row is the average over loci. The method of averaging is described in the Supplementary Material.

Locus	Africa	AusAb	Asian	Caucn	Hispn	InPak	NatAm	Inuit	Polyn
CSF1PO	0.0042	0.0020	0.0073	0.0031	0.0018	0.0064	0.0256	0.0066	0.0256
D1S1656	0.0047	—	0.0027	0.0039	0.0032	-0.0018	—	—	—
D2S441	0.0129	—	0.0082	0.0032	0.0212	0.1933	—	—	—
D2S1338	0.0101	0.0039	0.0129	0.0075	0.0097	0.0158	0.0661	—	0.0051
D3S1358	0.0042	0.0002	0.0104	0.0044	0.0111	0.0178	0.0548	0.0058	0.0014
D5S818	0.0044	0.0018	0.0090	0.0055	0.0125	0.0109	0.0553	0.0127	0.0073
D6S1043	—	—	0.0051	0.0005	—	—	—	—	—
D7S820	0.0055	0.0008	0.0117	0.0064	0.0070	0.0067	0.0282	0.0092	0.0047
D8S1179	0.0050	0.0005	0.0123	0.0059	0.0021	0.0122	0.0197	0.0273	0.0077
D10S124	0.0018	—	0.0137	0.0026	0.0036	-0.0018	—	—	—
D12S391	0.0090	—	0.0039	0.0016	0.0204	-0.0018	—	—	—
D13S317	0.0068	0.0008	0.0157	0.0071	0.0144	0.0062	0.0335	0.0220	0.0136
D16S539	0.0085	0.0015	0.0205	0.0047	0.0094	0.0049	0.0343	0.0402	0.0042
D18S51	0.0047	0.0003	0.0145	0.0050	0.0025	0.0053	0.0306	0.0076	0.0026
D19S433	0.0103	0.0011	0.0103	0.0091	0.0136	0.0002	0.0310	—	0.0143
D21S11	0.0151	0.0001	0.0140	0.0055	0.0064	0.0128	0.0530	0.0002	0.0184
D22S104	0.0270	—	0.0498	0.0017	-0.0022	-0.0018	—	—	—
FGA	0.0033	0.0002	0.0116	0.0038	0.0058	0.0076	0.0235	0.0147	0.0063
PENTAD	0.0075	—	0.0145	0.0107	0.0022	0.0171	0.0259	—	—
PENTAE	0.0015	—	0.0152	0.0056	0.0029	0.0115	0.0217	—	—
SE33	0.0062	—	0.0084	0.0264	-0.0009	-0.0018	—	—	—
TH01	0.0209	0.0006	0.0237	0.0176	0.0176	0.0128	0.0640	0.0366	0.0168
TPOX	0.0104	0.0871	0.0164	0.0076	0.0074	0.0174	0.0511	0.0359	0.0307
VWA	0.0042	0.0002	0.0114	0.0046	0.0076	0.0059	0.0198	0.0479	0.0045
All	0.0081	0.0064	0.0133	0.0066	0.0077	0.0171	0.0368	0.0199	0.0106

Table 4 Estimated values $\hat{\beta}_{Wl}$ of locus-specific β 's for each geographic region, and the value for all loci within a region. Each estimate is for θ_W relative to all pairs of populations in the survey. There are no estimates if a locus is scored in less than two populations for a region, or if less than 25 individuals were typed in a region. Values in the body of the table are for each population, averaged over populations in a region. The right-most column is the average over regions, and the bottom row is the average over loci. The method of averaging is described in the Supplementary Material.

	Africa	AusAb	Asian	Caucn	Hisp	IndPk	NatAm	Inuit	Polyn	World
CSF1PO	-0.0668	0.0130	0.0154	0.0127	0.0165	0.0197	0.0616	0.0406	0.0291	0.0117
D1S1656	0.0339	—	0.0658	-0.0018	0.0189	0.0316	—	0.0812	—	0.0157
D2S441	0.0153	—	0.0265	0.0316	0.1005	-0.0285	—	0.1625	—	0.0332
D2S1338	0.0029	0.0313	0.0319	0.0129	0.0234	0.0134	0.1210	0.1255	0.0035	0.0292
D3S1358	0.0145	0.0279	0.0578	-0.0345	0.0239	0.0227	0.2200	0.2196	0.0426	0.0254
D5S818	0.0102	-0.0229	-0.0132	0.0465	0.0474	0.0197	0.1192	0.0461	-0.0243	0.0337
D6S1043	-0.0006	—	0.0126	0.0669	0.0030	—	—	—	—	0.0233
D7S820	0.0244	0.0557	0.0345	0.0001	0.0165	0.0039	0.0842	0.0443	-0.0078	0.0222
D8S1179	0.0405	-0.0153	-0.0187	0.0169	0.0273	-0.0207	0.0885	0.1264	0.0227	0.0179
D10S1248	-0.0397	—	0.0383	0.0047	0.0473	-0.0195	—	0.1345	—	0.0102
D12S391	0.0317	—	0.0448	-0.0097	0.0745	0.0258	—	0.0522	—	0.0120
D13S317	0.1221	0.0806	0.0235	0.0445	0.0051	0.0093	0.0252	0.0990	0.0375	0.0384
D16S539	-0.0018	0.0597	0.0237	0.0288	0.0093	-0.0025	0.0720	0.1635	0.0227	0.0250
D18S51	-0.0012	0.0064	0.0345	0.0064	0.0026	0.0323	0.0503	0.0733	0.0538	0.0181
D19S433	-0.0095	0.1661	0.0226	0.0410	0.0053	0.0166	0.0132	-0.0013	0.0015	0.0254
D21S11	-0.0076	-0.0225	0.0422	0.0084	0.0126	0.0013	0.0702	0.0492	0.0393	0.0200
D22S1045	-0.0626	—	-0.0078	0.0300	0.0872	-0.0211	—	0.0836	—	0.0204
FGA	0.0027	0.0038	0.0183	0.0164	0.0011	0.0072	0.0226	0.0296	0.0655	0.0142
PENTAD	-0.0402	—	0.0567	0.0180	0.0015	0.0160	0.0380	—	—	0.0227
PENTAE	0.0185	—	0.0163	0.0235	0.0136	0.0137	0.0409	—	—	0.0202
SE33	0.0234	—	0.0138	0.0205	0.0152	0.0041	—	0.1081	—	0.0219
TH01	0.0731	0.0679	0.1465	0.0189	0.0369	0.0199	0.2084	0.5200	0.0464	0.0755
TPOX	-0.1336	-0.0031	0.0911	0.0578	0.0064	-0.0369	0.0736	0.0395	0.0412	0.0339
VWA	-0.0021	0.0246	0.0195	0.0087	0.0373	0.0055	0.0808	0.0231	0.0213	0.0198
All loci	0.0038	0.0328	0.0328	0.0193	0.0258	0.0065	0.0804	0.1050	0.0265	0.0244

See attached file: ParallelCoordinatesPlot.pdf

Figure 1 Parallel coordinate plot for first 10 principal coordinates for all populations with sample sizes at least 50. Each line in the plot represents one population. Color code: Black=African, Grey=AusAb, Yellow=Asian, Blue=Caucn, Purple=Hisp, Brown=IndPk, Red=NatAm, Orange=Inuit, Brown=Andam, Green=Polyn.

Figure 2 Estimated values of the population-specific β_i , ordered by size. Each vertical line in the plot represents one population, and the length of the line is the 95% confidence interval obtained by bootstrapping over loci.

Figure 3 Estimated values of the locus-specific β_{il} , ordered by locus. The black dots, and box-plots with whiskers, are for populations. The colored dots are for geographic groups. The box plots show the medians and interquartile ranges: the whiskers extend out from the box plots to 1.5 times the interquartile range.

Figure 4 Estimated values of the population-specific β 's, for each geographic group, using all loci. Each box plot indicates the inter-quartile variation among populations within the group, with whiskers extending out from the box plots by 1.5 times the interquartile range. For each geographic group, the left hand plot (white box plot) compares each within-population matching proportion to the average matching proportion among all pairs of populations. The right hand plot (grey box plot) compares each within-population matching proportion to the average matching proportion among all pairs of populations in that region.

APPENDIX

Combining Estimates over Loci and Populations

The within- and between-population-pair matching proportions for populations i and population-pairs i, j at locus l are $\tilde{M}_{il}$ and $\tilde{M}_{ijl}$. These have expectations $[1 - H_l(1 - \theta_{il})]$ and $[1 - H_l(1 - \theta_{ijl})]$ respectively and they provide the population/locus specific estimates

$$\hat{\beta}_{il} = \frac{\tilde{M}_{il} - \tilde{M}_{Bl}}{1 - \tilde{M}_{Bl}}$$

where $\tilde{M}_{Bl} = \sum_i \sum_{j \neq i} M_{ijl} / [r_l(1 - r_l)]$ if locus l is scored in r_l populations. These estimates have expectation $\beta_{il} = (\theta_{il} - \theta_{Bl}) / (1 - \theta_{Bl})$ if $\theta_{Bl} = \sum_i \sum_{j \neq i} \theta_{ijl} / [r_l(1 - r_l)]$.

Averaging over populations. Averaging matching proportions leads to average estimates over populations for locus l . If locus l is scored in r_l populations:

$$\begin{aligned} \tilde{M}_{Wl} &= \frac{1}{r_l} \sum_i \tilde{M}_{il} \\ \hat{\beta}_{Wl} &= \frac{1}{r_l} \sum_i \hat{\beta}_{il} = \frac{\tilde{M}_{Wl} - \tilde{M}_{Bl}}{1 - \tilde{M}_{Bl}} \\ \mathcal{E}(\hat{\beta}_{Wl}) &= \frac{\theta_{Wl} - \theta_{Bl}}{1 - \theta_{Bl}} \end{aligned}$$

Here the sums are over populations for which locus l is scored (e.g. $\theta_{Wl} = \sum_i \theta_{il} / r_l$), and the estimate is for the average over populations of the θ_{il} values relative to a reference value θ_{Bl} . These estimates allow a comparison among loci *if* the same populations are scored for each locus. In reality, this may not be the case and comparisons need to be interpreted carefully, although there is less of an issue as the number of populations increases.

Averaging over loci. Combining estimates over loci for population i is not as straightforward if the θ 's differ among loci. A unweighted average over loci is

$$\hat{\beta}_i^u = \frac{1}{L_i} \sum_l \hat{\beta}_{il}$$

where L_i is the number of loci scored in population i . This average has expectation

$$\mathcal{E}(\hat{\beta}_i^u) = \frac{1}{L_i} \sum_l \beta_{il} = \frac{1}{L_i} \sum_l \frac{\theta_{il} - \theta_{Bl}}{1 - \theta_{Bl}}$$

This is not the average over loci of the θ_{il} values relative to a reference value.

Variance is reduced by weighting estimates by their denominators:

$$\hat{\beta}_i^w = \frac{\sum_l (\tilde{M}_{il} - \tilde{M}_{Bl})}{\sum_l (1 - \tilde{M}_{Bl})} \quad , \quad \mathcal{E}(\hat{\beta}_i^w) = \frac{\sum_l H_l (\theta_{il} - \theta_{Bl})}{\sum_l H_l (1 - \theta_{Bl})}$$

If there is no variation among loci, $\theta_{il} = \theta_i$, and $\theta_{Bl} = \theta_B$, as might be expected for neutral loci with similar mutation rates, each of $\hat{\theta}_i^u, \hat{\theta}_i^w$ have the same expected value of $(\theta_i - \theta_B)/(1 - \theta_B)$.

A single estimate $\hat{\beta}_W$, for a set of populations, using all the scored loci, can be calculated as

$$\begin{aligned} \hat{\beta}_W^w &= \frac{\tilde{M}_W - \tilde{M}_B}{1 - \tilde{M}_B} \\ &= \frac{\sum_l (\tilde{M}_{Wl} - \tilde{M}_{Bl})}{\sum_l (1 - \tilde{M}_{Bl})} \\ \mathcal{E}(\hat{\beta}_W^w) &= \frac{\sum_l H_l (\theta_{Wl} - \theta_{Bl})}{\sum_l H_l (1 - \theta_{Bl})} \end{aligned}$$

An unweighted single estimate that does not have an expectation dependent on H_l is

$$\begin{aligned} \hat{\beta}_W^u &= \frac{1}{L} \sum_l \hat{\beta}_{Wl} = \frac{1}{L} \sum_l \frac{\tilde{M}_{Wl} - \tilde{M}_{Bl}}{1 - \tilde{M}_{Bl}} \\ \mathcal{E}(\hat{\beta}_W^u) &= \frac{1}{L} \sum_l \frac{\theta_{Wl} - \theta_{Bl}}{1 - \theta_{Bl}} \end{aligned}$$

This is not the average value over loci, θ_W , relative to a reference value. If $\theta_{il} = \theta_i$ for all l , as might be expected for neutral loci with equal mutation rates, each of $\hat{\beta}_W^w$ and $\hat{\beta}_W^u$ has the same expected value $(\theta_W - \theta_B)/(1 - \theta_B)$.

Which Estimate Should be Used?

The results shown in this paper show variation in $\hat{\beta}_{il}$, suggesting variation in θ_{il} , over both loci and populations. The estimates are ratios of quadratic forms and large variances are expected. The figures suggest more variation over populations than over loci and this is not unexpected. One goal of this paper is to suggest values of “theta” for the calculations of match probabilities with the “theta correction.” We have shown that the appropriate quantity is β_W , an average over populations in some relevant group of populations.

Our interest is primarily in the effect of population, rather than locus, on match probabilities. We would prefer to use a single theta, or β , value than separate values for each locus. [There is

a need for separate values for each locus if we wish, for example, to detect signatures of natural selection at individual loci.] Which value should be used? The mean or median, or some other percentile, of the collection of single-locus values for a population or group of populations are valid measures, and have been advocated by Balding and his colleagues. Collections of values are shown in Figures 2 and 3. Extensive sets of numerical values are shown in Supplementary Material section 4. We argue instead for the best single value calculated from all the loci, meaning the $\hat{\beta}_W^w$ values with smallest variance, and these are the values shown in Tables 3 and 4.

Figure 1 Parallel coordinate plot for first 10 principal coordinates for all populations with sample sizes at least 50. Each line in the plot represents one population. Color code: Black=African, Grey=AusAb, Yellow=Asian, Blue=Caucn, Purple=Hisp, Brown=IndPk, Red=NatAm, Orange=Inuit, Brown=Andam, Green=Polyn.

Figure

Table

	ShortName	Name for tree	Cluster	Reference
1	US Af. 10	Af Am Minnesota	Afric	[1]
2	Bah. Af.	Bahamas	Afric	[2]
2	Belarus 1	Belarusians	Caucn	[3]
2	Berber 2	Berber	Caucn	[4]
3	US Af. 3	Af Am FBI	Afric	[2]
4	US Af. 4	Af Am Florida	Afric	[2]
5	US Af. 5	Af Am Illinois	Afric	[2]
6	US Af. 6	Af Am Minnesota	Afric	[2]
7	Jam. Af.	Jamaica	Afric	[2]
8	US Af. 7	Af Am New York	Afric	[2]
9	US Af. 8	Af Am Virginia	Afric	[2]
10	Tunisia 1	Tunisia	Caucn	[5]
11	US Af. 9	Af Am US	Afric	[6]
12	China 1	Chinese California	Asian	[2]
13	Asia	Asian US	Asian	[2]
14	Japan 1	Japanese California	Asian	[2]
15	Japan 2	Japanese US	Asian	[2]
16	Korea 1	Korean California	Asian	[2]
17	Vietnam 2	Vietnamese Hanoi	Asian	[7]
18	US Cauc. 10	Cauc Minnesota	Caucn	[1]
19	US Cauc. 3	Cauc FBI	Caucn	[2]
20	Can. Cauc.	Cauc Canada	Caucn	[2]
21	US Cauc. 4	Cauc Florida	Caucn	[2]
22	US Cauc. 5	Cauc Michigan	Caucn	[2]
23	US Cauc. 6	Cauc Minnesota	Caucn	[2]
24	US Cauc. 7	Cauc New York	Caucn	[2]
25	US Cauc. 8	Cauc Virginia	Caucn	[2]

26	US Cauc. 9	Cauc US	Caucn	[6]
27	US Hisp. 10	Hisp Minnesota	Hisp	[1]
28	US Hisp. 3	Hisp FBI	Hisp	[2]
29	US Hisp. 4	Hisp Florida	Hisp	[2]
30	US Hisp. 5	Hisp Mexico	Hisp	[2]
31	US Hisp. 6	Hisp Michigan	Hisp	[2]
32	US Hisp. 7	Hisp Minnesota	Hisp	[2]
33	US Hisp. 8	Hisp New York	Hisp	[2]
34	US Hisp. 9	Hisp JUS	Hisp	[6]
35	Nat. Am. 1	Apache	NatAm	[2]
36	Nat. Am. 2	NA Michigan	NatAm	[2]
37	Nat. Am. 3	NA Minnesota	NatAm	[2]
38	Nat. Am. 4	Navajo	NatAm	[2]
39	Nat. Am. 5	NA Northern	NatAm	[2]
40	Nat. Am. 6	Salishan	NatAm	[2]
41	Nat. Am. 7	NA Saskatchewan	NatAm	[2]
42	Rwanda	Hutus	Afric	[8]
43	China 2	Singapore	Asian	[9]
44	Greece 1	Greek	Caucn	[10]
45	Spain 1	Spain Andalucia	Caucn	[11]
46	Turkey 2	Turkish	Caucn	[12]
47	Turkey 3	Turkish E Anatolia	Caucn	[13]
48	Inuit 1	NA Athabaskans	NatAm	[14]
49	Inuit 2	Inuit Inupiat	Inuit	[14]
50	Inuit 3	Inuit Yupik	Inuit	[14]
51	Colombia 1	Columbia Andean and Amazonian and Orinoquian	Afric	[15]
52	Colombia 2	Columbia Caribbean coast	Hisp	[15]
53	Colombia 3	Columbia North Colombian Pacific coast and	Hisp	[15]

		Caribbean Island		
54	Colombia 4	Columbia South Andean Occidental	Hispc	[15]
55	Brazil 1	Brazil	Hispc	[16]
56	Iceland	Iceland	Caucn	[17]
57	Angola 1	Angola	Afric	[18]
58	Serbia 1	Serbia	Caucn	[19]
59	Brazil 2	Brazil Espirito Santo	NatAm	[20]
60	Sweden 1	Sweden	Caucn	[21]
61	E. Slovak 1	Slovak Saris	Caucn	[22]
62	E. Slovak 2	Slovak Zemplin	Caucn	[22]
63	E. Slovak 3	Slovak Abov-Gemer	Caucn	[22]
64	E. Slovak 4	Slovak Spiš	Caucn	[22]
65	China 3	China Manchu	Asian	[23]
66	Brazil 3	Brazil Rio de Janerio	Hispc	[24]
68	Morocco 1	Morocans in Belgium	Caucn	[25]
69	Hungary 1	Hungary	Caucn	[26]
70	Nicaragua	Nicaragua	Afric	[27]
71	Italy 1	Italy	Caucn	[28]
72	Hungary 2	Hungary	Caucn	[29]
73	UK Cauc. 2	Cauc UK	Caucn	[30]
74	Afghanistan 1	Afghan Hazaras	IndPk	[31]
75	Afghanistan 2	Afghan Ouzbeks	IndPk	[31]
76	Afghanistan 3	Afghan Pachtouns	IndPk	[31]
77	Afghanistan 4	Afghan Tadjik	IndPk	[31]
78	Afghanistan 5	Afghan Turkmen	IndPk	[31]
79	Austria 1	Austria	Caucn	[32]
80	Argentina 1	Argentina	Hispc	[33]

81	China 4	China Du Long	Asian	[34]
82	China 6	China Hunan	Asian	[35]
83	Italy 2	Italy S	Caucn	[36]
84	Haiti	Haiti	Afric	[37]
85	S. Af. Colored	S Af Coloured	Afric	[38]
86	Khoe	Khoe	Afric	[38]
87	San	San	Afric	[38]
88	S. Africa 1	S Af SE African	Afric	[38]
89	S. Africa 2	S Af SW African	Afric	[38]
91	Macedonia 1	Macedonia	Caucn	[39]
92	Brazil 4	Brazil	Hispc	[40]
93	UK Cauc. 3	Cauc UK	Caucn	[41]
94	Romani 1	Romani SE Romania	Caucn	[42]
95	UK Af. 2	Af Caribb UK	Afric	[41]
96	Poland 1	Poland central	Caucn	[43]
97	Maldives	Maldives	IndPk	[44]
98	El Salvador	El Salvador	Hispc	[45]
99	Portugal 1	Portugal N	Caucn	[46]
100	China 7	China Uygur	Asian	[47]
101	Spain 2	Spain Galicia	Caucn	[48]
102	Italy 3	Italy Calabria	Caucn	[49]
103	China 8	China Gelao	Asian	[50]
104	Portugal 2	Portugal central	Caucn	[51]
105	Hungary 3	Hungarians in Serbia	Caucn	[52]
106	China 9	China Han	Asian	[50]
108	China 10	China Jing	Asian	[50]
109	China 11	China Shui	Asian	[50]

110	China 12	China Zhuang	Asian	[50]
111	Portugal 3	Portugal S	Caucn	[53]
112	China 13	China han Shanghai	Asian	[54]
113	Italy 4	Italy Apulia	Caucn	[55]
114	China 14	China Manchu Jilin Province	Asian	[56]
115	Spain 3	Spain	Caucn	[57, 58]
116	Korea 2	Korea	Asian	[59]
117	Italy 5	Italy Sicilian	Caucn	[60]
118	Nat. Am. 8	NA Cora	NatAm	[61]
119	Nat. Am. 9	NA Durango	NatAm	[61]
120	Nat. Am. 10	NA Jalisco	NatAm	[61]
121	Nat. Am. 11	NA Nayarit	NatAm	[61]
122	Nat. Am. 12	NA Huichol	NatAm	[61]
123	Nat. Am. 13	NA Mayos	NatAm	[61]
124	Nat. Am. 14	NA Mexicaneros	NatAm	[61]
125	Nat. Am. 15	NA Tarahumara	NatAm	[61]
126	Nat. Am. 16	NA Tepehuano	NatAm	[61]
127	Azores 1	Azores Gipuzkoa	Caucn	[51]
127	Basque 2	Basque	Caucn	[62]
127	Belgium 2	Belgium	Caucn	[63]
128	Azerbaijan	Azerbaijan	Caucn	[64]
128	Basque 1	Basque	Caucn	[65]
128	Belgium 1	Belgium	Caucn	[66]
128	Berber 5	Berber Morocco	Caucn	[62]
129	US Cauc. 1	Cauc Alabama	Caucn	[2]
130	US Hisp. 1	Hisp Arizona	Hisp	[2]
131	Vietnam 1	Vietnamese	Asian	[2]

132	Korea 3	Korea	Asian	[67]
133	China 15	China Ningxia Han	Asian	[68]
134	Lebanon	Lebanon	Caucn	[69]
135	Colombia 5	Colombia Department of Santander	Hispc	[70]
136	Italy 6	Italy	Caucn	[71]
137	China 16	China Bai	Asian	[72]
138	Italy 7	Italy N	Caucn	[73]
139	Somalia	Somalia	Afric	[74]
140	Afghanistan 6	Afghan	IndPk	[75]
141	NZ Asian	Asian NZ	Asian	[76]
142	NZ Cauc. 1	Cauc NZ	Caucn	[76]
143	NZ E. Polynesian	E Polynesian	Polyn	[76]
144	NZ W. Polynesian	W Polynesian	Polyn	[76]
145	Brazil 5	Brazil S	Hispc	[77]
146	Brazil 6	Brazil SE	Hispc	[77]
147	Brazil 7	Brazil C	Hispc	[77]
148	Brazil 8	Brazil NE	Hispc	[77]
149	Brazil 9	Brazil N	Hispc	[77]
150	Africa	S Af African	Afric	[78]
151	Angola 2	Angola Cadinba	Afric	[79]
152	Eq. Guinea 1	Equatorial Guineans - Fang - Bioko Island	Afric	[80]
153	Eq. Guinea 2	Equatorial Guineans - Spain - Madrid	Afric	[81]
154	US Af. 2	Af Am California	Afric	[2]
155	Mozambique	Mozambique	Afric	[82]
156	Tanzania	Tanzania	Afric	[83]
157	UK Af. 1	Af Caribb UK	Afric	[30]
158	Aust Abor 1	Aust Abor NT	AusAb	[84]

159	Aust Abor 2	Aust Abor NT	AusAb	[84]
160	China 17	China Eastern	Asian	[85]
161	China 18	China - Han - Henan region	Asian	[86]
162	China 19	China - Han - Jilin Province	Asian	[87]
163	China 20	China - Han - Min Nan	Asian	[88]
164	China 21	China - Han - Shaanxi	Asian	[89]
165	China 22	China Inner Mongolia	Asian	[90]
166	China 23	China Liaoning	Asian	[91]
167	China 24	China Tu	Asian	[92]
168	China 25	China Yi	Asian	[93]
169	China 26	China Dongxiang	Asian	[94]
170	E. Timor 1	E Timor	Asian	[95]
171	China 27	China Han Yunnan	Asian	[34]
172	China 28	China Han Yunnan	Asian	[34]
173	China 29	China Hui	Asian	[96]
174	Indonesia	Indonesia	Asian	[97]
175	China 30	China Lisu	Asian	[34]
176	Malaysia 1	Malay Kuala Lumpur	Asian	[98]
177	Malaysia 2	Malaysians - Bidayuh - Sarawak	Asian	[99]
178	Malaysia 3	Malaysians - Iban - Sarawak	Asian	[99]
179	Malaysia 4	Malaysians - Melanu - Sarawak	Asian	[99]
180	China 31	China Miao	Asian	[96]
181	Nepal	Nepal	Asian	[100]
182	China 32	China Nu	Asian	[34]
183	China 34	China Salar	Asian	[94]
184	China 35	Tibetan Yunnan Province China	Asian	[34]
185	Tibet 1	Tibetan Nepal	Asian	[101]

186	Tibet 10	Tibetian India	Asian	[102]
187	China 36	China Yi	Asian	[34]
188	Tibet 3	Tibetans - China - Tibet - Lassa	Asian	[103]
189	Tibet 4	Tibetans - China - Tibet - Luao	Asian	[104]
190	Tibet 5	Tibetans - China - Tibet - Naqu	Asian	[105]
191	Tibet 6	Tibetans - China - Tibet - Qinghai Province	Asian	[106]
192	Tibet 7	Tibetans - Nepal - Kathmandu	Asian	[107]
193	Tibet 8	Tibetans - Sherpas - Nepal - Namche Bazaar	Asian	[107]
194	Tibet 9	Tibeto Burmans - India - Arunachal Pradesh - Adi Minyong	Asian	[102]
195	Trin. Af.	Trinidad	Afric	[2]
196	Tibet 2	Tibetan China Changdu	Asian	[105]
197	China 37	China Liu	Asian	[96]
198	China 38	China Han Guangzhou	Asian	[108]
199	Albania	Albania	Caucn	[109]
202	Croatia 1	Croatia Zagreb	Caucn	[110]
203	E. Slovak 5	Slovak Romany	Caucn	[111]
204	S. Af. Cauc.	Cauc S Af	Caucn	[78]
205	Greece 2	Greece	Caucn	[112]
206	Iran 1	Iranian Dubai	Caucn	[113]
207	Iran 2	Iran	Caucn	[114]
208	Iraq	Iraq Shia and Sunni Arabs	Caucn	[115]
209	Lithuania	Lithuanians Poland - North	Caucn	[116]
210	Macedonia 2	Macedonia	Caucn	[117]
211	Malta	Malta	Caucn	[118]
212	US Cauc. 2	Cauc California	Caucn	[2]
213	Aust. Cauc.	Cauc Australia	Caucn	[84]
214	Oman	Omanis Dubai	Omitd	[113]

215	Poland 2	Poland N	Caucn	[119]
216	Poland 3	Poland SE	Caucn	[120]
217	Romani 2	Romani Transylvania - Csango	Caucn	[121]
218	Romani 3	Romani Transylvania - Csango	Caucn	[121]
219	Romani 4	Romani Transylvania - Csango	Caucn	[122]
220	Romani 5	Romani Croatia - North West	Caucn	[123]
221	Romani 6	Romani Macedonia	Caucn	[117]
222	Russia 1	Russia	Caucn	[124]
223	Russia 2	Russia	Caucn	[125]
224	Saudi Arabia	Saudi Arabians Dubai	Caucn	[113]
225	Serbia 2	Serbia Serbia Montenegro	Caucn	[126]
226	Siberia	Siberia	Asian	[124]
227	Spain 4	Spain Andalucia	Caucn	[127]
228	Spain 5	Spain	Caucn	[128]
230	Sweden 3	Sweden	Caucn	[129]
231	Tartar 1	Tartars Poland	Caucn	[130]
232	Thailand	Thailand	Asian	[131]
233	Yemen 2	Yemen Sanaa	Caucn	[132]
234	Mexico 1	Mexico central	HispC	[133]
235	Mexico 2	Mexico Metztlán	HispC	[134]
236	Mexico 3	Mexico Mestizos Jalisco	HispC	[135]
237	Mexico 4	Mexico Mestizos Puebla	HispC	[135]
238	Mexico 5	Mexico Mestizos Yucatan	HispC	[135]
239	US Hisp. 2	Hisp California	HispC	[2]
240	Mexico 6	Mexico valley of Mexico	HispC	[136]
241	Bangladesh 1	Bangladesh Dhaka	IndPk	[97]
241	Belarus 2	Belarus Poland - North East	Caucn	[137]

241	Berber 3	Berber Seni Snassen Tribe - Morocco - Bourhia	Caucn	[4]
242	Nat. Am. 17	NA Huastecos - Mexico - Hidalgo	NatAm	[138]
243	Nat. Am. 18	NA Otomies - Mexico - Hidalgo - Ixmiquilpan Valley	NatAm	[138]
244	Nat. Am. 19	NA - Otomies - Mexico - Hidalgo - Sierra Madre	NatAm	[138]
245	Nat. Am. 20	NA Minnesota	NatAm	[1]
249	Eygpt 1	Eygpt Copts	Caucn	[139]
250	Eygpt 2	Eygpt Muslims	Caucn	[139]
251	Morocco 2	Morocco	Caucn	[140]
252	Eygpt 3	Eygpt Upper	Caucn	[141]
253	Argentina 2	Argentina	Hispc	[142]
254	Brazil 10	Brazil Sao Paulo	Hispc	[143]
255	Brazil 11	Brazil Sao Paulo	Hispc	[144]
256	Brazil 12	Brazil Rio Grande do Sul	Hispc	[145]
257	Guinea-Bissua	Guinea-Bissua	Afric	[146]
258	China 39	China Sichuan	Asian	[147]
259	China 4	China Han S	Asian	[148]
260	China 40	China Baoan	Asian	[149]
261	W. Polynesian	W. Polynesian	Polyn	[150]
262	Azores 2	Azores	Hispc	[151]
262	Basque 3	Basque Spain - Vizcaya	Caucn	[152]
262	Berber 1	Berber Egypt - Siwa - NW Desert	Caucn	[139]
263	Uganda	Uganda	Afric	[153]
264	Austria 2	Austria Innsbruck	Caucn	[154]
266	Bosnia 1	Bosnia	Caucn	[155]
267	Bosnia 2	Bosnia Bosniacs Serbs Croats	Caucn	[156]
268	Croatia 2	Croatia Zagreb	Caucn	[154]
269	E. Slovak 6	E. Slovak	Caucn	[111]

270	Greece 3	Greek Cyprus	Caucn	[157]
271	Greece 4	Greece North	Caucn	[158]
272	Jewish 1	Jewish Ashkenazim - Hungary - Budapest	Caucn	[159]
273	Poland 4	Poland South	Caucn	[160]
274	Poland 5	Poland	Caucn	[161]
275	Portugal 4	Portugal North	Caucn	[162]
276	Romani 7	Romani Hungary	Caucn	[159]
277	Spain 6	Spain NE	Caucn	[163]
278	Syria	Syria Arabs	Caucn	[164]
279	Tunisia 2	Tunisia	Caucn	[165]
280	Turkey 1	Turkey	Caucn	[166]
281	UAE	UAE	Omitd	[167]
282	India 1	Indians - Agharia - Central	IndPk	[168]
283	India 2	Indians - Baniyas - Bihar State	IndPk	[169]
284	India 3	Indians - Dheria Gond - Central	IndPk	[168]
285	India 4	Indians - Kurmis - Bihar State	IndPk	[169]
286	India 5	Indians - Satnami - Central	IndPk	[168]
287	India 6	Indians - Teli - Central	IndPk	[168]
288	India 7	Indians - United Arab Emirates	IndPk	[167]
289	India 8	Indians - Yadavs - Bihar State	IndPk	[169]
290	Pakistan 1	Pakistan United Arab Emirates	IndPk	[167]
292	Egypt 4	Egypt El Minia	Caucn	[170]
293	Morocco 3	Morocco Arabs	Caucn	[164]
294	Argentina 3	Argentina Buenos Aires	Hispc	[171]
295	Argentina 4	Argentina Mendoza	Hispc	[171]
296	Argentina 5	Argentina Santa Fe	Hispc	[171]
297	Colombia 6	Colombia	Hispc	[172]

298	Cent. W. Af.	Cent. W. Af.	Afric	[173]
299	Aust Abor 3	Aust Abor Desert	AusAb	[174]
300	Aust Abor 4	Aust Abor E Arnhem	AusAb	
301	Aust Abor 5	Aust Abor Fitzmaurice	AusAb	
302	Aust Abor 6	Aust Abor Gulf	AusAb	
303	Aust Abor 7	Aust Abor Kimberley	AusAb	
304	Aust Abor 8	Aust Abor NE	AusAb	
305	Aust Abor 9	Aust Abor NW	AusAb	
306	Aust Abor 10	Aust Abor Rainforest	AusAb	
307	Aust Abor 11	Aust Abor Riverine	AusAb	
308	Aust Abor 12	Aust Abor SE	AusAb	
309	Aust Abor 13	Aust Abor Spencer	AusAb	
310	Aust Abor 14	Aust Abor Tiwi	AusAb	
311	Aust Abor 15	Aust Abor Torres Strait	AusAb	
312	Aust Abor 16	Aust Abor W Arnhem	AusAb	
313	Aust Abor 17	Aust Abor W Cape	AusAb	
314	Chammos	Chammos	Asian	[175]
315	China 42	China Henan	Asian	[176]
316	China 43	China Hong Kong	Asian	[177]
317	China 5	China	Asian	[178]
318	E. Timor 2	E. Timor NZ	Asian	[179]
319	Philippines	Philippines	Asian	[175]
320	Jewish 2	Jewish- Oriental - Israel f Iran Iraq	Caucn	[180]
321	Korea 4	Korean - China - North East	Asian	[181]
322	Australia	Cauc Australia	Caucn	[182]
323	Austria 3	Austrian- Salzburg	Caucn	[183]
326	Germany 1	German - Bavaria	Caucn	[184]
327	Greece 5	Greek Cypriot - Cyprus	Caucn	□[185]

328	Italy 8	Italians - North	Caucn	[186]
329	Italy 9	Italians - South	Caucn	[186]
330	Jewish 3	Jewish - Ashkenazi - Israel	Caucn	[180]
331	Jewish 4	Jewish - Sephardics - Israel f Turkey	Caucn	[180]
332	Poland 6	Polish - Podlasie - North East	Caucn	[187]
333	Portugal 5	Portuguese - North - Douro	Caucn	[188]
334	Russia 3	Russia	Caucn	[189]
335	Scotland	Scotland	Caucn	[190]
336	Spain 7	Spanish - Andalucia	Caucn	[62]
337	Spain 8	Spanish - Andalucia	Caucn	[191]
338	Spain 9	Spanish - Canary Islands	Caucn	[173]
339	Spain 10	Spanish - Catalonia - Girona	Caucn	[62]
340	Spain 11	Spanish - Majorca	Caucn	[192]
341	Spain 12	Spanish - Minorca	Caucn	[192]
342	Spain 13	Spanish - North - Cantabria	Caucn	[193]
343	Spain 14	Spanish - Valencia	Caucn	[192]
345	India 9	Andaman islanders - Nicobarese	Andam	[194]
346	India 10	Andaman islanders - Onge	Andam	[194]
347	India 11	Andaman islanders -Great Andamanese	Andam	[194]
348	India 12	Indians - Andhra Pradesh - Goud	IndPk	[195]
349	India 13	Indians - Andhra Pradesh - Padmashali	IndPk	[195]
350	India 14	Indians - Orissa and Nagaland - North East - Gabada	IndPk	[196]
351	India 15	Indians - Orissa and Nagaland - North East - Kuvi Kond	IndPk	[196]
352	India 16	Indians - Orissa and Nagaland - North East - Lotha Naga	IndPk	[196]
353	India 17	Indians - Orissa and Nagaland - North East -	IndPk	[196]

		Paroja		
355	Egypt 5	Egypt Cairo	Caucn	[197]
356	Jewish 5	Jewish Israel Nth Africa	Caucn	[180]
357	Argentina 6	Argentina NW Puna	Hispc	[198]
358	UK Cauc. 1	Cauc UK	Caucn	[199]
359	Gabon	Gabon	Afric	[200]
360	SE Asia	SE Asia NZ	Asian	[150]
361	Austria 4	Austria Vienna	Caucn	[201]
362	Austria 5	Austria	Caucn	[202]
364	UK Indo. Pak.	UK Indo. Pak.	IndPk	[41]
365	Czech	Czech	Caucn	[203]
366	Germany 2	Germany central	Caucn	[204]
367	Jewish 6	Jewish Israel	Caucn	[205]
368	NZ Cauc. 2	Cauc NZ	Caucn	[150]
369	Norway	Norway	Caucn	[206]
370	Poland 7	Polish - Central	Caucn	[207]
371	Poland 8	Polish - Lodz	Caucn	[208]
372	Poland 9	Polish - Pomerania Kujawy	Caucn	[209]
373	Poland 10	Polish - South	Caucn	
374	Poland 11	Polish - Warmia and Mazury - North East	Caucn	[210]
375	Poland 12	Polish - West	Caucn	[211]
376	Spain 15	Spanish - South East - Murcia	Caucn	[212]
377	Bangladesh 2	Bangladesh	IndPk	[213]
377	Belarus 3	Belarus	Caucn	[214]
377	Berber 4	BERBERS - Tunisia - Kesra	Caucn	[5]
378	US Af. 1	Af Am Alabama	Afric	[2]
379	Gujarat	Gujarat	IndPk	[215]
380	Pakistan 2	Pakistan	IndPk	[206]

381	E. Polynesians	E. Polynesians	Polyn	[150]
382	Yemen 1	YEMENITES - Dubai	Caucn	[113]
385	Slovak Vojvodina	Slovak Vojvodina	Caucn	[216]
383	Ruthenian Vojvodina	Ruthenian Vojvodina	Caucn	
384	Serbian Vojvodina	Serbian Vojvodina	Caucn	
386	Slovak w Slovakia	Slovak w Slovakia	Caucn	
391	Pozan	Poland Pozan	Caucn	[217]
394	Wroclaw	Poland Wroclaw	Caucn	
393	Warszawa	Poland Warszawa	Caucn	
392	Szczecin	Poland Szczecin	Caucn	
390	Lubin	Poland Lubin	Caucn	
389	Krazow	Poland Krazow	Caucn	
388	Gdansk	Poland Gdansk	Caucn	
387	Bialystok	Poland Bialystok	Caucn	
395	Brazil	Brazil	Omitd	[219]
396	Turkish	Turkish	Caucn	[220]
397	Han Shanghai	China Han	Asian	[221]
398	China Mongolian	China Mongolian	Asian	[222]
399	Dutch	Dutch	Caucn	[223]
400	Mexico	Mexico	Omitd	[224]
403	Somalis	Somalis	Afric	[225]
401	Danes	Danes	Caucn	
402	Greenland	Inuit Greenland	Inuit	
404	Kuwait	Kuwait	Caucn	[226]
405	Cauca	Columbia Cauca	Omitd	[227]
406	E Timor	E Timor	Asian	[229]

407	Central Portugal	Central Portugal	Caucn	[230]
408	Tianjin, North China.	China Tianjin	Asian	[231]
409	Ceara, Brazil	Brazil Ceara	Omitd	[232]
410	Fujian, China	China Fujian	Asian	[233]
411	Japan	Japan	Asian	[234]
412	W Romanian	W Romanian	Caucn	[235]
413	Japan	Japan	Asian	[236]
414	Mar del Plata	Argentina Mar del Plata	Omitd	[238]
415	Bahia Blanca	Argentina Bahia Blanca	Omitd	
416	Swiss	Swiss	Caucn	[239]
417	Mayan	NA Mayan	NatAm	[240]
418	Lacandon	NA Lacandon	NatAm	
419	Tzeltal	NA Tzeltal	NatAm	
420	Tojolabal	NA Tojolabal	NatAm	
421	Tzotzil	NA Tzotzil	NatAm	
422	Mazatec	NA. Mazatec	NatAm	
423	Nahuas from Puebla	NA Nahuas from Puebla	NatAm	
424	h.west-Nahua	NA.west-Nahua	NatAm	
425	i.Nahua from Chapala river region	NA.Nahua Chapala	NatAm	
426	Paraná (Southern Brazil)	Brazil Paraná	Omitd	[241]
427	SA White	S Af White	Caucn	[242]
428	SA Black	S Af Black	Afric	
429	SA Indian	S Af Indian	IndPk	
430	SA Coloured	S Af Coloured	Afric	
433	Ojibwe	Ojibwe	NatAm	[243]
432	Cree	Cree	NatAm	

434	Salishan	Salishan	NatAm	
431	Canadian CaucAsians	Cauc Canadian	Caucn	
435	Afghanastan	Afghanastan	Caucn	[244]
436	Italian	Italian	Caucn	[245]
437	Turkish Cypriots	Turkish Cypriots	Caucn	[246]
438	Bangladeshi	Bangladeshi	IndPk	[248]
439	Derung	China Tibeto Burman Derung	Asian	[249]
440	Nu	China Tibeto Burman Nu	Asian	
441	Yunnan Tibetan	China Tibeto Burman Yunnan Tibetan	Asian	
442	Buenos Aires	Argentina Buenos Aires	Omitd	[250]
445	Neuquén	Argentina Neuquén	Omitd	
444	Cordoba	Argentina Cordoba	Omitd	
443	Chubut	Argentina Chubut	Omitd	
446	Santa Cruz	Argentina Santa Ccruz	Omitd	

1. Gross, A.M. and B. Budowle, *Minnesota Population Data on 15 STR Loci Using the Identifiler® Kit* Journal of Forensic Sciences, 2006. **51**(6): p. 1410-1413.
2. Budowle, B., et al., *CODIS STR loci data from 41 sample populations*. . Journal of Forensic Sciences, 2001. **46**(3): p. 453-489.
3. Rebala, K., et al., *Belarusian population genetic database for 15 autosomal STR loci*. Forensic Science International, 2007. **173**(2-3): p. 235-237.
4. Coudray, C., et al., *Population genetic data of 15 tetrameric short tandem repeats (STRs) in Berbers from Morocco*. Forensic Science International, 2007. **167**(1): p. 81-86.
5. Cherni, L., et al., *Data for 15 autosomal STR markers (Powerplex 16 System) from two Tunisian populations: Kesra (Berber) and Zriba (Arab)*. Forensic Science International, 2005. **147**(1): p. 101-106.
6. Hill, C.R., et al., *U.S. population data for 29 autosomal STR loci*. Forensic Science International: Genetics, 2013. **7**(3): p. e82-e83.
7. Shimada, I., et al., *Allele frequency data for 16 STR loci in the Vietnamese population*. International Journal of Legal Medicine, 2002. **116**: p. 246-248.
8. Tofanelli, S., et al., *Variation at 16 STR loci in Rwandans (Hutu) and implications on profile frequency estimation in Bantu-speakers*. International Journal of Legal Medicine, 2003. **117**: p. 121-126.
9. Syn, C.K.C., et al., *Genetic data for the 13 CODIS STR loci in Singapore China*. Forensic Science International, 2005. **152**(2-3): p. 285-288.
10. Skitsa, I., et al., *STR-CODIS typing in Greece*. Forensic Science International, 2003. **137**(1): p. 104-106.
11. Sanz, P., et al., *Population data of 13 STRS in southern Spain (Andalusia)*. Forensic Science International, 2001. **119**(1): p. 113-115.
12. Akbasak, B.S., et al., *Turkish population data with the CODIS multiplex short tandem repeat loci*. Forensic Science International, 2001. **123**(2-3): p. 227-229.
13. Hadi Çakır, A., et al., *Allele frequencies for 15 STR loci in Van-Ağrı districts of the Eastern Anatolia region of Turkey*. Forensic Science International, 2003. **135**(1): p. 60-63.
14. Budowle, B., et al., *Population studies on three Native Alaska population groups using STR loci*. Forensic Science International, 2002. **129**(1): p. 51-57.
15. Paredes, M., et al., *Analysis of the CODIS autosomal STR loci in four main Colombian regions*. Forensic Science International, 2003. **137**(1): p. 67-73.
16. Poiares, L.d.A., et al., *15 STR loci frequencies in the population from Paraná, Southern Brazil*. Forensic Science International: Genetics, 2009. **4**(1): p. e23-e24.
17. Andreassen, R., et al., *Icelandic population data for the STR loci in the AMPFISTR®SGM Plus(TM) system and the PowerPlex® Y-system*. Forensic Science International: Genetics, 2010. **4**(4): p. e101-e103.
18. Melo, M.M., et al., *Genetic study of 15 STRs loci of Identifiler system in Angola population*. Forensic Science International: Genetics, 2010. **4**(5): p. e153-e157.
19. Novković, T., et al., *Genetic polymorphisms of 15 AmpFISTR Identifiler loci in a Serbian population*. Forensic Science International: Genetics, 2010. **4**(5): p. e149-e150.
20. Wolfgramm, E.d.V., et al., *Genetic analysis of 15 autosomal and 12 Y-STR loci in the Espírito Santo State population, Brazil*. Forensic Science International: Genetics, 2011. **5**(3): p. e41-e43.

21. Albinsson, L., et al., *Swedish population data and concordance for the kits PowerPlex® ESX 16 System, PowerPlex® ESI 16 System, AmpFISTR® NGM™, AmpFISTR® SGM Plus™ and Investigator ESSplex*. Forensic Science International: Genetics, 2011. **5**(3): p. e89-e92.
22. Soták, M., et al., *Population database of 17 autosomal STR loci from the four predominant Eastern Slovakia regions*. Forensic Science International: Genetics, 2011. **5**(3): p. 262-263.
23. Xing, J., et al., *Genetic polymorphism of 15 STR loci in a Manchu population in Northeast China*. Forensic Science International: Genetics, 2011. **5**(3): p. e93-e95.
24. Rodrigues, E.L., et al., *Genetic data on 15 STR autosomal loci for a sample population of the Northern Region of the State of Rio de Janeiro, Brazil*. Forensic Science International: Genetics, 2009. **4**(1): p. e25-e26.
25. Mertens, G., et al., *Population genetic analysis of Moroccans residing in Belgium using 16 autosomal STRs of the PowerPlex ESI 17 multiplex*. Forensic Science International: Genetics, 2011. **5**(4): p. 352-353.
26. Rak, S.Á., et al., *Population genetic data on 15 STR loci in the Hungarian population*. Forensic Science International: Genetics, 2011. **5**(5): p. 543-544.
27. Gutiérrez, C.C., et al., *Population genetic data for 15 STR loci (PowerPlex 16 kit) in Nicaragua*. Forensic Science International: Genetics, 2011. **5**(5): p. 563-564.
28. Berti, A., et al., *Allele frequencies of the new European Standard Set (ESS) loci in the Italian population*. Forensic Science International: Genetics, 2011. **5**(5): p. 548-549.
29. Molnár, A., et al., *Allele distribution of the new European Standard Set (ESS) loci in the Hungarian population*. Forensic Science International: Genetics, 2011. **5**(5): p. 555-556.
30. Tucker, V.C., et al., *UK population data generated with the PowerPlex® ESI 16 system*. Forensic Science International: Genetics, 2012. **6**(4): p. e112-e118.
31. Di Cristofaro, J., et al., *Genetic data of 15 STR loci in five populations from Afghanistan*. Forensic Science International: Genetics, 2012. **6**(1): p. e44-e45.
32. Hatzer-Grubwieser, P., et al., *Allele frequencies and concordance study of 16 STR loci – including the new European Standard Set (ESS) loci – in an Austrian population sample*. Forensic Science International: Genetics, 2012. **6**(1): p. e50-e51.
33. Muñoz, A., et al., *Allele frequencies of 15 STRs in the Calchaqui Valleys population (North-Western Argentina)*. Forensic Science International: Genetics, 2012. **6**(1): p. e58-e60.
34. Kraaijenbrink, T., et al., *Allele frequency distribution of 21 forensic autosomal STRs in 7 populations from Yunnan, China*. Forensic Science International: Genetics, 2014. **3**(1): p. e11-e12.
35. Chang, Y.-f., et al., *Genetic polymorphism of 17 STR loci in China population from Hunan province in Central south China*. Forensic Science International: Genetics, 2012. **6**(5): p. e151-e153.
36. Barbaro, A., et al., *Allele frequencies of the new European Standard Set (ESS) loci in a population of Southern Italy (Calabria)*. Forensic Science International: Genetics, 2012. **6**(1): p. e37-e38.
37. Branham, A., R. Wenk, and F. Chiafari, *Allele frequencies of fifteen STR loci in U.S. immigrants from Haiti compared with African Americans and Afro-Caribbeans*. Forensic Science International: Genetics, 2012. **6**(1): p. e3-e4.

38. Schlebusch, C.M., H. Soodyall, and M. Jakobsson, *Genetic variation of 15 autosomal STR loci in various populations from southern Africa*. Forensic Science International: Genetics, 2012. **6**(1): p. e20-e21.
39. Jakovski, Z., et al., *Allele frequencies of the new European Standard Set (ESS) loci plus SE33 locus in a population from the Republic of Macedonia*. Forensic Science International: Genetics, 2012. **6**(3): p. e90-e92.
40. Aguiar, V.R.d.C., et al., *Updated Brazilian STR allele frequency data using over 100,000 individuals: An analysis of CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, Penta D, Penta E, TH01, TPOX and vWA loci*. Forensic Science International: Genetics, 2012. **6**(4): p. 504-509.
41. Thomson, J.A., et al., *Validation of short tandem repeat analysis for the investigation of cases of disputed paternity*. Forensic Science International, 1999. **100**(1-2): p. 1-16.
42. Stanciu, F., O.R. Popescu, and I.M. Stoian, *STR data for the AmpFISTR Identifiler from Dobruja region (SE Romania)*. Forensic Science International: Genetics, 2009. **3**(2): p. 146-147.
43. Jedrzejczyk, M., et al., *Genetic population studies on 15 NGM™ STR loci in central Poland population*. Forensic Science International: Genetics, 2012. **6**(4): p. e119-e120.
44. Ali, I., et al., *Allele frequencies of 15 STR loci using AmpF/STR Identifiler kit in the Maldivian population*. Forensic Science International: Genetics, 2012. **6**(5): p. e136.
45. Monterrosa, J.C., et al., *Population genetic data for 16 STR loci (PowerPlex ESX-17 kit) in El Salvador*. Forensic Science International: Genetics, 2012. **6**(5): p. e134.
46. Pontes, M.L. and M.F. Pinheiro, *Population data of the AmpFISTR® NGM™ STR loci in a North of Portugal sample*. Forensic Science International: Genetics, 2012. **6**(5): p. e127-e128.
47. Chen, J.-g., et al., *Population genetic data of 15 autosomal STR loci in Uyghur ethnic group of China*. Forensic Science International: Genetics, 2012. **6**(6): p. e178-e179.
48. Fernandez-Formoso, L., et al., *Allele frequencies of 20 STRs from Northwest Spain (Galicia)*. Forensic Science International: Genetics, 2012. **6**(5): p. e149-e150.
49. Barbaro, A., et al., *Distribution of allele frequencies of 20 STRs loci in a population sample from Calabria, Southern Italy*. Forensic Science International: Genetics, 2012. **6**(5): p. e137-e138.
50. Yang, L., et al., *Allele frequencies of 15 STRs in five ethnic groups (Han, Gelao, Jing, Shui and Zhuang) in South China*. Forensic Science International: Genetics, 2013. **7**(1): p. e9-e14.
51. Lopes, V., et al., *Allelic frequency distribution of 17 STRs from Identifiler and PowerPlex-16 in Central Portugal area and the Azores archipelago*. Forensic Science International: Genetics, 2009. **4**(1): p. e1-e7.
52. Petrić, G., et al., *Genetic variation at 15 autosomal STR loci in the Hungarian population of Vojvodina Province, Republic of Serbia*. Forensic Science International: Genetics, 2012. **6**(6): p. e163-e165.
53. Ribeiro, T., et al., *Population data of the AmpFISTR® NGM™ loci in South Portuguese population*. Forensic Science International: Genetics, 2013. **7**(2): p. e37-e39.

54. Huang, Q., et al., *Genetic polymorphism of 15 STR loci in China Han population from Shanghai municipality in East China*. Forensic Science International: Genetics, 2013. **7**(2): p. e31-e34.
55. Piglionica, M., et al., *Allele frequencies of the new European Standard Set (ESS) loci in a population of Apulia (Southern Italy)*. Forensic Science International: Genetics, 2013. **7**(2): p. e35-e36.
56. He, J. and F. Guo, *Genetic variation of fifteen autosomal STR loci in a Manchu population from Jilin Province, Northeast China*. Forensic Science International: Genetics, 2013. **7**(2): p. e45-e46.
57. García, O., et al., *Population genetic data and concordance study for the kits Identifiler, NGM, PowerPlex ESX 17 System and Investigator ESSplex in Spain*. Forensic Science International: Genetics, 2012. **6**(2): p. e78-e79.
58. Aguiar, V.R.d.C., et al., *Corrigendum to "Updated Brazilian STR allele frequency data using over 100,000 individuals: An analysis of CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, Penta D, Penta E, TH01, TPOX and vWA loci" [Forensic Sci. Int.: Genet. 6 (4) (2012) 504–509]*. Forensic Science International: Genetics, 2013. **7**(2): p. 321-325.
59. Hong, S.B., et al., *Korean population genetic data and concordance for the PowerPlex® ESX 17, AmpFISTR® Identifiler®, and PowerPlex® 16 systems*. Forensic Science International: Genetics, 2013. **7**(3): p. e47-e51.
60. González-Liñán, A., et al., *Population data of the 16 autosomal STRs loci of the Powerplex ESI 17 System in Sicilian Population (Italy)*. Forensic Science International: Genetics, 2013. **7**(3): p. e93-e94.
61. Rangel-Villalobos, H., et al., *Forensic parameters for 15 STRs in eight Amerindian populations from the north and west of Mexico*. Forensic Science International: Genetics, 2013. **7**(3): p. e62-e65.
62. Perez-Lezaun, A., et al., *Allele frequencies of 13 short tandem repeats in population samples from the Iberian Peninsula and Northern Africa*. International Journal of Legal Medicine, 2000. **113**: p. 208-214.
63. Decorte, R., et al., *Belgian population data for 15 STR loci (AmpFISTR® SGM Plus and AmpFISTR(TM) profiler PCR amplification kit)*. Forensic Science International, 2004. **139**(2-3): p. 211-213.
64. Nasibov, E., et al., *Allele frequencies of 15 STR loci in Azerbaijan population*. Forensic Science International: Genetics, 2013. **7**(3): p. e99-e100.
65. Yurrebaso, I., et al., *Allele frequencies and concordance study between the Identifiler and the PowerPlex ESX17 systems in the Basque Country population*. Forensic Science International: Genetics, 2011. **5**(3): p. e79-e80.
66. Dognaux, S., et al., *Allele frequencies for the new European Standard Set (ESS) loci and D1S1677 in the Belgian population*. Forensic Science International: Genetics, 2012. **6**(2): p. e75-e77.
67. Park, J.-H., et al., *Genetic variation of 23 autosomal STR loci in Korean population*. Forensic Science International: Genetics, 2013. **7**(3): p. e76-e77.
68. Wang, H.-d., et al., *Allelic diversity distributions of 21 new autosomal short tandem repeat loci in China Ningxia Han population*. Forensic Science International: Genetics, 2013. **7**(3): p. e78-e79.
69. El Andari, A., et al., *Population genetic data for 23 STR markers from Lebanon*. Forensic Science International: Genetics, 2013. **7**(4): p. e108-e113.
70. Castillo, A., et al., *Genetic variation for 20 STR loci in a northeast Colombian population (Department of Santander)*. Forensic Science International: Genetics Supplement Series, (0).

71. Rocchi, A., et al., *Italian data of 23 STR loci amplified in a single multiplex reaction*. Forensic Science International: Genetics, 2012. **6**(6): p. e157-e158.
72. Shen, C.-m., et al., *Allelic polymorphic investigation of 21 autosomal short tandem repeat loci in a China Bai ethnic group*. Legal Medicine, 2013. **15**(2): p. 109-113.
73. Turrina, S., et al., *Evaluation of genetic parameters of 22 autosomal STR loci (PowerPlex® Fusion System) in a population sample from Northern Italy*. International Journal of Legal Medicine, 2013: p. 1-3.
74. Tillmar, A.O., G. Bäckström, and K. Montelius, *Genetic variation of 15 autosomal STR loci in a Somali population*. Forensic Science International: Genetics, 2009. **4**(1): p. e19-e20.
75. Älgenäs, C. and A. Tillmar, *Population genetics of 29 autosomal STRs and 17 Y-chromosomal STRs in a population sample from Afghanistan*. International Journal of Legal Medicine, 2013: p. 1-2.
76. Bright, J.-A., J.S. Buckleton, and C.E. McGovern, *Allele frequencies for the four major sub-populations of New Zealand for the 15 Identifiler loci*. Forensic Science International: Genetics, 2010. **4**(2): p. e65-e66.
77. de Assis Poiares, L., et al., *Allele frequencies of 15 STRs in a representative sample of the Brazilian population*. Forensic Science International: Genetics, 2010. **4**(2): p. e61-e63.
78. Kido, A., et al., *Population data on the AmpFISTR Identifiler loci in Africans and Europeans from South Africa*. Forensic Science International, 2007. **168**(2-3): p. 232-235.
79. Beleza, S., et al., *17 STR data (AmpF/STR Identifiler and Powerplex 16 System) from Cabinda (Angola)*. Forensic Science International, 2004. **141**(2-3): p. 193-196.
80. Calzada, P., et al., *The Fang population of Equatorial Guinea characterised by 15 STR-PCR polymorphisms*. International Journal of Legal Medicine, 2005. **119**: p. 107-110.
81. Alves, C., et al., *STR allelic frequencies for an African population sample (Equatorial Guinea) using AmpFISTR Identifiler and Powerplex 16 kits*. Forensic Science International, 2005. **148**(2-3): p. 239-242.
82. Alves, C., et al., *Contribution for an African autosomic STR database (AmpF/STR Identifiler and Powerplex 16 System) and a report on genotypic variations*. Forensic Science International, 2004. **139**(2-3): p. 201-205.
83. Forward, B.F., et al., *AMPFISTR® Identifiler™ STR Allele Frequencies in Tanzania, Africa*. Journal of Forensic Sciences, 2008. **53**(1): p. 245-247.
84. Eckhoff, C., S.J. Walsh, and J.S. Buckleton, *Population data from sub-populations of the Northern Territory of Australia for 15 autosomal short tandem repeat (STR) loci*. Forensic Science International, 2007. **171**(2-3): p. 237-249.
85. Gao, Y., et al., *Genetic data of 15 STR forensic loci in eastern China population*. Forensic Science International, 2005. **154**(1): p. 78-80.
86. Zeng, Z.S., et al., *Population genetic data of 15 STR loci in Han population of Henan province (central China)*. Legal Medicine, 2007. **9**(1): p. 30-32.
87. Tang, J., et al., *Genetic analyzing of 15 STR loci in a Han population of Jinan (northern China)*. Legal Medicine. **11**(3): p. 144-146.
88. Hu, S.P., et al., *Genetic profile of 15 STR loci in the Min Nan population in Southeast China*. Forensic Science International, 2005. **152**(2-3): p. 263-265.

89. Wu, Y.-M., et al., *Genetic polymorphisms of 15 STR loci in China Han population living in Xi'an city of Shaanxi Province*. Forensic Science International: Genetics, 2008. **2**(2): p. e15-e18.
90. Chen, L., He, Y., and L. Shengbin, *Genetic Analysis of 15 STR Loci of China Ewenki Ethnic Population*. Journal of Forensic Sciences, 2006. **51**(6): p. 1408-1409.
91. Wang, R., et al., *Genetic distribution on 15 STR loci from a population of Southern Liaoning in northeast of China*. Forensic Science International: Genetics, 2008. **2**(2): p. e25-e26.
92. Zhu, B., et al., *Population genetic analysis of 15 STR loci of China Tu ethnic minority group*. Forensic Science International, 2008. **174**(2-3): p. 255-258.
93. Zhu, B.-F., et al., *Population data of 15 STR loci of China Yi ethnic minority group*. Legal Medicine, 2008. **10**(4): p. 220-224.
94. Deng, Y., et al., *Genetic polymorphisms of 15 STR loci of China Dongxiang and Salar ethnic minority living in Qinghai Province of China*. Legal Medicine, 2007. **9**(1): p. 38-42.
95. Souto, L., et al., *Population data on 15 autosomal STRs in a sample from East Timor*. Forensic Science International, 2005. **155**(1): p. 77-80.
96. Liu, C., C. Liu, and H. Wang, *STR data for the 15 loci from three minority populations in Guangxi municipality in South China*. Forensic Science International, 2006. **162**(1-3): p. 49-52.
97. Dobashi, Y., et al., *STR data for the AmpFLSTR Identifiler loci in Bangladeshi and Indonesian populations*. Legal Medicine, 2005. **7**(4): p. 222-226.
98. Maruyama, S., et al., *Population data on 15 STR loci using AmpF/STR Identifiler kit in a Malay population living in and around Kuala Lumpur, Malaysia*. Legal Medicine, 2008. **10**(3): p. 160-162.
99. Suadi, Z., et al., *STR Data for the AmpFISTR® Identifiler® Loci from the Three Main Ethnic Indigenous Population Groups (Iban, Bidayuh, and Melanau) in Sarawak, Malaysia*. Journal of Forensic Sciences 2007. **52**(1): p. 231-234.
100. Kraaijenbrink, T., et al., *Allele frequency distribution for 21 autosomal STR loci in Nepal*. Forensic Science International, 2007. **168**(2-3): p. 227-231.
101. Kido, A., et al., *STR data for 15 AmpFLSTR Identifiler loci in a Tibetan population (Nepal)*. International Congress Series, 2006. **1288**: p. 349-351.
102. Krithika, S., et al., *Genotype Profile for Fifteen Tetranucleotide Repeat Loci in Two Tibeto-Burman Speaking Tribal Populations of Arunachal Pradesh, India*. Journal of Forensic Sciences, 2007. **52**(1): p. 239-241.
103. Kang, L.L., et al., *Allele frequencies of 15 STR loci of Tibetan lived in Tibet Lassa*. Forensic Science International, 2007. **168**(2-3): p. 236-240.
104. Kang, L. and S. Li, *Allele frequencies of 15 STR loci of Luoba ethnic group in Tibet (Southwestern China)*. Forensic Science International, 2005. **155**(2-3): p. 219-221.
105. LongLi, K., et al., *Genetic polymorphisms of 15 STR loci in two Tibetan populations from Tibet Changdu and Naqu, China*. Forensic Science International, 2007. **169**(2-3): p. 239-243.
106. Yan, J., et al., *Genetic analysis of 15 STR loci on China Tibetan in Qinghai Province*. Forensic Science International, 2007. **169**(1): p. e3-e6.
107. Ota, M., et al., *Allele frequencies for 15 STR loci in Tibetan populations from Nepal*. Forensic Science International, 2007. **169**(2-3): p. 234-238.
108. Liao, G., et al., *Population Genetic Study of 15 STR Loci in a China Population*. Journal of Forensic Sciences, 2008. **53**(1): p. 252-253.
109. Kubat, M., et al., *Population genetics of the 15 AmpFISTR Identifiler loci in Kosovo Albanians*. International Journal of Legal Medicine, 2004. **118**: p. 115-118.

110. Projić, P., et al., *Allele Frequencies for 15 Short Tandem Repeat Loci in Representative Sample of Croatian Population*. Croatian Medical Journal, 2007. **48**: p. 473-7.
111. Soták, M., et al., *Genetic variation analysis of 15 autosomal STR loci in Eastern Slovak Caucasian and Romany (Gypsy) population*. Forensic Science International: Genetics, 2008. **3**(1): p. e21-e25.
112. Sánchez-Diz, P., et al., *16 STR data of a Greek population*. Forensic Science International: Genetics, 2008. **2**(4): p. e71-e72.
113. Alshamali, F., et al., *STR population diversity in nine ethnic populations living in Dubai*. Forensic Science International, 2005. **152**(2-3): p. 267-279.
114. Shepard, E.M. and R.J. Herrera, *Iranian STR variation at the fringes of biogeographical demarcation*. Forensic Science International, 2006. **158**(2-3): p. 140-148.
115. Barni, F., et al., *Allele frequencies of 15 autosomal STR loci in the Iraq population with comparisons to other populations from the middle-eastern region*. Forensic Science International, 2007. **167**(1): p. 87-92.
116. Pepinski, W., et al., *Allele distribution of 15 STR loci in a population sample of the Lithuanian minority residing in the Northeastern Poland*. Forensic Science International, 2004. **144**(1): p. 65-67.
117. Havas, D., et al., *Population genetics of 15 AmpfSTR Identifiler loci in Macedonians and Macedonian Romani (Gypsy)*. Forensic Science International, 2007. **173**(2-3): p. 220-224.
118. Cassar, M., C. Farrugia, and C. Vidal, *Allele frequencies of 14 STR loci in the population of Malta*. Legal Medicine, 2008. **10**(3): p. 153-156.
119. Szczerkowska, Z., et al., *Northern Polish population data and forensic usefulness of 15 autosomal STR loci*. Forensic Science International, 2004. **144**(1): p. 69-71.
120. Koziol, P., et al., *Genetic data on 19 STR loci in south-east Poland*. Forensic Science International, 2004. **139**(1): p. 89-92.
121. Egyed, B., S. Füredi, and Z. Padar, *Population genetic study in two Transylvanian populations using forensically informative autosomal and Y-chromosomal STR markers*. Forensic Science International, 2006. **164**(2-3): p. 257-265.
122. Marian, C., et al., *STR data for the 15 AmpFISTR identifiler loci in the Western Romanian population*. Forensic Science International, 2007. **170**(1): p. 73-75.
123. Novokmet, N. and Z. Pavcec, *Genetic polymorphisms of 15 AmpFISTR identifiler loci in Romani population from Northwestern Croatia*. Forensic Science International, 2007. **168**(2-3): p. e43-e46.
124. Zhivotovsky, L.A., et al., *A reference data base on STR allele frequencies in the Belarus population developed from paternity cases*. Forensic Science International: Genetics, 2009. **3**(3): p. e107-e109.
125. Malyarchuk, B.A., *Variation of 15 Autosomal Microsatellite DNA Loci in the Russian Population*. Molecular Biology, 2007. **41**(1): p. 1-4.
126. Veselinovic, I., et al., *Allele frequencies of the 15 AmpFISTR Identifiler loci in the population of Vojvodina Province, Serbia and Montenegro*. International Journal of Legal Medicine, 2004. **118**: p. 184-186.
127. Coudray, C., et al., *Allele frequencies of 15 tetrameric short tandem repeats (STRs) in Andalusians from Huelva (Spain)*. Forensic Science International, 2007. **168**(2-3): p. e21-e24.
128. Camacho, M.V., C. Benito, and A.M. Figueiras, *Allelic frequencies of the 15 STR loci included in the AmpFISTR® Identifiler(TM) PCR Amplification Kit in an autochthonous sample from Spain*. Forensic Science International, 2007. **173**(2-3): p. 241-245.

129. Montelius, K., A.O. Karlsson, and G. Holmlund, *STR data for the AmpFISTR Identifiler loci from Swedish population in comparison to European, as well as with non-European population*. Forensic Science International: Genetics, 2008. **2**(3): p. e49-e52.
130. Pepinski, W., et al., *Genetic data on 15 STR loci in the ethnic group of Polish Tatars residing in the area of Podlasie (Northeastern Poland)*. Forensic Science International, 2005. **149**(2-3): p. 263-265.
131. Rerkamnuaychoke, B., et al., *Thai population data on 15 tetrameric STR loci--D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818 and FGA*. Forensic Science International, 2006. **158**(2-3): p. 234-237.
132. Klintschar, M., N. Al-Hammadi, and B. Reichenpfader, *Significant differences between Yemenite and Egyptian STR profiles and the influence on frequency estimations in Arabs*. International Journal of Legal Medicine, 2001. **114**: p. 211-214.
133. Hernández-Gutiérrez, S., et al., *STR data for 15 loci in a population sample from the central region of Mexico*. Forensic Science International, 2005. **151**(1): p. 97-100.
134. Gorostiza, A., et al., *Allele frequencies of the 15 AmpF/Str Identifiler loci in the population of Metztitlán (Estado de Hidalgo), México*. Forensic Science International, 2007. **166**(2-3): p. 230-232.
135. Rubi-Castellanos, R., et al., *Genetic data of 15 autosomal STRs (Identifiler kit) of three Mexican Mestizo population samples from the States of Jalisco (West), Puebla (Center), and Yucatan (Southeast)*. Forensic Science International: Genetics, 2009. **3**(3): p. e71-e76.
136. Luna-Vazquez, A., et al., *Population data for 15 loci (Identifiler® Kit) in a sample from the Valley of Mexico*. Legal Medicine, 2005. **7**(5): p. 331-333.
137. Pepinski, W., et al., *Allele distribution of 15 STR loci in a population sample of Byelorussian minority residing in the northeastern Poland*. Forensic Science International, 2004. **139**(2-3): p. 265-267.
138. Barrot, C., et al., *Characterisation of three Amerindian populations from Hidalgo State (mexico) by 15 STR-PCR polymorphisms*. International Journal of Legal Medicine, 2005. **119**: p. 111-115.
139. Coudray, C., et al., *Allele frequencies of 15 short tandem repeats (STRs) in three Egyptian populations of different ethnic groups*. Forensic Science International, 2007. **169**(2-3): p. 260-265.
140. Bouabdellah, M., et al., *STR data for the 15 AmpFISTR Identifiler loci in the Moroccan population*. Forensic Science International: Genetics Supplement Series, 2008. **1**(1): p. 306-308.
141. Omran, G.A., G.N. Ratty, and M.A. Jobling, *Genetic variation of 15 autosomal STR loci in Upper (Southern) Egyptians*. Forensic Science International: Genetics, 2009. **3**(2): p. e39-e44.
142. Borosky, A., L. Catelli, and C. Vullo, *Analysis of 17 STR loci in different provinces of Argentina*. Forensic Science International: Genetics, 2009. **3**(3): p. e93-e95.
143. Fridman, C., et al., *Brazilian population profile of 15 STR markers*. Forensic Science International: Genetics, 2008. **2**(2): p. e1-e4.
144. São-Bento, M., et al., *STR data for the 15 AmpFISTR® Identifiler(TM) loci in the Brazilian population of São Paulo State*. Forensic Science International: Genetics Supplement Series, 2008. **1**(1): p. 367-369.
145. Chula, F.G.L., et al., *15 STR loci frequencies with mutation rates in the population from Rio Grande do Sul, Southern Brazil*. Forensic Science International: Genetics, 2009. **3**(2): p. e35-e38.

146. Gonçalves, R., et al., *Genetic profile of a multi-ethnic population from Guiné-Bissau (west African coast) using the new PowerPlex® 16 System kit*. Forensic Science International, 2002. **129**(1): p. 78-80.
147. Zhang, H., et al., *Analysis of 15 STR loci in China population from Sichuan in West China*. Forensic Science International, 2007. **171**(2-3): p. 222-225.
148. Tong, D., et al., *Polymorphism analysis of 15 STR loci in a large sample of the Han population in southern China*. Forensic Science International: Genetics, 2009. **4**(1): p. e27-e29.
149. Chen, T., et al., *Population genetics of nine STR loci from Baoan population in NW China*. Forensic Science International, 2006. **157**(2-3): p. 218-220.
150. Walsh, S.J., J.R. Cullen, and S.A. Harbison, *Allele frequencies for the four major sub-populations of New Zealand at the 10 AMPFISTR SGM Plus loci*. Forensic Science International, 2001. **122**(2-3): p. 189-195.
151. Velosa, R.G., A.T. Fernandes, and A. Brehm, *Genetic profile of the Açores Archipelago population using the new PowerPlex®16 System Kit*. Forensic Science International, 2002. **129**(1): p. 68-71.
152. Pérez-Miranda, A.M., et al., *Allelic frequencies of 13 STR loci in autochthonous Basques from the province of Vizcaya (Spain)*. Forensic Science International, 2005. **152**(2-3): p. 259-262.
153. Gomes, V., et al., *Population data defined by 15 autosomal STR loci in Karamoja population (Uganda) using AmpF/STR Identifier kit*. Forensic Science International: Genetics, 2009. **3**(2): p. e55-e58.
154. Ross, J., et al., *Multiplex PCR amplification of eight STR loci in Austrian and Croatian populations*. International Journal of Legal Medicine, 2001. **115**: p. 57-60.
155. Konjhodzic, R., M. Kubat, and J. Skavic, *Bosnian population data for the 15 STR loci in the Power Plex 16 kit*. International Journal of Legal Medicine, 2004. **118**: p. 191-121.
156. Marjanovic, D., et al., *Allele frequencies for 15 short tandem repeat loci in a representative sample of Bosnians and Herzegovinians*. Forensic Science International, 2006. **156**(1): p. 79-81.
157. Cariolou, M.A., et al., *Allele distribution of 15 STR loci used for human identity purposes in the Greek Cypriot population of the island of Cyprus*. Forensic Science International, 2006. **164**(1): p. 75-78.
158. Kovatsi, L., et al., *Genetic variation for 15 autosomal STR loci (PowerPlex 16) in a population sample from northern Greece*. Forensic Science International, 2006. **159**(1): p. 61-63.
159. Egyed, B., et al., *Analysis of the population heterogeneity in Hungary using fifteen forensically informative STR markers*. Forensic Science International, 2006. **158**(2-3): p. 244-249.
160. Wolanska-Nowak, P., W. Branicki, and T. Kupiec, *STR data for SGM Plus and penta E and D loci in a population sample from south Poland*. Forensic Science International, 2002. **127**(3): p. 237-239.
161. Soltyszewski, I., et al., *Genetic variation of STR loci D3S1358, TH01, D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D, vWA, D8S1179, TPOX and FGA by GenePrint PowerPlex 16 in a Polish population*. Forensic Science International, 2006. **159**(2-3): p. 241-243.
162. Pinheiro, M.F., et al., *Allele frequencies of sixteen STRs in the population of Northern Portugal*. Forensic Science International, 2005. **148**(2-3): p. 221-223.

163. Paredes, M., et al., *STR frequencies for the PowerPlex® 16 System Kit in a population from Northeast Spain*. Forensic Science International, 2003. **135**(1): p. 75-78.
164. Abidin, L., et al., *Analysis of 15 short tandem repeats reveals significant differences between the Arabian populations from Morocco and Syria*. Legal Medicine, 2003. **5**(Supplement 1): p. S150-S155.
165. Brandt-Casadevall, C., et al., *Tunisian population data on 15 PCR-based loci*. Forensic Science International, 2002. **126**(3): p. 272-274.
166. Aslcoglu, F., et al., *Allele distribution data of nine short tandem repeat loci for Turkish population: D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820*. Forensic Science International, 2002. **129**(1): p. 75-77.
167. Mohammed, A.A.A., et al., *STR data for the GenePrint(TM) PowerPlex(TM) 1.2 system loci from three United Arab Emirates populations*. Forensic Science International, 2001. **119**(3): p. 328-329.
168. Sarkar, N. and V.K. Kashyap, *Genetic diversity at two pentanucleotide STR and thirteen tetranucleotide STR loci by multiplex PCR in four predominant population groups of central India* Forensic Science International 2002. **128**(3): p. 196-201
169. Ashma, R. and V.K. Kashyap, *Genetic polymorphism at 15 STR loci among three important subpopulation of Bihar, India*. Forensic Science International, 2002. **130**(1): p. 58-62.
170. Ahmed, A., et al., *STR population data for 10 STR loci including the GenePrint(TM) PowerPlex(TM) 1.2 kit from El-Minia (Central Egypt)*. Forensic Science International, 2001. **117**(3): p. 233-234.
171. Marino, M., A. Sala, and D. Corach, *Population genetic analysis of 15 autosomal STRs loci in the central region of Argentina*. Forensic Science International, 2006. **161**(1): p. 72-77.
172. Sánchez-Diz, P., et al., *Population data on 15 autosomal STRs in a sample from Colombia*. Forensic Science International: Genetics, 2009. **3**(3): p. e81-e82.
173. Gamero, J.J., et al., *A study on ten short tandem repeat systems: African immigrant and Spanish population data*. Forensic Science International, 2000. **110**(3): p. 167-177.
174. Walsh, S.J. and J.S. Buckleton, *Autosomal microsatellite allele frequencies for 15 regionally defined Aboriginal Australian population datasets*. Forensic Science International, 2007. **168**(2-3): p. e29-e42.
175. Budowle, B., D.A. Defenbaugh, and K.M. Keys, *Genetic variation at nine short tandem repeat loci in Chamorros and Filipinos from Guam*. Legal Medicine, 2000. **2**(1): p. 26-30.
176. Zheng, X., et al., *Genetic data of 9 STR loci from Henan Province (central China)*. Forensic Science International, 2007. **169**(2-3): p. 244-246.
177. Wong, D.M., et al., *Population data for 12 STR loci in Honk Kong China*. International Journal of Legal Medicine, 2001. **114**: p. 281-284.
178. Fan, S.-l., et al., *Population genetics polymorphisms on 17 autosomal STRs from China Bai ethnic minority group*. Forensic Science International: Genetics, 2012. **6**(1): p. e22-e23.
179. Buckleton, J., S. Walsh, and J. Mitchell, *Autosomal microsatellite diversity within the Australian population*. Report of the National Institute of Forensic Sciences Standing Committee on Sup-Population Data, 2007.
180. Picornell, A., et al., *Jewish population genetic data in 20 polymorphic loci*. Forensic Science International, 2002. **125**(1): p. 52-58.

181. Zhang, Y., et al., *Allele frequencies for nine STR loci in a Korean ethnic group from northeast China*. Forensic Science International, 2005. **155**(2-3): p. 216-218.
182. Walsh, S.J. and J.S. Buckleton, *Autosomal microsatellite allele frequencies for a nationwide dataset from the Australian CaucasAsian sub-population*. Forensic Science International, 2007. **168**(2-3): p. e47-e50.
183. Neuhuber, F., et al., *Nine STR markers plus amelogenin (AmpFISTR Profiler Plus): a forensic study in an Austrian population*. International Journal of Legal Medicine, 1999. **113**: p. 60-62.
184. Schmid, D., K. Anslinger, and B. Rolf, *Allele frequencies of the ACTBP2 (=SE33), D18S51, D8S1132, D12S391, D2S1360, D3S1744, D5S2500, D7S1517, D10S2325 and D21S2055 loci in a German population sample*. Forensic Science International, 2005. **151**(2-3): p. 303-305.
185. Bashiardes, E., et al., *Data on nine STR loci used for forensic and paternity testing in the Greek Cypriot population of Cyprus*. Forensic Science International, 2001. **123**(2-3): p. 225-226.
186. Andreini, E., et al., *Allele frequencies for nine STR loci (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820) in the Italian population*. Forensic Science International, 2007. **168**(1): p. e13-e16.
187. Pepinski, W., et al., *Population genetics of 15 STR loci in the population of Podlasie (NE Poland)*. Forensic Science International, 2001. **124**(2-3): p. 226-227.
188. Amorim, A., L. Gusmão, and C.n. Alves, *STR data (AmpFISTR Profiler Plus) from north Portugal*. Forensic Science International, 2001. **115**(1-2): p. 119-121.
189. Kornienko, I.V., D.I. Vodolazhsky, and P.I. Ivanov, *Genetic variation of the nine Profiler Plus loci in Russians*. International Journal of Legal Medicine, 2002. **116**: p. 309-311.
190. Goodwin, W., C. Scoular, and A. Linacre, *13 STR loci frequency data from a Scottish population*. Forensic Science International, 2001. **116**(2-3): p. 187-188.
191. Entrala, C., et al., *Fluorescent multiplex analysis of nine STR loci: Spanish population data*. Forensic Science International, 1998. **98**(3): p. 179-183.
192. Tomàs, C., et al., *STR data for the AmpFISTR profiler plus loci from Majorcan, Minorcan and Valencian populations (Eastern Spain)*. Forensic Science International, 2001. **121**(3): p. 201-204.
193. Zarrabeitia, M.a.T. and J.A. Riancho, *Population data on nine STRs from Cantabria, a mountainous region in northern Spain*. Forensic Science International, 2001. **122**(2-3): p. 175-177.
194. Thangaraj, K., et al., *Autosomal STR data on the enigmatic Andaman Islanders*. Forensic Science International, 2007. **169**(2-3): p. 247-251.
195. Pandu, G., et al., *Genetic profile of nine STR loci among Goud and Padmashali populations of Andhra Pradesh, India*. Forensic Science International, 2006. **157**(2-3): p. 201-205.
196. Mastana, S.S., et al., *Genetic variation of 13 STR loci in the four endogamous tribal populations of Eastern India*. Forensic Science International, 2007. **169**(2-3): p. 266-273.
197. Klitschar, M., N. Al-Hammadi, and B. Reichenpfader, *Population genetic studies on the tetrameric short tandem repeat loci D3S1358, VWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820 in Egypt*. Forensic Science International, 1999. **104**(1): p. 23-31.

198. Albeza, M.V., et al., *Genetic variability at 14 STR loci in the Puna population of north western Argentina*. International Journal of Legal Medicine, 2002. **116**: p. 126-132.
199. Tucker, V.C., et al., *Developmental validation of the PowerPlex® ESI 16 and PowerPlex® ESI 17 Systems: STR multiplexes for the new European standard*. Forensic Science International: Genetics, 2011. **5**(5): p. 436-448.
200. Steinlechner, M., et al., *Gabon black population data on the ten short tandem repeat loci D3S1358, vWA, D16S539, D2S1338, D8S1179, D21S11, D18S51, D19S433, TH01 and FGA*. International Journal of Legal Medicine, 2002. **116**: p. 176-178.
201. Nussbaumer, C., et al., *STR data for the AmpFISTR SGM plus from a regional population of Austria*. Forensic Science International, 2001. **122**(2-3): p. 181-183.
202. Steinlechner, M., et al., *Population genetics of ten STR loci (AmpFISTR SGM plus) in Austria*. International Journal of Legal Medicine, 2001. **114**: p. 288-290.
203. Vanek, D., R. Hradil, and B. Budowle, *Czech population data on 10 short tandem repeat loci of SGM Plus STR system kit using DNA purified in FTA™ cards*. Forensic Science International, 2001. **119**: p. 107 - 108.
204. Reichenpfader, B., U. Immel, and M. Klitschar, *Population data on the AmpFISTR SGM plus PCR amplification kit in Germans and Austrians*. Forensic Science International, 2003. **132**(1): p. 84-86.
205. Korostishevsky, M., et al., *Erroneous identification in a mixed population: simulation using Israeli STR data*. International Journal of Legal Medicine, 2004. **118**: p. 111-114.
206. Andreassen, R., S. Jakobsen, and B. Mevaag, *Norwegian population data for the 10 autosomal STR loci in the AMPFISTR® SGM Plus(TM) system*. Forensic Science International, 2007. **170**(1): p. 59-61.
207. Kuzniar, P., E. Jastrzebska, and R. Ploski, *Validation of nine non-CODIS STR loci for forensic use in a population from Central Poland*. Forensic Science International, 2006. **159**(2-3): p. 258-260.
208. Jacewicz, R., et al., *Genetic diversity and forensic evaluation of 10 STR loci in Lodz region of Poland*. Forensic Science International, 2003. **137**(1): p. 94-96.
209. Czarny, J., *Population genetics of the STRs D3S1358, FGA, D2S1338, D8S1179, D21S11, D18S51 and D19S433 in the Pomerania-Kujawy region of Poland*. Forensic Science International, 2002. **125**(1): p. 90-92.
210. Soltyszewski, I., et al., *STR data for the AmpFISTR SGM Plus loci from Warmia and Mazury (NE Poland)*. Forensic Science International, 2004. **141**(1): p. 69-71.
211. Soltyszewski, I., et al., *Genetic data on 10 STR loci a population of western Poland*. Forensic Science International, 2006. **161**(1): p. 69-71.
212. Torío, A., et al., *STR data for the AmpFISTR Profiler Plus loci from Murcia region (SE Spain)*. Forensic Science International, 2002. **129**(2): p. 140-141.
213. Ferdous, A., et al., *Genetic data on 10 autosomal STR loci in the Bangladeshi population*. Legal Medicine, 2006. **8**(5): p. 297-299.
214. Zhivotovsky, L.A., et al., *A comprehensive population survey on the distribution of STR frequencies in Belarus*. Forensic Science International, 2007. **172**(2-3): p. 156-160.
215. Clark, D., et al., *STR data for the AmpFISTR® SGM Plus® loci from two South Asian populations*. Legal Medicine, 2009. **11**(2): p. 97-100.

216. Rębała, K., et al., *Northern Slavs from Serbia do not show a founder effect at autosomal and Y-chromosomal STRs and retain their paternal genetic heritage*. Forensic Science International: Genetics, 2014. **8**(1): p. 126-131.
217. Soltyszewski, I., et al., *Polish population data on 15 autosomal STRs of AmpFISTR NGM PCR kit*. Forensic Science International: Genetics, 2014. **9**(0): p. 142-149.
218. Ali, N., et al., *Genetic variation comparison of 15 autosomal STR loci in an immigrant population living in the UK (British Pakistanis) with an ancestral origin population from Pakistan*. Forensic Science International: Genetics, 2014. **9**(0): p. e10-e12.
219. Pinto, L.M., et al., *Molecular characterization and population genetics of non-CODIS microsatellites used for forensic applications in Brazilian populations*. Forensic Science International: Genetics, 2014. **9**(0): p. e16-e17.
220. Bulbul, O., et al., *Allele frequencies of the five new European Standard Set (ESS) STRs and 15 established STRs in a Turkish population*. Forensic Science International: Genetics, 2014. **9**(0): p. e26.
221. Xie, J., et al., *Genetic distribution on 20 STR loci from the Han population in Shanghai, China*. Forensic Science International: Genetics, 2014. **9**(0): p. e30-e31.
222. Zha, L., et al., *Genetic polymorphism of 21 non-CODIS STR loci in the China Mongolian ethnic minority*. Forensic Science International: Genetics, 2014. **9**(0): p. e32-e33.
223. Westen, A.A., et al., *Comparing six commercial autosomal STR kits in a large Dutch population sample*. Forensic Science International: Genetics, 2014. **10**(0): p. 55-63.
224. Ramírez-Flores, E., et al., *Genetic variation of 24 STR loci in a Mexican Mestizo population from Mexico D.F.* Forensic Science International: Genetics, 2014. **10**(0): p. e4-e6.
225. Tomas, C., et al., *Concordance study and population frequencies for 16 autosomal STRs analyzed with PowerPlex® ESI 17 and AmpFℓSTR® NGM SElect™ in Somalis, Danes and Greenlanders*. Forensic Science International: Genetics, 2014. **11**(0): p. e18-e21.
226. Al-enizi, M., et al., *Population data on 25 autosomal STRs for 500 unrelated Kuwaitis*. Forensic Science International: Genetics, 2014. **12**(0): p. 126-127.
227. Builes, J.J., et al., *Genetic characterization of a population from Cauca Department with 15 STRs markers*. Forensic Science International: Genetics Supplement Series. **4**(1): p. e81-e82.
228. Hashiayda, M., et al., *Genetic variation in a Japanese population, using the multiplex 24 STRs analysis system*. Forensic Science International: Genetics Supplement Series. **4**(1): p. e41-e42.
229. Souto, L., et al., *Evaluation of the PowerPlex Fusion System in a sample from East Timor*. Forensic Science International: Genetics Supplement Series. **4**(1): p. e105-e106.
230. Bento, A.M., et al., *Population data for Central Portugal population with NGM amplification kit*. Forensic Science International: Genetics Supplement Series. **4**(1): p. e152-e153.
231. Wang, B., et al., *Genetic polymorphisms of 20 STR loci in China Han population in Tianjin, North China*. Forensic Science International: Genetics Supplement Series. **4**(1): p. e192-e193.

232. Farias, D.A.B., et al., *Population genetics data on 15 STR markers in the population of Ceará State, Northeast Region of Brazil*. Forensic Science International: Genetics Supplement Series. **4**(1): p. e250-e251.
233. Yuan, L., et al., *Population genetics analysis of 38 STR loci in the She population from Fujian Province of China*. Legal Medicine, 2014. **16**(5): p. 314-318.
234. Fujii, K., et al., *Allele frequencies for 22 autosomal short tandem repeat loci obtained by PowerPlex Fusion in a sample of 1501 individuals from the Japanese population*. Legal Medicine, 2014. **16**(4): p. 234-237.
235. Anghel, A., et al., *Genetic polymorphism data on 15 autosomal STR markers in a Western Romanian population sample*. Legal Medicine, 2014. **16**(4): p. 238-240.
236. Ozeki, M. and K. Tamaki, *Allele frequencies of 37 short tandem repeat loci in a Japanese population*. Legal Medicine, 2013. **15**(6): p. 342-346.
237. Romani, C., et al., *Allele frequencies for 26 STR loci in a population of Tuscany (Central Italy)*. International Journal of Legal Medicine, 2014. **128**(3): p. 455-456.
238. Parolin, M., et al., *Analysis of 15 autosomal STR loci from Mar del Plata and Bahia Blanca (Central Region of Argentina)*. International Journal of Legal Medicine, 2014. **128**(3): p. 457-459.
239. Gehrig, C., et al., *Allelic proportions of 16 STR loci—including the new European Standard Set (ESS) loci—in a Swiss population sample*. International Journal of Legal Medicine, 2014. **128**(3): p. 461-465.
240. Rangel-Villalobos, H., et al., *Forensic evaluation of the AmpFℓSTR Identifiler kit in nine Mexican native populations from the pre-Columbian Mesoamerican region*. International Journal of Legal Medicine, 2014. **128**(3): p. 467-468.
241. Alves, H., et al., *STR data for 15 autosomal STR markers from Paraná (Southern Brazil)*. International Journal of Legal Medicine, 2014. **128**(2): p. 269-270.
242. Lucassen, A., et al., *Allele frequency data of 15 autosomal STR loci in four major population groups of South Africa*. International Journal of Legal Medicine, 2014. **128**(2): p. 275-276.
243. Laurin, N. and E. Milot, *Population genetic data of the AmpFℓSTR® Identifiler® Plus and PowerPlex® 16 HS STR loci in four Canadian populations*. International Journal of Legal Medicine, 2014. **128**(2): p. 277-278.
244. Älgenäs, C. and A. Tillmar, *Population genetics of 29 autosomal STRs and 17 Y-chromosomal STRs in a population sample from Afghanistan*. International Journal of Legal Medicine, 2014. **128**(2): p. 279-280.
245. Turrina, S., et al., *Evaluation of genetic parameters of 22 autosomal STR loci (PowerPlex® Fusion System) in a population sample from Northern Italy*. International Journal of Legal Medicine, 2014. **128**(2): p. 281-283.
246. Gurkan, C., et al., *Population genetic data for 15 autosomal STR markers in Turkish Cypriots from Cyprus*. Forensic Science International: Genetics, 2015. **14**(0): p. e1-e3.
247. Song, F., H.-b. Luo, and Y.-p. Hou, *Population data of 21 non-CODIS STR loci in the China Uyghur ethnic minority*. Forensic Science International: Genetics, 2014. **13**(0): p. e1-e2.
248. Hossain, T., et al., *Population genetic data on 15 autosomal STR loci in Bangladeshi population*. Forensic Science International: Genetics, 2014. **13**(0): p. e4-e5.

- 249. Zhai, D., et al., *The allele frequency of 15 STRs among three Tibeto-Burman-speaking populations from the southwest region of mainland China*. Forensic Science International: Genetics, 2014. **13**(0): p. e22-e24.
- 250. Borosky, A., et al., *Forensic population data for 20 STR loci in Argentina*. Forensic Science International: Genetics, 2014. **13**(0): p. e20-e21.

Table

	CSF1PO	D10S1248	D12S391	D13S317	D16S539	D18S51	D19S433
USAf.1	-0.0837	NA	NA	0.1480	-0.0027	-0.0032	NA
Bah.Af.	-0.0891	NA	NA	0.1131	-0.0124	-0.0019	NA
USAf.2	-0.0648	NA	NA	0.0997	-0.0090	-0.0050	NA
USAf.3	-0.0646	NA	NA	0.1520	-0.0101	-0.0007	NA
USAf.4	-0.0622	NA	NA	0.1107	-0.0136	NA	NA
USAf.5	-0.0687	NA	NA	0.1036	-0.0130	-0.0008	NA
Jam.Af.	-0.0574	NA	NA	0.1488	-0.0009	0.0029	NA
USAf.6	-0.0729	NA	NA	0.1034	NA	-0.0016	NA
USAf.7	-0.0718	NA	NA	0.1566	-0.0113	0.0042	NA
Trin.Af.	-0.0799	NA	NA	0.0389	-0.0369	-0.0097	NA
USAf.8	-0.0594	NA	NA	0.1353	-0.0230	0.0021	NA
Rwanda	-0.0790	NA	NA	0.0949	-0.0257	0.0034	NA
Colombia.1	0.0154	NA	NA	0.0023	-0.0058	-0.0003	NA
Angola.1	-0.0820	NA	NA	0.1709	0.0364	0.0204	-0.0157
Haiti	-0.0769	NA	NA	0.1391	0.0019	0.0002	NA
S.Af.Colored	-0.0621	NA	NA	0.0676	0.0025	-0.0370	-0.0441
Khoe	-0.1331	NA	NA	0.2181	-0.0146	-0.0527	-0.0476
San	-0.1209	NA	NA	0.0994	-0.0145	-0.0106	-0.0307
S.Africa.1	-0.0438	NA	NA	0.1060	0.0480	0.0135	0.0355
UKAf.1	NA	-0.0407	0.0219	NA	-0.0080	-0.0043	-0.0044
USAf.9	-0.0601	-0.0483	0.0230	0.1326	-0.0054	-0.0022	-0.0332
Somalia	-0.0233	NA	NA	0.0882	0.0132	-0.0271	0.0141
Africa	-0.0635	NA	NA	0.1100	-0.0092	-0.0157	0.0747
Angola.2	-0.0641	NA	NA	0.1359	0.0287	0.0114	-0.0507
Eq.Guinea.1	-0.0671	NA	NA	0.1592	0.0001	0.0215	-0.0425
Eq.Guinea.2	-0.0720	NA	NA	0.1410	-0.0215	0.0363	-0.0304
USAf.10	-0.0694	NA	NA	0.1032	0.0095	-0.0003	-0.0002
Mozambique	-0.0274	NA	NA	0.1130	0.0130	0.0034	0.0333
Tanzania	-0.0763	NA	NA	0.1369	0.0119	0.0056	-0.0222
Uganda	-0.0441	NA	NA	0.1391	-0.0198	0.0090	-0.0020
Guinea-Bissua	-0.0862	NA	NA	0.1799	0.0019	0.0059	NA
UKAf.2	-0.0787	NA	NA	0.1629	0.0136	-0.0025	NA
Gabon	NA	NA	NA	NA	-0.0129	0.0295	-0.0543
Somalis.225	NA	-0.0300	0.0501	NA	-0.0126	-0.0271	0.0267
SABlack.242	-0.0479	NA	NA	0.0953	0.0399	-0.0063	0.0125
β_w	-0.0668	-0.0397	0.0317	0.1221	-0.0018	-0.0012	-0.0095
Aust.Abor.1	0.0079	NA	NA	0.0759	0.0600	0.0025	0.1725
Aust.Abor.2	0.0181	NA	NA	0.0852	0.0593	0.0102	0.1596
β_w	0.0130	NA	NA	0.0806	0.0597	0.0064	0.1661
China.1	-0.0041	NA	NA	-0.0008	0.0346	0.0163	NA
Asia	0.0174	NA	NA	0.0226	0.0078	0.0092	NA
Japan.1	0.0025	NA	NA	0.0179	0.0066	0.0313	NA
Japan.2	0.0014	NA	NA	0.0102	0.0192	0.0138	NA
Korea.1	0.0075	NA	NA	-0.0044	0.0098	0.0246	NA
Vietnam.1	0.0239	NA	NA	0.0489	0.0000	0.0412	NA