Population Genetic Issues for Forensic DNA Profiles.
NIJ 2011-DN-BX-K541
January 1, 2014 - December 31, 2014
Final Report

Bruce S. Weir

May 22, 2015

The specific aims for this award were to develop and apply new population genetic theory to aid the interpretation of DNA profiles, with an emphasis on population structure, lineage markers and mixtures. Good progress was made, as is now described.

Estimation of Population Structure Parameter

A DNA match probability is the probability an untyped person has a DNA profile given that a typed person has the profile, and this depends on the genetic structure of the population to which these two people belong. The key result, for a single-locus homozygote AA for example, for the match probability is

$$\Pr(AA|AA) = \frac{[2\theta + (1-\theta)p_A][3\theta + (1-\theta)p_A]}{(1+\theta)(1+2\theta)}$$

where p_A is the population frequency for allele A and θ is the population structure parameter. There is some doubt as to the appropriate value for θ and we have developed new theory and applied that to a survey of published STR frequencies.

The problem with structured populations arises when the people of interest belong, or are assumed to belong, the same subpopulation but data are available from only the whole population. The number and nature of subpopulations is generally unknown. It is helpful to introduce θ_i for a random pair of alleles in the ith subpopulation and $\theta_{ii'}$ for a random pair of alleles, one from the ith and one from the i'th subpopulation. These θ's are often regarded to be probabilities of identity by descent, although more generally they are correlations for pairs of alleles. Averaging over subpopulations and pairs of subpopulations gives the within- and between-subpopulation quantities θ_W and θ_B. We have shown that the matching probability P_M for an allele, where the two sources of the allele are in the same subpopulation, is

$$P_M = \theta_W + (1-\theta_W)H \tag{1}$$

where H is the sum of squares of population allele frequencies. This result is an average over all subpopulations and over all alleles. If H is replaced by $\tilde{H}$, its value in a sample from the whole population, then the match probability is estimated by

$$\hat{P}_M = \beta_W + (1-\beta_W)\tilde{H} \tag{2}$$

Here, $\beta_W = (\theta_W - \theta_B)/(1-\theta_B)$ is usually written as θ or as F_{ST}. A moment estimate is

$$\hat{\beta}_W = \frac{\tilde{M}_W - \tilde{M}_B}{1-\tilde{M}_B} \tag{3}$$

The quantities $\tilde{M}_W, \tilde{M}_B$ are the proportions of pairs of alleles that match, within subpopulations or between pairs of subpopulations, averaged over single and pairs or subpopulations. Specifically, for population i, if a sample of n_i alleles from that population (i.e. $n_i/2$ individuals) has n_{iu} copies of allele type u then the sample matching proportion is $\tilde{M}_i = \sum_u n_{iu}(n_{iu}-1)/[n_i(n_i-1)]$ and the average over r populations is $\tilde{M}_W - \sum_i \tilde{M}_i/r$. Similarly, the proportion of pairs of alleles, one from population i and one from population j that match is $\tilde{M}_{ij} = \sum_u n_{iu}n_{ju}/(n_i n_j)$ and the average over all pairs of populations is $\tilde{M}_B = \sum_{i\neq j} \tilde{M}_{ij}/[r(r-1)]$. There may be interest in values obtained at a locus for each of the populations, and the appropriate moment estimates are $\hat{\beta}_i = (\tilde{M}_i - \tilde{M}_B)/(1-\tilde{M}_B)$. The quantity $\hat{\theta}_W$ is the average of these: $\hat{\beta}_W = \sum_i \hat{\beta}_i/r$.

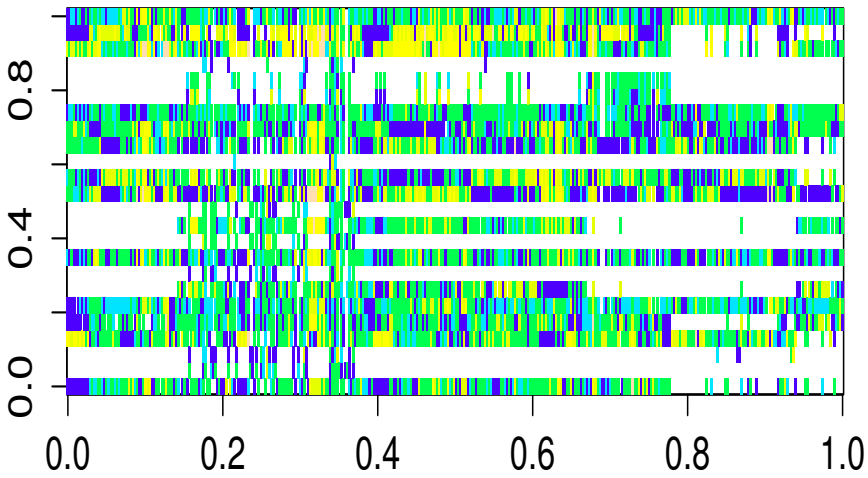

Figure 1: Moment estimates of β_{il}.

Moment estimators are designed to have low bias but they can have substantial variances. In practice, a common value is assigned to all loci and then the appropriate estimator is

$$\hat{\beta}_W = \frac{\sum_l (\tilde{M}_{Wl} - \tilde{M}_{Bl})}{\sum_l (1 - \tilde{M}_{Bl})}$$

where $\tilde{M}_{Bl}, \tilde{M}_{Wl}$ are the observed matching proportions at locus l within subpopulations and between pairs of subpopulations.

Moment estimates make few assumptions, other than the parametric form of allele frequency means and variances, and they are very easy to calculate. If more assumption are made then estimates with smaller variances can be obtained, and we have followed the example of Balding and colleagues (e.g. Steele and Balding, 2014; Steele et a., 2014) in assuming allele frequencies have a Dirichlet distribution over populations, assuming a beta distribution for θ and using Bayesian methods to find a posterior distribution for θ. We have found it convenient to use the BayeScan software (Foll et al., 2010) in practice

We have surveyed the forensic science literature for published allele frequencies for autosomal STR loci. We have used data from 378 populations, with frequency data on up to 24 loci. For each locus-population combination we estimated β_W as above. In Figure 1 we summarize these results. The 378 populations are displayed on the X-axis, and the 24 loci on the Y-axis. White cells indicate that data were not available for that combination. Blue values are negative, green values are small and positive, and brown values are larger and positive. The extreme variability of these estimates reduces their value.

In Table 1, we show the β_{il} and β_{Wl} estimates for both the collection of populations within

Table 1: Moment Estimates of β.

Locus	African	AustAb.	Asian	Caucas.	Hispanic	Indo-Pak	Nat.Am.	Polynes.	Average
VWA	-0.064	0.018	0.014	0.012	0.021	0.019	0.005	0.023	0.006
CSF1PO	0.028	–	0.058	-0.014	0.009	0.023	–	–	0.021
D1S1656	0.018	–	0.014	0.017	0.082	0.056	–	–	0.037
D2S441	-0.008	0.029	0.021	0.003	0.011	0.005	0.119	-0.000	0.022
D2S1338	0.004	-0.002	0.039	-0.049	0.004	0.003	0.074	0.034	0.013
D3S1358	0.032	0.008	0.003	0.066	0.060	0.032	0.061	-0.000	0.032
D5S818	-0.002	–	0.004	0.060	0.001	–	–	–	0.015
D6S1043	0.026	0.043	0.028	-0.002	0.014	0.017	0.046	-0.005	0.021
D7S820	0.044	-0.006	-0.025	0.012	0.030	-0.034	0.083	0.021	0.015
D8S1179	-0.034	–	0.031	0.011	0.055	-0.010	–	–	0.010
D10S1248	0.005	–	0.042	-0.029	0.045	0.008	–	–	0.014
D12S391	0.121	0.068	0.011	0.039	-0.007	0.001	-0.007	0.032	0.032
D13S317	-0.007	0.062	0.003	0.024	0.003	-0.004	0.044	0.022	0.018
D16S539	-0.005	0.004	0.022	0.004	0.002	0.038	0.022	0.054	0.018
D18S51	-0.024	0.166	0.013	0.031	-0.007	0.016	0.004	-0.008	0.024
D19S433	-0.012	-0.010	0.031	0.009	0.013	-0.008	0.018	0.032	0.009
D21S11	-0.068	–	0.006	0.072	0.122	0.016	–	–	0.030
D22S1045	0.003	0.004	0.009	0.013	-0.002	0.005	-0.001	0.062	0.011
FGA	-0.024	–	0.068	0.026	0.021	0.023	-0.005	–	0.018
PENTAD	0.022	–	0.008	0.022	0.015	0.008	0.014	–	0.015
PENTAE	0.017	–	0.001	-0.002	0.010	-0.002	–	–	0.004
SE33	0.056	0.065	0.121	-0.003	0.018	0.003	0.190	0.031	0.060
TH01	-0.070	0.031	0.146	0.119	-0.019	0.026	0.094	0.102	0.053
TPOX	-0.006	0.032	0.004	0.007	0.031	-0.000	0.053	0.023	0.018
Average	0.002	0.021	0.028	0.019	0.022	0.010	0.034	0.017	0.019

each of eight continental-ancestry groups, and for all 378 populations. Using more loci or more populations clearly decreases variation in the estimates. The "Caucas." values, for example, show β_W values for each locus and then for all 24 loci for the 152 populations we classified as Caucasian. For these estimates the $\tilde{M}_B$ quantities were for all pais of the the Caucasian populations. The value of 0.019 for all loci is the value we would suggest using for a Caucasian subpopulation when data were available from only a larger, maybe national, database. We do note, however, the large variation in values among loci even when those use all the populations.

We note that the group-specific estimates in Table 1 are generally larger than the commonly accepted value of 0.01, and our peer-reviewed publication will containg the recommendation that a more appropriate value is 0.03. The practical implications for profiles made from many loci, such as the CODIS set, will not be of great significance: match probabilities will remain small if "θ" is changed from 0.01 to 0.03. There is a small effect from our use of unweighted estimators, following the recommendation of Bhatia et al. (2013), instead of the weighted analyses of Weir and Cockerham (1984), but the main reason our estimates are larger than those reported by some other authors (e.g. Budowle et al., 2001) is that we have used a wider collection of sampled populations. We believe our more conservative estimates are appropriate for the usual situation where the exact ancestral background of an unknown contributor to an evidence profile is not known.

In Figure 2 we contrast the simple moment estimates, using all loci but for each of the 378

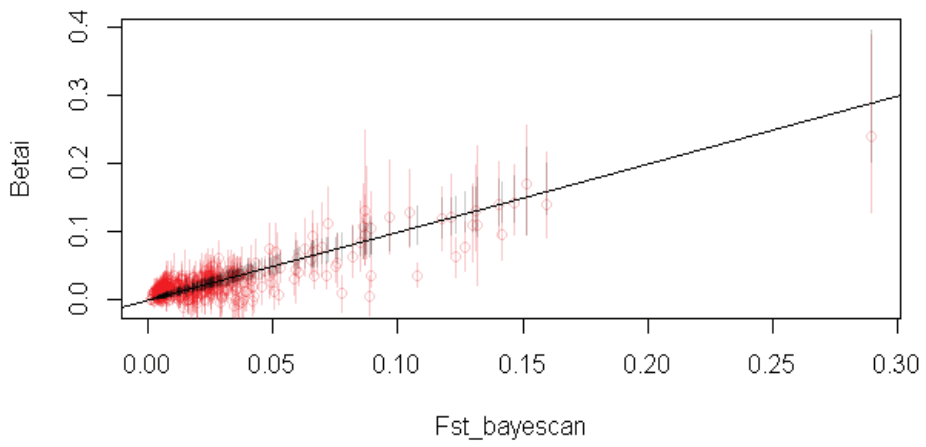

Figure 2: Moment (red) vs Bayesian (black) Estimates.

populations with the Bayesian estimates. Specifically, we show the 95% confidence intervals for the moment estimates, obtained by bootstrapping over loci, and the 95% credible intervals from the Bayesian posterior distributions. We believe the reason for the two intervals not to overlap in about one third of the populations is because the Bayesian approach ignores the correlation in allele frequencies between populations.

Two publications, one containing the theoretical development and one containing the applications to the survey data, are about to be submitted.

Y-STR Match Probabilities

There is growing interest in the use of Y-STR profiles for forensic purposes. Issues have arisen on how to determine match probabilities and how to combine Y-STR and autosomal match probabilities. We have addressed each of these issues.

Match probabilities follow the same logic as shown in the last section for autosomal alleles, except that now they apply to Y-haplotypes rather than separate alleles at each locus because of the lack or recombination on the Y chromosome. Specifically, for haplotype A, the match probability is

$$\Pr(A|A) \quad = \quad \theta_{Yi} + (1 - \theta_{Yi})p_A$$

where θ_{Yi} is the value for the Y-markers in this haplotype for the ith subpopulation and p_A is the haplotype frequency in the whole population. Averaging over subpopulations and haplotypes, the match probability is estimated as

$$\hat{P}_M \quad = \quad \frac{\tilde{M}_W - \tilde{M}_B}{1 - \tilde{M}_B} + \frac{1 - \tilde{M}_W}{1 - \tilde{M}_B}\tilde{H}$$

as in Equations 2, 3. In Table 2 we show data from the NISTpop.htm page of STRBase. The values shown are for matching averaged over within and between-pairs of the four groups African

Table 2: Matching Proportions for Y-STR Loci in NIST Database

Locus	$\tilde{M}_W$	$\tilde{M}_B$	$\hat{\beta}_W$
DYS19	0.32571062	0.24309148	0.10915340
DYS385a/b	0.07982377	0.04427420	0.03719640
DYS389I	0.41279418	0.38319082	0.04799436
DYS389II	0.26072434	0.23741323	0.03056847
DYS390	0.28981997	0.18813203	0.12525182
DYS391	0.52191425	0.48517426	0.07136392
DYS392	0.39961865	0.35168087	0.07394164
DYS393	0.50285122	0.48769253	0.02958906
DYS437	0.46400112	0.38595032	0.12710828
DYS438	0.36817530	0.23212655	0.17717601
DYS439	0.35507469	0.34990863	0.00794667
DYS448	0.30091326	0.22640195	0.09631787
DYS456	0.33444029	0.32578009	0.01284478
DYS458	0.21642167	0.19701369	0.02416976
DYS481	0.18867019	0.14121936	0.05525373
DYS533	0.39365769	0.37177174	0.03483757
DYS549	0.33976578	0.30691346	0.04740003
DYS570	0.21298105	0.20775666	0.00659442
DYS576	0.20955290	0.18125443	0.03456321
DYS635	0.27720127	0.20653182	0.08906400
DYS643	0.28394262	0.20058158	0.10427710
Y-GATA-H4	0.40667782	0.39899963	0.01277568

American, Caucasian, Hispanic, and Asian. There is variation among loci. We do not have data from populations within each of these four groups, so the β_W estimates are larger than they would be for use within one group. The estimates in Table 2 were produced as in Equation 3, where matching now refers to the alleles at each locus.

A more helpful indication of θ values is provided in Figure 3, where all possible haplotypes of 1 to 23 loci are used to estimate β_W for the same NIST data. The red line is the median value for all sets of the specified number of loci, the blue lines delineate the central 95% of the values and the black lines show the maxima and minima. Note that the estimates are shown on a logarithmic scale, and that independence of mutation across loci would suggest a linear dependence on the number of loci. (For STR loci undergoing stepwise mutations that change the number of repeat units by 1, Kimura and Ohta (1975) showed that $\beta_W = 1/\sqrt{(1+8N\mu)}$ for populations of size N and mutation rate μ. In other words $-\ln(\beta_W) = 4N\mu$ and we might assume the haplotype mutation rate is proportional to the number of loci.) This is clearly not the case, showing that as more loci match, the greater the chance that additional loci will also match. The same phenomenon was noted for autosomal loci by Laurie and Weir (2003).

Walsh et al. (2008) used a coalescent approach to address the effect of Y-STR matching on autosomal matching, and their work was followed up by Buckleton and Myers (2014). We are finding it helpful to introduce θ_{AY} as the probability that, for two men in the same population, their Y-haplotypes are identical by descent and so are a pair of autosomal alleles, one taken randomly from each. Our interest will center on the conditional identity probabilities: $\theta_{Y|A} = \theta_{AY}/\theta_A$ for Y identity given autosomal identity and $\theta_{A|Y} = \theta_{AY}/\theta_Y$ for autosomal identity given Y identity.

There are five possible arrangements $R_i, i = 1, 2, ..., 5$ of two autosomal and two Y alleles. In

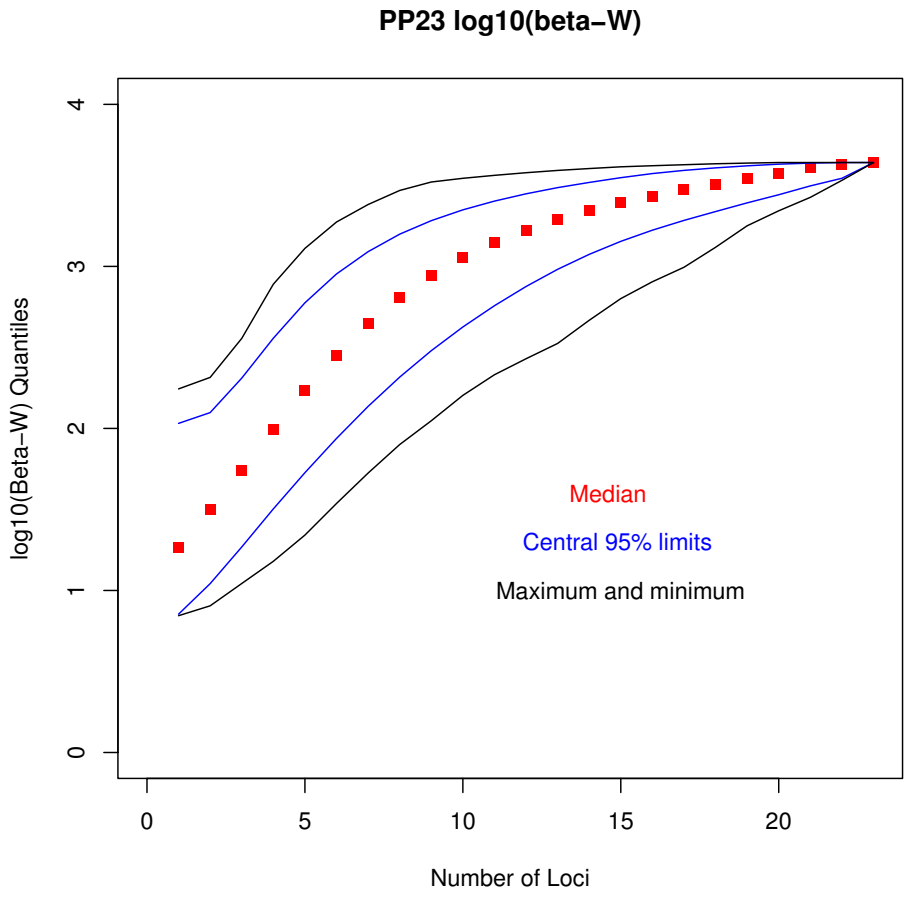

Figure 3: Dependence of β_W on the Number of Y-STR Loci.

the following list of these arrangements, a, a' are two autosomal alleles and y, y' are two Y alleles (or haplotypes), and brackets enclose alleles in the same individual:

$$\begin{aligned} R_1 &: [ay], [a'y'] \\ R_2 &: [aa'y], [y'] \\ R_3 &: [ay], [a'], [y'] \\ R_4 &: [aa'], [y], [y'] \\ R_5 &: [a], [a'], [y], [y'] \end{aligned}$$

To establish the transition equations for the five probabilities each of the four alleles is traced back to an individual in the previous generation.

Numerical iteration of the transition equations for a range of values of $N = N_M = N_F$ and μ (with $\nu_A = (1-\mu)^2$, $\nu_Y = (1-\mu)^{40}$ for 20 Y loci) are shown in Table 3. This shows that Y-matching has little effect on autosomal coancestry when θ_A, θ_Y are large but the effect can be substantial when they are small.

Another view of these results is shown in Figure 4. Only for moderately large values of θ_A can $\theta_{A|Y}$ be equated to θ_A and match probabilities for autosomal and Y profiles be multiplied.

Table 3: Equilibrium Values of Joint and Conditional Identities

Table 1

N	μ	$\hat{\theta}_Y$	$\hat{\theta}_{AY}$	$\hat{\theta}_A$	$\hat{\theta}_{A\|Y}$	$\hat{\theta}_{AY}/(\hat{\theta}_A\hat{\theta}_Y)$
10^4	10^{-2}	0.00040	0.00001270	0.00123	0.03143	25.5580
10^4	10^{-3}	0.00447	0.00007101	0.01233	0.01587	1.2878
10^4	10^{-4}	0.04343	0.00483898	0.11110	0.11142	1.0029
10^5	10^{-2}	0.00004	0.00000123	0.00012	0.03036	246.6184
10^5	10^{-3}	0.00045	0.00000217	0.00125	0.00483	3.8785
10^5	10^{-4}	0.00452	0.00005742	0.01234	0.01271	1.0293
10^6	10^{-2}	0.00000	0.00000012	0.00001	0.03025	2457.2222
10^6	10^{-3}	0.00004	0.00000017	0.00012	0.00372	29.7852
10^6	10^{-4}	0.00045	0.00000073	0.00125	0.00161	1.2928

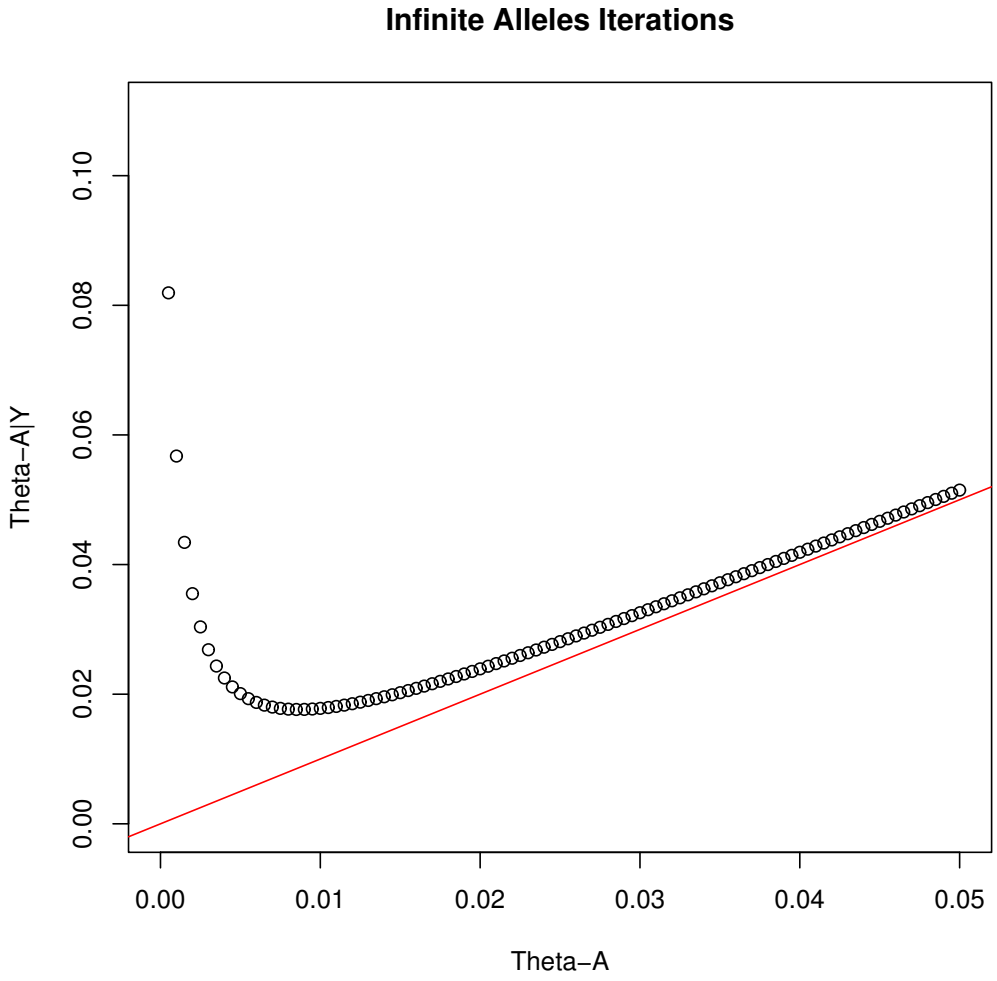

Figure 4: Conditional vs Unconditional Autosomal θ values.

Both the theoretical work and applications to published Y-STR data are about to be submitted for publication.

Continuous Model for Mixtures

Over the past three years we have made substantial contributions to the literature on providing numerical characterization of the evidentiary strength of DNA evidence. Our work assumes the applicability of likelihood ratios, and it has been designed to avoid problems with the "binary model" where decision rules on allelic presence in a profile rest on detection or analysis thresholds. These problems have been described by Evett et al. (1998) and others. Apart from the difficulty of assigning values to thresholds and not attaining a conservative interpretation there is the real danger of ignoring relevant information in the electropherograms for STR markers (Perlin et al., 2011) with the binary model.

There have been studies (e.g. Butler, 2006) showing a divergence in interpreting mixtures by different laboratories when, for example, alleles are called depending on peak heights in relation to a threshold. A consequence of rigid thresholds is that profiles that differ in the most minor way, say replicates from a single extraction, can lead to opposing interpretations. A counterexample to the apparent "conservativeness" of not calling alleles below a threshold was described by Lohmueller and Rudin (2013): a potential contributor, whose profile inclusion in a mixed profile would have favored the defense, was excluded by the binary model. Another concern is the widespread use of the "$2p$" rule in cases of allele dropout as this can be quite non-conservative (Buckleton and Triggs, 2006). A growing literature (Cowell et al., 2008; Balding and Buckleton, 2009; Haned, 2011; Perlin et al., 2011; Lohmueller and Rudin, 2013) preceded our own basic paper (Taylor et al., 2013) on continuous models that avoid thresholds: these papers have reviewed difficulties with the binary model.

It is convenient to describe our work as having three stages. Firstly there is modeling of the complexities of STR electropherograms to account for heterozygote imbalance, allelic dropout and stutter peaks. In Bright et al. (2013) we described models for allele and stutter peak heights and we referred to our empirical studies. We confirmed the dependence of the ratio of stutter to parent peak height on the longest uninterrupted sequence of repeat units (LUS) as opposed to the allele size or total number of repeat units. We modeled allele plus stutter peak heights as a function of molecular weight with three "mass variables" a locus effect, a replicate effect, and the slope of the regression line (Bright et al., 2013a). We subsequently allowed for a non-linear relationship (Bright et al., 2013b).

The heart of our approach is contained in Taylor et al. (2013). We consider alternative hypotheses $H_m, m = 1, 2$ for an STR profile G. Usually H_1 denotes the prosecution hypothesis and H_2 that of the defense. For each H_m we consider all sets S_j of multi-locus genotypes consistent with that hypothesis. Once the genotypes are specified the hypotheses are not needed and we work with genotype weights $w_j = \Pr(G|S_j)$. The likelihood ratio is

$$\text{LR} = \frac{\Pr(G|H_1)}{\Pr(G|H_2)} = \frac{\sum_j \Pr(G|S_j)\Pr(S_j|H_1)}{\sum_j \Pr(G|S_j)\Pr(S_j|H_2)} = \frac{\sum_j w_j \Pr(S_j|H_1)}{\sum_j w_j \Pr(S_j|H_2)}$$

Note that the binary model assigns every set of genotypes the weight of 0 or 1 depending on whether the profile G is deemed impossible or possible to have originated from the genotypes specified by S_j under H_m. The collection $(\{S_j\})$ of sets of genotypes will be different under H_1 and H_2. The continuous approach avoids the procedure sometimes employed under the binary model of omitting loci if certain criteria are not met. This is conservative only if the LR based on the approach

described here is greater than one. Alleles that fail to meet a threshold may well have low or zero probability under one or other hypothesis.

An exact analytical approach that would account for all the complexities of electropherograms is not possible so we have adopted a Markov chain Monte Carlo (MCMC) method. Briefly, we choose an S_j and this specifies the molecular weights for each allele in the profile. These, and our other parameters and model, lead to expected allelic and stutter peak heights E_i, where i ranges over all peaks. Bright et al. (2013a) showed that ratio of the observed peak heights O_i to these expected values has a log-normal distribution: $\ln(O_i/E_i) \sim N(0, \sigma^2)$. This allows probabilities to be attached to the $\Pr(O_i|S_j)$'s: the genotype S_j is then changed at a randomly-chosen locus and the probability of the profile $\{O_i\}$ is re-calculated with the updated probabilities. If the profile now has a higher probability the new set S_j is "accepted" and becomes the new profile, otherwise it is accepted with a probability that is less than one. This Metropolis-Hastings algorithm leads to all profiles S_j being visited by the process with a frequency that depends on the probabilities of the profile given the S_j and the procedure provides numerical values of the weights w_j in an efficient way. The probabilities $\Pr(S_j|H_m)$ just use standard methods: if S_j was the set (AA, BC) of genotypes at one locus that under H_1 accounted for the alleles observed in the evidentiary profile then $\Pr(S_j|H_1) = p_A p_B p_C (1-\theta)^2[\theta + (1-\theta)p_A]/[(1+\theta)(1+2\theta)(1+3\theta)]$ (Balding and Nichols, 1994). Sufficient details are given by Taylor et al. (2013) to allow other investigators to write their own computer code to implement our method.

The final step is to attach probabilities to the likelihood ratios. Although there is merit in calculating and reporting only a point estimate of the likelihood ratio, using the MCMC-derived weights Ww_j and the conventional profile probabilities $\Pr(S_j|H_m)$, there is still the difficulty of interpreting this value. When does a likelihood ratio indicate compelling evidence: is a million sufficiently large, or is a billion necessary? Other authors have discussed this. We made an initial attempt to address this question (Beecham and Weir, 2011). For a multi-locus situation, we regarded the logarithm ln(LR) as being normally distributed and constructed confidence intervals of the form $\mathrm{LR}/C, \mathrm{LR} \times C)$ where C depended on the variance of the estimated LR, taking into account sampling variation for the allele frequency database and the variation among populations (i.e. the "theta" effects).

In Taylor et al. (2014) we considered several sources of variation that affect the distribution of LR values. Specifically, we incorporated uncertainty in allele frequencies, uncertainty in θ, uncertainty in genotype weights w_j, and uncertainty in relatedness amongst hypothesized unknown contributors. In each of these four directions, distributions were assumed for the appropriate parameters and sampling from these distributions was added to the LR calculations. The procedure does add to the computational burden but it has the advantage of being able to attach probabilities: the probability of the calculated LR or some more extreme value. It could be argued that there is still an element of subjectivity, but it seems less so than does a verbal scale (e.g. "weak, moderate, strong, very strong") for LR values.

Outreach Activities

The work on population structure and Y-STR matching played a large part in our contributions to the "SWGDAM Interpretation Guidelines for Y-Chromosome STR Typing by Forensic DNA Laboratories" approved by SWGDAM on January 9, 2014. Weir hosted a meeting of the SWGDAM Working Group, that included Buckleton, at the University of Washington in February 2013, Weir attended a SWGDAM meeting, and Buckleton and Weir took part in several conference calls. They both attended the July, 2014 SWGDAM meeting. The Guidelines contain explicit language about the distinction between profile probabilities (p_u) and profile match probabilities $P_{u|u}$:

> "10.3 Theta (θ) is used in the following equation for the match probability, $\Pr(A|A) = \theta + (1-\theta)p_A$, (3) where A is the haplotype of interest and $\Pr(A|A)$ is the probability of observing A given that it has already been seen once in another individual of the same subpopulation. p_A is the profile probability.
> 10.3.1 Equation (3) is a match probability. It is the haplotype analog of the formula described in National Research Council (1996) Recommendation 4.2."

The same language was used by Weir et al. (2014). Weir (2007) had already discussed match and profile probabilities, "Among the many advantages of adopting this approach to comparing competing hypotheses is the clarification that it is match probabilities $\Pr(G_S|G_C)$ for profiles from two people that are relevant rather than profile probabilities $\Pr(G_S)$." Buckleton et al. (2011) had also said "Note that the match probability within a particular subpopulation is also greater than the haplotype frequency in the whole population since $\theta + (1-\theta)p_A > p_A$." This crucial distinction between match and profile probabilities has recently been stressed by Brenner (2014).

The methods described above were used by SWGDAM to estimate "θ" (actually β_W) for the US Y-STR data as well as the NIST data (not shown in the Guidelines).

Weir is participating in a new SWGDAM Working Group to examine software for continuous approaches to interpreting STR profiles, and Buckleton is on an ISFG Commission looking at validation of such software.

Weir and Curran serve on the ad-hoc Committee for Forensic Science established by the American Statistical Association.

Weir was on the advisory committee for the 9th International Conference on Forensic Inference and Statistics held in Leiden in 2014: he and Curran presented papers at that conference.

Weir was a member of the Working Group of the United Nations Office of Drugs and Crime that met in 2013 to establish guidelines for the identification of seized ivory.

Literature Cited

- Balding DJ, Buckleton J. 2009. Interpreting low template DNA profiles. Forensic Science International: Genetics 4: 1-10.
- Beecham GW, Weir BS. 2011. Confidence intervals for DNA evidence likelihood ratios. Journal of Forensic Sciences 56 Supplement 1:S166-S171.
- Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting F_{ST}: the impact of rare variants. Genome Research 23:1514-1521.
- Brenner CH. 2014. Understanding Y haplotype matching probability. Forensic Science International: Genetics 8:233-243.
- Bright J, Taylor D, Curran JM, Buckleton J. 2013a. Developing allelic and stutter peak height models for a continuous method of DNA interpretation. Forensic Science International: Genetics 7:296-304.
- Bright J-A, Taylor D, Curran JM, Buckleton JS. 2013b. Degradation of forensic DNA profiles. Australian Journal of Forensic Science 45: 445-449.
- Buckleton J, Myers S. 2014. Combining autosomal and Y chromosome match probabilities using coalescent theory. Forensic Science International: Genetics 11:52-55.
- Budowle B, Shea B, Niezgoda S, Chakraborty R. 2001. CODIS STR loci data from 41 sample populations. Journal of Forensic Sciences 46:453-489.
- Butler JM. 2006. Mixture Interpretation: Lessons from Interlab Study MIX05. National CODIS Conference; Arlington, VA2006.
- Cowell RG, Lauritzen SL, Mortera J. 2008. Probabilistic modelling for DNA mixture analysis. Forensic Science International: Genetics 1:640-642.
- Evett IW, Gill PD, Lambert JA. 1998. Taking account of peak areas when interpreting mixed DNA profiles. Journal of Forensic Sciences 43:62-69.
- Foll M, Fischer MC, Heckel G, Excoffier L. 2010. Estimating population structure from AFLP amplification intensity. Molecular Ecology 19: 4638-4647.
- Haned H. 2011. Forensim: an open-source initiative for the evaluation of statistical methods in forensic genetics. Forensic Science International: Genetics 4:265-268.
- Kimura M, Ohta T. 1975. Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. Proceedings of the National Academy of Sciences (USA) 72:2761-2764.
- Laurie C, Weir BS. 2003. Dependency effects in multi-locus match probabilities. Theoretical Population Biology 63:207-219.
- Lohmueller KE, Rudin N. 2013. Calculating the weight of evidence in low-template forensic DNA casework. Journal of Forensic Sciences 58:S243-S249.
- Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW. 2011. Validating TrueAllele^R DNA mixture interpretation. Journal of Forensic Sciences 56:1430-1447.

- Steele CD, Balding J. 2014. Statistical Evaluation of Forensic DNA Profile Evidence. Annual Review of Statistics and Its Applications 1:361-384.
- Steele CD, Syndercombe Court D, Balding DJ. 2014. Worldwide FST estimates relative to five continental-scale populations. Genetics (in press)
- Taylor D, Bright J, Buckleton J. 2013. The interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics 7:516-528.
- Taylor D, Bright J-A, Buckleton J, Curran J. 2014. An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations. Forensic Science International: Genetics 11: 5663.
- Walsh B, Redd AJ, Hammer MF. 2008. Joint match probabilities for Y chromosomal and autosomal markers. Forensic Science International 174:234-238.
- Weir BS. 2007. The rarity of DNA profiles. Annals of Applied Statistics 1:358-370.
- Weir BS, Ballantyne J, Bright JA, Buckleton JS, Curran JM, Laurie CA, Moretti T, Myers S. 2014. A population-genetic perspective for Y-chromosomal matching profiles. International Society for Forensic Genetics.
http://isfg2013.org/wp=content/uploads/2013/09/Thu-P3-1230-B-Weir-Y.pdf.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

Scientific Presentations on Topics Related to this Award

1. Bright, J-A.,Taylor, D., Buckleton, J.S. 2013. Matching mixtures against DNA databases. 25th International Society of Forensic Genetics Congress, Melbourne, Australia.

2. Buckleton, J.S. 2013. How do we interpret DNA evidence properly? 25th International Society of Forensic Genetics Congress, Melbourne, Australia.

3. Buckleton, J.S. 2013. Interpretation of DNA Mixtures with a Continuous Model. USACIL, Atlanta, GA.

4. Buckleton, J.S. 2013. Interpretation of DNA Mixtures with a Continuous Model. California Association of Criminalists, San Francisco, CA.

5. Curran, J. M. 2012. Is Forensic Science the last bastion of resistance against Statistics? 6th European Academy of the Forensic Sciences, The Hague, Netherlands.

6. Curran, J.M. 2012. Design and Analysis of Experiments. Presented at Workshop on Modern Methods for DNA Evidence, Department of Forensic Medicine. Faculty of Health Sciences. University of Copenhagen.

7. Curran, J.M. 2012. Modern methods for estimating θ. Workshop on Modern Methods for DNA Evidence, Department of Forensic Medicine. Faculty of Health Sciences. University of Copenhagen.

8. Curran, J.M. 2012. What is Fst or θ? Workshop on Fst. Department of Forensic Medicine. Faculty of Health Sciences. University of Copenhagen.

9. Curran, J.M., Bright, J-A., Buckleton, J.S., Taylor, D., Kelly, H. 2013. Statistical building blocks for continuous DNA interpretation systems. 25th International Society of Forensic Genetics Congress, Melbourne, Australia.

10. Weir, B.S. 2012. Interpretation of lineage markers. 8th Y-Chromosomal User Workshop, Innsbruck, Austria. September.

11. Weir, B.S. 2012. Population Genetic Issues for Forensic DNA Profiles. NIJ Grantees' Meeting, Atlanta, GA. (Repeated twice online in the "Live Seminar Series" hosted by the Forensic Sciences Center at Research Triangle Institute.)

12. Weir, B.S. 2012. Haplotype Frequency Estimation. SWGDAM, Fredericksburg, VA.

13. Weir, B.S. 2013. Population Genetic Issues for Forensic DNA Profiles. FBI DNA Unit, Quantico, VA.

14. Weir, B.S. 2013. Y-STR matching: A population-genetic perspective. 59th Congress of the Brazilian Genetics Society, Aguas de Lindoia, Brazil

15. Weir, B.S. 2013. Characterizing the genetic structure of populations: Application to Y-STR profiles. Department of Biostatistics, Harvard University.

16. Weir, B.S. 2013. Unweighted estimation for Fst. Impact of Large-scale Genomic Data on Statistical and Quantitative Genetics Conference, University of Washington.

17. Weir, B.S. 2014. Population structure and parentage calculations. Ribeiro Preto SP, Brazil.

18. Weir, B.S. 2014. Estimating F-statistics:Updating Weir and Cockerham Evolution 38:1538-1570 (1984). Annual Meeting of the Society for the Study of Evolution, Raleigh, NC.

19. Weir, B.S. 2014. Using match probabilities to characterize the effects of population structure on the strength of DNA evidence. International Conference on Forensic Inference and Statistics, Leiden, The Netherlands.

20. Weir, B.S. 2014. Characterizing population structure with F-statistics: relatedness on an evolutionary time scale. International Center for Mathematical Statistics, Edinburgh, United Kingdom.

21. Weir, B.S., Ballantyne, J., Bright, J-A., Buckleton, J.S., Curran, J., Laurie, C.A., Moretti, T. and Myers S. 2013. Population genetic theory for lineage markers. 25th International Society of Forensic Genetics Congress, Melbourne, Australia.

22. Weir, B.S., J.S. Buckleton and J. Curran. 2012. Incorporating uncertainty into likelihood ratios for DNA evidence. Joint Statistical Meetings, San Diego, CA.

Teaching on Topics Related to this Award

1. Weir, B.S. 2012. DNA Mixture Interpretation. Las Vegas Metropolitan Police Department Forensic Laboratory.

2. Weir, B.S. 2014. Forensic Genetics. University of Washington, PHG302.

3. Weir, B.S., Buckleton J.S. 2013. Statistical Genetics for Forensic Scientists. Summer Institute in Statistical Genetics, University of Washington.

4. Weir, B.S., Gittelson, S. 2014. Statistical Genetics for Forensic Scientists. Summer Institute in Statistical Genetics, University of Washington.

5. Weir, B.S. 2014. Statistical Genetics for Forensic Scientists. University of Central Florida, CHS6356.

Memberships Relevant to this Award

1. Buckleton, J.S. 2012-2014. SWGDAM Committee on Interpretation Guidelines for Y-chromosome STR typing.

2. Buckleton, J.S. 2014. ISFG Commission on Software Validation.

3. Curran, J.M. 2012-2104. Ad hoc Advisory Committee on Forensic Science, American Statistical Association.

4. Weir, B.S. 2012-2014. SWGDAM Committee on Interpretation Guidelines for Y-chromosome STR typing.

5. Weir, B.S. 2012-2104. Ad hoc Advisory Committee on Forensic Science, American Statistical Association.

6. Weir, B.S. 2014. Biology/DNA Scientific Area Committee within the Organization of Scientific Area Committees (OSAC) of NIST-DOJ.

Participants & Other Collaborators for this Award

- Bruce S. Weir, Principal Investigator, University of Washington
- Jo-Ann Bright, Collaborator, ESR New Zealand
- John S. Buckleton, Co-investigator, ESR New Zealand
- James M. Curran, Co-investigator, University of Auckland
- Jérôme Goudet, Collaborator, University of Lausanne
- Cecelia A. Laurie, Co-investigator, University of Washington
- Lisa Brown, Graduate Research Assistant, University of Washington
- Samuel Wasser, Collaborator, University of Washington.

Scientific Publications Related to this Award, 1/1/12-12/31/14

1. Balding DJ, Krawczak M, Buckleton JS, Curran JM. 2012. Decision-making in familial database searching: K1 alone or not alone? Forensic Science International: Genetics. 7:52-54.

2. Berger CEH, Buckleton J, Champod C, Evett IW, Jackson G. 2012. Re: Response to Jamieson regarding "More on the Bayesian Approach and the LR." Science & Justice 52:203-203.

3. Berger CEH, Vergeer, Buckleton JS. 2014. A more straightforward derivation of the LR for a database search. Forensic Science International: Genetics (in press)

4. Bille T, Bright J-A, Buckleton JS. 2013. Application of random match probability calculations to mixed STR profiles. Journal of Forensic Sciences 58:474-485.

5. Booth AM, Curran JM, Travas-Sejdic J, Harbison S, Vogel R.2013. Detection of target-probe oligonucleotide hybridization using synthetic nanopore resistive pulse sensing. Biosensors and Bioelectronics 45(1):136–140.

6. Bright J, Buckleton JS, Curran JM. 2013. Investigation into the performance of different models for predicting stutter. Forensic Science International: Genetics 7:422-427.

7. Bright J, Buckleton JS, Curran JM. 2013. Investigation into the performance of different models for predicting stutter (vol 7, pg 422, 2013). Forensic Science International: Genetics 7:564.

8. Bright J, Curran JM, Buckleton J. 2013. Relatedness calculations for linked loci incorporating subpopulation effects. Forensic Science International: Genetics 7:380-383.

9. Bright J, Curran JM, Buckleton J. 2013. Relatedness calculations for linked loci incorporating subpopulation effects (vol 7, pg 380, 2013). Forensic Science International: Genetics 7:565.

10. Bright J-A, Curran JM, Buckleton JS. 2014. Modelling PowerPlex Y stutter and artefacts. Forensic Science International: Genetics 11:126-136.

11. Bright J-A, Curran JM, Buckleton JS. 2014. The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation. Forensic Science International: Genetics 12:208-214.

12. Bright J-A, Curran JM, Buckleton JS. 2014. Investigation into stutter ratio variance. Australian Journal of Forensic Sciences 46:313-316.

13. Bright J, Curran JM, Hopwood AJ, Puch-Solis R, Buckleton J. 2012. Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts II. Science and Justice 53:371.

14. Bright J-A, Evett IW, Taylor D, Curran JM, Buckleton J. 2014. A series of recommended tests when validating probabilistic DNA profile interpretation software. Forensic Science International: Genetics (in press)

15. Bright J-A, Gill P, Buckleton J. 2012. Composite profiles in DNA analysis. Forensic Science International: Genetics 6:317-321.

16. Bright J-A, Kelly H, Taylor D, Tvedebrink T, Curran JM. 2014. Utlizing allelic dropout probabilities estimated by logistic regression in casework. Forensic Science International: Genetics 9:9-11.

17. Bright J-A, McManus K, Harbison S, Gill P, Buckleton J. 2012. A comparison of stochastic variation in mixed and unmixed casework and synthetic samples. Forensic Science International: Genetics 6:180-184.

18. Bright J-A, Neville S, Curran JM , Buckleton JS. 2014. Variability of mixed DNA profiles separated on a 3130 and 3500 capillary electrophoresis instrument. Australian Journal of Forensic Sciences 46:304-312.

19. Bright J-A, Stevenson KE, Coble MD, Hill CR, Curran JM, Buckleton JS. 2014. Characterising the STR locus D6S1043 and examination of its effect on stutter rates, Forensic Science International: Genetics. 8:20-23.

20. Bright J, Taylor D, Curran JM, Buckleton J. 2012. Developing allelic and stutter peak height models for a continuous method of DNA interpretation. Forensic Science International: Genetics 7:296-304.

21. Bright J-A, Taylor D, Curran JM, Buckleton JS. 2013. Degradation of forensic DNA profiles. Australian Journal of Forensic Science 45: 445-449.

22. Bright J-A, Taylor D, Curran JM, Buckleton J. 2014. Searching mixed DNA profiles directly against profile databases. Forensic Science International: Genetics 9:102-110.

23. Brookes C, Bright J-A, Harbison S, Buckleton J. 2012. Characterising stutter in forensic STR multiplexes. Forensic Science International: Genetics 6:58-63.

24. Buckleton J, Bright J-A, Taylor D, Evett I, Hicks T, Jackson G, Curran JM. 2014. Helping formulate propositions in forensic DNA analysis. Science and Justice 54:258-261

25. Buckleton J, Evett I, Weir B. 2014. Letter to the Editor in response to editorial by Risinger et al. 2014. Science and Justice (in press)

26. Buckleton J, Kelly H, Bright J-A, Taylor D, Tvedebrink T, Curran JM. 2014. Utilising allelic dropout probabilities estimated by logistic regression in casework, Forensic Science International: Genetics. 9: 9-11.

27. Buckleton J, Myers S. 2014. Combining autosomal and Y chromosome match probabilities using coalescent theory. Forensic Science International: Genetics 11:52-55.

28. Cockerton S, McManus K, Buckleton J. 2013. Interpreting lineage markers in view of subpopulation effects. Forensic Science International: Genetics 6: 393-397.

29. Curran JM. 2013. Is Forensic Science the last bastion of resistance against Statistics? Science and Justice 53:251-252.

30. Curran JM, Buckleton J. 2014. Uncertainty in the number of contributors for the European Standard Set of loci. Forensic Science International: Genetics 11:205-206.

31. Gill P, Phillips C, McGovern C, Bright J-A, Buckleton J. 2012. An evaluation of potential allelic association between the STRs vWA and D12S391: Implications in criminal casework and applications to short pedigrees. Forensic Science International: Genetics 6:477-486.

32. Gill L, Gusmão H, Mayr WR, Morling N, Parson W, Prieto L, Prinz M, Schneider H, Schneider PM, Weir BS. 2012. DNA Commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. Forensic Science International: Genetics 6:679-688.

33. Hopwood AJ, Puch-Solis R, Tucker VC, Curran JM, Skerrett J, Pope S, Tully G. 2012. Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts. Science and Justice 52: 185–190.

34. Jackson G, Evett I, Champod C, Buckleton J. 2014. Re "Perception problems of the verbal scale." Science & Justice 54:180-180.

35. Kelly H, Bright J, Curran JM, Buckleton J. 2012. Modeling heterozygote balance in forensic DNA profiles. Forensic Science International: Genetics 6: 729-734. doi:10.1016/j.fsigen.2012.08.002

36. Kelly H, Bright J-A, Curran J, Buckleton J. 2012. The interpretation of low level DNA mixtures. Forensic Science International: Genetics, 6: 191-197.

37. Kelly H, Bright J-A, Buckleton JS, Curran JM. 2013. Identifying and modelling the drivers of stutter in forensic DNA profiles, Australian Journal of Forensic Sciences 46:194-203.

38. Kelly H, Bright J-A, Buckleton JS, Curran JM. 2014. A comparison of statistical models for the analysis of complex forensic DNA profiles. Science and Justice 54:66-70.

39. Menotti-Raymond M, David VA, Weir BS, O'Brien SJ. 2012. A population genetic database of cat breeds developed in coordination with a domestic cat STR multiplex. Journal of Forensic Science 57:596-601.

40. Puch-Solis R, Rodgers L, Balding DJ, Evett IW, Pope S, Mazumder A, Curran JM. Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters. Forensic Science International: Genetics (in press).

41. Robertson B, Vignaux T, Buckleton J. 2013. Flawed reasoning in court. 2013. Criminal Law Journal

42. Rohlfs RV, Fullerton SM, Weir BS. 2012. Familial identification: Population structure and relationship distinguishability. PLoS Genetics 8:e1002469.

43. Taylor D, Bright J, Buckleton J. 2014. The interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics 7:564.

44. Taylor D, Bright J-A, Buckleton J, Curran J. 2014. An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations. Forensic Science International: Genetics 11:56-63.

45. Taylor D, Bright J-A, Buckleton J. 2014. Considering relatives as alternate sources of DNA to mixed DNA profiles. Forensic Science International: Genetics (in press)

46. Tvedebrink T, Eriksen PS, Curran JM, Mogensen HS, Morling N. 2012. Analysis of matches and partial-matches in a Danish STR data set. Forensic Science International: Genetics, 6 (3), 387–392. doi:10.1016/j.fsigen.2011.08.001

47. Weir BS. 2012. Estimating F-Statistics: A Historical View. Philosophy of Science 79:637-643.

48. Zheng XW, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326-3328.