Title	**Population Genetic Issues for Forensic DNA Profiles.**
Award Number	NIJ 2014-DN-BX-K028
Draft Final Report	January 1, 2015 - December 31, 2017
Submitting Official	Carol Rhodes, Director, Office of Sponsored Programs
	osp@uw.edu
	(206) 543-4043
Recipient Organization	Office of Sponsored Programs
	University of Washington
	4333 Brooklyn Avenue NE
	Seattle, WA 09195-9472
	DUNS:
	EIN:

Submitted by

(University of Washington)

Accomplishments

Major Goals of the Project

This project was planned to conduct research on the population genetic issues affecting the interpretation of forensic DNA profiles. The particular topics identified were: (1) Characterizing population structure; (2) Interpreting lineage marker evidence; Extending a continuous approach. To date, we have published 27 papers addressing these three topics, and they are now summarized.

Population Structure

The interpretation of matching between DNA profiles of a person of interest and an item of evidence can be undertaken with 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 θ and this quantity affects matching proportions.

Worldwide Survey The widely-adopted match probability equations of Balding and Nichols (1994) refer to θ, the probability two alleles in a population are identical by descent. In Buckleton et al. (2016) we clarified that this probability can be estimated only as a comparison to the probability for alleles taken from different populations. If the within-population identity probability is written as θ_W and the between-population-pair probability as θ_B, then we estimate $= (\theta_W - \theta_B)/(1 - \theta_B)$. We showed that it is , rather than θ, needed for match probabilities for a subpopulation when allele frequencies are available only from a database representing the whole population. In Wasser et al. (2015) we implemented this new   estimation procedure to characterize the population structure for elephants in Africa, as part of work to identify the sources of seized ivory and combat transnational crime.

In Buckleton et al. (2016) we presented results based on our survey of 250 publications that reported allele frequencies at 24 STR loci from 446 populations around the world, representing information from nearly a half-million people. The immediate forensic motivation was to present values of the population structure parameter θ for use in match probability calculations. An estimate that used all the data we examined was 0.02, a little higher than is generally used by US forensic agencies. The single value of 0.02, however, ignores the concept of reference population we laid out in our paper. We found it useful to phrase our work in terms of allelic matching proportions: the proportion of alleles taken randomly from a population that have the same type, compared to the proportion of allele pairs, one allele from each of two populations, that match. The pairs of populations are from a set of populations that include the target population. They may represent the ethnicity of the target population, in the way that the FBI Caucasian database represents US populations of European ancestry, or they may represent a broad array of populations as would be appropriate when the target ethnicity was unknown. The contrast between the former regional reference set of populations and the latter global set, was shown in Buckleton et al. (2016).

SNPs and SNVs The amount of genetic data collected by forensic scientists worldwide is now substantial, with short tandem repeat (STR) frequencies having been reported in this and other journals from close to one million people. National o ender databases now have a total of about 50 million profiles, although the possibility of conducting numerical experiments on those data in order to address issues such as the dependencies of matching probabilities among loci is presently limited to very few countries. Forensic science is turning attention to large-scale genetic data, such as those provided for single nucleotide polymorphisms (SNPs) from chip-array technology or single nucleotide variants (SNVs) from next-generation sequencing (NGS). The 1000 Genomes project, www.1000genomes.org, has already published whole-genome sequence data that includes over 78 million SNPs. The data will accumulate rapidly from several projects: in the US the National Human Genome Research Institute plans to sequence 200,000 people and the National Heart Lung and Blood Institute anticipates sequencing another 100,000 people. Similar projects are underway in other countries. The implications for forensic science are likely to be substantial, and will include the characterization of biogeographical ancestry (BGA) and externally visible characters (EVC) as well as enhanced deconvolution of mixtures. Weir and Zheng(2015) explored the use of SNPs or SNVs to characterize relatedness on the evolutionary time scale and the immediate family time scale: this theme was expanded upon by Weir and Goudet (2017).

Relatedness and Population Structure Many population genetic activities, including forensic identification, rely on appropriate estimates of population structure or relatedness. Weir and Goudet (2017) re-cast existing treatments of population structure, relatedness and inbreeding to make explicit that the parameters of interest involve di erences of probabilities of identity by descent in the target and the reference sets of alleles and so can be negative. They provided simple moment estimates of these parameters, phrased in terms of allele matching within and between individuals for relatedness and inbreeding, or within and between populations for population structure. A multi-level hierarchy of alleles within individuals, alleles between individuals within populations, and alleles between populations allows a unified treatment of relatedness and population structure. This particular paper was not supported by NIJ but it is informing our current NIJ work on the forensic uses of genetic SNP data being generated by next generation sequencing.

Lineage Markers

Y-STR profiling makes up a small but important proportion of forensic DNA casework. Often Y-STR profiles are used when autosomal profiling has failed to yield an informative result. Consequently Y-STR profiles are often from the most challenging samples. In addition to these points, Y-STR loci are linked, meaning that evaluation of haplotype probabilities are either based on overly simplified counting methods or computationally costly genetic models, neither of which extend well to the evaluation of mixed Y-STR data. For all of these reasons Y-STR data analysis has not seen the same advances as autosomal STR data. In Hall (2016) we examined three datasets of Y-STR profiles from a population-genetic standpoint, and in Taylor et al. (2016) we provided the basis for a continuous analysis.

Y-STR Database Analyses Hall (2016) analyzed three publicly-available datasets of Y-STR profiles: the Y-Chromosome Haplotype Reference Database, YHRD, (Willuweit and Roewer, 2013), the Human Genome Diversity Project HGDP (Cann et al., 2002), and the data published by Xu

et al. (2015). These three databases of Y-STR profiles were used to provide estimates of a profile probability: the probability an unknown person has that profile. The problem is that, with 10 or more alleles at a locus, the number of possible profiles quickly exceeds the database size as the number of loci increases. It is not uncommon for a profile of interest not to appear in a database, and eventually the number of loci makes it likely that a profile occurs once or not all in a database unless profiles from close relatives such as father-son or brothers are present. Binomial-theory confidence limits can provide some guidance to population probabilities from database proportions: the upper 95% confidence limit for a profile not seen in a database of size n, for example, is close to $3/n$, but the same result holds if the profile is based on 7 or 17 or 27 loci.

The solution to the problem of unobserved autosomal profiles has been to assume the constituent loci are independent and multiply their probabilities together. The lack or recombination on the Y chromosome as been regarded as a reason not to assume independence among loci. However, mutation may be assumed to act independently at each locus, so that the dependence among loci is not absolute. The usual check on locus independence has been to test for linkage disequilibrium for pairs of loci, but Hall (2016) found that an appreciable number of pairs of loci have haplotype frequencies not significantly different from the product of the two single-locus frequencies.

The two-locus approach does not address the issue of multi-locus profile probabilities and we noted the recent introduction of entropy theory into Y-STR calculations (Caliebe et al., 2015; Siegert et al., 2015). The entropy of an L-locus profile, where $\tilde{p}_u$ is the sample frequency of haplotype u, is $H_L = -\sum_u \tilde{p}_u \ln(\tilde{p}_u)$. The extent of inter-locus dependencies can be addressed by first choosing the locus with the largest single-locus entropy. Subsequent loci are added, the $(L+1)$th one chosen to have the largest conditional entropy $H_{L+1|L} = H_{L+1} - H_L$. This ensures that loci are added according to their information content, and the orders in which they are added for the YHRD database are shown in Table 3. If locus $L + 1$ is independent of loci $1, 2, \ldots L$, then $H_{L+1|L} = H_L$. We found that none of the Y-STR loci are truly independent but that the dependencies have little effect on entropy once the number of loci reaches about 10. There is little additional discriminating power beyond 10 loci, suggesting that 10-locus matches indicate membership in a common male lineage and, absent mutation, all further loci will match.

Y-STR profile interpretation   Taylor, Bright and Buckleton (2016) introduced a discussion of probabilistic genotyping methods for Y-STR profiles. They used models developed for autosomal STR loci to determine the probability of an observed Y-STR profile given potential contributing haplotypes. In doing this, the models allow a "weight" to be given to each potential contributor haplotype set that acts as an indication of how well the proposed haplotypes describe the observed data. This then lends itself to development of interpretational guidelines. As for autosomal work (e.g. Taylor et al., 2013), the Y-STR model employs these mass parameters:

- Template amount for each of the contributors.

- Degradation, which models the decay with respect to molecular weight in template for each of the contributors.

- Amplification efficiency, to allow for the observed amplification levels of each locus.

- A replicate amplification multiplier. This effectively scales all peaks up or down between replicates.

The total allelic product from an allele is split between back stutter, forward stutter and allelic peak. Given the mass parameters the height of allelic peaks are expected to be independent within and between loci. Stutter peaks are dependent on their parent peak heights, but given this they are also expected to be independent within and between loci. This allows the deconvolution to occur in a locus by locus manner as described by Taylor et al. (2013), rather than having to consider the entire haplotype or haplotypic mixture as a whole entity. In this way the deconvolution of Y-STR data becomes very similar to that of autosomal STR data.

Continuous Approach

We have made several advances to the interpretation of STR profiles when they are regarded as being continuous indicators of DNA abundance as opposed to a binary present/absent call for specific alleles. A probabilistic approach enhances the interpretation of multiple-contributor profiles, and our methodological research has informed the commercial STRmix$^{\text{TM}}$ software package that is being implemented by several US forensic agencies. Probabilistic genotyping refers to the use of software and computer algorithms to apply biological modeling, statistical theory, and probability distributions to infer the probability of the profile from single source and mixed DNA typing results given different contributor genotypes. The software weighs potential genotypic solutions for a mixture by utilizing more DNA typing information and accounting for uncertainty in random variables within the model, such as peak heights, rather than a stochastic or dropout threshold.

Low-template DNA The sensitivity and resolution of modern DNA profiling hardware is such that forensic laboratories generate more data than they have resources to analyze. One coping mechanism is to set a threshold, above the minimum required by instrument noise, so that weak peaks are screened out. In binary interpretations of forensic profiles, the impact of this threshold (sometimes called an analytical threshold - AT) was minimal as interpretations were often limited to a clear major component. With the introduction of continuous typing systems, the interpretation of weak minor components of mixed DNA profiles is possible and consequently the consideration of peaks just above or just below the analytical threshold becomes relevant. Taylor, Buckleton et al. (2017) investigated the occurrence of low-level DNA profile information, specifically that which falls below the analytical threshold. We investigated how it can be dealt with and the consequences of each choice in the framework of continuous DNA profile interpretation systems. Where appropriate we illustrated how these can be implemented using the probabilistic interpretation software STRmix$^{\text{TM}}$. We demonstrated a feature of STRmix$^{\text{TM}}$ that allows the analyst to guide the software, using human observation that there is a low-level contributor present, through user-designated prior distributions for contributor mixture proportions.

A set of low template mixed DNA profiles with known ground truths was examined by Taylor and Buckleton (2015) using software that utilized peak heights (STRmix$^{\text{TM}}$ V2.3) and an adapted version that did not use peak heights and mimicked models based on a drop-out probability (known as semi-continuous or 'drop' models) (STRmix$^{\text{TM}}$ lite). The use of peak heights increased the LR when H_p was true in the vast majority of cases. The effect was most notable at moderate template levels but was also present at quite low template levels. There is no level at which we can say that height information is totally uninformative. Even at the lowest levels the bulk of the data show some improvement from the inclusion of peak height information.

Peak Height Variability   In forensic DNA analysis a DNA extract is amplified using polymerase chain reaction (PCR), separated using capillary electrophoresis and the resulting DNA products are detected using fluorescence. Sampling variation occurs when the DNA molecules are aliquotted during the PCR setup stage and this translates to variability in peak heights in the resultant electropherogram or between electropherograms generated from a DNA extract. Beyond the variability caused by sampling variation it has been observed that there are factors in generating the DNA profile that can contribute to the magnitude of variability observed, most notably the number of PCR cycles. Taylor, Buckleton and Bright (2016) investigated a number of factors in the generation of a DNA profile to determine which contribute to levels of peak height variability.

Robustness of Results   Cooper et al. (2015) examined the concordance in profile interpretation of three crime samples by twenty di erent analysts across twelve di erent international laboratories using STRmix^TM. The three profiles selected for this study exhibited a range of template and complexity. Although the use of probabilistic software has compelled a level of concordance between di erent analysts, there remain di erences within profile interpretation, particularly with the objective assignment of the number of contributors to profiles.

Software Validation   Bright, Evett at al. (2015) described a set of experiments which may be used to internally validate in part probabilistic interpretation software. These experiments included both single source and mixed profiles calculated with and without dropout and drop-in and studies to determine the reproducibility of the software with replicate analyses. The experiments used three software packages: STRmix^TM, LRmix, and Lab Retriever. to demonstrate profile examples where the expected answer may be calculated and we provided all calculations.

In 2015 the Scientific Working Group on DNA Analysis Methods published the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems. STRmix^TM is probabilistic genotyping software that employs a continuous model of DNA profile interpretation. Bright et al. (2016) described the developmental validation activities of STRmix^TM following the SWGDAM guidelines. They addressed the underlying scientific principles, and the performance of the models with respect to sensitivity, specificity and precision and results of interpretation of casework type samples. This work demonstrated that STRmix^TM is suitable for its intended use for the interpretation of single source and mixed DNA profiles.

STRmix^TM is one of several software packages to employ continuous analyses, and we note the comments by the President's Council of Advisors on Science and Technology (PCAST) in 2016: "These probabilistic genotyping software programs clearly represent a major improvement over purely subjective interpretation." The PCAST report went on to say "However, they still require careful scrutiny to determine (1) whether the methods are scientifically valid, including defining the limitations on their reliability (that is, the circumstances in which they may yield unreliable results) and (2) whether the software correctly implements the methods. This is particularly important because the programs employ di erent mathematical algorithms and can yield di erent results for the same mixture profile." Accordingly, we coordinated a large multi-laboratory validation study, and published results (Moretti et al., 2017) that should allay the reservations in the PCAST report. Dr Moretti presented some of these results at the 2017 International Symposium on Human Identification (ISHI).

The Moretti et al. (2017) study used lab-specific parameters and more than 300 single-source

and mixed contributor profiles. Simulated forensic specimens, including constructed mixtures that included DNA from two to five donors across a broad range of template amounts and contributor proportions, were used to examine the sensitivity and specificity of the system via more than 60,000 tests comparing hundreds of known contributors and non-contributors to the specimens. Conditioned analyses, concurrent interpretation of amplification replicates, and application of an incorrect contributor number were also performed to further investigate software performance and probe the limitations of the system. In addition, the results from manual and probabilistic interpretation of both prepared and evidentiary mixtures were compared.

Using weights assigned to the various genotypes or genotype sets, STRmix^{TM}calculates LRs, which are the probability of the DNA evidence under two opposing hypotheses referred to as H_1 and H_2. An LR greater than 1 provides support for a specified person of interest as a contributor to the DNA evidence (H_1), whereas an LR less than 1 provides support that the person of interest is not a contributor (H_2). An LR of 1 provides no greater support for either proposition. To indicate the e ects of variation in the various processed leading to electropherogram, STRmix^{TM}produces a distribution of values and the (say) 5th percentile of this distribution can be reported (analogous to a lower confidence limit).

The findings in this study show that STRmix^{TM}is sufficiently robust for implementation in forensic laboratories, o ering numerous advantages over historical methods of DNA profile analysis and greater statistical power for the estimation of evidentiary weight, and can be used reliably in human identification testing. With few exceptions, likelihood ratio results reflected intuitively correct estimates of the weight of the genotype possibilities and known contributor genotypes. This comprehensive evaluation provides a model in accordance with SWGDAM recommendations for internal validation of a probabilistic genotyping system for DNA evidence interpretation.

Variability of Software Results Some of probabilistic genotyping software solutions utilize Markov chain Monte Carlo techniques (MCMC). They will not produce an identical answer after repeat interpretations of the same evidence profile because of the Monte Carlo aspect. This is a new source of variability within the forensic DNA analysis process. In Bright, Stevenson et al. (2015) we explored the size of the MCMC variability within the interpretation software STRmix^{TM}compared to other sources of variability in forensic DNA profiling including PCR, capillary electrophoresis load and injection, and the makeup of allele frequency databases. The MCMC variability within STRmix^{TM}was shown to be the smallest source of variability in this process.

Other Topics

Database Matches Matching DNA profiles of an accused person and a crime scene trace are one of the most common forms of forensic evidence. A number of years ago the so-called DNA controversy was concerned with how to quantify the value of such evidence. Given its importance, the lack of understanding of such a basic issue was quite surprising and concerning. Deriving the equation for the likelihood ratio of a DNA database match in a much more direct and simple way is the topic covered in Berger et al. (2015). As it is much easier to follow it is hoped that this derivation will contribute to the understanding.

DNA Interpretation Methods Curran and Weir (2016) provided a review of recent work in forensic DNA interpretation for the statistics community, ending with a description of modeling work that informs commercial software products such as STRmix™ and TrueAllele®. Such continuous interpretation software reduces interpretation variability by removing many binary decisions from the interpretation process. The software works with the raw epg information, and analysts are not required to specifically designate alleles, pair genotypes, or decide whether peaks are stutters or not. Sophisticated Bayesian modeling and significant computing is used to provide an LR. A natural consequence of this is that the ability to see the steps that went into the calculation of the LR has been removed. There is a natural fear of the "black box" nature of modern methods. However, realistically, there are plenty of other scientific procedures that require significant scientific knowledge and training to fully understand how they work. For example, the gene sequencers that produce the epg required knowledge of biochemistry, mass spectrometry, physics, quantum optics, signal processing for their development but users of such equipment are not expected to be familiar with all these fields. We trust these machines because they have been subjected to many studies which have been published in peer-reviewed scientific literature. We use them because they are the best methods we have. The same is true for advanced statistical methods for the interpretation of DNA.

Expanded CODIS Core Allele distributions for twenty-three autosomal short tandem repeat (STR) loci (D1S1656, D2S441, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, CSF1PO, FGA, Penta D, Penta E, SE33, TH01, TPOX and vWA) were reported in Moretti et al. (2016) for samples of Caucasians, Southwestern Hispanics, Southeastern Hispanics, African Americans, Bahamians, Jamaicans, Trinidadians, Chamorros, Filipinos, Apaches, and Navajos. The data are included in the FBI PopStats software for calculating statistical estimates of DNA typing results and cover the expanded CODIS Core STR Loci required of U.S. laboratories that participate in the National DNA Index System (NDIS).

Factor of 10 Gittleson et al. (2017) gave an update of the classic experiments that led to the view that profile probability assignments are usually within a factor of 10 of each other. The data used in this study consisted of 15 Identifiler loci collected from a wide range of forensic populations. Following Budowle et al. (1994), the terms cognate and non-cognate are used. The cognate database is the database from which the profiles are simulated. The profile probability assignment was usually larger in the cognate database. In 44%-65% of the cases, the profile probability for 15 loci in the non-cognate database was within a factor of 10 of the profile probability in the cognate database. This proportion was between 60% and 80% when the FBI and NIST data were used as the non-cognate databases. A second experiment compared the match probability assignment using a generalized database and recommendation 4.2 from NRC II (the Balding-Nichols match probability) with a proxy for the matching proportion developed using subpopulation allele frequencies and the product rule. The findings support that the 4.2 assignment has a large conservative bias. These results are in agreement with previous research results.

GlobalFiler™ Frequencies Taylor, Bright et al. (2017) reported GlobalFiler™ assigned autosomal allele proportions for Caucasian, Asian, self-declared Aboriginal and pure Aboriginal pop-

ulations from Australia and Caucasian and Eastern and Western Polynesian populations from New Zealand. Population sample sizes vary from 122 to 528. All populations underwent tests for the presence of allelic dependencies (i.e. departures from the expectations of Hardy-Weinberg and linkage equilibrium) and some large dependencies were observed in the Australian Aboriginal populations. We provided allele frequency files for all populations examined.

H_d-**true Testing**   The performance of any model used to analyze DNA profile evidence should be tested using simulation, large scale validation studies based on ground-truth cases, or alignment with trends predicted by theory. Taylor, Buckleton and Evett (2015) investigated a number of diagnostics to assess the performance of the model using defense hypothesis H_d true tests. Of particular focus in this work is the proportion of comparisons to non-contributors that yield a likelihood ratio (LR) higher than or equal to the likelihood ratio of a known contributor (LRPOI), designated as p, and the average LR for H_d true tests. Theory predicts that p should always be less than or equal to 1/LRPOI and hence the observation of this in any particular case is of limited use. A better diagnostic is the average LR for H_d true which should be near to 1.

H_d-true testing is a way of assessing the performance of a model, or DNA profile interpretation system. These tests involve simulating DNA profiles of non-donors to a DNA mixture and calculating a likelihood ratio (LR) with one proposition postulating their contribution and the alternative postulating their non contribution. Following Turing it is possible to predict that "The average LR for the H_d-true tests should be one". This suggests a way of validating software.

A limitation with H_d-true tests, when non-donor profiles are generated at random (or in accordance with expectation from allele frequencies), is that the number of tests required depends on the discrimination power of the evidence profile. If the H_d-true tests are to fully explore the genotype space that yields non-zero LRs then the number of simulations required could be in the 10 s of orders of magnitude (well outside practical computing limits). We describe here the use of importance sampling, which allows the simulation of rare events to occur more commonly than they would at random, and then adjusting for this bias at the end of the simulation in order to recover all diagnostic values of interest. Importance sampling, although having been employed by others for H_d true tests, is largely unknown in forensic genetics. Taylor, Curran and Buckleton (2017) took time to explain how importance sampling works, the advantages of using it and its application to H-d true tests. They concluded by showing that employing an importance sampling scheme brings H_d-true testing ability to all profiles, regardless of discrimination power.

ISFG Recommendations on Software Validation   The use of biostatistical software programs to assist in data interpretation and calculate likelihood ratios is essential to forensic geneticists and part of the daily case work flow for both kinship and DNA identification laboratories. Previous recommendations issued by the DNA Commission of the International Society for Forensic Genetics (ISFG) covered the application of biostatistical evaluations for STR typing results in identification and kinship cases, and this is now being expanded to provide best practices regarding validation and verification of the software required for these calculations. With larger multiplexes, more complex mixtures, and increasing requests for extended family testing, laboratories are relying more than ever on specific software solutions and sufficient validation, training and extensive documentation are of upmost importance.

In Coble et al. (2016) we presented recommendations for the minimum requirements to validate

biostatistical software to be used in forensic genetics. We distinguished between developmental validation and the responsibilities of the software developer or provider, and the internal validation studies to be performed by the end user. Recommendations for the software provider address, for example, the documentation of the underlying models used by the software, validation data expectations, version control, implementation and training support, as well as continuity and user notifications. For the internal validations the recommendations include: creating a validation plan, requirements for the range of samples to be tested, Standard Operating Procedure development, and internal laboratory training and education. To ensure that all laboratories have access to a wide range of samples for validation and training purposes the ISFG DNA commission encourages collaborative studies and public repositories of STR typing results.

Likelihood Ratio Variability Curran (2016) argued that, given our current state of knowledge, reporting uncertainty in the likelihood ratio is best practice. This may in time be replaced by reporting a Bayes factor, but we are currently unable to do this in all but the simplest of examples.

Mixture Interpretation The evaluation and interpretation of forensic DNA mixture evidence faces greater interpretational challenges due to increasingly complex mixture evidence. Such challenges include: casework involving low quantity or degraded evidence leading to allele and locus dropout; allele sharing of contributors leading to allele stacking; and di erentiation of PCR stutter artifacts from true alleles. There is variation in statistical approaches used to evaluate the strength of the evidence when inclusion of a specific known individual(s) is determined, and the approaches used must be supportable. There are concerns that methods utilized for interpretation of complex forensic DNA mixtures may not be implemented properly in some casework. Similar questions are being raised in a number of U.S. jurisdictions, leading to some confusion about mixture interpretation for current and previous casework.

Key elements necessary for the interpretation and statistical evaluation of forensic DNA mixtures were described by Bieber et al. (2016). Given the most common method for statistical evaluation of DNA mixtures in many parts of the world, including the USA, is the Combined Probability of Inclusion/Exclusion (CPI/CPE). Exposition and elucidation of this method and a protocol for use is the focus of this article. Formulae and other supporting materials were provided. This description should help reduce the variability of interpretation with application of this methodology and thereby improve the quality of DNA mixture interpretation throughout the forensic community.

Multiplexes In response to requests from the forensic community, commercial companies are generating larger, more sensitive, and more discriminating STR multiplexes. These multiplexes are now applied to a wider range of samples including complex multi-person mixtures. In parallel there is an overdue reappraisal of profile interpretation methodology. Aspects of this reappraisal include: 1. The need for a quantitative understanding of allele and stutter peak heights and their variability; 2. An interest in reassessing the utility of smaller peaks below the often used analytical threshold; 3. A need to understand not just the occurrence of peak drop-in but also the height distribution of such peaks; 4. A need to understand the limitations of the multiplex-interpretation strategy pair implemented.

Taylor, Bright, McGovern et al. (2015) presented a full scheme for validation of a new multiplex that is suitable for informing modern interpretation practice. They predominantly used

GlobalFiler™ as an example multiplex but suggested that the aspects investigated here are fundamental to introducing any multiplex in the modern interpretation environment.

Number of Contributors The probability that multiple contributors are detected within a forensic DNA profile improves as more highly polymorphic loci are analyzed. The assignment of the correct number of contributors to a profile is important when interpreting the DNA profiles. Coble et al. (2015) investigated the probability of a mixed DNA profile appearing as having originated from a fewer number of contributors for the African American, Asian, Caucasian and Hispanic US populations. They investigated a range of locus configurations from the proposed new CODIS set. These theoretical calculations were based on allele frequencies only and ignore peak heights. They showed that the probability of a higher order mixture (five or six contributors) appearing as having originated from one less individual is high. This probability decreases as the number of loci tested increases.

Wildcard Designations Forensic DNA databases are powerful tools used for the identification of persons of interest in criminal investigations. Typically, they consist of two parts: (1) a database containing DNA profiles of known individuals and (2) a database of DNA profiles associated with crime scenes. The risk of adventitious or chance matches between crimes and innocent people increases as the number of profiles within a database grows and more data is shared between various forensic DNA databases, e.g. from di erent jurisdictions.

The DNA profiles obtained from crime scenes are often partial because crime samples may be compromised in quantity or quality. When an individual's profile cannot be resolved from a DNA mixture, ambiguity is introduced. A wild card, F, may be used in place of an allele that has dropped out or when an ambiguous profile is resolved from a DNA mixture.

Variant alleles that do not correspond to any marker in the allelic ladder or appear above or below the extent of the allelic ladder range are assigned the allele designation R for rare allele. R alleles are position specific with respect to the observed/unambiguous allele. The F and R designations are made when the exact genotype has not been determined. The F and R designation are treated as wild cards for searching, which results in increased chance of adventitious matches. In Tvedbrink et al. (2015) we investigated the probability of adventitious matches given these two types of wild cards.

References: *acknowledge this award

- *Berger CEH, Vergeer P, Buckleton JS. 2015. A more straightforward derivation of the LR for a database search. Forensic Science International: Genetics 14:156-160.

- *Bieber FR, Buckleton JS, Budowle B, Butler JM, Coble MD. 2016. Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. BMC Genetics. 2016;17:1-15.

- *Bright JA, Evett IW, Taylor D, Curran JM, Buckleton J. 2015. A series of recommended tests when validating probabilistic DNA profile interpretation software. Forensic Science International: Genetics 14:125-131.

- *Bright JA, Stevenson KE, Curran JM, Buckleton JS. 2015. The variability in likelihood ratios due to di erent mechanisms. Forensic Science International: Genetics 14:187-190.

- *Bright JA, Taylor D, McGovern C, Cooper S, Russell L, Abarno D, Buckleton J. 2016. Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles. Forensic Science International: Genetics 23:226-239.

- *Buckleton JS, Curran JM, Goudet J, Taylor D, Thiery A, Weir BS. 2016. Population-specific F_{ST} values: A worldwide survey. Forensic Science International: Genetics 23:191-100.

- Budowle B, Monson KL, Guisti AM. 1994. A reassessment of frequency estimates of PVUII-generated VNTR profiles in a Finnish, an Italian, and a general Caucasian database – no evidence of ethnic subgroups a ecting forensic estimate. American Journal of Human Genetics 55:533-539.

- Caliebe A, Jochens A, Willuweit S, Roewer L, Krawczak M. 2015, No shortcut solution to the problem of Y-STR match probability calculation. Forensic Science International: Genetics 15:6975.

- Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piou re L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu JY, Carcassi C, Contu L, Du FR, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang XY, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian YP, Shu QF, Xu JJ, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL. 2002. A human genome diversity cell line panel. Science 296:61-262.

- *Coble MD, Bright JA, Buckleton J, Curran JM. 2015. Uncertainty in the number of contributors in the proposed new CODIS set. Forensic Science International: Genetics 19:207-211.

- *Coble MD, Buckleton J, Butler JM, Egeland T, Fimmers R, Gill P, Gusmao L, Guttman B, Krawczak M, Morling N, Parson W, Pinto N, Schneider PM, Sherry ST, Willuweit S, Prinz M. 2016. DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications. Forensic Science International: Genetics 25: 91-197.

- *Cooper S, McGovern C, Bright JA, Taylor D, Buckleton J. 2015. Investigating a common approach to DNA profile interpretation using probabilistic software. Forensic Science International: Genetics 16: 121-131.

- *Curran JM. 2016. Admitting to uncertainty in the LR. Science and Justice 56:380-382.

- *Curran JM, Weir BS. 2016. Modern methods of DNA interpretation. Chance 29:17-26.

- *Gittelson S, Kalafut T, Myers S, Taylor D, Hicks T, Taroni F, Evett IW, Bright JA, Buckleton J. 2016. A practical guide for the formulation of propositions in the Bayesian approach to DNA evidence interpretation in an adversarial environment. Journal of Forensic Sciences 61:186-195.

- *Gittelson S, Moretti TR, Onorato AJ, Budowle B, Weir BS, Buckleton J. 2017. The factor of 10 in forensic DNA match probabilities. Forensic Science International: Genetics 28:178-187.

- Hall TO. 2016. The Y chromosome in forensic and public health genetics. Dissertation, University of Washington.

- *Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright JA, Taylor DA, Onorato AJ. 2017. Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics 29:126-144.

- *Moretti TR, Moreno LI, Smerick JB, Pinone ML, Hizon R, Buckleton JS, Bright JA, Onorato AJ. 2016. Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. Forensic Science International: Genetics 25:175-181.

- Siegert S, Roewer L, Nothnagel M. S2015. Shannon's equivocation for forensic Y-STR marker selection. Forensic Science International: Genetics 16:216225.

- *Taylor D, Bright JA, Buckleton J. 2013. The interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics 7:516-528.

- *Taylor D, Bright JA, Buckleton J. 2016. Using probabilistic theory to develop interpretation guidelines for Y-STR profiles. Forensic Science International: Genetics 21:22-34.

- *Taylor D, Bright JA, McGovern C, He ord C, Kalafut T, Buckleton J. 2016. Validating multiplexes for use in conjunction with modern interpretation strategies. Forensic Science International: Genetics 20:6-19.

- *Taylor D, Bright JA, McGovern C, Neville S, Grover D. 2017. Allele frequency database for GlobalFiler™ STR loci in Australian and New Zealand populations. Forensic Science International: Genetics 28:E38-E40.

- *Taylor D, Buckleton J. 2015. Do low template DNA profiles have useful quantitative data? Forensic Science International: Genetics 16: 13-16.

- *Taylor D, Buckleton J, Bright JA. 2016. Factors a ecting peak height variability for short tandem repeat data. Forensic Science International: Genetics 21:126-133.

- *Taylor D, Buckleton J, Bright JA. 2017. Does the use of probabilistic genotyping change the way we should view sub-threshold data? Australian Journal of Forensic Sciences 49:78-92.

- *Taylor D, Buckleton J, Evett I. 2015. Testing likelihood ratios produced from complex DNA profiles. Forensic Science International: Genetics 16:165-171.

- *Taylor D, Curran JM, Buckleton J. 2017. Importance sampling allows H-d true tests of highly discriminating DNA profiles. Forensic Science International: Genetics 27:74-81.

- *Tvedebrink T, Bright JA, Buckleton JS, Curran JM, Morling N. 2015. The e ect of wild card designations and rare alleles in forensic DNA database searches. Forensic Science International: Genetics 16:98-111.

- *Wasser SK, Brown L, Mailand C, Mondol S, Clark W, Laurie C, Weir BS. 2015. Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. Science 349:84-87.

- Weir BS, Goudet J. 2017. A unified characterization for population structure and relatedness. Genetics 206:2085-2103.

- *Weir BS, Zheng X. 2015. SNPs and SNVs in Forensic Science. Forensic Science International: Genetics Supplement Series 5:e267–e268.

- Willuweit S, Roewer L. 2013. Y Chromosome Haplotype Reference Database. Forensic Science International: Genetics 15:4348.

- Xu HY, Wang CC, Shrestha R, Wang LX, Zhang MF, He YG, Kidd JR, Kidd KK, Jin L, Li H. 2015. Inferring population structure and demographic history using Y-STR data from worldwide populations. Molecular Genetics and Genomics 290:141-150.