NCJRS
National Criminal Justice Reference Service
Office of Justice Programs
BJA BJS NIJ OJJDP OVC SMART

The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	**Enhancing Mixture Interpretation with Highly Informative STRs**
Author(s):	**Bruce Budowle**
Document Number:	**253067**
Date Received:	**July 2019**
Award Number:	**2015-DN-BX-K067**

This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Federal Agency: Office of Justice Programs, National Institute of Justice, Department of Justice

Federal Grant: Award Number 2015-DN-BX-K067

Project Title: Enhancing Mixture Interpretation with Highly Informative STRs

PI: Bruce Budowle, University of North Texas Health Science Center at Fort Worth, Center for Human Identification, Department of Microbiology, Immunology, and Genetics, Fort Worth, TX 76107
tel: 817-735-2979; email: bruce.budowle@unthsc.edu

Submission Date: June 12, 2018

DUNS Number:

EIN:

Recipient Organization: University of North Texas Health Science Center
3500 Camp Bowie Blvd., Fort Worth, TX 76107

Institutional Profile Number: 6108502

Project/Grant Period: 01-01-2016 - 06-30-2018

Reporting Period: 01-01-2016 - 06-30-2018

Report: **Final Summary Report**

Signature: 

Bruce Budowle, Ph.D.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Summary

The Final Technical Report contains the details of the study and thus is not described in detail herein. Those interested in more explicit technical aspects should consult the Technical Report (dated 06-30-2017 and accepted by NIJ). Instead, a summary of the findings is provided to briefly capture the topic of the research.

Forensic DNA typing technologies for human identity testing are powerful tools to support investigations. The general methodology of PCR and detection of fluorescently-labeled amplicons during capillary electrophoresis (CE) is the mainstay approach of forensic DNA typing and is robust and reliable. The primary markers used for identity testing are short tandem repeats (STRs). However, complex DNA mixture profiles (i.e., three or more contributors, inhibited and degraded DNA profiles, and/or situations where stochastic effects are exacerbated) present challenges for analysts interpreting the profile(s) (Coble 2014). As the number of contributors increases and the quality of the DNA decreases, mixture interpretation becomes increasingly more difficult. Even with several publications providing interpretation guidance (Budowle et al 2009, Gill et al 2006, Gill et al 2008, Schneider et al 2009, SWGDAM 2010) and statistical approaches with supporting software to facilitate interpretation (Balding and Buckleton 2009, Balding 2013, Bright et al 2015, Perlin et al 2011, Prieto et al 2014, Inman et al 2015, Bleka et al 2016, Moretti et al 2017,) in some cases good results are being ignored and in some other cases inappropriate interpretations have been made (Bieber and Budowle 2015, Coble 2014). An important limitation is that mixtures are complex. The current set of markers may not be able to resolve all contributors of a mixture.

Since STRs are the main currency of forensic DNA typing, applying STRs better suited for mixture analysis would improve the overall process. The desired characteristics of additional STRs are: high heterozygosity, narrow spread of alleles (length based), and compatibility with diagnostic chemistry. The first two criteria are important for discrimination power and robust performance during PCR, respectively. The latter is not readily controlled by the analyst (as the sequence dictates downstream typing compatibility), but does impact the ability to research candidate markers.

With the advent of massively parallel sequencing (MPS), the actual sequence of each STR allele (repeat region and flanking regions) can be determined. Beyond the size of a STR allele, intra-allelic sequence variants will contribute to greater diversity and may provide information to distinguish stutter products from minor contributor alleles and shared or stacked alleles from multiple contributors in mixtures. It is now possible to select STRs with high heterozygosity and limited allele spread that may be better suited for mixture deconvolution.

There are other advantages to the use of MPS. With increased throughput, more markers (and more samples) can be analyzed at one time without consuming more precious evidence. Because size separation is not required for detection as it is with CE methods, all PCR-generated STR amplicons could be of more similar size and overall shorter in length than current constructs (Fordyce et al 2015, Zeng et al 2014). It is entirely possible that amplicon size of all STRs in a multiplex panel could be approximately 250 base pairs or less. This size reduction would result in a more efficient PCR and a more robust assay for typing challenged samples. MPS can provide data with a wide dynamic range, which makes the delineation of true sequence from artifacts more effective.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

The data from MPS studies indicate that there is substantial hidden variation in some STR loci (see Novroski et al 2016). Given the substantial number of STRs residing in the human genome that have not been investigated and the power of MPS, novel STRs could be discovered that meet the above desired criteria to support better mixture deonvolution capabilities. This project sought to identify STRs (autosomal and Y chromosome) that may increase the resolving power of DNA typing, as well as characterize the underlying sequence variation of the current core STRs.

Prior to the initiation of this project it was not possible to analyze STR variation from MPS data. Therefore, Warshauer et al (2013) developed STRait Razor to readily extract STR data from MPS data. STRait Razor is a software program that detects autosomal, X-chromosome, Y-chromosome, or all STRs simultaneously from MPS generated data. Subsequently, STRait Razor v2.0 (Warshauer et al 2015) was developed and incorporated several new features and improvements upon the original software, such as a larger default locus configuration file that increases the number of detectable loci, an enhanced custom locus list generator, a novel output sorting method that highlights unique sequences for intra-repeat variation detection, and a genotyping tool that emulates traditional electropherogram data. The software has supported our laboratory and other laboratories (such as FBI, NIST, AFDIL) performing MPS research and validation studies.

At the initiation of this project, commercial software was not capable of analyzing both the repeat motif and flanking region variation. In other words SNP variation in the flanking region near the repeats was ignored and thus some intra-allelic variation was not being considered. Intra-allelic variation in this context is defined as SNPs and repeat motif variation within the repeat region and sequence variation in the near flanking regions. Moreover, INDELs in the flanking region affect the length of an amplicon and hence the allele call by CE. Thus, the size of an allele by CE may not be due solely to the number of repeats it contains. Flanking variation will need to be captured to increase backwards compatibility of MPS data with CE data. STRait Razor v2S was developed to be able to detect this additional intra-allelic variation and provide the community with a freely accessible tool. This version of STRait Razor facilitated STR allele calling that is compatible with CE-based STR typing nomenclature, captured intra-repeat sequence variation and flanking variation. In addition, a set of features were added to capture SNPs and INDELs that reside in the flanking areas near the repeats and report the results in a fashion similar to variation within the repeats. STRait Razor 2vS (King et al 2017).

While STRait Razor v2S was a successful product, the software was restricted to unix-type environments and was still inordinately slow. A new redesign v3.0 (Woerner et al 2017) provides both a stable code-base that operates on all major operating systems including Microsoft Windows and an indexing strategy tailored to the identification of sequence variants based on anchor sequences. STRait Razor v3.0 is considerably faster than its predecessor, with single-core processing times decreasing by ~660x, going from an average of ~2 hours (7,627 seconds) to 11 seconds for a single high-coverage sample.

Current standard methodology in forensic DNA typing relies on amplification of STR markers by the PCR, and subsequently allele sizes (which are length-based) are determined for each locus by CE. MPS allows for detection of nominal length-based genetic variation as well haplotype sequence variation. The increased effective number of alleles per marker for some STR loci

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

improves power of discrimination. To enable use of MPS data in the forensic setting, population data are needed (especially for core marker systems) to calculate appropriate statistics. A population study was performed in this project using the ForenSeq™ DNA Signature Prep Kit (Illumina, San Diego, CA) to characterize sequence-based alleles (flanking and repeat regions) of 27 autosomal, 7 X chromosome and 24 Y chromosome STR markers in 780 individuals in four populations (African American, Caucasian, Hispanic, and Chinese). Consistent with other studies (see Novroski et al 2016 for citations) the diversity of some STR loci increased notably due to sequence variation.

Given these findings, it stands to reason that other STRs exist in the human genome that may be suitable for enhancing forensic analyses. Because of the online availability of human genome data through such resources as the 1000 Genomes Project (2015) and the STR Catalog Viewer (http://strcat.teamerlich.org/) (Willems et al 2014), STRs were searched using fundamental criteria of high heterozygosity, tetra-, penta-, or hexanucleotide repeat length, and overall relative narrow allele spread (based on length). All candidates were further scrutinized for chemistry compatibility. The resulting candidate STRs were multiplexed and sequenced in a limited sample population set. Each candidate STR was evaluated for analytical performance and desired biological properties. The findings resulted in a refined set of 53 potentially highly polymorphic STR markers (high sequence diversity/heterozygosity; reduced allele spread) that may be suitable to supplement the current core marker set(s) for possible enhanced characterization of complex DNA profiles.

Preliminary mining of Y STR data in the publicly available data (1000 Genomes Project 2015) identified ~1300 highly polymorphic STRs on the Y chromosome. Of these loci, the preliminary higher diversity 290 of these STRs were sequenced on a small sample population. The Y chromosome contains less diversity than its autosomal counterparts and thus many candidates will not be useful for enhancing forensic applications. Of the 176 loci that met the calling criteria, 109 showed no evidence of heterozygosity within individuals. The remaining candidates (plus a few other potential markers) formed a pool of 96 candidate markers with potential high diversity.

Stutter is an artifact of STR typing commonly occurring due to strand slippage during amplification. Understanding the rate and variance of stutter is an essential part of mixture interpretation (particularly for resolving minor and trace contributions from that of stutter). In addition, the modeling of stutter is an important aspect of probabilistic genotyping. A major predictor of the rate of stutter is the longest uninterrupted stretch (LUS) of tandem repeats (Vilsen et al 2017, Brookes et al 2012, Walsh et al 1996, Bright et al 2013, Bright et al 2014). Generally, the LUS should be a better predictor of rates of stutter compared with the parental allele length (PAL) (Vilsen et al 2017, Bright et al 2014, Aponte et al 2015) which is determined by CE. However, because of INDELs in the flanking region or sequence variation within the repeat motifs, LUS and PAL may not always be concordant. Sequence data allow for better elucidation of the variation within alleles and thus a better estimate of true LUS which in turn provides better modeling of stutter.

Rates of stutter were evaluated by both the LUS and the PAL on differing haplotypic backgrounds in four population groups that were analyzed by Novroski et al (2016). Simple repeats were studied primarily as there are less confounding factors than compound and complex repeats. Different flanking haplotypes within loci were identified and characterized, and the types and locations of

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

the variants involved were described. Longer repeating alleles tend to be associated with higher rates of stutter which is consistent with stutter ratios with length-based CE generated data. Analyzing stutter on different haplotypic backgrounds allows for a better understanding of stutter variance associated with particular alleles and may allow for better modeling for probabilistic genotyping. Overall the rates of stutter can vary across loci and on different haplotypic backgrounds within loci, even in the simplest types of repeats. Thus there is no single approach that will apply to all loci to model the stutter ratio in simple repeats.

Compound repeats are more complex for determining the causes and consequences of stutter. As compound repeats have two repeating motifs, each motif may have a different rate of stutter; thus as the repeat within the locus is compound, so is the resultant stutter. Further, the rates of stutter formation between these motifs may not be independent. This lack of independence may complicate modeling strategies, thus contributing to the challenges of mixture interpretation with CE but the user will be better informed with MPS data. Because of the complexity of modeling this study explored two compound repeat loci, wherein two types of stutter are apparent: one that, in the samples evaluated, lacks flanking variation (D2S1338) and another with extant flanking variation (D12S391). Consistent with expectations longer alleles have more relative stutter (with the longer repeat stretch contributing more so to stutter) and some flanking variation affects stutter while other does not.

Overall these data support that stutter rates and models will be locus specific. These stutter analyses indicate that sequence data provide better insight into stutter modeling than CE data. Thus, probabilistic genotyping can be improved by moving to sequence data.

Account of the Activities

The primary goals of this research were to: (1) develop software to identify and characterize all the potential variants within a STR amplicon; (2) determine and characterize allele sequence variation of those STRs of the CODIS core set and typically used Y STRs in several major population groups, which included identification and description of the type of variation(s) that exist at these STRs; (3) Identify novel autosomal STRs and Y-STRs by mining from public data sets (such as 1000 Genomes Project) that may be better suited for analysis of complex sample. Criteria for selection were high diversity and limited allele length spread; and (4) model simple and compound locus stutter, such that stutter may be better defined and provide data to contribute to the development of probabilistic genotyping of allelic sequence data.

Accomplishments

The project was highly successful in all activity areas.

A suite of software tools collectively named STRait Razor was developed. The software is freely available and provides any user the capability of characterizing sequence variation (for STRs, SNPs, INDELs, and microhaplotypes) from data generated by MPS. The software is used by several laboratories worldwide and the accompanying publications have been well cited. The download package for STRait Razor v2s may be found at https://www.unthsc.edu/graduate-school-of-biomedical-sciences/molecular-and-medical-genetics/laboratory-faculty-and-staff/strait-razor. STRait Razor v3.0 also is open source and freely available at

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

https://github.com/Ahhgust/STRaitRazor and at https://www.unthsc.edu/graduate-school-of-biomedical-sciences/molecular-and-medical-genetics/laboratory-faculty-and-staff/strait-razor, along with standard configuration files that will enable it to be used on a variety of MPS systems.

Second, the underlying genetic variation was described for the commonly used autosomal and Y STRs. Increased genetic variation was observed for a majority of the loci. These population data can support statistical calculations with MPS generated data. Moreover, the data point to the increased diversity that can be obtained with MPS. The complete population data are available in Novroski et al (2016) and at https://www.unthsc.edu/graduate-school-of-biomedical-sciences/molecular-and-medical-genetics/laboratory-faculty-and-staff/.

Third, publicly available sequence databases were mined to identify STRs that may be better suited for forensic casework and in particular mixture analyses. The main criteria for the autosomal STRs were high heterozyosity and narrow length allele spread. There were 53 STRs, after filtering and refining (inlcuding chemistry compatibility), that meet these criteria. Y STRs show less variability compared with their autosomal counterparts as predicted. None the less 96 candidates were discovered. Empirical data were generated on a limited population sample to determine if the candidate loci should meet the selection criteria. These data have been submitted for publication and will be made available at our website in the near future.

Fourth, although sample data are limited, modeling of stutter from simple and two example compound STRs was performed. The data show that stutter rates are more complex than just the length of the repeat region. Flanking regions may have an influence on stutter. There was no overarching model that could be applied to all loci. Therefore, stutter will have to be modeled specific to each locus. These stutter results can be used as a basis for probabilistic genotyping with MPS data.

Lastly, the overall data support that MPS will provide substantial improvements in data analysis (greater variation detected) and interpretation (higher resolution of mixture components and better modeling of stutter for probabilistic genotyping).

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

References

Aponte RA, Gettings KB, Duewer DL, Coble M, Vallone PM. 2015. Sequence-based analysis of stutter at STR loci: Characterization and utility. Forensic Sci Int Genet Suppl Ser 5:e456–e458.

Balding DJ. 2013. Evaluation of mixed-source, low-template DNA profiles in forensic science. Proc Natl Acad Sci USA 110(30):12241-12246.

Balding DJ, Buckleton J. 2009. Interpreting low template DNA profiles. Forensic Sci Int Genet 4(1):1-10.

Bieber FR, Budowle B. 2015. Final Report on Review of Mixture Interpretation in Selected Casework of the DNA Section of the Forensic Science Laboratory Division, Department of Forensic Sciences, District of Columbia. http://dfs.dc.gov/page/usao-report-april-20152015.

Bleka Ø, Storvik G, Gill P. 2016. EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. Forens Sci Int Gen 21:35-44.

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

Bright JA, Stevenson KE, Coble MD, Hill CR, Curran JM, Buckleton JS. 2014. Characterising the STR locus D6S1043 and examination of its effect on stutter rates. Forens Sci Int Genet 8:20–23.

Bright JA, Taylor D, Curran JM, Buckleton JS. 2013.Developing allelic and stutter peak height models for a continuous method of DNA interpretation. Forens Sci Int Genet 7:296–304.

Brookes C, Bright JA, Harbison S, Buckleton J. 2012. Characterising stutter in forensic STR multiplexes, Forens Sci Int Genet 6:58–63.

Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL. 2009. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. J Forens Sci 54:810-821.

Coble M. 2014. MIX13: An Interlaboratory Study on the Present State of DNA Mixture Interpretation in the U.S. The American Bar Association's Fifth Annual Prescriptions for Criminal Justice Forensics, Fordham Law School, New York, NY.

Fordyce SL, Mogensen HS, Børsting C, Lagacé RE, Chang CW, Rajagopalan N, and Morling N. 2015. Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™. Forens Sci Int Genet 14, 132-140.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS. 2006. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Sci Int 160(2-3):90-101.

Gill P, Brown RM, Fairley M, Lee L, Smyth M, Simpson N, Irwin B, Dunlop J, Greenhalgh M, Way K, Westacott EJ, Ferguson SJ, Ford LV, Clayton T, Guiness J. 2008. Technical UK DNA working group. National recommendations of the Technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. Forensic Sci Int Genet 2(1):76-82.

Inman K, Rudin N, Cheng K, Robinson C, Kirschner A, Inman-Semerau L, et al. 2015. Lab Retriever: a software tool for calculating likelihood ratios incorporating a probability of drop-out for forensic DNA profiles. BMC Bioinformatics 16:298.

King JL, Wendt FR, Sun J, Budowle B. 2017. STRait Razor v2s: Advancing sequence-based STR allele reporting and beyond to other marker systems. Forens Sci Int Genet 29:21-28.

Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright J-A, et al. 2017. Internal validation of STRmix; for the interpretation of single source and mixed DNA profiles. Forens Sci Int Genet 29:126-144.

Novroski NM, King JL, Churchill JD, Seah LH, Budowle B. 2016. Characterization of genetic sequence variation of 58 STR loci in four major population groups. Forens Sci Int Genet 25:214–226.

Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL. 2011. Validating TrueAllele® DNA mixture interpretation. J Forensic Sci 56 (6):1430-1434.

Prieto L, Haned H, Mosquera A, Crespillo M, Alemañ M, Aler M, et al. 2014. Euroforgen-NoE collaborative exercise on LRmix to demonstrate standardization of the interpretation of complex DNA profiles. Forens Sci Int Genet 9:47-54.

Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, Rothämel T, Schmitter H, Schneider H, Brinkmann B. 2009. The German Stain Commission: recommendations for the interpretation of mixed stains. Int J Legal Med 123(1):1-5.

SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. 2010. At: http://www.fbi.gov/about-us/lab/biometric-analysis/codis/swgdam.pdf.

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation, Nature 526:68-74.

Vilsen SB, Tvedebrink T, Mogensen HS, Morling N. 2017. Statistical modelling of Ion PGM HID STR 10-plex MPS data. Forens Sci Int Genet 28:82–89.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Walsh PS, Fildes NJ, Reynolds R. 1996. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus Vwa. Nucleic Acids Res 24:2807–2812.

Warshauer DH, King JL, Budowle B. 2015. STRait Razor v2.0: the improved STR allele identification tool – razor. Forens Sci Int Genet 14:182–186.

Warshauer DH, Lin D, Hari K, Jain R, Davis C, LaRue B, King J, Budowle B. 2013. STRait Razor: A length-based forensic STR allele-calling tool for use with second generation sequencing data. Forens Sci Int Genet 7:409–417.

Willems T, Gymrek M, Highnam G. 2014. The 1000 Genomes Project. Genome Res 24(11):1894-904.

Woerner AE, King JL, Budowle B. 2017. Fast STR allele identification with STRait Razor 3.0. Forens Sci Int Genet 29:21-28.

Zeng X, King JL, Stoljarova M, Warshauer DH, LaRue BL, Sajantila A, Patel J, Storts DR, Budowle B. 2014. High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing. Forensic Sci Int Genet 16C:38-47.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Products Produced

In addition to the research results obtained, numerous presentations and a peer-review published paper have been produced that document the work.

Presentations at National and International Meetings that were supported by this work

1. Novroski, N., Churchill, J., King, J., and Budowle, B.: What's hiding between the primers? Using massively parallel sequencing to capture STR repeat region and flanking region sequence variation, 27th International Symposium on Human Identification, Minneapolis, MN, 2016.

2. Novroski, N.M., Churchill, J.D., King, J. and Budowle, B.: The application of short tandem repeat (STR) sequence variation for the selection of novel STR markers to enhance DNA mixture deconvolution: what do we know and where are we headed?, American Academy of Forensic Sciences, New Orleans, LA, 2017.

3. Budowle, B.: Massively parallel sequencing can advance forensic genetics capabilities, 2nd Saudi International Conference of Forensic Medicine and Sciences, Riyadh, Saudi Arabia, 2017.

4. Budowle, B.: Massively parallel sequencing is NGS - that is now generation sequencing, 14th Annual DNA Conference – Bode West, Phoenix, AZ, 2016.

5. Budowle, B.: The adoption process of MPS into US forensic genetics laboratories: US perspective, Human Identification Solutions Conference, Vienna, Austria, 2017.

6. Novroski, N., Woerner, A.E., King, J., and Budowle, B.: Upping the mixture game: newly-adopted STR markers for enhanced DNA mixture de-convolution, 28th International Symposium on Human Identification, Seattle, WA, 2017.

7. Woerner, A.E., King, J., and Budowle, B.: Quality scores in MPS data: what are they good for? 28th International Symposium on Human Identification, Seattle, WA, 2017.

8. Budowle, B., King, J.L., Novroski, N.M.M., Takahashi, M., Wendt, F.R., and Woerner, A.E.: The research and development progress of enhancing mixture interpretation with highly informative STRs, National Institute of Justice Forensic Science Research & Development Symposium, Pittcon, Orlando, FL, 2018.

Publications

Novroski, N.M, King, J.L., Churchill, J.D., Seah, L.H., and Budowle, B.: Characterization of genetic sequence variation of 58 STR loci in four major population groups. Forens. Sci. Int. Genet. 25:214-226, 2016.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

King, J.L., Wendt, F.R., Sun, J., and Budowle, B.: STRait Razor v2s: Advancing sequence-based STR allele reporting and beyond to other marker systems. Forens. Sci. Int. Genet. 29:21-28, 2017.
Woerner, A.E., King, J.L., and Budowle, B.: Fast STR allele identification with STRait Razor 3.0. Forensic Sci. Int. Genet. 30:18-23, 2017.

Woerner, A.E., King, J.L., and Budowle, B.: Flanking variation influences rates of stutter in simple repeats. Genes 8(11):329, 2017.

Churchill, J.D., Novroski, N.M.M., King, J.L., Seah, L.H., and Budowle, B.: Population and performance analyses of four major populations with Illumina's FGx Forensic Genomics System. Forens. Sci. Int. Genet. 30:81-92, 2017.

Novroski N.M.M., Woerner A.E., and Budowle, B.: Insertion within the flanking region of the D10S1237 Locus. Forens. Sci. Int. Genet. (in press).

Novroski NMM, Woerner AE, Budowle B. 2018. Potential highly polymorphic short tandem repeat markers for enhanced forensic identity testing. Forens Sci Int Genet (submitted).

Woerner AE, King JL, Budowle B. 2018. Compound stutter in D2S1338 and D12S391. Forens Sci Int Genet (submitted).

Invention Report
Two software packages were developed and are freely available

STRait Razor v2S
Available at: https://www.unthsc.edu/graduate-school-of-biomedical-sciences/molecular-and-medical-genetics/laboratory-faculty-and-staff/strait-razor

STRait Razor 3.0
Available at: https://github.com/Ahhgust/STRaitRazor and at https://www.unthsc.edu/graduate-school-of-biomedical-sciences/molecular-and-medical-genetics/laboratory-faculty-and-staff/strait-razor

No patents were submitted related to this project

Participating Scientists and Collaborators

What individuals have worked on the project?

Name: Bruce Budowle
Project role: PI
Nearest person month worked: 1 year

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Contribution to Project: Dr. Budowle provided the overall direction and management of the project as well as participated in the technical research, data analysis, and manuscript preparation.

Name: August Woerner
Nearest person month worked: 1 year
Contribution to Project: Dr. Woerner is a Research Assistant Professor who performed bioinformatics analyses of STR sequence variation data, performed simulation studies to model stutter, and redesigned STRait Razor (v 3.0) to a C++ format for faster speed and utility. He also participated in manuscript preparation.

Name: Nicole Novroski
Nearest person month worked: 1 year
Contribution to Project: Nicole Novroski is a graduate student who performed the main effort of the work on autosomal STRs. She also participated in manuscript preparation.

Name: Maiko Takahashi
Nearest person month worked: 1 year
Contribution to Project: Maiko Takahashi is a postdoctoral fellow who performed part of the effort of the work on Y chromosome STRs.

Name: Jonathan King
Nearest person month worked: 1 year
Contribution to Project: Mr. King is the laboratory manager who performed data analyses and evaluation of supporting data accumulation, assisted in laboratory work data analysis, and updated STRait Razor (v2S). He also participated in manuscript preparation.

What other organizations have been involved as partners? none
Have other collaborators or contacts been involved? no

Impact

What is the impact of the project on the criminal justice system?

The software tool has been shared with ADFDIL, FBI, NIST, and others worldwide and has been placed on the internet for free access. As cited in the text, it is expected that the forensic community will continue to make use of the new versions of STRait Razor (v2s and 3.0). Other contributions are identification of informative STRs, description of the alleles and their frequencies of such loci in relevant populations, and submission of allelic data to STRSeq (supported by NIST). Candidate markers better suited for mixture analyses have been identified. Industry has shown an interest in evaluating the markers for enhanced MPS kits. Overall, the data should be useful to improve interpretation of mixture evidence and augment kinship analyses.

b. How has it contributed to crime laboratories?

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

The current impact of this project on the criminal justice system can be assessed in part as the STRait Razor software has facilitated analysis of MPS data for the other major laboratories that are testing and validating MPS. Without this software progress on eventual implementation would have been slowed. These enhanced software for STR analyses from MPS data are freely available to the forensic science and greater scientific communities. There has been some impact in that the software tool STRait Razor 3.0 is being used by some entities (such as AFDIL, NIST and the FBI). Thus, the benefit of the software is immense. A set of currently used autosomal and Y STRs are fully described so that MPS population data now are available for statistical analyses. As far as the novel STRs for better mixture interpretation, contribution cannot be assessed yet. Although markers have been identified, commercial kits are necessary to enable end users to employ these markers. The potential impact of this project, however, will be substantial. The outcome of this effort will be improved capabilities to interpret DNA mixture evidence and identification of novel STRs with greater discrimination power. A novel set of STRs are available now to enhance mixture interpretation. Due to the increased power of discrimination, these markers also will be useful for other identity testing applications, such as missing persons cases. The outcome is that more biological evidence will be analyzed successfully, which in turn will result in more and better investigative leads to solve crimes. In addition, this project helps support the utility of MPS and will promote its application. These data will be completely available and results should be able to be uploaded into national databases (such as STRSeq) to augment investigative lead value that ultimately will solve more cases, better data sharing among laboratories, and generate innovation in the development of better systems for analyzing a greater variety of forensic evidence. In addition, this work defines some of the advantages and limitations of MPS which policy and decision makers can use for long term planning and funding directions.

c. What is the impact on technology transfer?

Software to the community (freely available)

Changes/Problems

There were no challenges that could not be met. The only changes were a no cost extension to ensure that the work could be completed and in a change in the salary of allotment to Dr. Budowle (reduced) to ensure sufficient reagents for MPS.

Proprietary Information

There was no proprietary information related to this work.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.