UNDERSTANDING THE BEHAVIOR OF STUTTER THROUGH THE SEQUENCING OF STR ALLELES

P-176

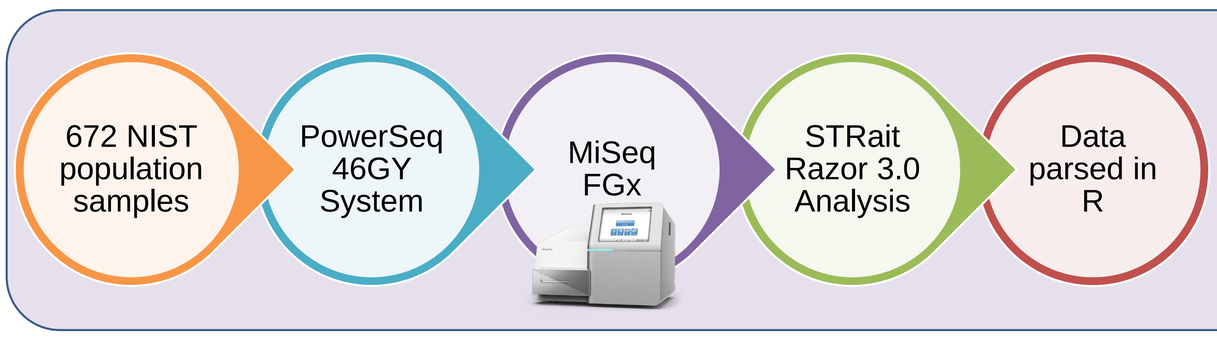
Sarah Riman[1], Hariharan Iyer[2], Lisa A. Borsuk[1], Peter M. Vallone[1]

peter.vallone@nist.gov
sarah.riman@nist.gov

[1] U.S. National Institute of Standards and Technology, 100 Bureau Drive, Applied Genetics Group, Gaithersburg, MD 20899-8314, USA
[2] U.S. National Institute of Standards and Technology, 100 Bureau Drive, Statistical Design, Analysis, and Modeling Group, Gaithersburg, MD 20899-8980, USA

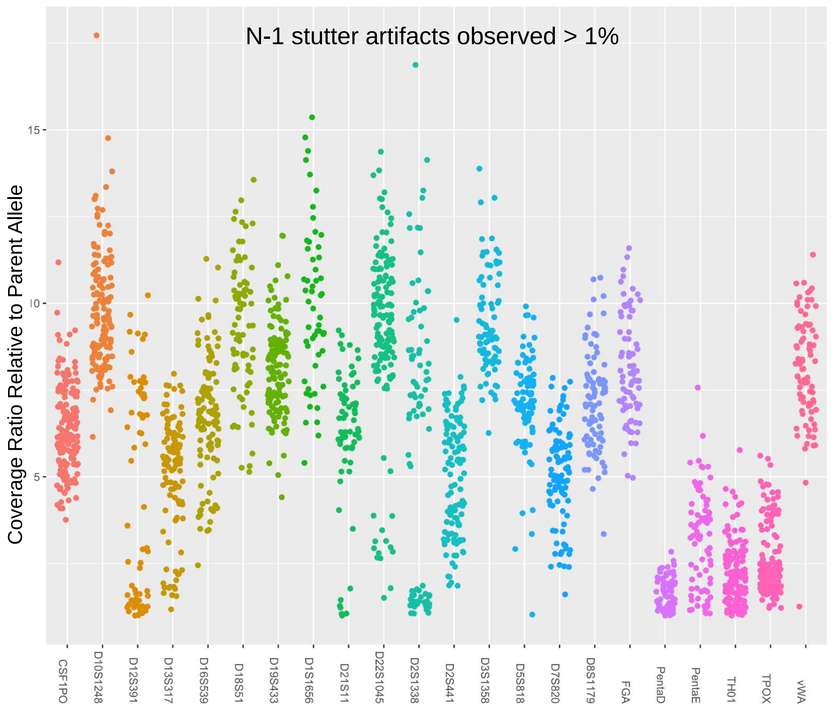
This work explores the influence of several variables on stutter formation across sequenced autosomal STR loci (simple, compound, and complex motifs) and different alleles within each locus. The variables are sequence variations within the repeating motifs and flanking region [1, 2]; longest uninterrupted stretch (LUS) [3]; parental allele length [3]; and base pair content and length value of each repeating motif from which the stutter has generated [3, 4]. Over six hundred unrelated individuals from different populations were amplified with the prototype PowerSeq 46GY System and sequenced on the Illumina MiSeq platform. Raw FASTQ files were analyzed with STRait Razor v3.0 [5]. Stutter ratio was calculated for motifs that exhibited stutter using the ratio of the observed coverage of the stutter sequence at (N-1) position to the observed coverage of the allelic sequence.

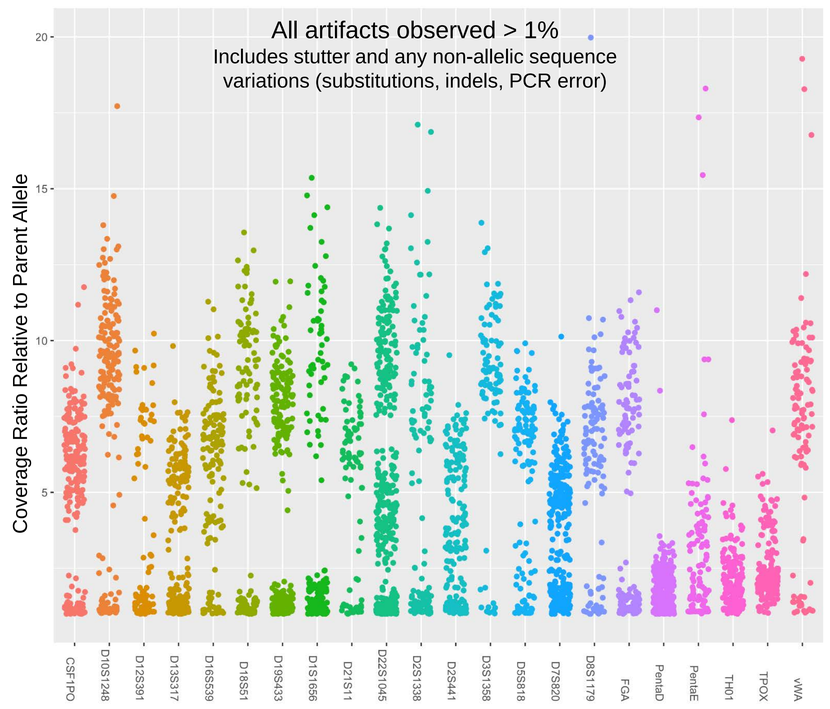
Materials and Methods

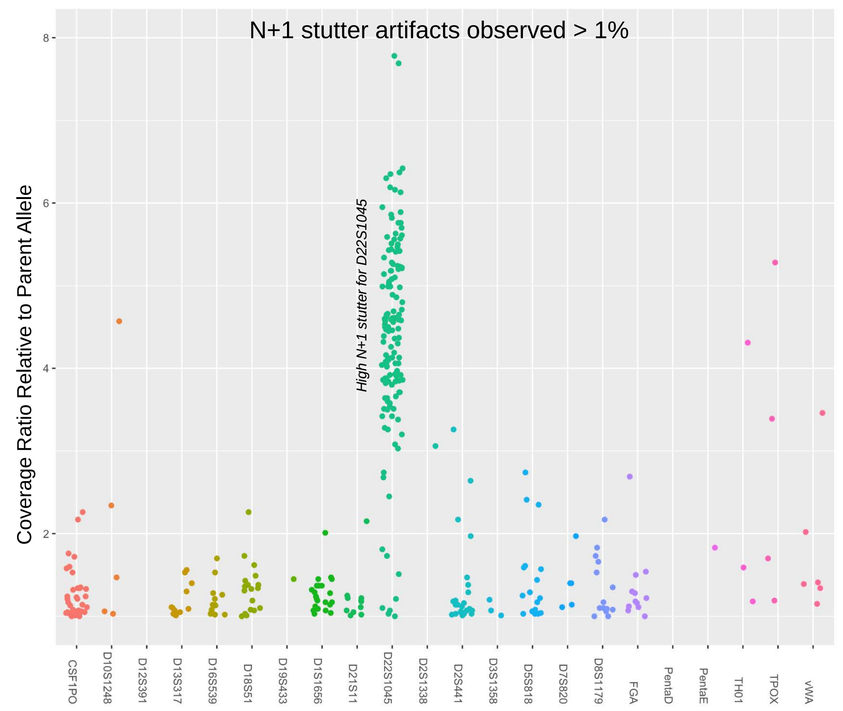


- U.S. Caucasian, Hispanic, African American samples (n=672)
- Library preparation performed using prototype PowerSeq 46GY System (Promega)
  - 1 ng template
  - 22 autosomal STRs, 23 Y STR markers, and Amelogenin
- Sequenced on the Illumina MiSeq platform (v3-300)
- FASTQ files parsed using STRait Razor 3.0 [5]
- **Initial analysis**: Homozygous alleles for the autosomal STR markers (n=2197 alleles)
- Sequences observed at > 1% of the total locus coverage were characterized

Artifacts Observed Across All Autosomal STR Loci

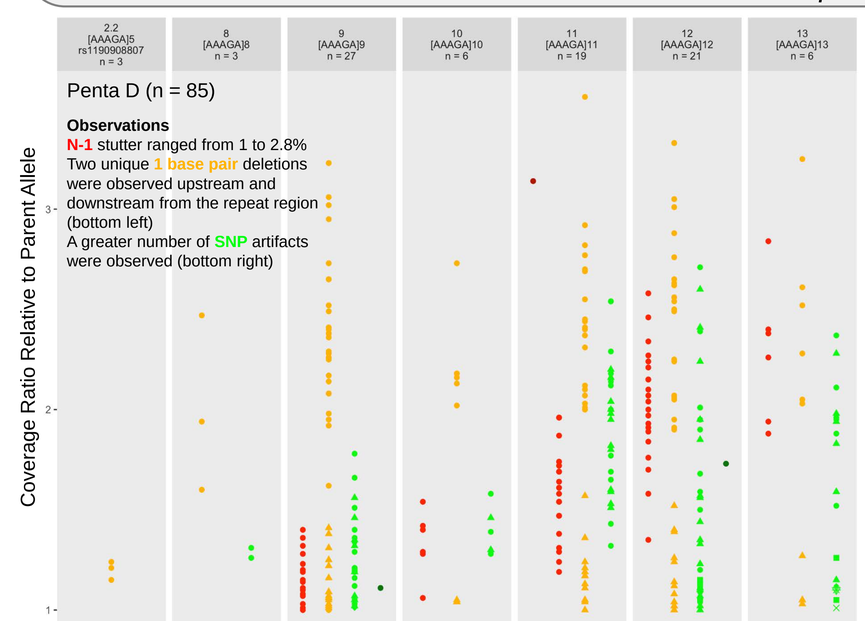




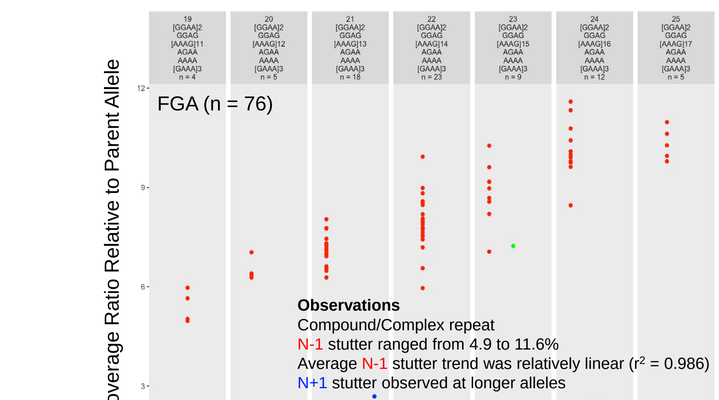


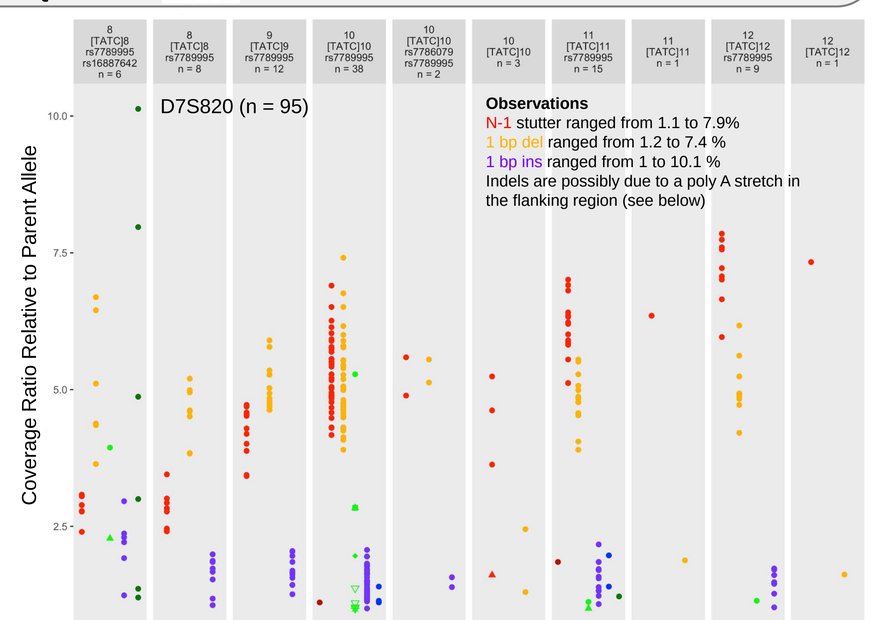
Interesting Observations for Six of the 22 Autosomal STRs

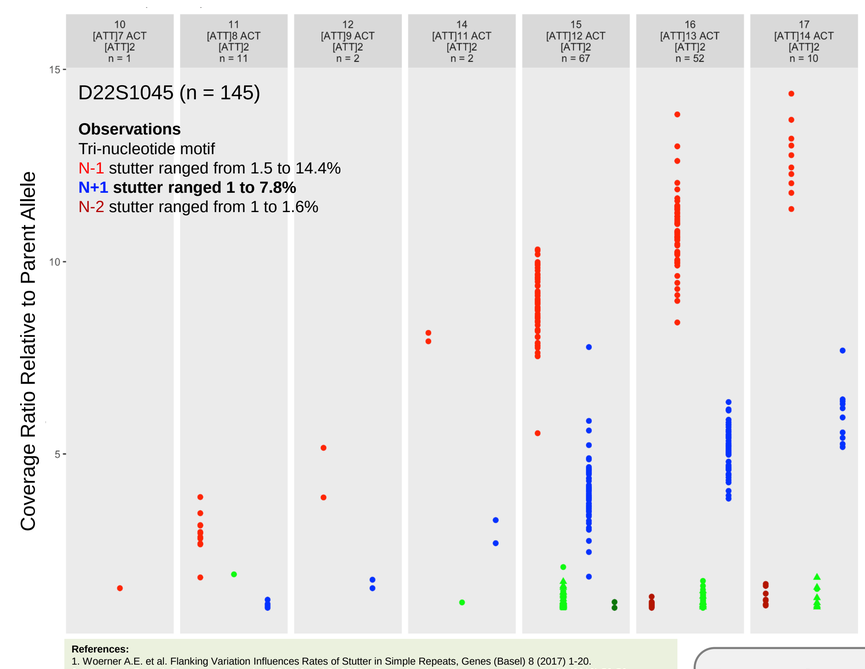
This poster and plots for all 22 autosomal loci can be downloaded using the QR code or https://strbase.nist.gov/NISTpub.htm#Presentations

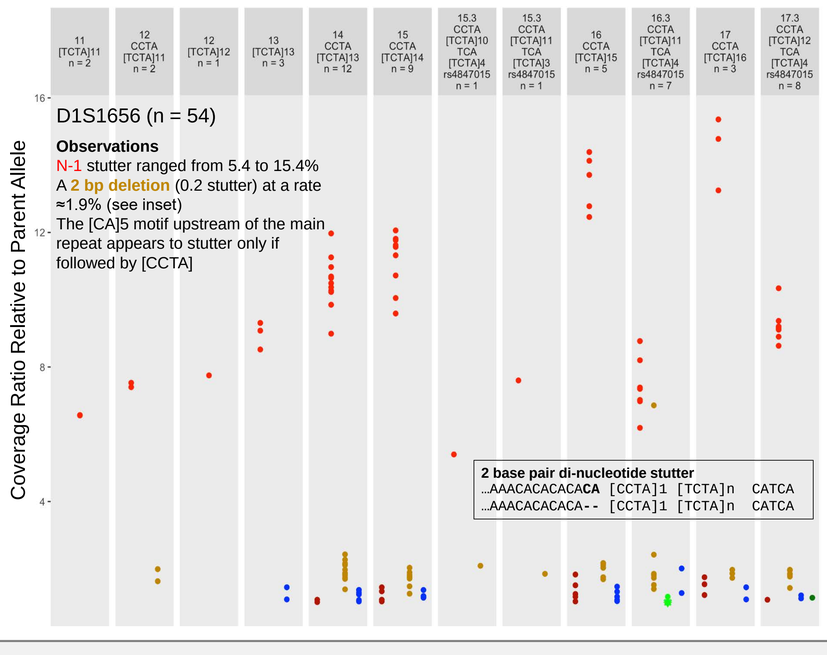


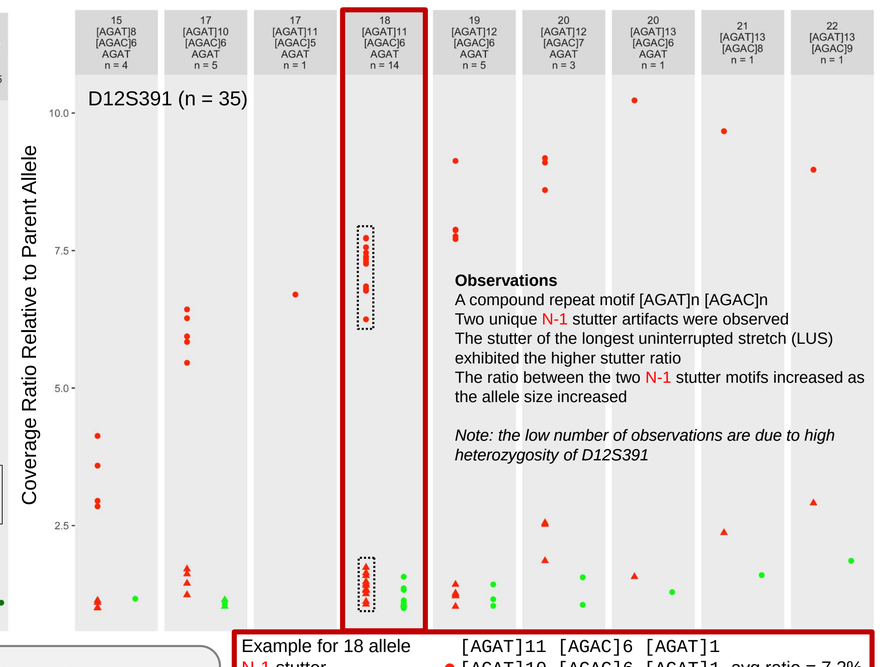
Artifact Type	Legend Key	Comment
	N-1, N-2, N+1	Stutter where the integer represents the motif length (tri, tetra, penta)
	Other	Other stutter motifs (N-3, N+2, etc)
	-1bp, -2bp, +1bp	Insertion or deletions in the flanking region of the sequence
	SNP	A single base substitution observed in the parent allele
	Unique Sequence ● ▲ ■ ◆ ▽ ✳ ✕	Symbols indicate unique sequence types within a category











References:
1. Woerner A.E. et al. Flanking Variation Influences Rates of Stutter in Simple Repeats, Genes (Basel) 8 (2017) 1-20.
2. Woerner A.E. et al. Compound stutter in D2S1338 and D12S391, Forensic science international. Genetics 39 (2019) 50-56.
3. Brookes C. et al. Characterising stutter in forensic STR multiplexes, Forensic science international. Genetics 6 (2012) 58-63.
4. Vilsen S.B. et al. Stutter analysis of complex STR MPS data, Forensic science international. Genetics 35 (2018) 107-112.
5. Woerner A.E. et al. Fast STR allele identification with STRait Razor 3.0, Forensic science international. Genetics 30 (2017) 18-23.

Acknowledgements: We wish to thank Doug Storts and Promega for access to the PowerSeq 46GY chemistry, **Becky Steffen and Kevin Kiesler** of the Applied Genetics Group (NIST) for performing the PowerSeq sequencing…and Min Ra for edits and design!

Funding: NIST Special Programs Office: Forensic DNA and the FBI Biometric Center of Excellent Unit: DNA as a Biometric

Future plans: Incorporate the heterozygous genotypes in the analysis. Understanding the behavior **(abundance, reproducibility, sequence context)** of non-allelic artifacts will help in establishing probabilistic models for the prediction of stutter rate and interpretation of sequence-based STR profiles.

Example for 18 allele		[AGAT]11 [AGAC]6 [AGAT]1
N-1 stutter	●	[AGAT]10 [AGAC]6 [AGAT]1 avg ratio = 7.2%
N-1 stutter	▲	[AGAT]11 [AGAC]5 [AGAT]1 avg ratio = 1.4%

Plots for all 22 autosomal STRs