An Exploration of a Sequence-Specific Stutter Model for STR Analysis in Next Generation Sequencing

Damani Johnson, BS; Tim Kalafut, PhD

Department of Forensic Science
Sam Houston State University
Huntsville, TX 77340

FSF FORENSIC SCIENCES FOUNDATION
EMERGING FORENSIC SCIENTIST AWARD

MEMBER THE TEXAS STATE UNIVERSITY SYSTEM

Disclaimer

- The views and opinions expressed in this presentation are those of the presenter/author, and does not represent any official views or opinions of the American Academy of Forensic Sciences.
- Certain commercial equipment, instruments, software, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the authors or the National Institute of Justice nor does it imply that any of the materials, instruments, software, or equipment identified are necessarily the best available for the purpose.

Don’t Let My Relaxed Demeanor Fool You – This Is Exciting

- Being able to identify and filter stutter effectively is important
- Next Generation Sequencing is incredibly powerful and informative, and addressing obstacles can help ease its implementation in crime labs
- Development of software tools can make stutter validation easier and more efficient
- Before and after photos of graph lines getting slightly more linear are very satisfying

Paul

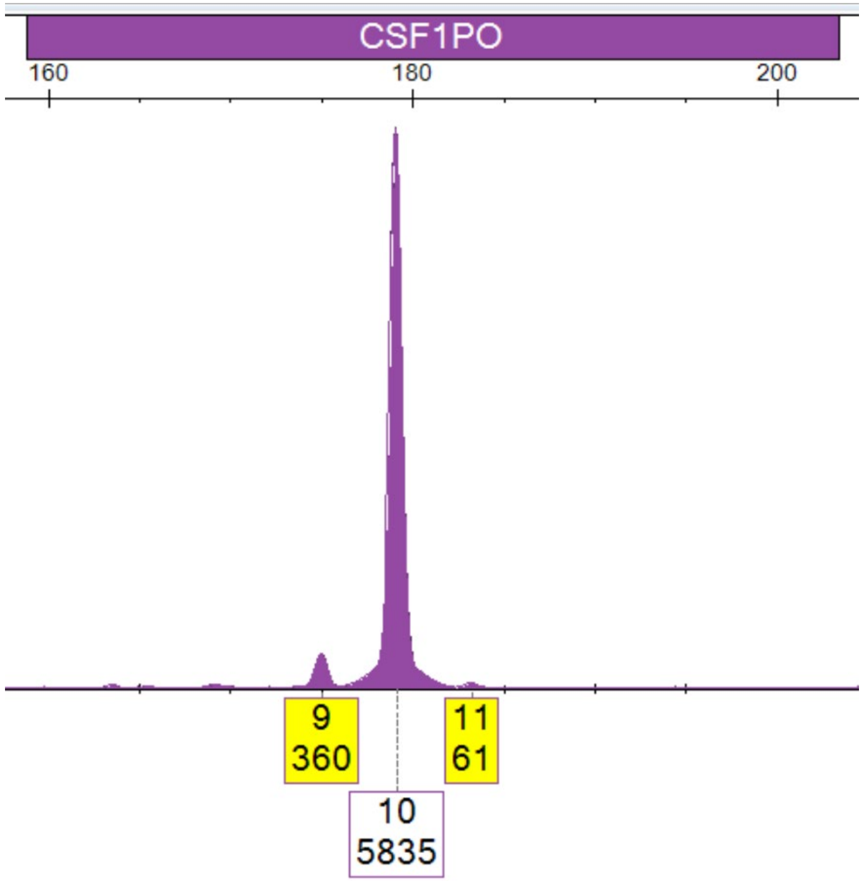
Methods

	CE	NGS
Samples	96 single-source buccal swabs	
Chemistry	Investigator 24plex QS	ForenSeq MainstAY SE
Instrument	3500	Miseq FGx
Data Analysis	OSIRIS/ArmedXpert	MixtureAce

- Additional software tools currently in development were used for calculating and plotting stutter ratios

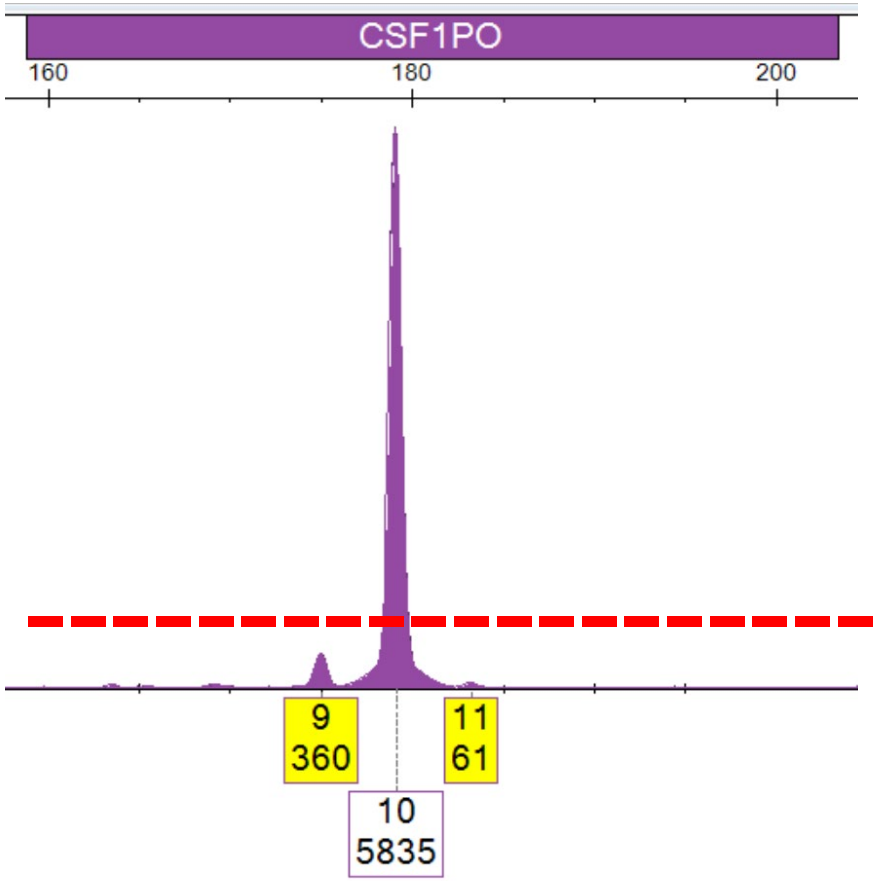
The Issue with Stutter

- Short tandem repeat (STR) markers are susceptible to strand slippage during PCR
- On CE this most commonly results in a stutter peak appearing one repeat unit short of a true allele peak
    - Also seen in a forward +1 or double back -2 position



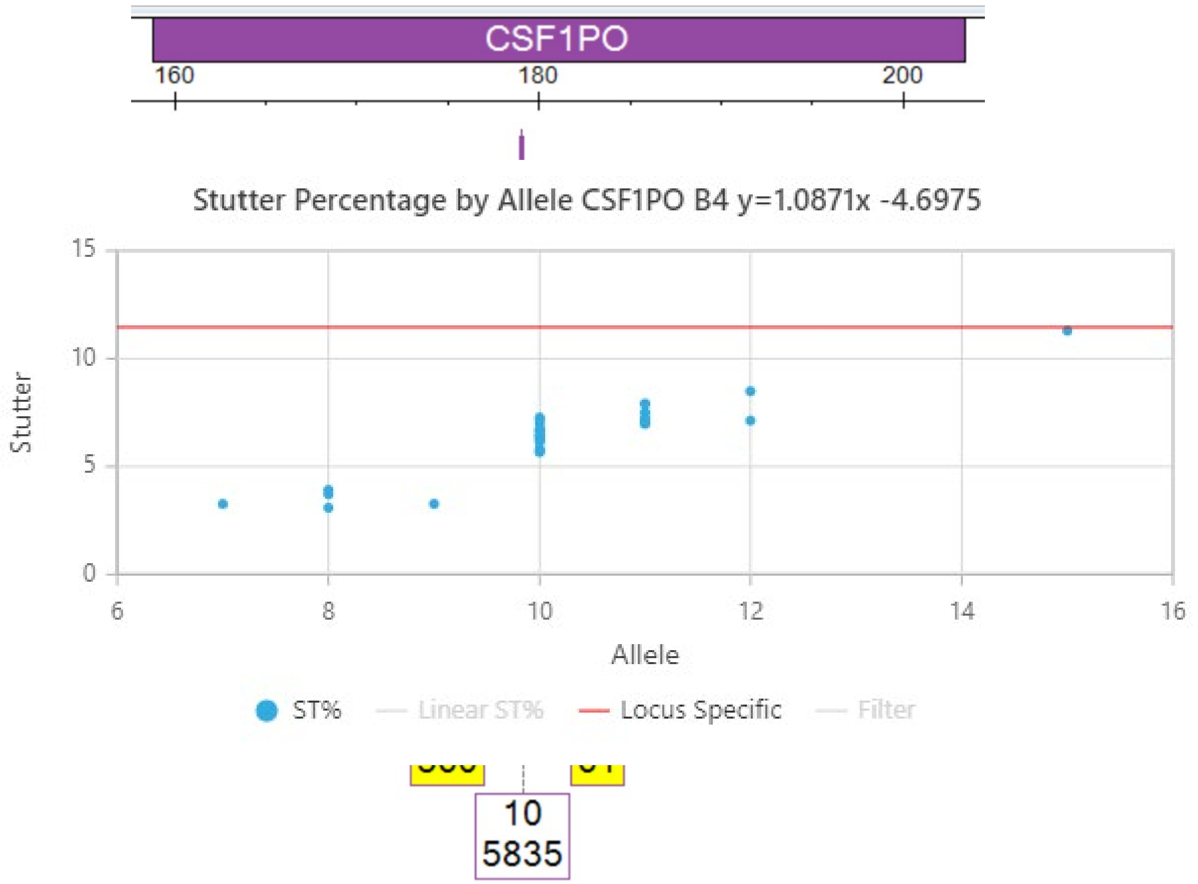
How We Address Stutter

- We use stutter filters in analysis software to help identify stutter peaks automatically
  - Ideally, the better the filter, the less manual editing is needed
- Filters are set through validation of a large single-source sample set
  - Need to validate each kit
  - Need enough samples to have as many alleles and stutter represented as possible



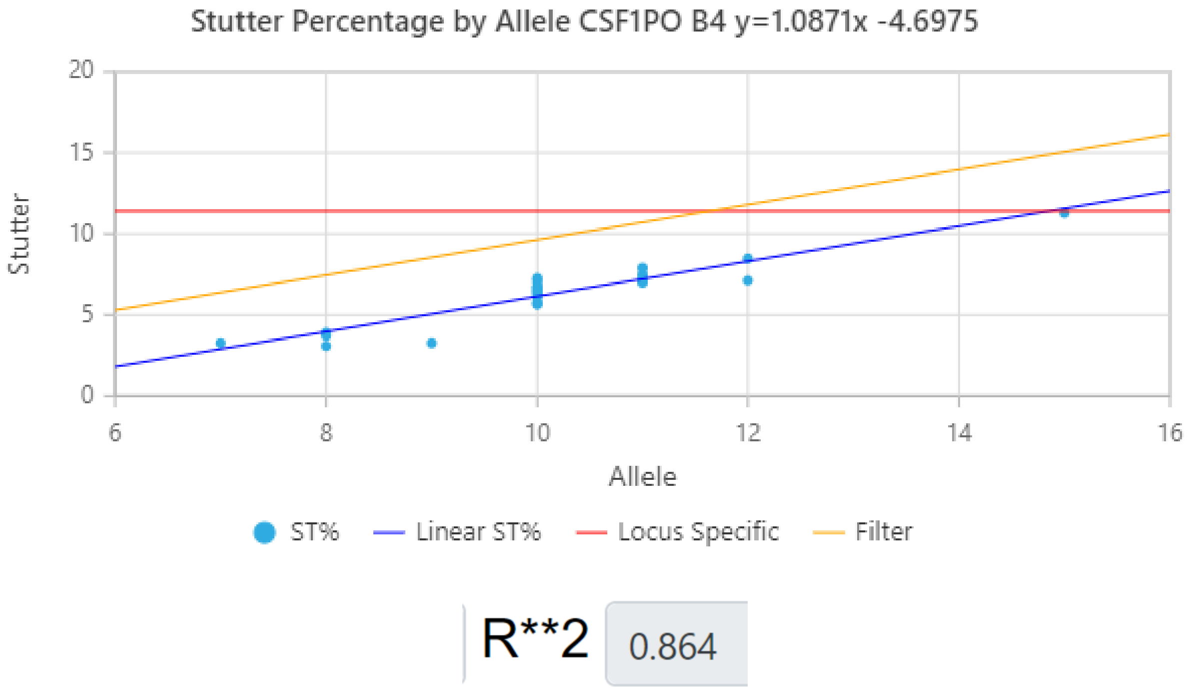
Stutter Filters

- Locus-Specific
  - Static threshold per locus
  - Based on stutter ratio mean and standard deviation
  - Risks over-filtering shorter alleles or under-filtering longer alleles



Stutter Filters

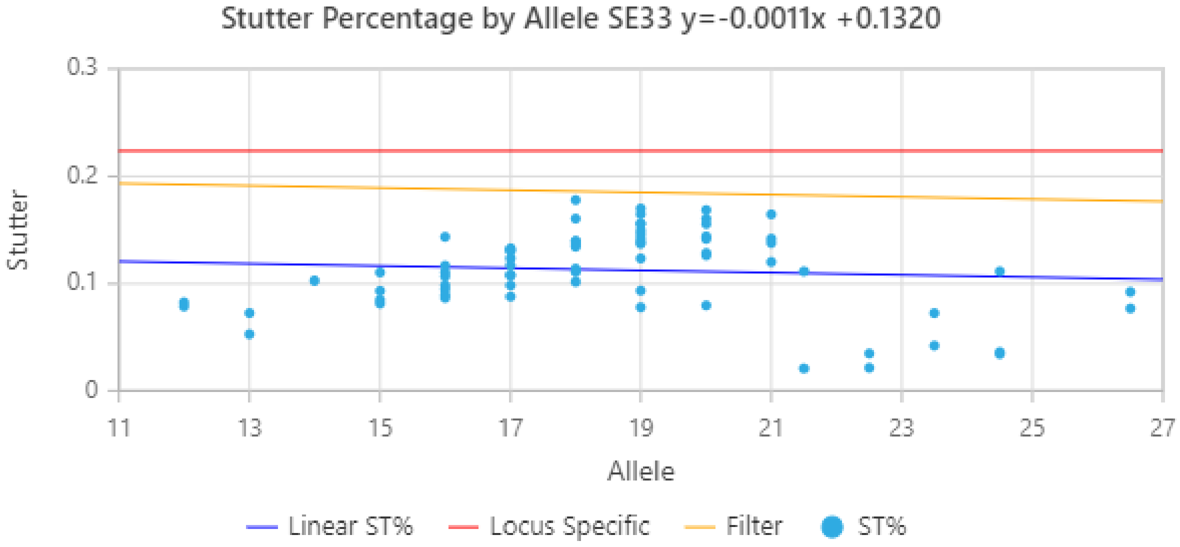
- Locus-Specific
  - Static threshold per locus
  - Based on stutter ratio mean and standard deviation
  - Risks over-filtering shorter alleles or under-filtering longer alleles
- Allele-Specific
  - Based on slope of linear trend line of stutter ratio vs. allele
  - Works very well for loci with simple repeat motifs



R**2 0.864

Allele-Specific Stutter Filter

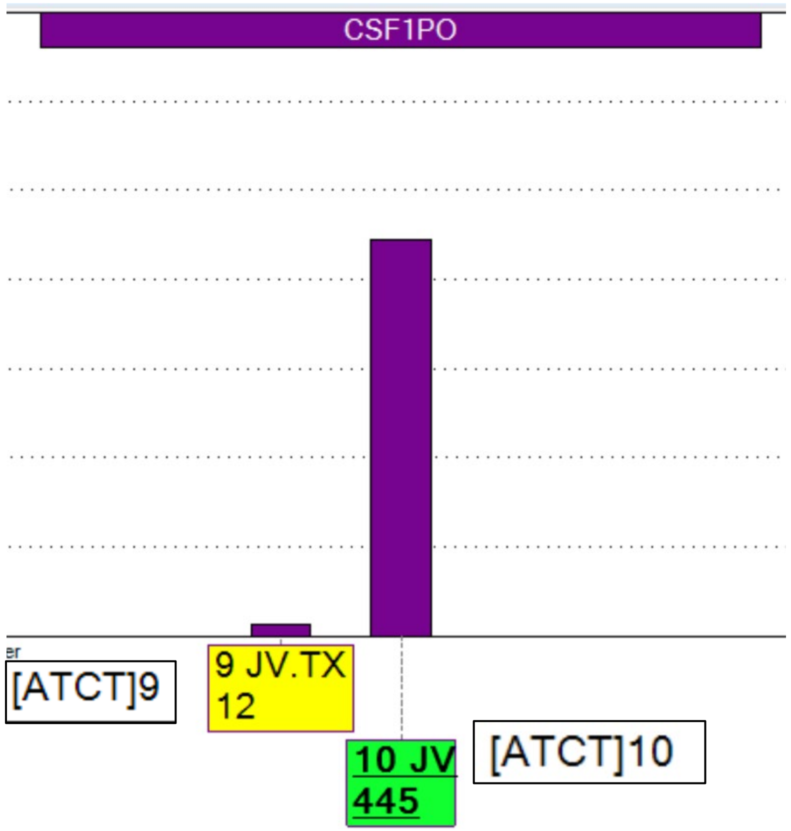
- Current standard for CE based stutter filtering
- Has trouble modeling more complex loci
    - Often appear to have multiple slopes
    - Some loci seem to stutter considerably less at “partial” alleles (.2, .3)
- CE does not give us the information needed to model these more accurately



R**2 0.008

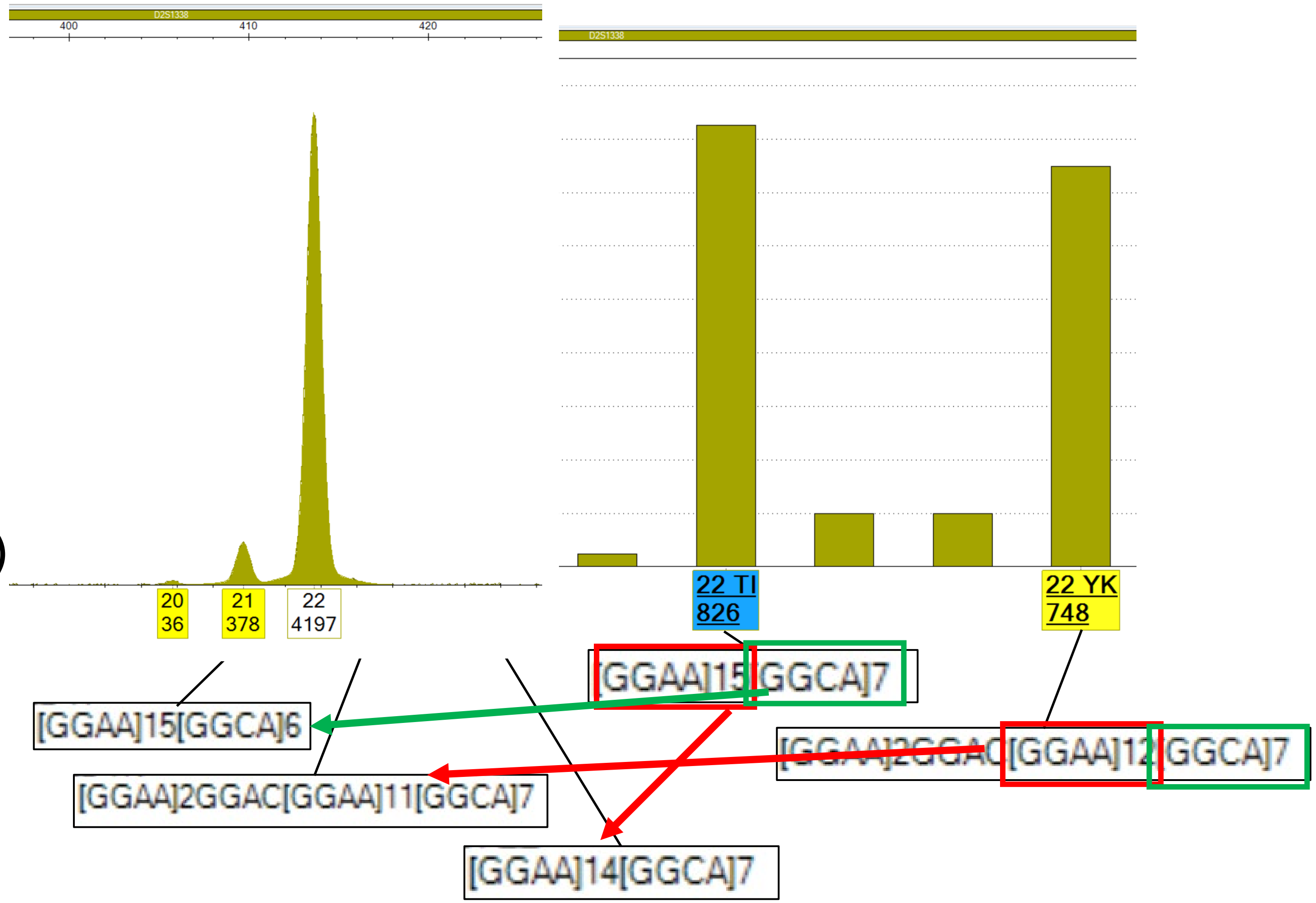
Next Generation Sequencing

- Allows us to view sequence data rather than just length
- When viewed in MixtureAce, alleles are assigned a sequence ID (SID) label that corresponds to a full or bracketed sequence
  - Stutter sequences are associated with their parent allele



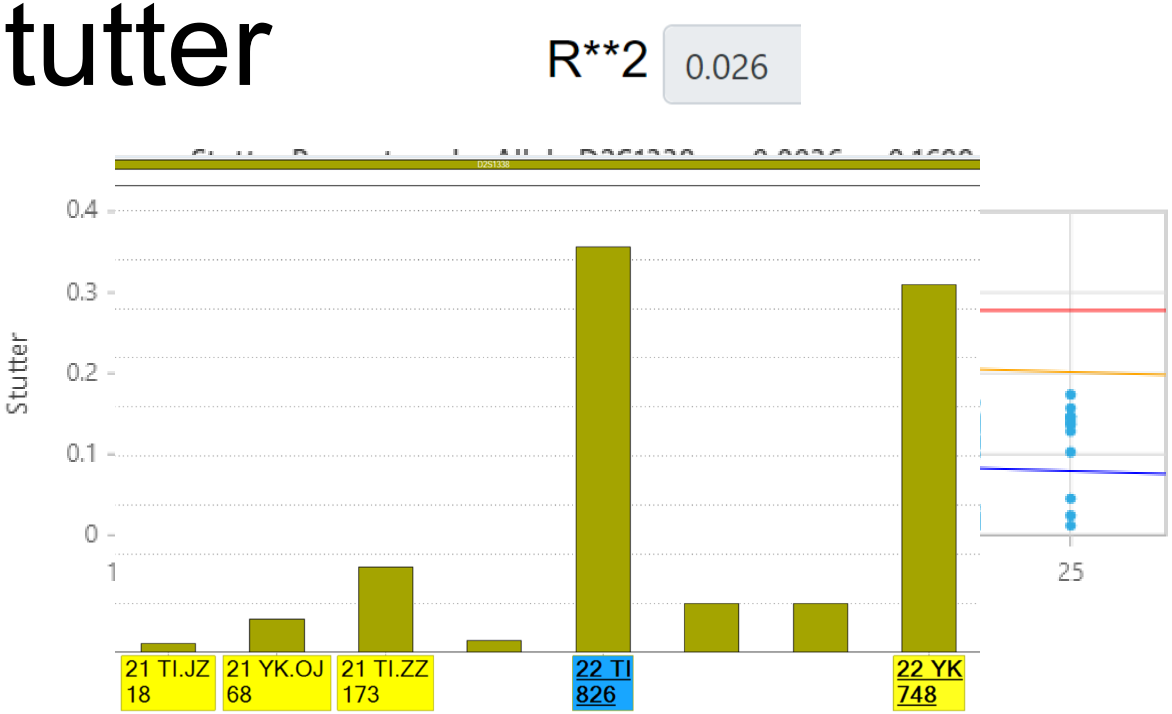
Next Generation Sequencing

- Isoalleles allow us to distinguish alleles of the same length with different sequences
  - Longest Uninterrupted Stretch (LUS)
  - Second Longest Uninterrupted Stretch (SLUS)
- Sequence can be used to associate stutter peaks with correct parent isoallele



Sequence Based Stutter

- These isoalleles from the same sample stutter at very different rates
- Using an allele-specific stutter model treats these all the same

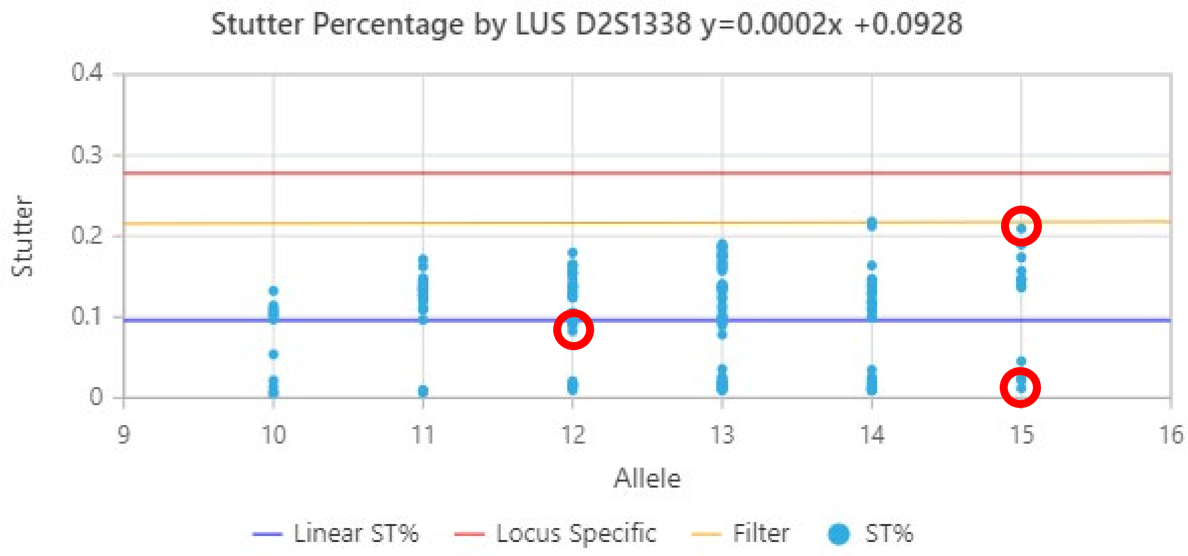


Parent Allele	Stutter Allele	Stutter Ratio	Sequence
22 YK	21 YK.OJ	0.090909	[GGAA]2GGAC[GGAA]12[GGCA]7
22 TI	21 TI.ZZ	0.209443	[GGAA]15[GGCA]7
22 TI	21 TI.JZ	0.021792	[GGAA]15[GGCA]7

Sequence Based Stutter

R**2 0.000

- Plotting by LUS rather than allele number improves slope slightly
- Still not ideal
    - SLUS stutter appears consistently lower
    - R^2 went down

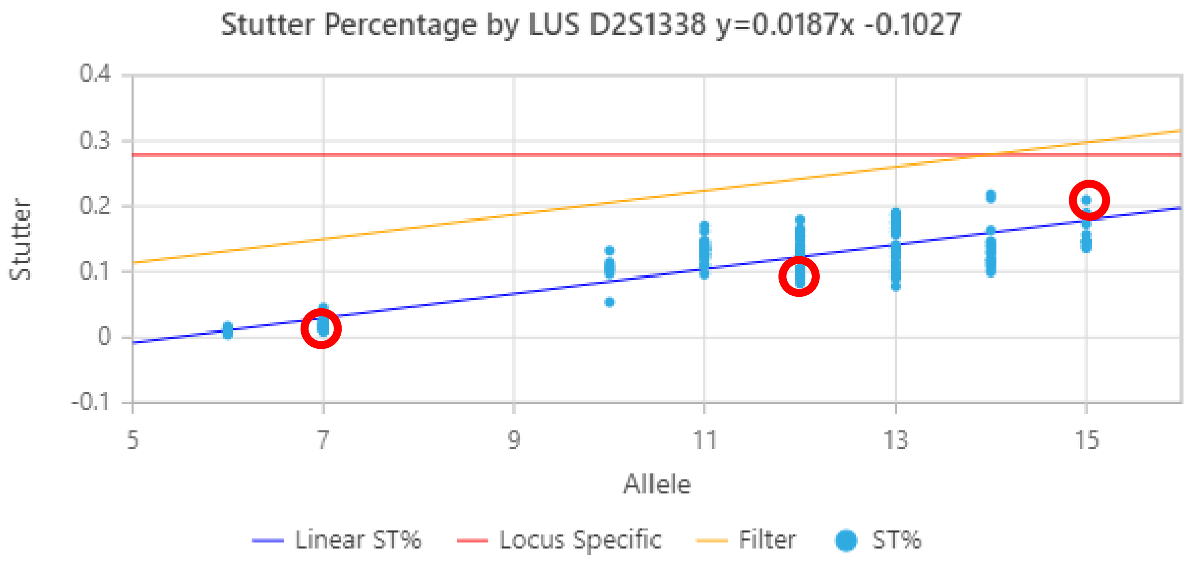


Parent Allele	Stutter Allele	Stutter Ratio	Sequence
22 YK	21 YK.OJ	0.090909	[GGAA]2GGAC[GGAA]12[GGCA]7
22 TI	21 TI.ZZ	0.209443	[GGAA]15[GGCA]7
22 TI	21 TI.JZ	0.021792	[GGAA]15[GGCA]7

Sequence Based Stutter

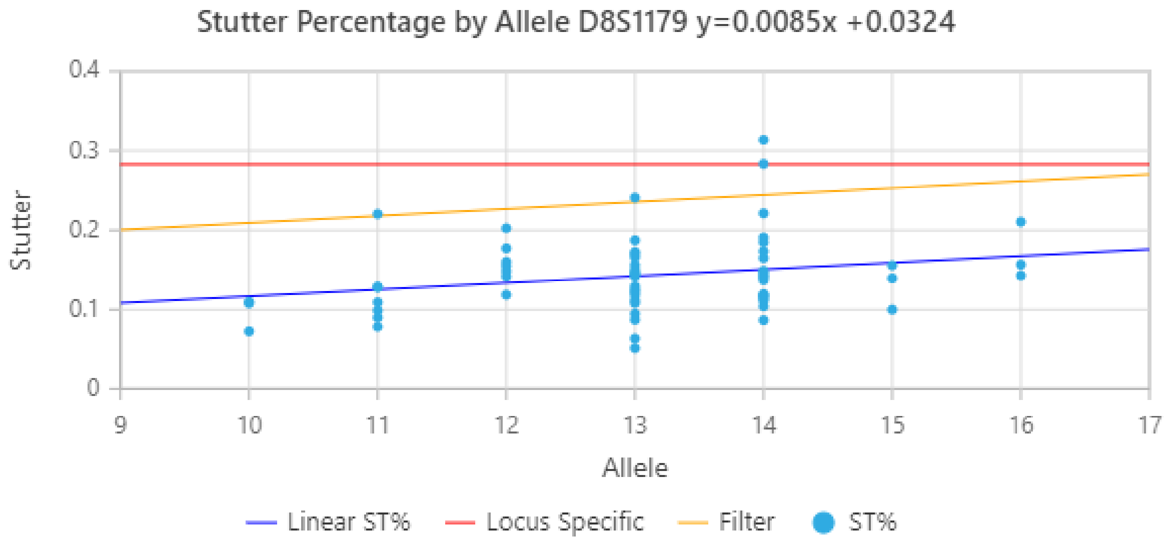
R**2 0.795

- Plotting each stutter ratio by the number of the stuttering repeat unit produced the best linear fit

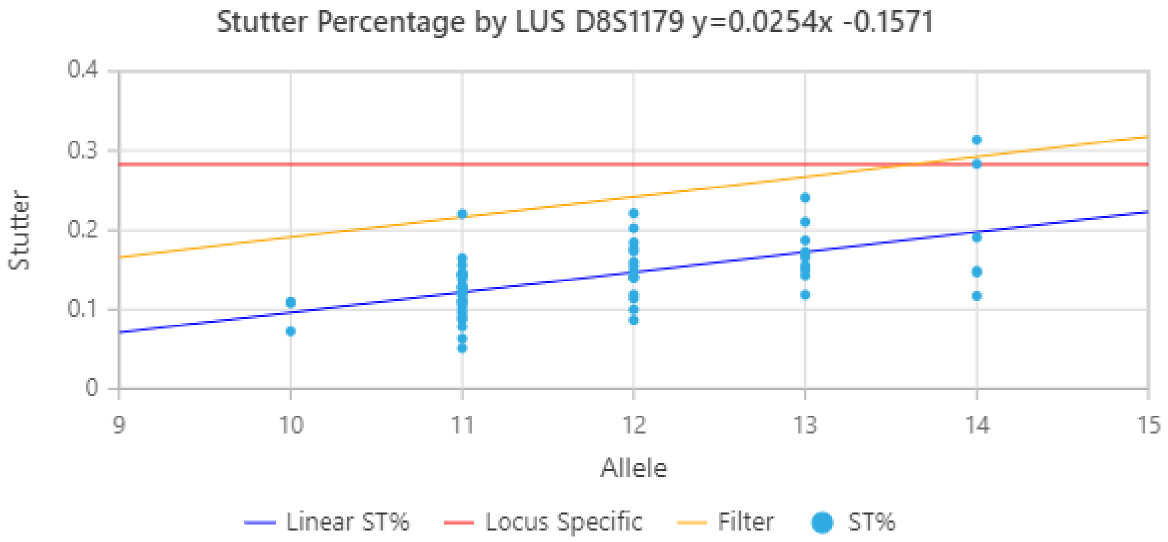


Parent Allele	Stutter Allele	Stutter Ratio	Sequence
22 YK	21 YK.OJ	0.090909	[GGAA]2GGAC[GGAA]12[GGCA]7
22 TI	21 TI.ZZ	0.209443	[GGAA]15[GGCA]7
22 TI	21 TI.JZ	0.021792	[GGAA]15[GGCA]7

Allele vs Sequence

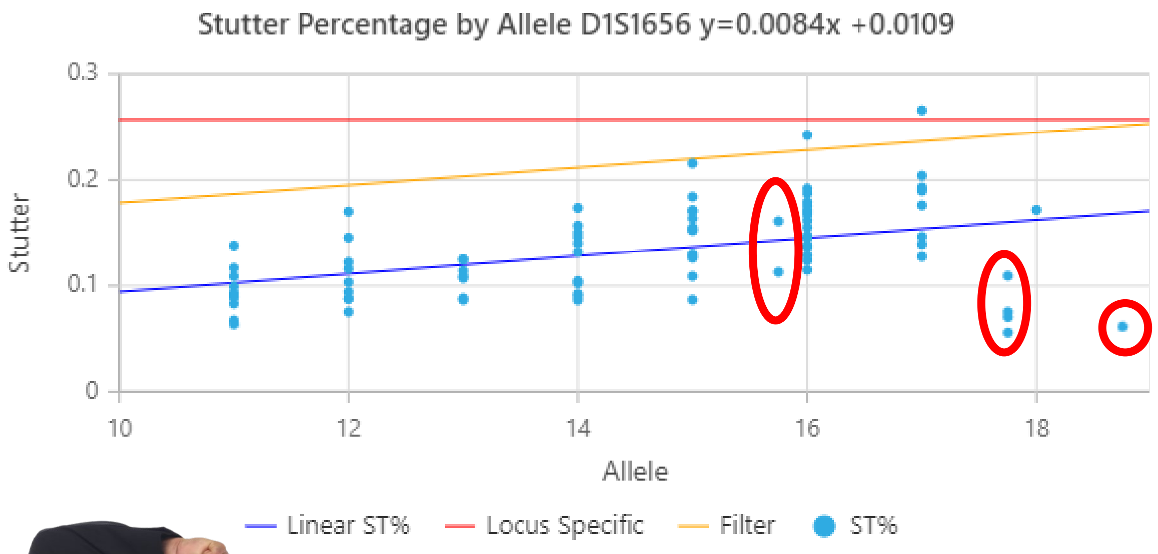

Stutter Percentage by Allele D8S1179 y=0.0085x +0.0324

R**2 0.065

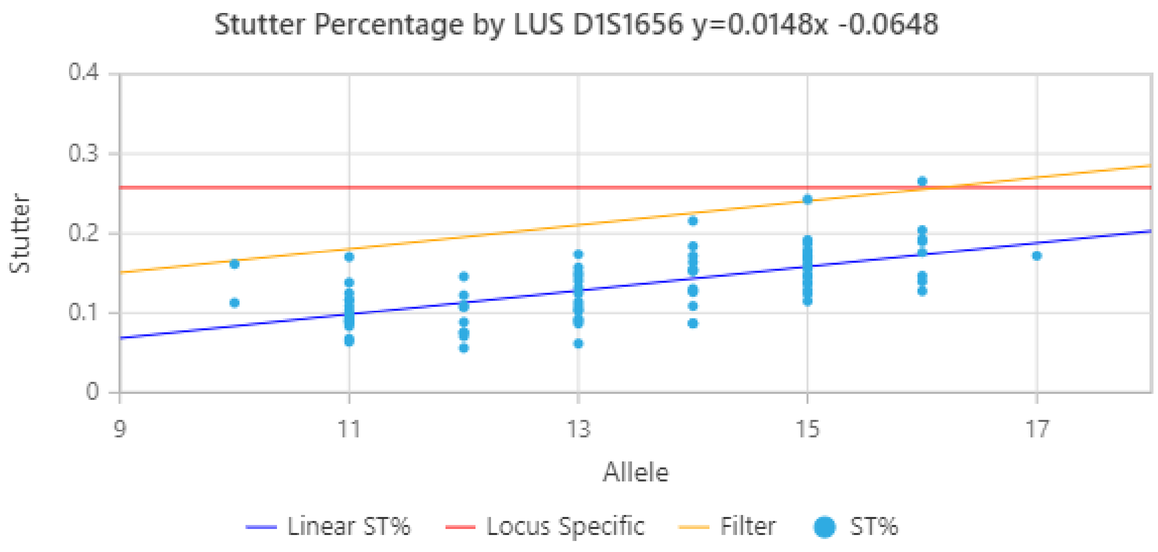

Stutter Percentage by LUS D8S1179 y=0.0254x -0.1571

R**2 0.337

Allele vs Sequence

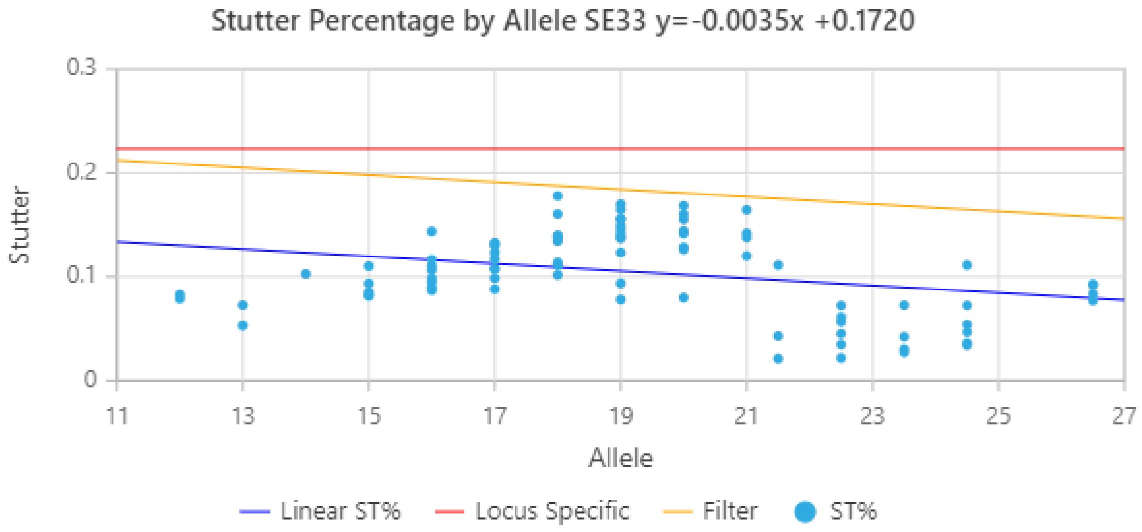


R**2 0.182



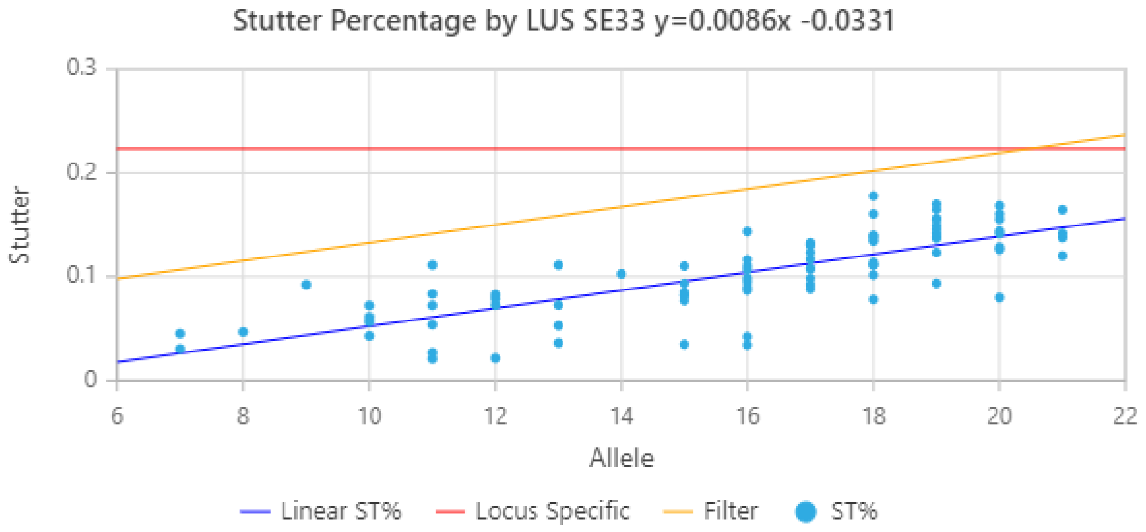
R**2 0.397

Allele vs Sequence



Stutter Percentage by Allele SE33 y=-0.0035x +0.1720

R**2 0.091

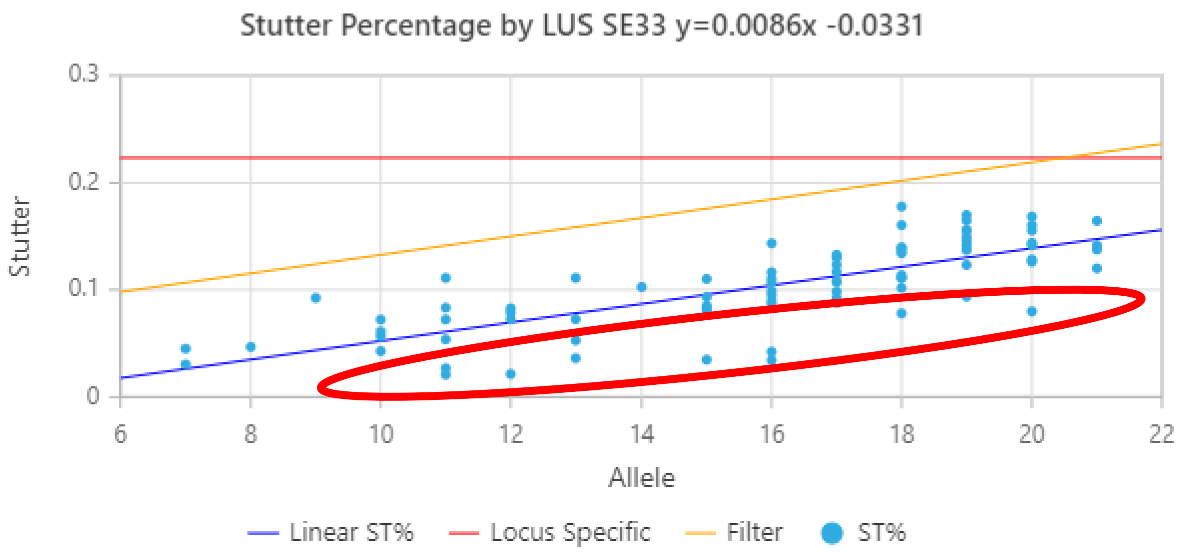


Stutter Percentage by LUS SE33 y=0.0086x -0.0331

R**2 0.588

Additional Considerations

- Small insertion sequences within the STR seem to have further impact on stutter ratio
- SE33 has many ".2" alleles which appear to reduce stutter rates



Allele	Stutter Ratio	Sequence	LUS
16 WD	0.143411	[CTTT]16	16
23.2 PK	0.042157	[CTTT]16TT[CTTT]7	16

R**2 0.506

Final Thoughts

- Plotting stutter by repeat unit rather than allele length improves linearity at many complex loci
- There do appear to be more factors on stutter ratio that will require further study
    - Small insertions within the STR
- Effective stutter validation requires as many observed alleles and stutter as possible
- There is still work to do to make this model and validation tools implementable

Don’t let my relaxed demeanor fool you – This is exciting

Acknowledgements

This project was partially funded by:
NIJ Grant #15PNIJ-21-GG-02710-SLFO
Center for Advanced Research in Forensic Science (CARFS)

Thank you to:
NicheVision Forensics
SHSU Department of Forensic Science
American Academy of Forensic Sciences

Questions?

Damani Johnson

dtj012@shsu.edu