


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC



NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Sequence-based Analysis of Stutter at Forensic STR Loci

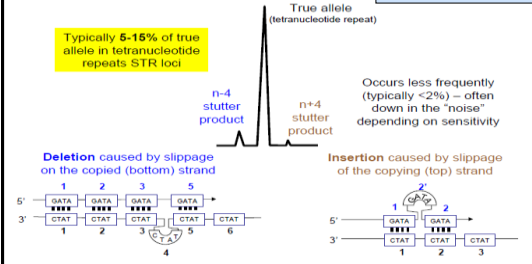
R.A. Aponte^a, D.L. Duwerf^b, P. M. Vallone^b, and K.B. Gettings^b

^a Department of Forensic Sciences, The George Washington University, Washington, DC, 20007-1150
^b National Institute of Standards and Technology, Gaithersburg, MD, 20899-8314

Stutter Product Formation

PRINCIPLES

Stutter % =
RFU of stutter peak
RFU of true allele peak



Typically 5-15% of true allele in tetranucleotide repeats STR loci

True allele (tetranucleotide repeat)

n-4 stutter product n+4 stutter product


Occurs less frequently (typically <2%) – often down in the “noise” depending on sensitivity

Deletion caused by slippage on the copied (bottom) strand Insertion caused by slippage of the copying (top) strand

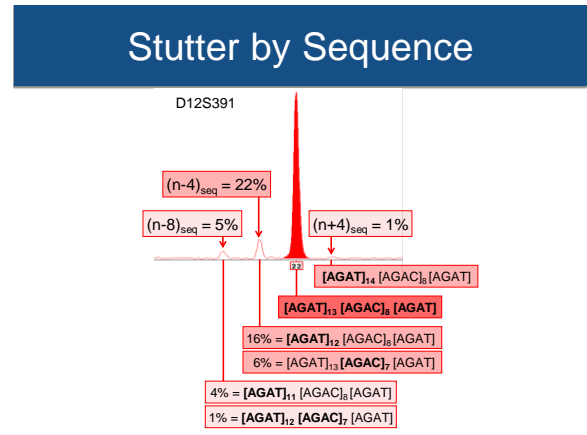
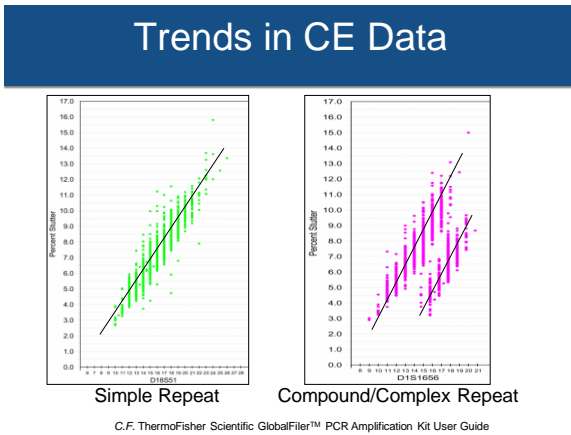
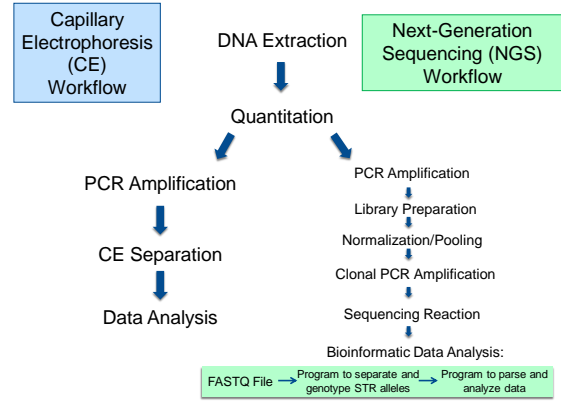
http://www.cstl.nist.gov/strbase/mixture/4%20-%20Stutter.pdf

Why Analyze Stutter?

- Validation to aid in sample analysis and interpretation
 - Allows software to ignore stutter, improving efficiency in data analysis
 - Distinguishing minor contributors from artifact in mixtures
- Improved models of stutter characterization by understanding what influences stutter formation
 - Allele length
 - STR sequence motif
 - AT/GC content
 - Flanking region
 - Amount of DNA



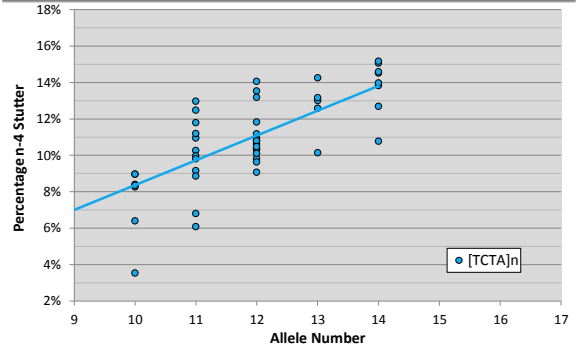
http://www.cstl.nist.gov/strbase/pub_prel/AF42016_NGS%20Workshop_Gettings.pdf



D8S1179 Sequence Motifs

D8S1179	
Allele	Repeat Structure
	[TCTA] ₁₀₋₁₄
10	(TCTA) ₁₀
11	(TCTA) ₁₁
12	(TCTA) ₁₂
13	(TCTA) ₁₃
14	(TCTA) ₁₄

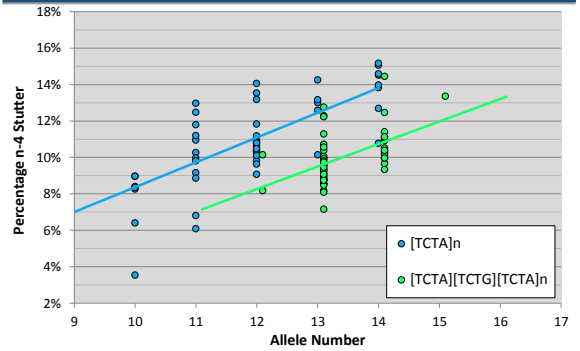
D8S1179 n-4 Stutter by Allele



D8S1179 Sequence Motifs

D8S1179	
Allele	Repeat Structure
	[TCTA] ₁₀₋₁₄
10	(TCTA) ₁₀
11	(TCTA) ₁₁
12	(TCTA) ₁₂
13	(TCTA) ₁₃
14	(TCTA) ₁₄
	[TCTA][TCAG][TCTA] ₁₀₋₁₄
12	(TCTA)(TCAG)(TCTA) ₁₀
13	(TCTA)(TCAG)(TCTA) ₁₁
14	(TCTA)(TCAG)(TCTA) ₁₂
16	(TCTA)(TCAG)(TCTA) ₁₄

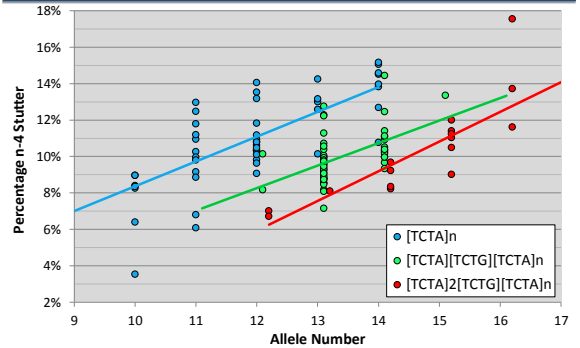
D8S1179 n-4 Stutter by Allele



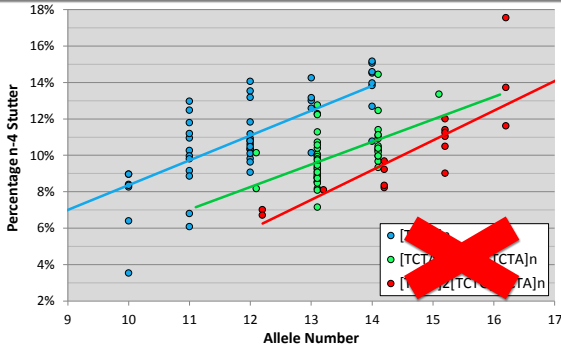
D8S1179 Sequence Motifs

D8S1179	
Allele	Repeat Structure
	[TCTA] ₁₀₋₁₄
10	(TCTA) ₁₀
11	(TCTA) ₁₁
12	(TCTA) ₁₂
13	(TCTA) ₁₃
14	(TCTA) ₁₄
	[TCTA][TCAG][TCTA] ₁₀₋₁₄
12	(TCTA)(TCAG)(TCTA) ₁₀
13	(TCTA)(TCAG)(TCTA) ₁₁
14	(TCTA)(TCAG)(TCTA) ₁₂
16	(TCTA)(TCAG)(TCTA) ₁₄
	[TCTA][TCTA][TCAG][TCTA] ₈₋₁₅
11	(TCTA)(TCTA)(TCAG)(TCTA) ₈
12	(TCTA)(TCTA)(TCAG)(TCTA) ₉
13	(TCTA)(TCTA)(TCAG)(TCTA) ₁₀
14	(TCTA)(TCTA)(TCAG)(TCTA) ₁₁
15	(TCTA)(TCTA)(TCAG)(TCTA) ₁₂
16	(TCTA)(TCTA)(TCAG)(TCTA) ₁₃
17	(TCTA)(TCTA)(TCAG)(TCTA) ₁₄
18	(TCTA)(TCTA)(TCAG)(TCTA) ₁₅

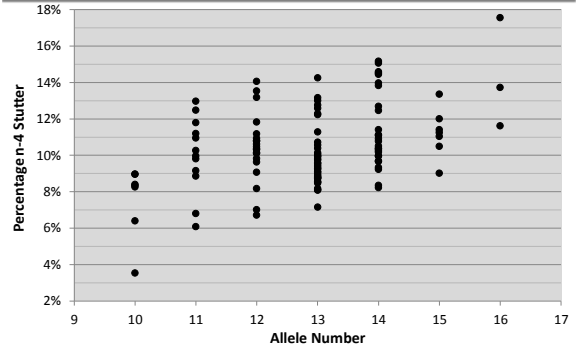
D8S1179 n-4 Stutter by Allele



D8S1179 n-4 Stutter by Allele



D8S1179 n-4 Stutter by Allele



Longest Uninterrupted Stretch (LUS)

Forensic Science International: Genetics

Characterising the STR locus D6S1043 and examination of its effect on stutter rates

Ju-Anne Bright^{1,2,3*}, Kate E. Stevenson¹, Michael D. Coble¹, Carolyn R. Hill¹, James M. Curran², John S. Buckleton³

¹CSIRO, Private Bag 4800, Sturtville QLD 4057, Australia
²Department of Forensic Science, Victoria University of Wellington, Private Bag 600, Wellington 6140, New Zealand
³CSI National Institute of Standards and Technology, Gaithersburg, MD, USA

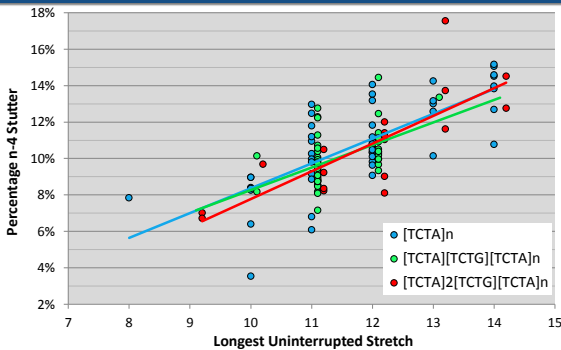
D8S1179 Allele	Repeat Structure
14	[TCTA]14
14	[TCTA][TCFG][TCTA]12
14	[TCTA][TCTA][TCFG][TCTA]11

D8S1179 Sequence Motifs

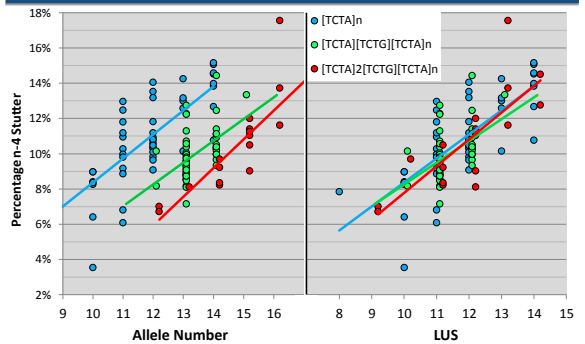
LONGEST UNINTERRUPTED STRETCH

Allele	Repeat Structure
10	[TCTA]10-14
10	[TCTA]10
11	[TCTA]11
12	[TCTA]12
13	[TCTA]13
14	[TCTA]14
12	[TCTA][TCFG][TCTA]10-14
12	[TCTA][TCFG][TCTA]10
13	[TCTA][TCFG][TCTA]11
14	[TCTA][TCFG][TCTA]12
16	[TCTA][TCFG][TCTA]14
16	[TCTA][TCTA][TCFG][TCTA]8-15
11	[TCTA][TCTA][TCFG][TCTA]9
12	[TCTA][TCTA][TCFG][TCTA]9
13	[TCTA][TCTA][TCFG][TCTA]10
14	[TCTA][TCTA][TCFG][TCTA]11
15	[TCTA][TCTA][TCFG][TCTA]12
16	[TCTA][TCTA][TCFG][TCTA]13
17	[TCTA][TCTA][TCFG][TCTA]14
18	[TCTA][TCTA][TCFG][TCTA]15

D8S1179 n-4 Stutter by LUS



D8S1179 n-4 Stutter by Allele vs LUS

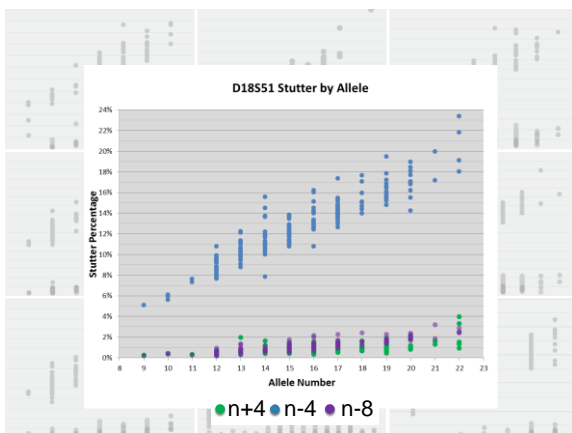
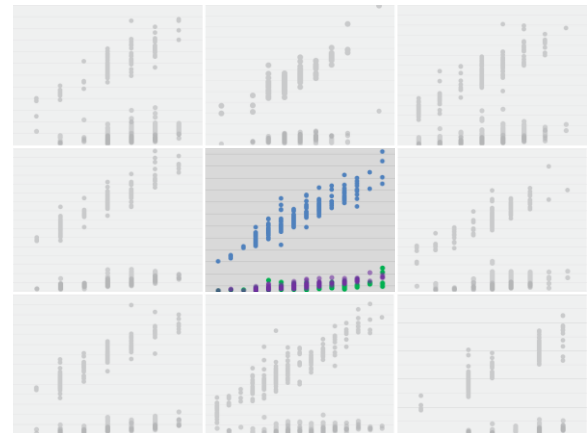
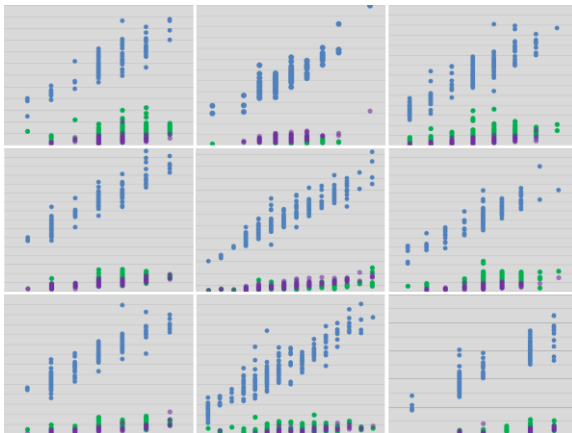
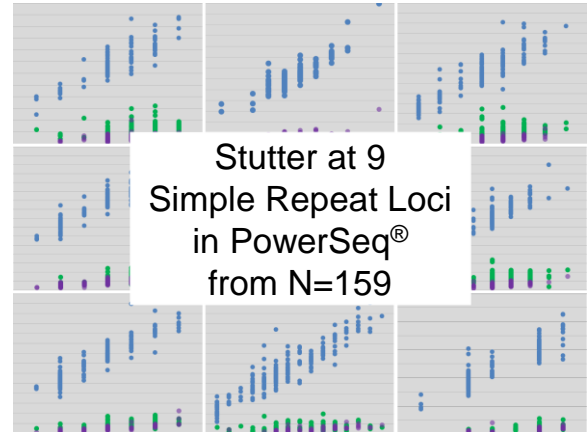


Next Generation Sequencing Data

- Samples
 - N = 159 (Caucasian, African American, and Hispanic individuals)
- Assay and Instrument
 - Promega PowerSeq® Auto
 - Illumina MiSeq
- Loci Amplified
 - AMEL, D1S16S56, D2S1338, D2S441, D3S1359, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, CSF1PO, FGA, Penta D, Penta E, TH01, TPOX, vWA, DYS391
- Analysis Parameters
 - FASTQ File -> STRait Razor Perl based software -> NGS Stutter Filter
 - 10X minimum coverage



C.F. K.B. Gettings et al. / Forensic Science International: Genetics 21 (2016) 15–21



Compound and Complex Loci

1. Allele Length

- D22S1045

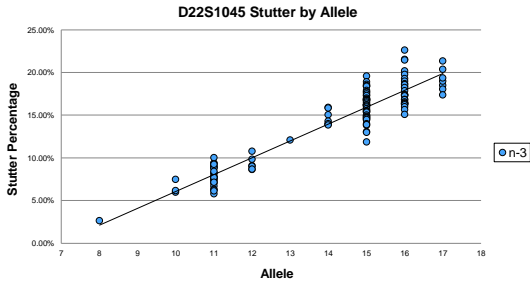
2. Longest Uninterrupted Stretch

- D19S443, D3S1358, D8S1179

3. Unique to Sequence Variants

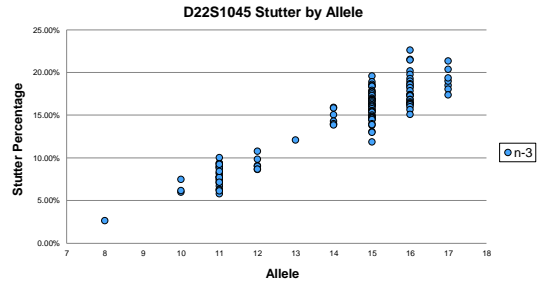
- D1S1656, D12S391, D21S11, D2S1338, D2S441, FGA, vWA

Stutter by Allele Length

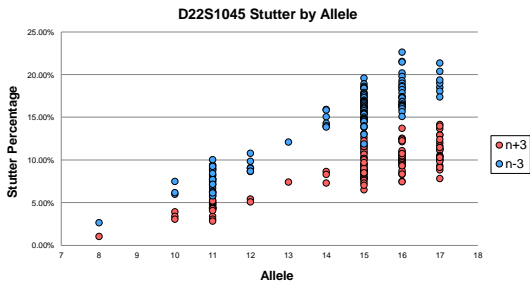


D22S1045 sequence motif: $[ATT]_n[ACT][ATT]_2$

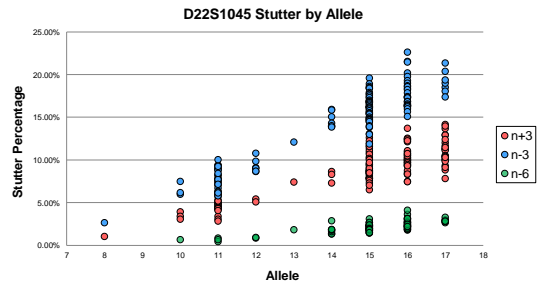
Stutter by Allele Length



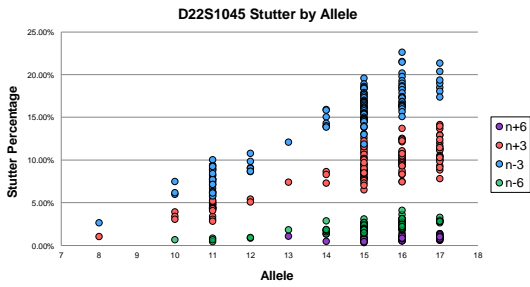
Stutter by Allele Length



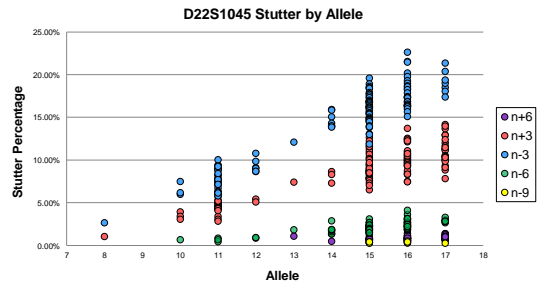
Stutter by Allele Length



Stutter by Allele Length



Stutter by Allele Length



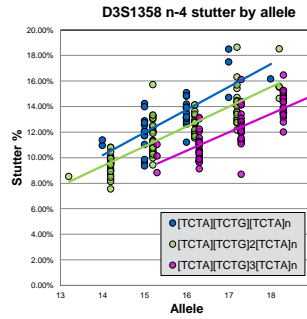
D3S1358 Sequence Motifs

$[TCTA][TCTG][TCTA]_n$

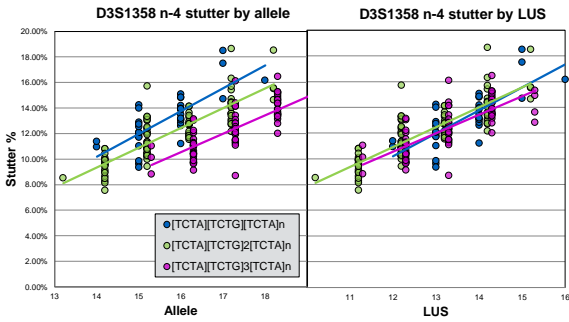
$[TCTA][TCTG]_2[TCTA]_n$

$[TCTA][TCTG]_3[TCTA]_n$

Stutter by LUS



Stutter by LUS



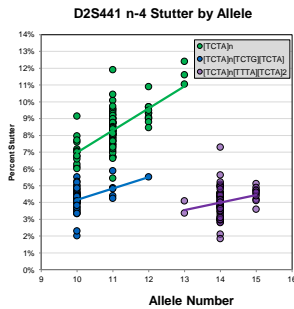
D2S441 Sequence Motifs

$[TCTA]_n$

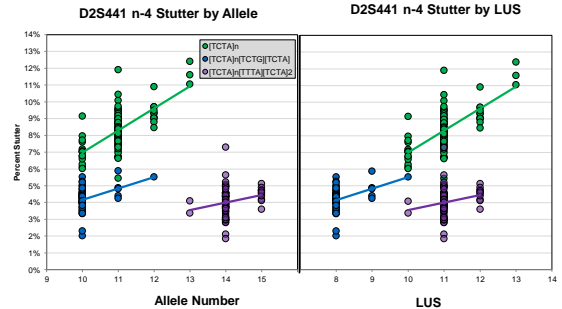
$[TCTA]_n[TCTG][TCTA]$

$[TCTA]_n[TTTA][TCTA]_2$

Stutter Unique to Sequence Variants



Stutter Unique to Sequence Variants

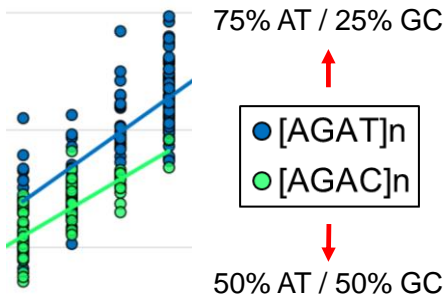
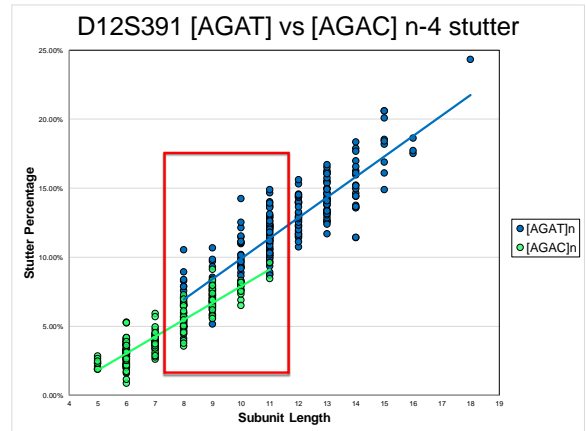
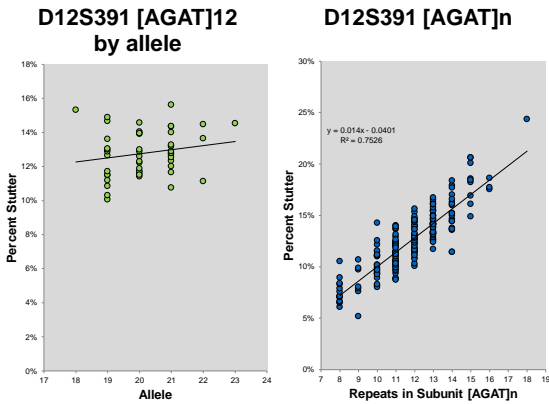
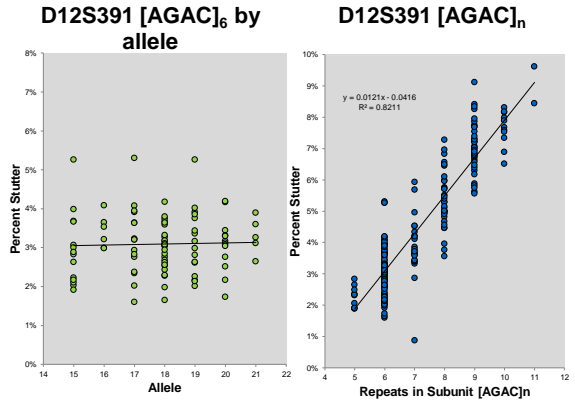


D12S391 Sequence motifs

[AGAT]_n[AGAC]_n

[AGAT]_n[AGAC]_n[AGAT]

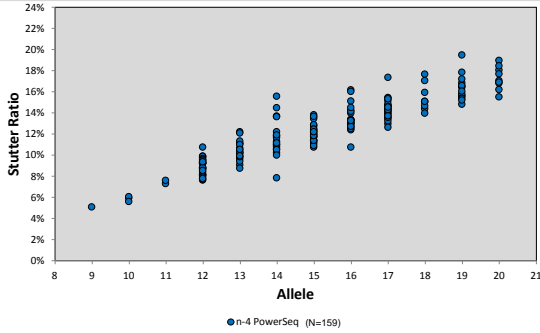
[AGAT] GAT [AGAT]_n[AGAC]_n[AGAT]



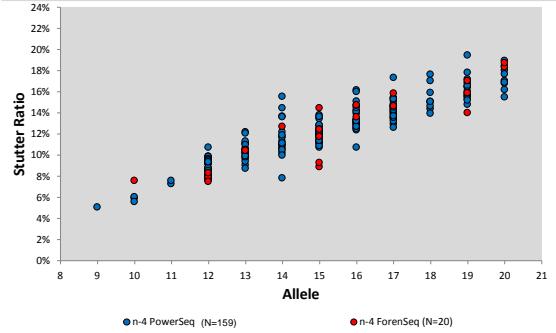
Stutter Comparison Across Assays and Platforms

Assay	Method	Amplification Cycles	Input DNA (ng)	Samples Analyzed	Data Analysis Threshold
Promega PowerSeq® Auto	NGS on Illumina MiSeq	30	0.5	159	10X Coverage
Illumina® ForenSeq DNA Signature Prep Kit	NGS on Illumina MiSeq FGX	18	1	20	10X Coverage
GlobalFiler™	CE on ABI 3500xL	29	1	20	10 RFU
Promega PowerPlex® Fusion	CE on ABI 3500xL	30	0.5	20	10 RFU

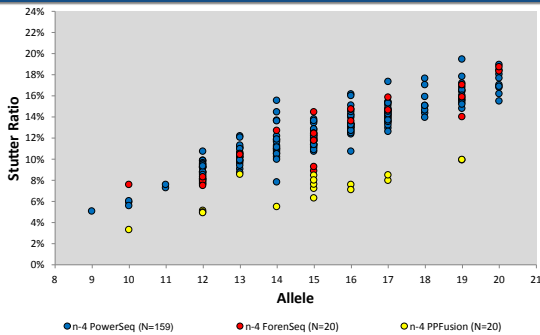
D18S51 n-4 Stutter by Kit



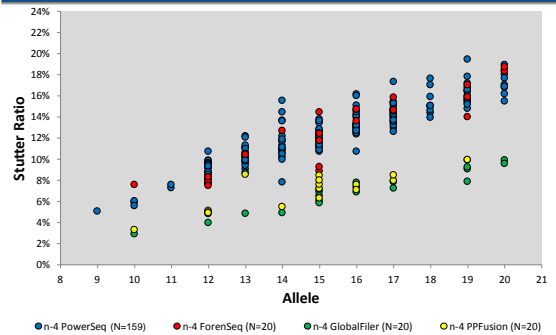
D18S51 n-4 Stutter by Kit



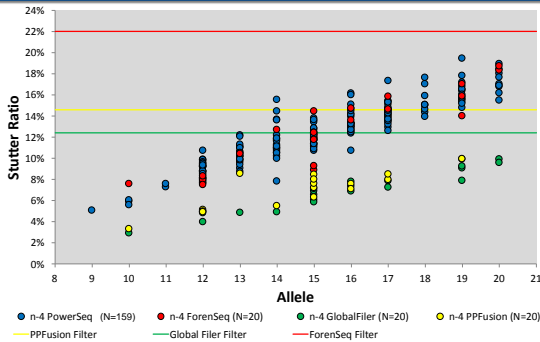
D18S51 n-4 Stutter by Kit



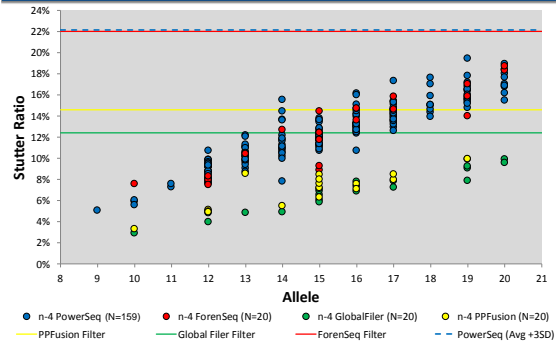
D18S51 n-4 Stutter by Kit



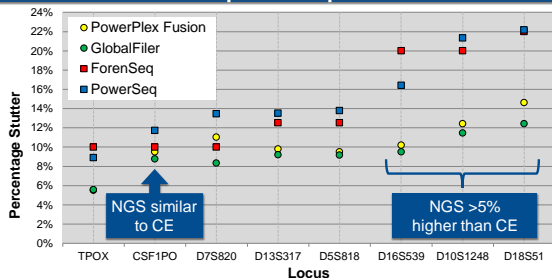
D18S51 n-4 Stutter by Kit



D18S51 n-4 Stutter by Kit



n-4 Stutter Rates by Assay at 8 Simple Repeat Loci



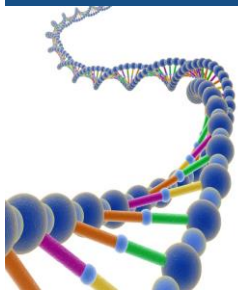
Stutter rates obtained from:

PowerPlex Fusion: K. Oestdik et al. / Forensic Science International: Genetics 12 (2014) 69-76
 GlobalFiler: ThermoFisher Scientific GlobalFiler™ PCR Amplification Kit User Guide
 ForenSeq: ForenSeq™ Universal Analysis Software Guide, Part # 15053876 Rev. C, June 2015
 PowerSeq: Analysis of NIST population samples (N=159), Average + 3 Standard Deviations

Conclusions

- Conclusions
 - NGS technology benefits characterization of stutter artifacts
 - Development of analysis software for NGS data with the ability to include sequence specific stutter thresholds
 - Incorporate allele based thresholds in CE analysis software
- Considerations
 - Not every population group is included
 - Kit specificity

Acknowledgements



NIST
 Katherine Gettings
 Pete Vallone
 David Duerer
 Lisa Borsuk
 Mike Coble
 Kevin Kiesler
 Becky Steffen

George Washington University
 Daniele Podini

NIST Disclaimer: Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

Contact Information
 raponte@gwmail.gwu.edu