


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC



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

Sequence-based Analysis of Stutter at STR Loci: Implementation and Utilization

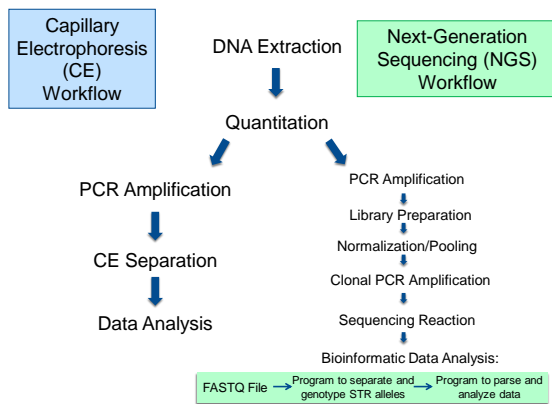
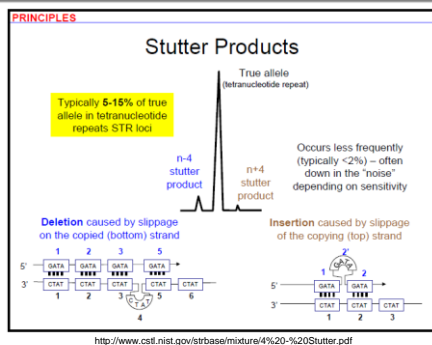
R.A. Aponte¹, K.B. Gettings², D.L. Duwer², C.R. Steffen², M.D. Coble², and P. M. Vallone²

¹ Department of Forensic Sciences, The George Washington University, Washington, DC, 20007-1150
² National Institute of Standards and Technology, Gaithersburg, MD, 20899-8314

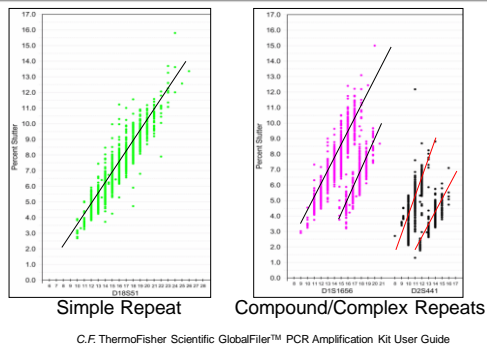
Why Analyze Stutter?

- Validation to aid in sample analysis and interpretation
 - Allows software to remove stutter, improving efficiency in data analysis
 - Distinguishing minor contributors from artifact in mixtures
- Improved models of stutter characterization by understanding what influences stutter formation
 - STR sequence, allele length, flanking region, and other factors
 - Amount of DNA

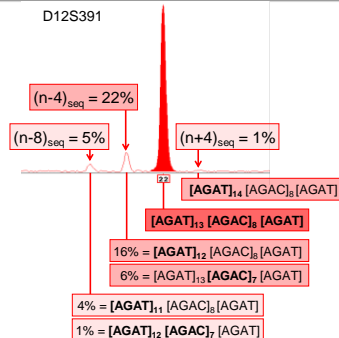
Stutter Product Formation



Trends in CE Data



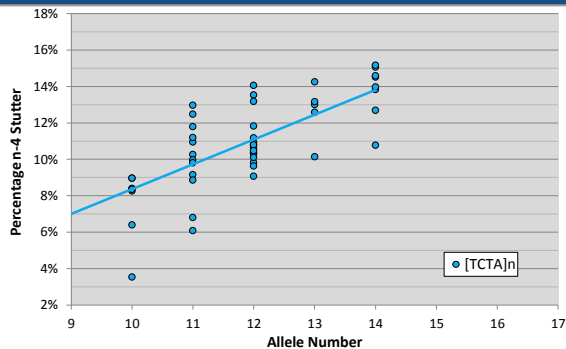
Stutter by Sequence



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

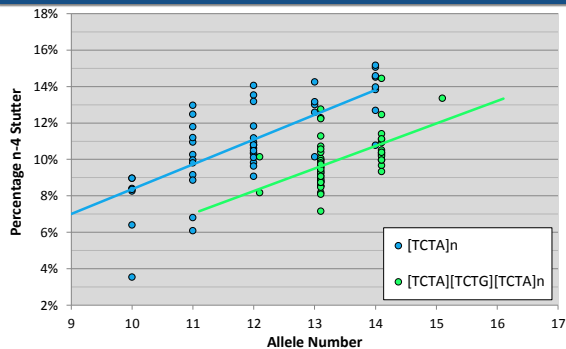
D8S1179 n-4 Stutter by Allele



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

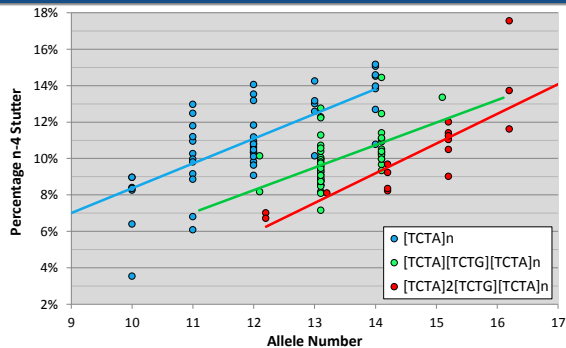
D8S1179 n-4 Stutter by Allele



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



Longest Uninterrupted Stretch (LUS)



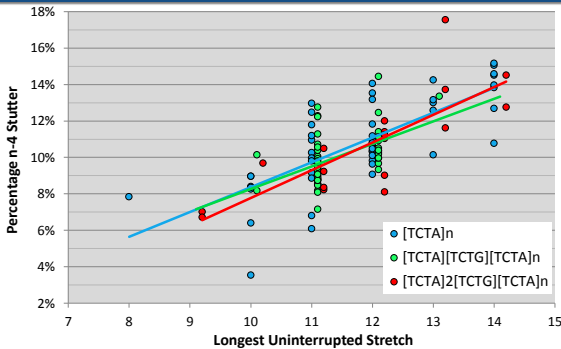
D8S1179	Repeat Structure
14	[TCTA]14
14	[TCTA] [TC TG] [TCTA]12
14	[TCTA] [TCTA] [TC TG] [TCTA]11

D8S1179 Sequence Motifs

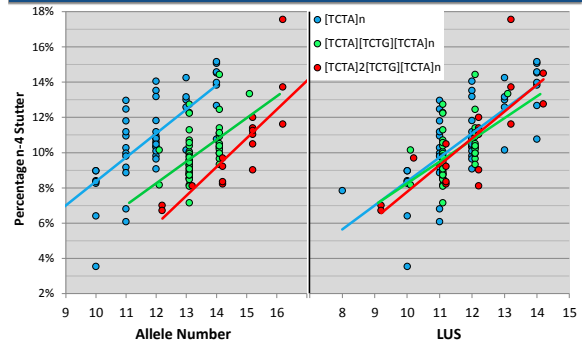
LONGEST UNINTERRUPTED STRETCH

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

D8S1179 n-4 Stutter by LUS

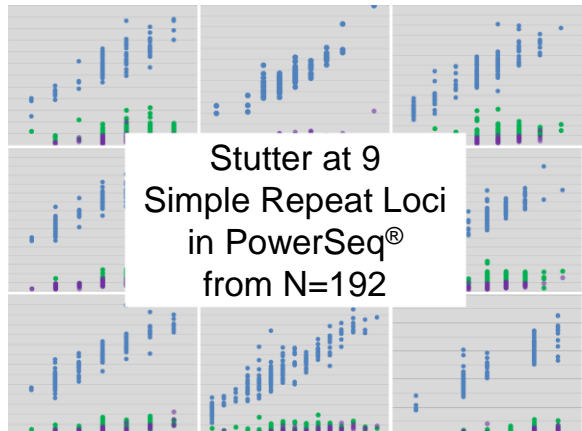


D8S1179 n-4 Stutter by Allele vs LUS

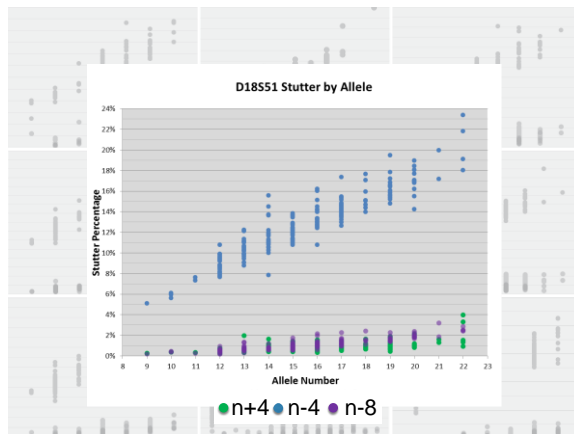
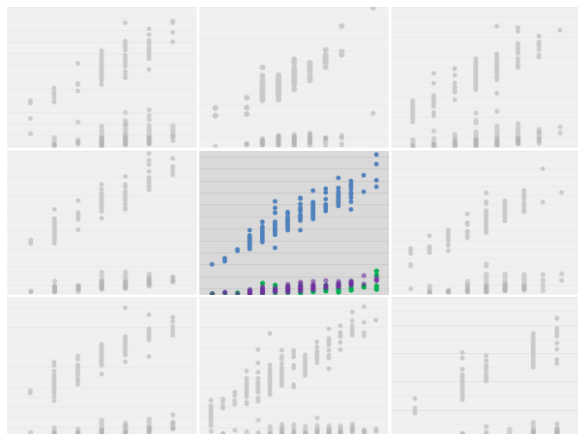


Next Generation Sequencing Data

- Samples
 - N = 192 (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



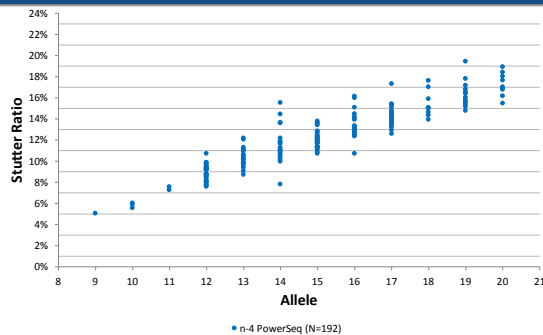
C.F. K.B. Gettings et al. / Forensic Science International: Genetics 21 (2016) 15–21



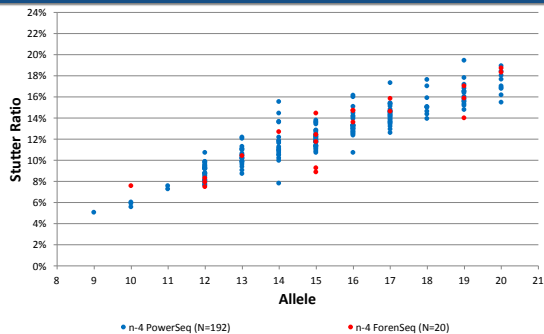
Stutter Comparison Across Kits and Platforms

Assay	Method	Amplification Cycles	Input DNA (ng)	Samples Analyzed	Data Analysis Threshold
Promega PowerSeq® Auto	NGS on Illumina MiSeq	30	0.5	192	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

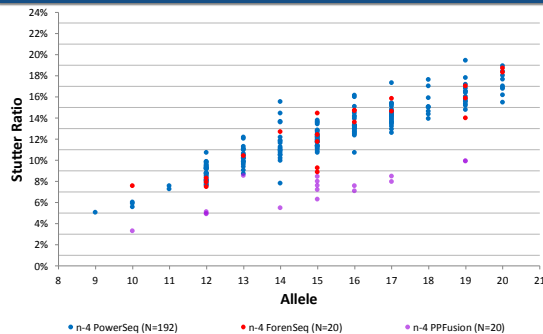
D18S51 n-4 Stutter by Kit



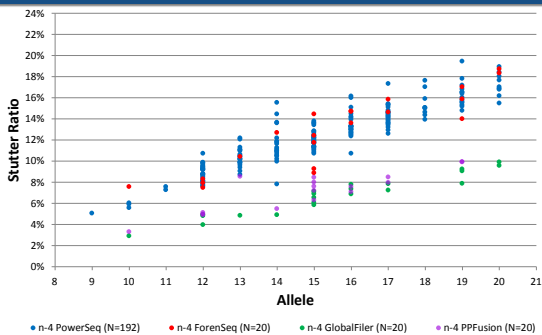
D18S51 n-4 Stutter by Kit



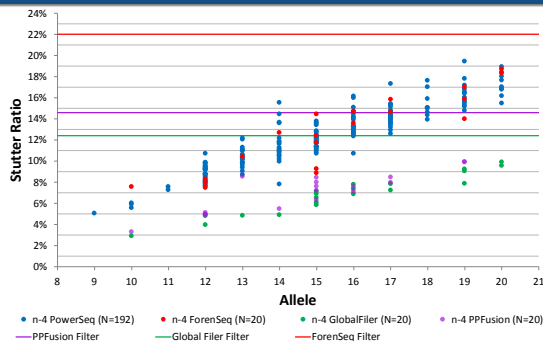
D18S51 n-4 Stutter by Kit



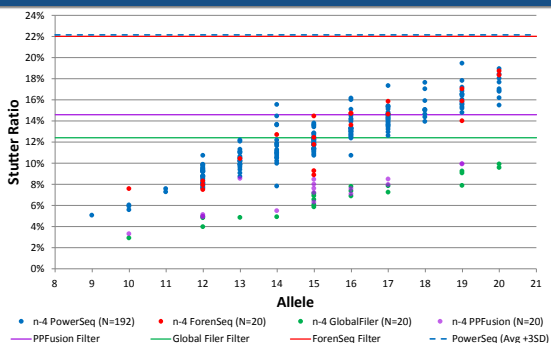
D18S51 n-4 Stutter by Kit



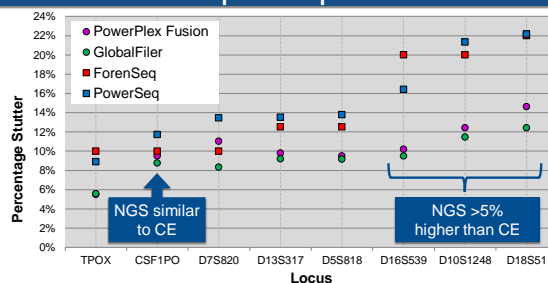
D18S51 n-4 Stutter by Kit



D18S51 n-4 Stutter by Kit



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

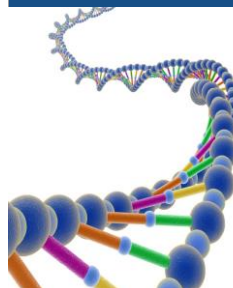


Stutter rates obtained from:
 PowerPlex Fusion: K. Oostdik et al. / Forensic Science International: Genetics 12 (2014) 69-76
 GlobalFiler: ThermoFisher Scientific GlobalFiler™ PCR Amplification Kit User Guide
 ForenSeq: ForenSeq™ Universal Analysis Software Guide, Part # 15053876 Rev. C, June 2015
 PowerSeq: Analysis of NIST population samples (N=192), Average + 3 Standard Deviations

Conclusions

- Conclusions and Recommendations
 - Better differentiation between stutter artifacts and minor contributors with NGS technology
 - Development of analysis software for NGS data with the ability to include sequence specific stutter thresholds
 - Incorporate allele based thresholds in CE analysis software
- Future Directions
 - Complete the stutter analysis for all 22 loci in PowerSeq® Auto data
 - Analyze sequence motifs to improve prediction of stutter and establishment of robust stutter ratios

Acknowledgements



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

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