Email: sarah.riman@nist.gov

Impact of Sample Complexity on STR Stutter Ratios

WASHINGTON, DC
ISFG 2022
INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
P035

Sicen Liu[1], Peter M. Vallone[2], Sarah Riman[2]

[1]Johns Hopkins University Whiting School of Engineering, Department of Computer Science, Baltimore, MD 21218, USA
[2]National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA

Introduction

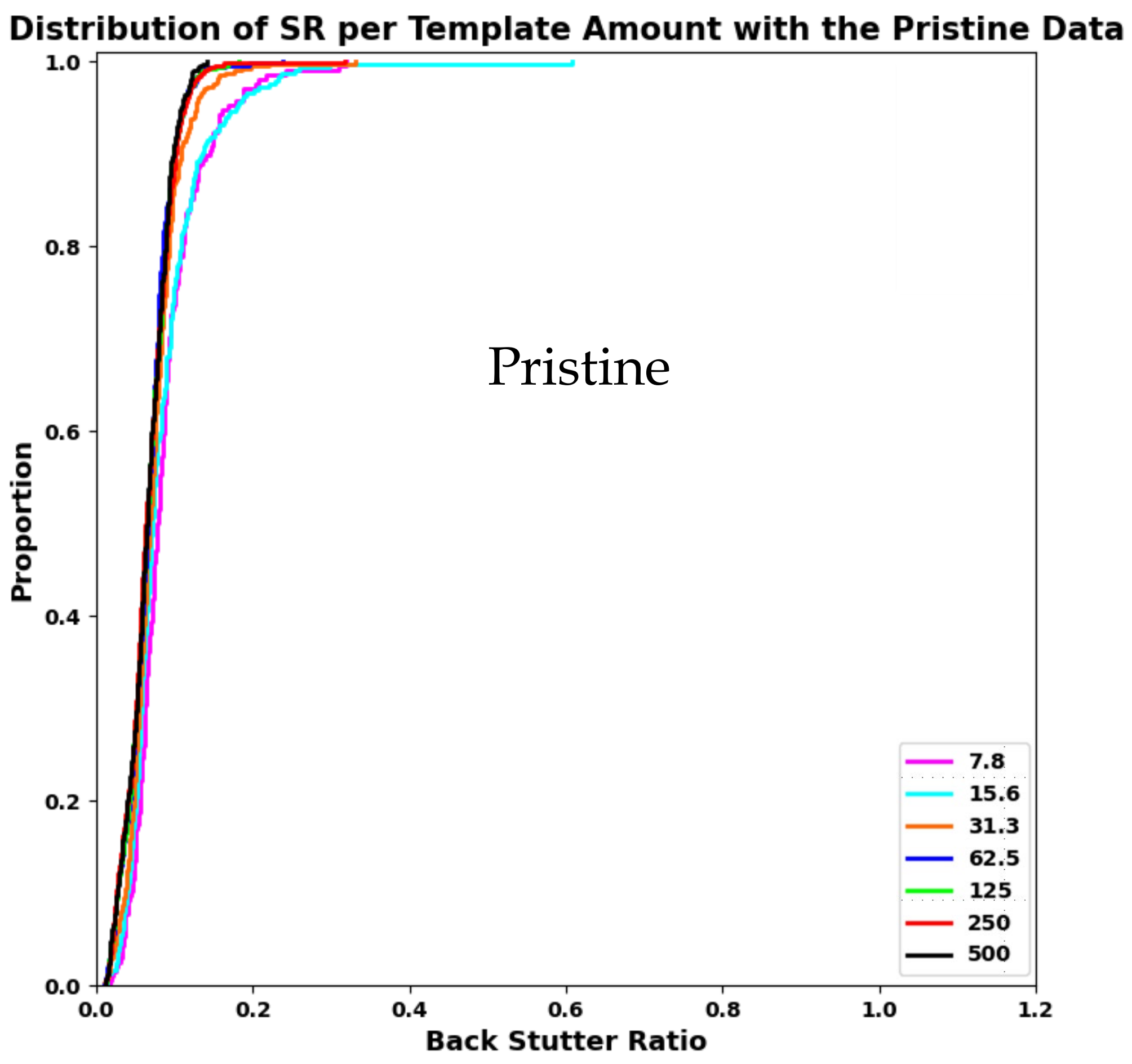
Stutter products are PCR artifacts generated due to strand slippage during the amplification of STR loci resulting in deletion(s) or insertion of base pairs on the newly synthesized DNA strand. Different stutter variants are commonly observed: back stutter, double-back stutter, forward stutter, and half back stutter. In this study we explore the impact of sample complexity (different DNA template amounts and treatments used to compromise DNA) on the distribution of allelic peak heights, stutter peak heights, and stutter ratios (SR).

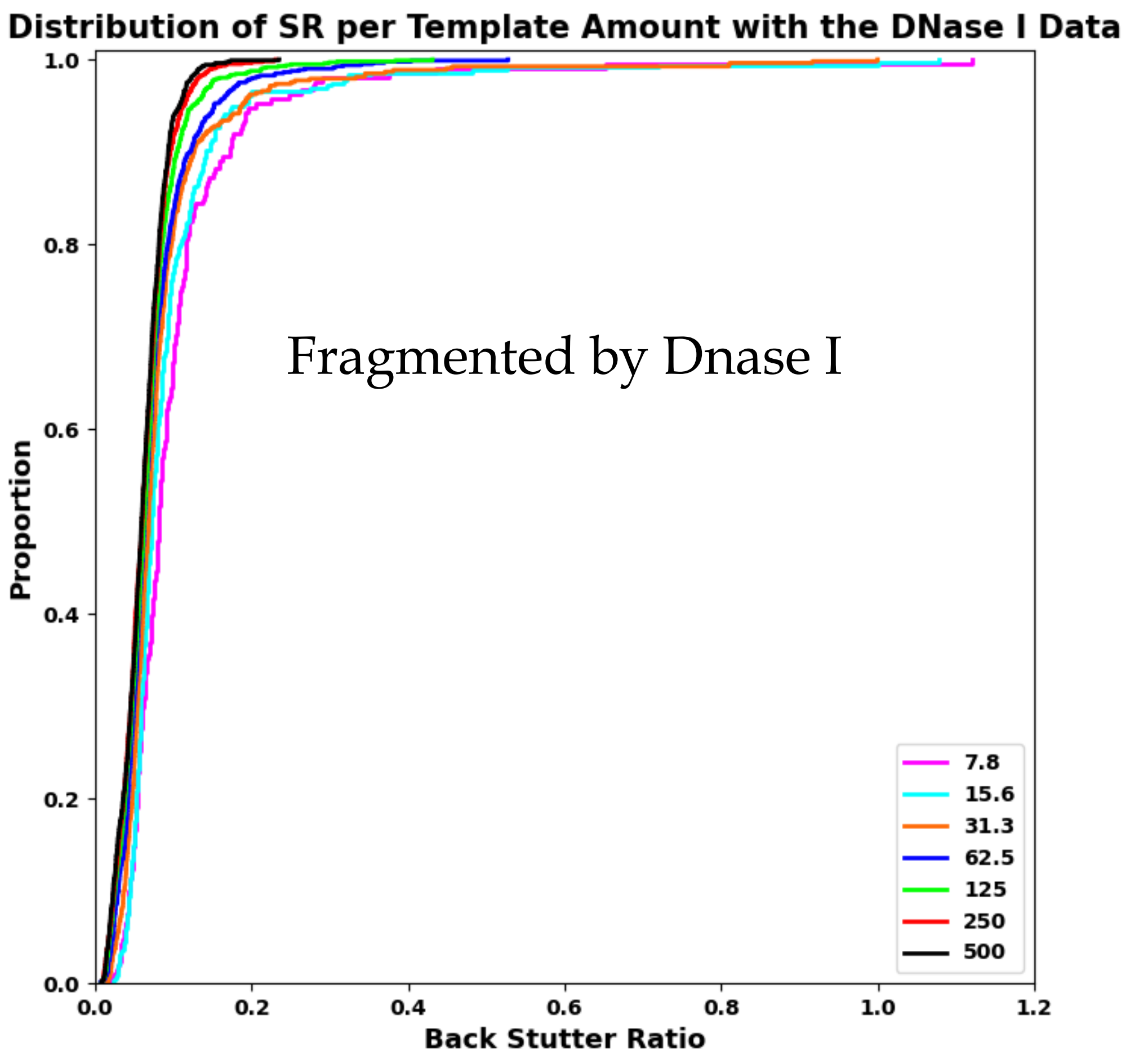
Methods

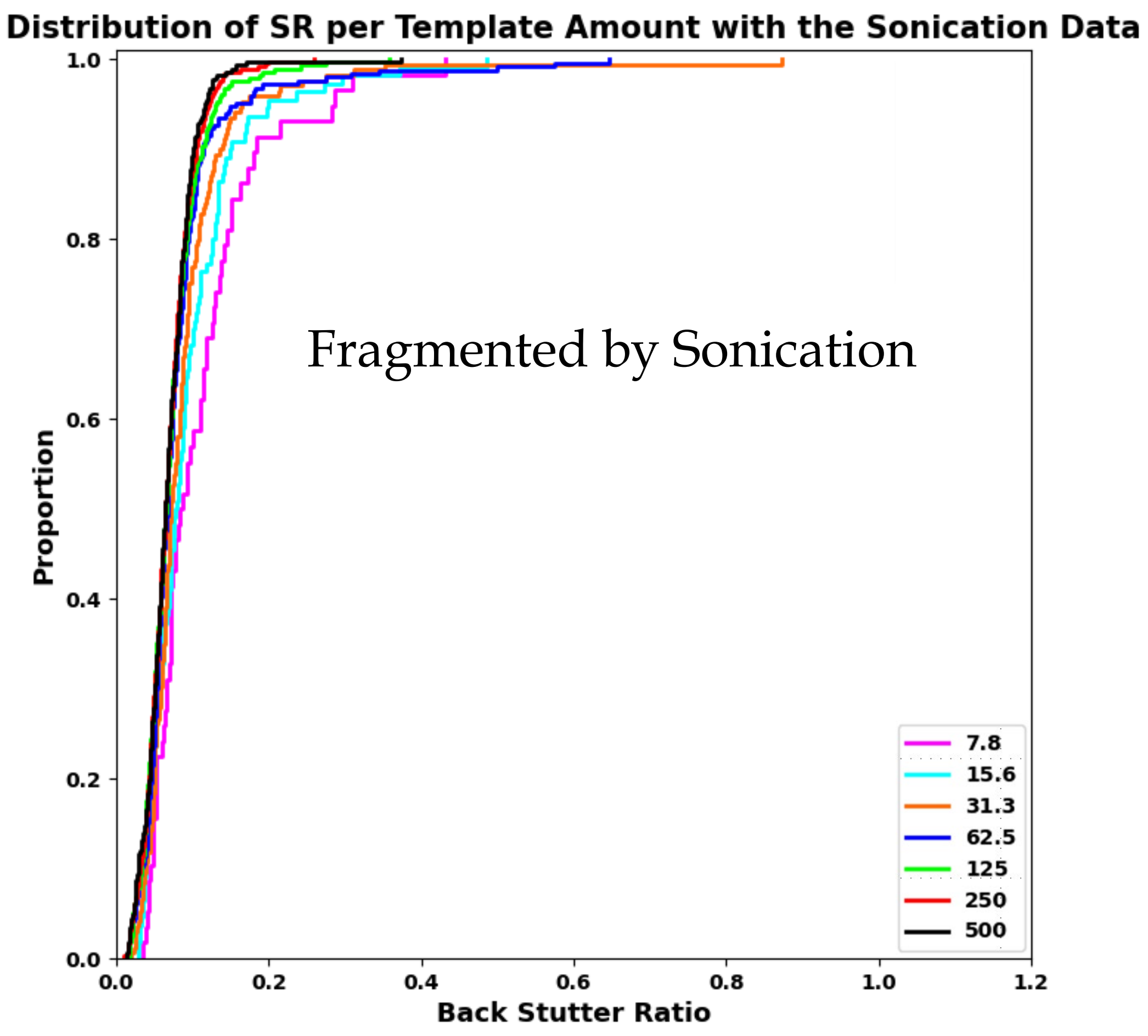
Single-source samples with varying DNA quality and DNA quantity were selected from the publicly available PROVEDIt database (GlobalFiler 29 cycles 25s). Raw HID files were analyzed in FaSTR DNA software. Based on the known ground truth, each observed peak in the exported CSV files was categorized as either noise, allele, or stutter. Allelic and stutter calls were then extracted with their associated heights and SR were calculated (for the 22 autosomal STR loci) after applying a global analytical threshold of 10 RFU. Only data points attributed to non-overlapping stutter-stutter or stutter-allele signals were considered and parsed in python.

Results

1. Impact of different template amounts on the distribution of back stutter (N-1) ratios for three DNA treatments



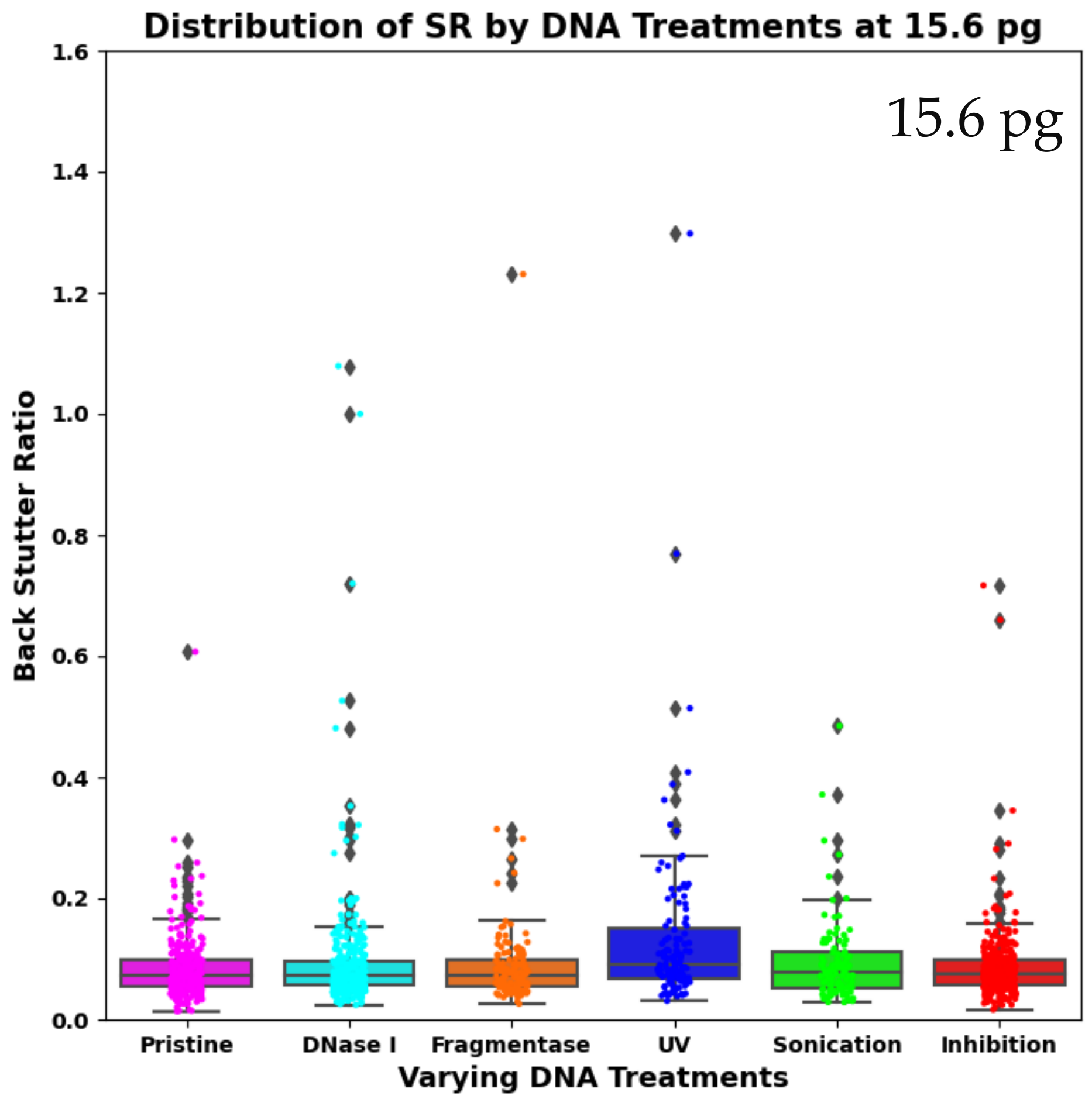


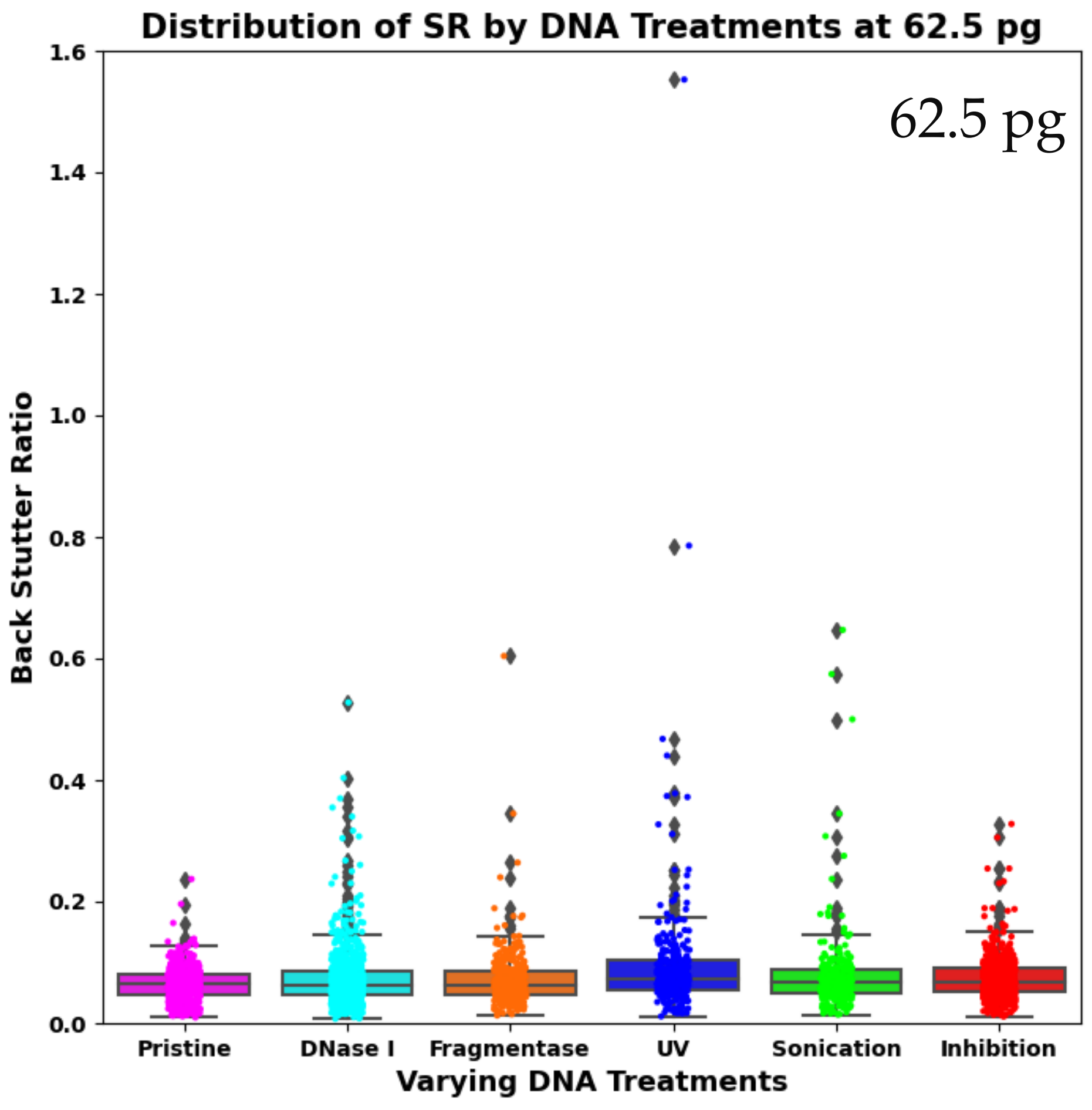


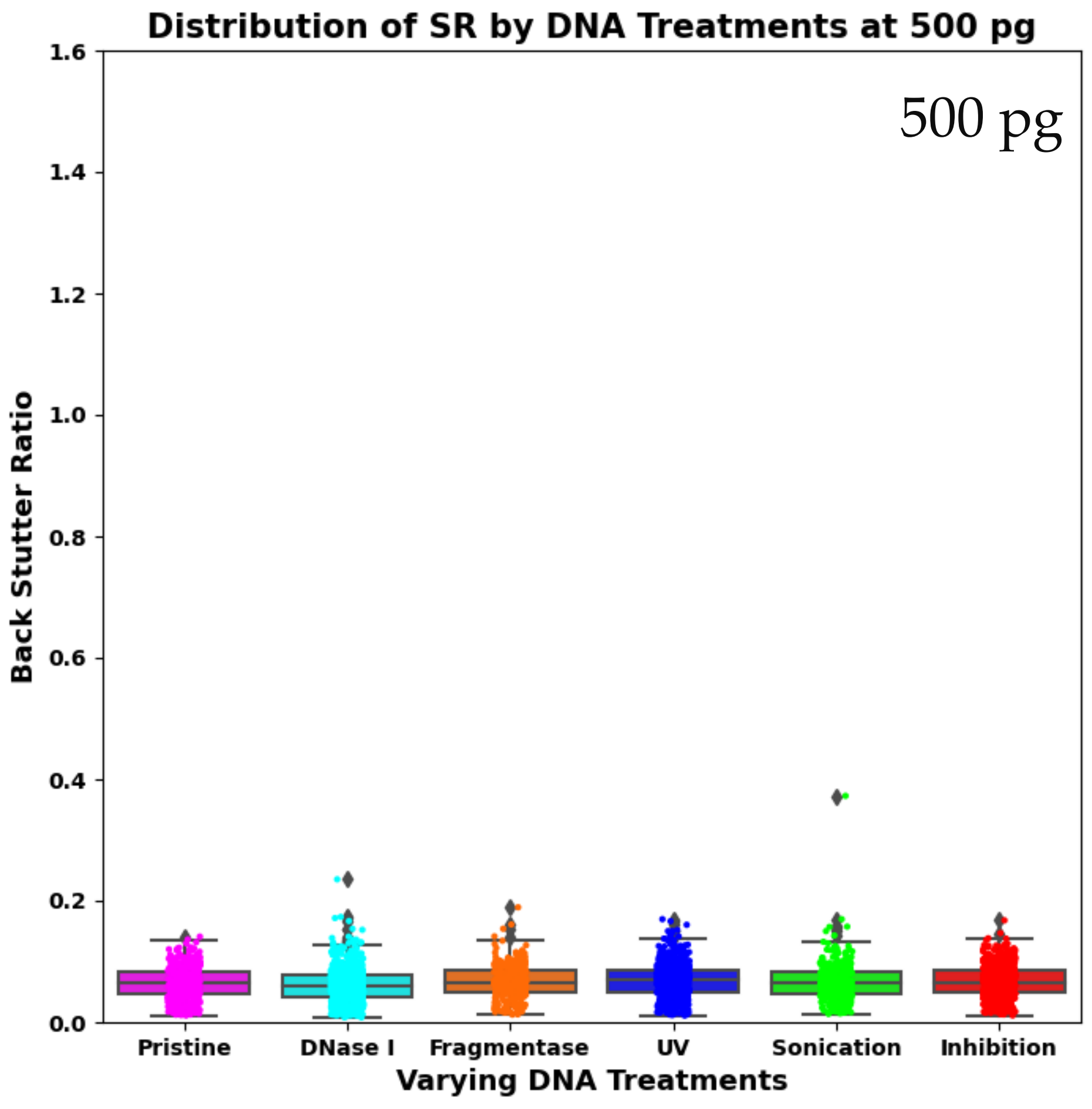
Comparison Group	***P-values***		
	Pristine	DNase I	Sonication
7.8 pg vs 500 pg	7.343E-06*	1.665E-15*	9.597E-06*
15.6 pg vs 500 pg	3.363E-06*	1.125E-10*	0.000683*
31.3 pg vs 500 pg	0.125	1.156E-06*	0.0089
62.5 pg vs 500 pg	0.569	5.274E-05*	0.355
125 pg vs 500 pg	0.982	0.044	0.851
250 pg vs 500 pg	0.928	0.864	0.998

- Based on the two-sample Kolmogorov-Smirnov (KS) test and depending on the DNA treatment dataset, the empirical cumulative distribution functions (ECDFs) of the low template amounts (7.8 pg, 15.6 pg, 31.3 pg, and 62.5 pg) are non-identical and statistically significant than the ECDFs of higher template amounts. The p-values were adjusted by applying a Bonferroni correction for multiple comparisons (α′ = 0.05/34=~0.0015).
- *:p-value <0.0015

2. Impact of DNA quality on the distribution of back stutter (N-1) ratios at three DNA template amounts







- Per each template amount, SR appear to be more variable with compromised/degraded DNA vs pristine DNA.
- As expected, more variations are observed with SR at lower template amounts versus higher ones (e.g., 15.6 pg and 62.5 pg vs 500 pg).

Future work

- Study the effect of template amounts and DNA treatments on the ratios of other stutter types.
- Understand the impact of sample complexity per each locus on a per allele basis.
- Understand the impact of different distributions/variations of stutter ratios on the (1) stutter filters set to assist in assigning NOCs and (2) models for stutter peaks.

Acknowledgment

The authors would like to thank Hari Iyer (Statistical Engineering Division, ITL, NIST) for meaningful discussions on the project and feedback on the statistical analysis.

Funding: This work was in part supported by the NIST Special Programs Office: Forensic Genetics.

Disclaimer: Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. 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 endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose. All work presented has been reviewed and approved by the NIST Human Subjects Protections Office. All work has been reviewed and approved by the National Institute of Standards and Technology Research Protections Office. This study was determined to be "not human subjects research" (often referred to as research not involving human subjects) as defined in U. S. Department of Commerce Regulations, 15 CFR 27, also known as the Common Rule (45 CFR 46, Subpart A), for the Protection of Human Subjects by the NIST Human Research Protections Office and therefore not subject to oversight by the NIST Institutional Review Board.