INVESTIGATING THE EFFECTS OF DIFFERENT LIBRARY PREPARATION PROTOCOLS ON STR SEQUENCING

Sarah Riman[1], H. Iyer[2], L. Borsuk[1], S. Lund[2], K. Gettings[1], P.M. Vallone[1]

[1] U.S. National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8314, USA

[2] U.S. National Institute of Standards and Technology, Statistical Design, Analysis, and Modeling Group, Gaithersburg, MD 20899-8314, USA

National Institute of Standards and Technology Technology Administration, U.S. Department of Commerce

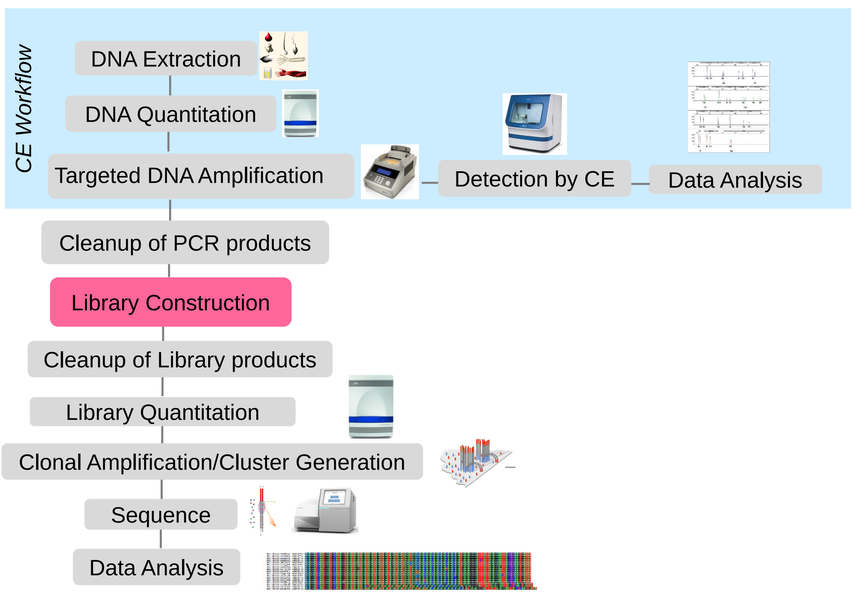
Email: sarah.riman@nist.gov

P01-30

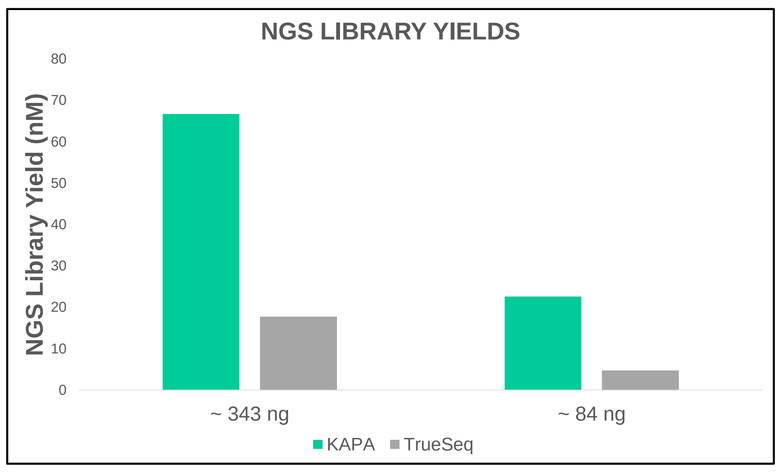
Next Generation Sequencing (NGS) is transforming the landscape of the Short Tandem Repeat (STR) genotyping. NGS determines the length as well as the sequence of each allele, identifies polymorphisms in the repeats or adjacent DNA regions, allows for a greater degree of multiplexing, and generates Gigabases of reads in a single run [1-3]. An additional step introduced into the DNA typing process with NGS methods is library preparation. The PCR amplicons of the targeted loci are further purified and modified with adapters and sample specific indices prior to sequencing.

The influence of different library preparation protocols and reagents on the consistency of allele calling, depth of coverage, and allele coverage ratio was examined. In this study, the general workflow consisted of amplifying STR loci by PCR using the Prototype PowerSeq Auto System; amplicon purification; library construction and cleanup; cluster generation on the MiSeq platform; and post-sequencing data processing. Two commercially library kits were used to perform library construction: Illumina TruSeq DNA HT Sample Prep and Kapa Hyper Prep. The selection of various methods for amplicon purification and library cleanup (e.g. columns vs. different concentration of beads) were assessed. The goal was to optimize library construction protocols to reduce bias in allele signal and allow for fast and efficient library preparation. These optimizations could lead to high-quality STR data by generating maximum number of sequencing reads and ameliorating the likelihood of alleles/loci from being poorly represented, especially from mixed forensic samples and those with low amount of DNA material.

General Workflow for NGS and CE



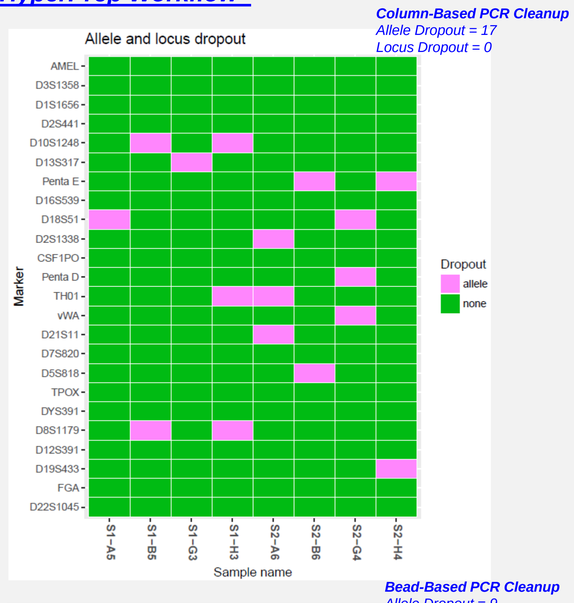
Conversion Rate for Different Levels of Input DNA

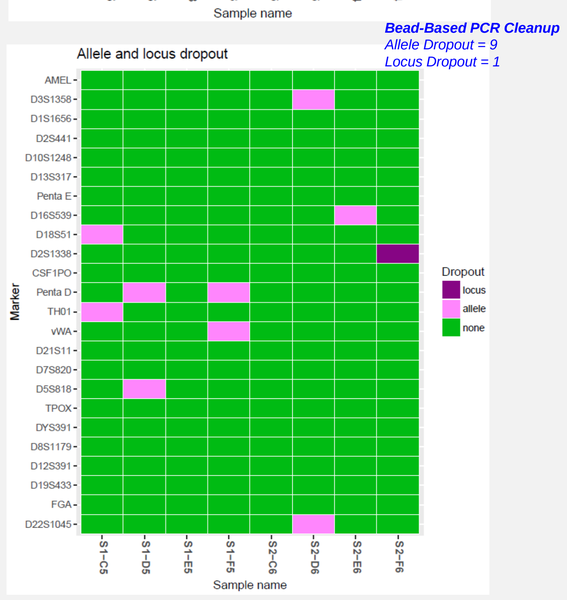


The amount of adaptor ligated STR amplicons after the ligation step was measured by qPCR-based quantification using KAPA Library Quantification Kit. The percentage of library input DNA converted to adapter ligated libraries was greater with the KAPA HyperPrep workflow.

Allele and Loci Drop-Out Events with 30 pg of DNA Input

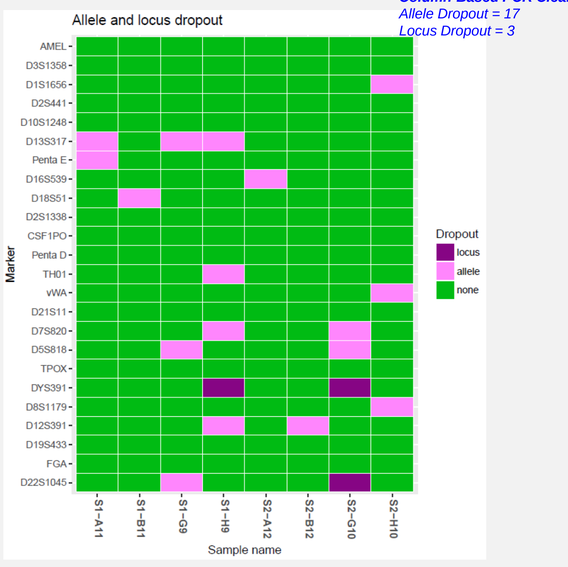
KAPA HyperPrep Workflow

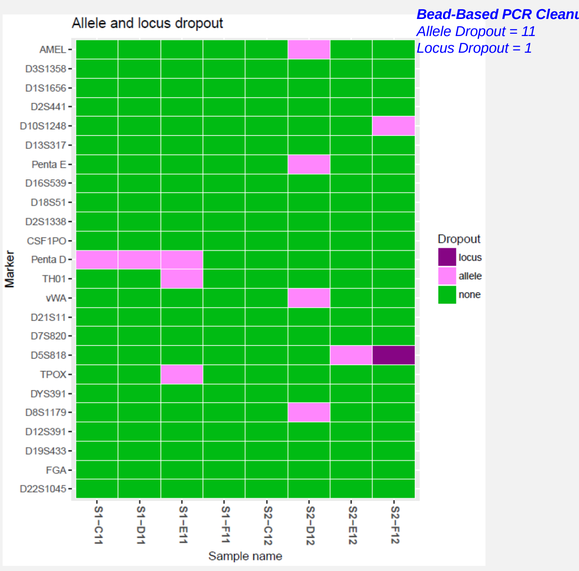




Bead-Based PCR cleanup had the lowest dropout events with KAPA HyperPrep workflow.

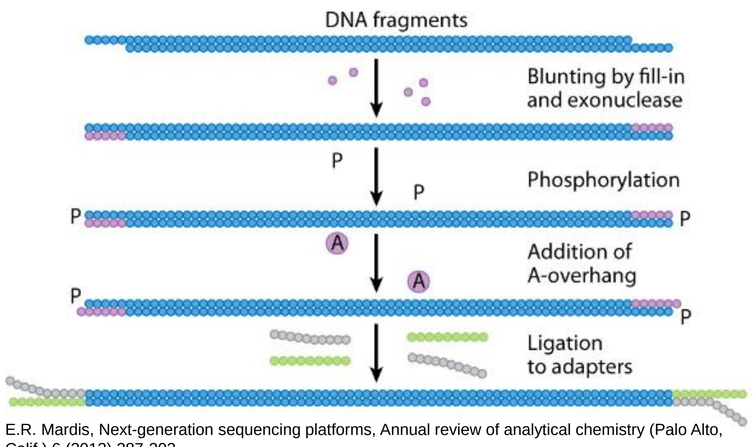
TruSeq Workflow





Bead-Based PCR cleanup had also the lowest dropout events with TruSeq workflow.

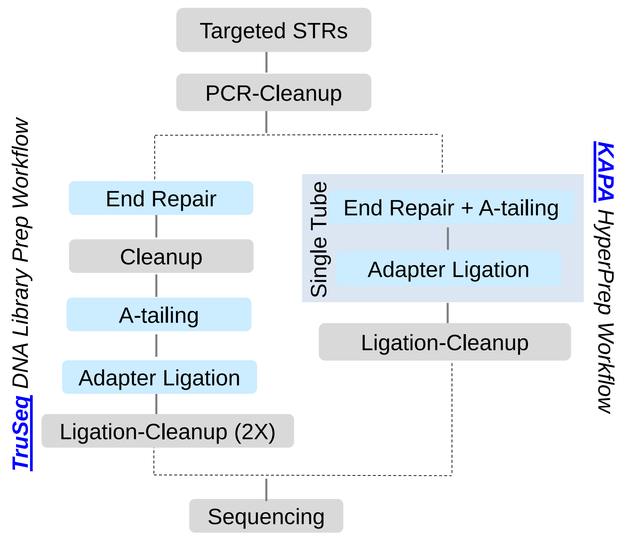
Ligation-Based Library Preparation Workflow



E.R. Mardis, Next-generation sequencing platforms, Annual review of analytical chemistry (Palo Alto, Calif.) 6 (2013) 287-303.

The aim of sequencing library preparation is to flank the amplified STR amplicons with adapters. Adapters contain important information for cluster amplification, sequencing, and multiplexing. The three core enzymatic steps in preparing ligation-based libraries for Illumina sequencing are: (i) end repair; (ii) A-tailing; and (iii) adapter ligation.

Different Ligation-Based Library Preparation Kits

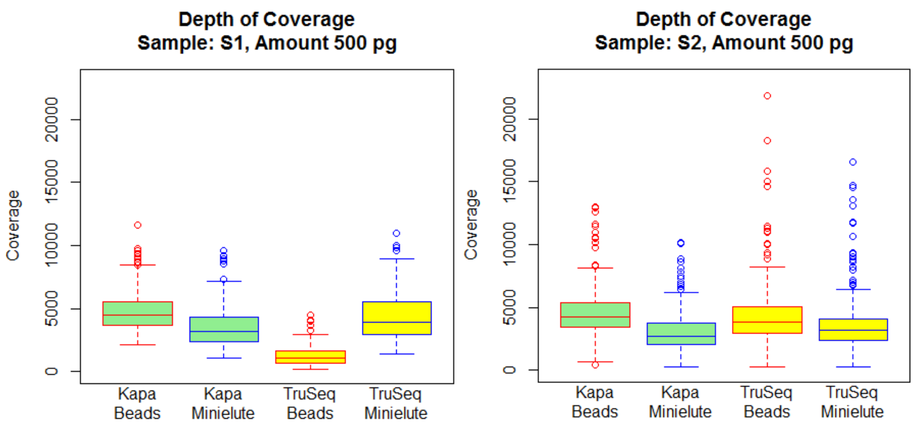


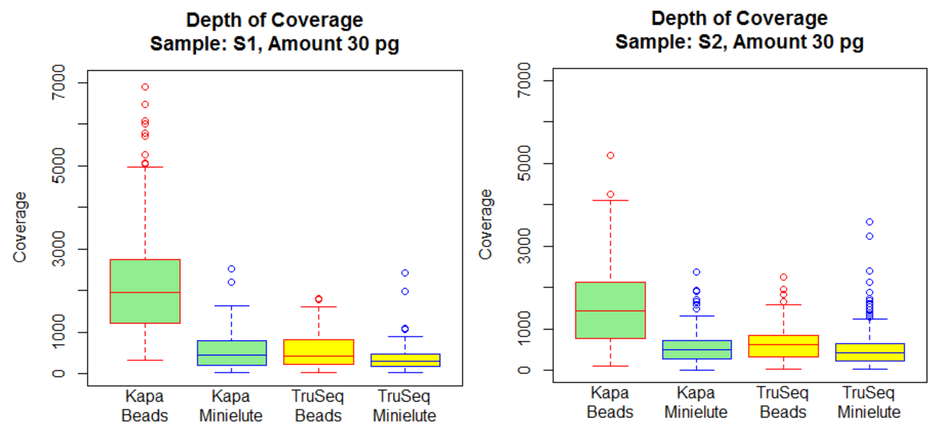
The two library preparations kits used in this study, KAPA HyperPrep and TruSeq DNA PCR-Free library preparations kits vary in their reagent supplier and in their protocols. KAPA HyperPrep workflow combines the two enzymatic steps (end repair and A-tailing) into a single one and removes the clean-up step in between, reducing library preparation time and improving DNA recovery particularly for low amount of DNA input.

Material and Methods

- STR loci were amplified from two purified DNA samples, (S1) and (S2), with the prototype PowerSeq Auto System (22 autosomal STR loci + AMEL + DYS391) according to the manufacturer's protocol.
- All amplification reactions were processed in duplicate and contained either 500 pg or 30 pg of genomic DNA template.
- Amplified PCR products were purified using either columns or Solid Phase Reversible Immobilization (SPRI) beads (3X).
- Purified PCR products were subjected to DNA library preparation using two kits according to the manufacturers recommended protocol: Truseq DNA PCR-free from Illumina or KAPA Hyper from KAPA Biosystems.
- During post-ligation clean-up step, different bead:DNA ratios, (0.7X; 0.8X; 1.1X), were tested.
- 96 libraries were pooled and sequenced on an Illumina Miseq instrument with the V3 chemistry.
- STR sequence region were extracted from the FASTQ files using STRait Razor.
- Analysis of the data was performed in R and STR validator.

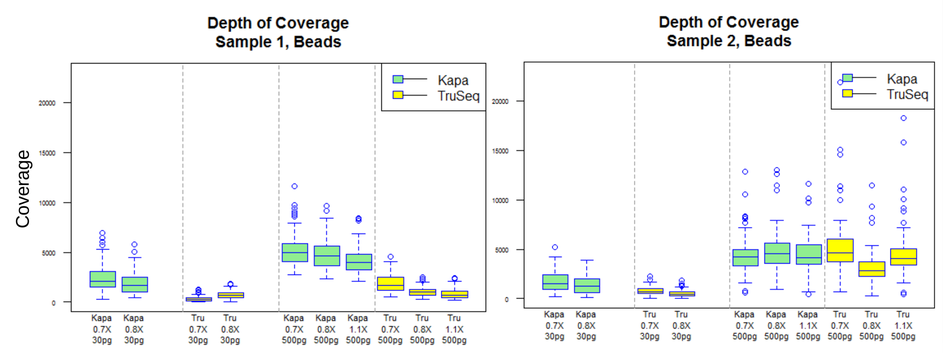
Effects of DNA-Input, Amplicon Purification, and Library Preparation kit on Depth of Coverage

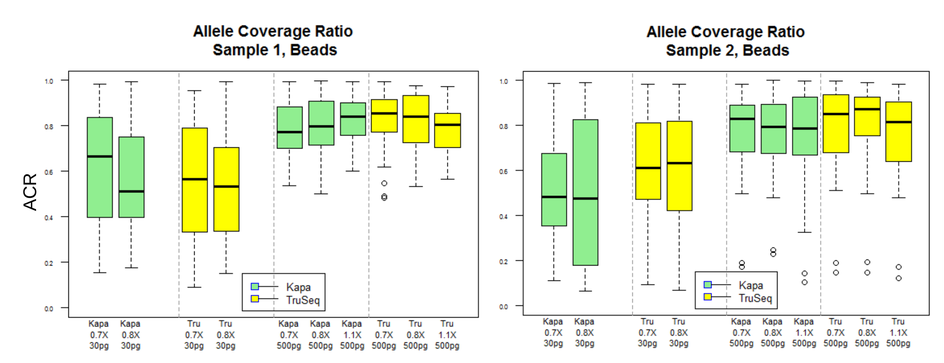




PCR reactions of either 500 pg or 30 pg of DNA amount purified by bead-based purification and subjected to library preparation through KAPA HyperPrep Workflow had the highest depth of coverage.

Effects of Different Bead Cleanup Concentrations on Depth of Coverage and Allele Coverage Ratio (ACR)





Lower SPRI:DNA ratio recovers larger amplicons while a higher ratio recovers smaller amplicons. A good depth of coverage and heterozygote balance were observed even when decreasing the ratio to 0.7X with both PCR DNA input. Lower SPRI:DNA ratio, minimizes adapter-dimer formation and preserves more space on flow cell for sequencing STR libraries.

References:

[1] C. Gelardi, E. Rockenbauer, S. Dalsgaard, C. Borsting, N. Morling, Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in Danes and a new nomenclature for sequenced STR alleles, Forensic science international. Genetics 12 (2014) 38-41.

[2] K.B. Gettings, K.M. Kiesler, S.A. Faith, E. Montano, C.H. Baker, B.A. Young, R.A. Guerrieri, P.M. Vallone, Sequence variation of 22 autosomal STR loci detected by next generation sequencing, Forensic science international. Genetics 21 (2016) 15-21.

[3] N.M. Novroski, J.L. King, J.D. Churchill, L.H. Seah, B. Budowle, Characterization of genetic sequence variation of 58 STR loci in four major population groups, Forensic science international. Genetics 25 (2016) 214-226.

Summary of Observations:

- KAPA HyperPrep kit had a higher overall library yield as measured by qPCR.
- Purifying PCR reactions with beads preserved more DNA and had lower drop out events as compared to column-based purification.
- Using low SPRI:DNA ratio (0.7X) showed a good depth of coverage and heterozygote balance.

Funding: The FBI Biometrics Center of Excellence (BCOE): *Forensic DNA Typing as a Biometric tool*.

A copy of this presentation is available at: http://strbase.nist.gov/NISTpub.htm#Presentations