The Answer You Seq: NGS Workflows for Forensic Genetics, Peter Vallone

3/17/2020

1

Outline

• Brief introduction - why sequence?
• Let's examine the wetlab workflow for the ForenSeq DNA Signature Prep Kit (Versagen) and MiSeq FGX
• Illumina sequencing

2

Sequencing in a Forensic Workflow

• There is an interest in sequencing for forensic analyses
  • More markers and marker types – higher multiplexing capability than CE
  • More information → sequence level resolution for STRs
  • The promise: access to this additional information will support forensic casework applications

• Differs from the traditional PCR-CE workflow
  • How?
  • What is the same?
  • What is different?
Why sequence STRs?

- **Sequencing of STRs**
  - STR motif sequence variation; flanking region variation (more polymorphisms)
  - Further understand simple versus complex repeat motifs
  - Characterize stutter by sequence

- **Applications**
  - **One to one matching?** – RMPs with 20 STR markers are already quite low (<10^{-16})
  - Partial profiles
  - Kinship
  - Length-based allele calls are back compatible with existing databases

- **Mixtures**
  - Resolve alleles identical by length, but differ by sequence
  - Separate stutter from low level contributors (based on sequence)
  - A sequenced allele may have a lower frequency (resulting in a higher LR)

- **Will still need an interpretation framework** e.g. probabilistic genotyping

---

Compound/Complex autosomal STRs

Increase in observed alleles through sequencing

<table>
<thead>
<tr>
<th>Compound/Complex autosomal STRs</th>
<th>Increase in observed alleles through sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2S1338</td>
<td>Length</td>
</tr>
<tr>
<td>D2S441</td>
<td>Motif sequence</td>
</tr>
<tr>
<td>D5S818</td>
<td>Flanking region</td>
</tr>
<tr>
<td>D5S388</td>
<td></td>
</tr>
<tr>
<td>D13S388</td>
<td></td>
</tr>
<tr>
<td>D16S539</td>
<td></td>
</tr>
<tr>
<td>D2S1338</td>
<td></td>
</tr>
<tr>
<td>D2S441</td>
<td></td>
</tr>
<tr>
<td>D5S818</td>
<td></td>
</tr>
<tr>
<td>D5S388</td>
<td></td>
</tr>
<tr>
<td>D13S388</td>
<td></td>
</tr>
<tr>
<td>D16S539</td>
<td></td>
</tr>
</tbody>
</table>

# of unique alleles

Sequencing other marker systems: applications

- Mitochondrial DNA sequencing
  - Control region and/or full genome
  - Higher throughput than Sanger methods
  - Measure lower levels of heteroplasmy
  - Easier workflow?

- SNPs
  - Ancestry, Identity, Phenotype, Microhaplotypes (closely linked SNPs)

- In the future
  - Non-targeted sequencing, RNA targets, metagenomics, epigenetics
  - New technologies, methods, marker systems

Targeted Sequencing

- We wish to sequence ‘our’ markers (STRs, SNPs, mtGenome)

- You might ‘see’ these markers in whole genome sequencing (shotgun)
  - Low coverage
  - Issues with STR regions
  - Can be more bioinformatically challenging
  - Inefficient use of the sequencing ‘space’

- For now...it seems that forensic genetics will implement a targeted approach
Is Sequencing more sensitive?

- More sensitive than what? CE-based testing?
- Targeted sequencing is PCR-based
  - Still using PCR, stochastic effects are encountered at low amounts of DNA
  - Expect similar levels of sensitivity (in terms of generating PCR products)
- ‘Sensitive’ can also relate to: more markers, more information, improved matching statistics
- Sequencing methods may allow for a deeper understanding of S/N and artifacts

Comparing workflows — targeted sequencing

Library preparation

Add sequence adapters
PCR amplify
Purify
Quantify
Normalize
Pool

General...we’ll get into specifics
Goal: prepare PCR products for sequencing
Select listing of commercial sequencing workflows

<table>
<thead>
<tr>
<th>Assay</th>
<th>Platform</th>
<th>Associated Software</th>
<th>Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>ForenSeq DNA Signature Prep Kit</td>
<td>MiSeq</td>
<td>UAS</td>
<td>auSTRs, Y STRs, X STRs and SNPs</td>
</tr>
<tr>
<td>ForenSeq mtDNA Control Region Solution</td>
<td>MiSeq</td>
<td>UAS</td>
<td>Mitochondrial control region (WG soon?)</td>
</tr>
<tr>
<td>PowerSeq HAP System</td>
<td>MiSeq</td>
<td>Open</td>
<td>auSTR and Y STRs</td>
</tr>
<tr>
<td>PowerSeq CRM Mixed System, Custom</td>
<td>MiSeq</td>
<td>Open</td>
<td>Mitochondrial control region (and WG?)</td>
</tr>
<tr>
<td>Precision ID MP Ancestry Panel</td>
<td>S5</td>
<td>Converge</td>
<td>Ancestry SNPs</td>
</tr>
<tr>
<td>Precision ID STR Globaliser NSG STR Panel</td>
<td>S5</td>
<td>Converge</td>
<td>Ancestry STRs</td>
</tr>
<tr>
<td>Precision ID STR Globaliser Whole Genome</td>
<td>S5</td>
<td>Converge</td>
<td>Whole mitochondrial genome</td>
</tr>
<tr>
<td>Precision ID mtDNA Control Region Panel</td>
<td>S5</td>
<td>Converge</td>
<td>Mitochondrial control region</td>
</tr>
<tr>
<td>Precision ID SNP Phenotype Panel</td>
<td>S5</td>
<td>Converge</td>
<td>SNPs</td>
</tr>
<tr>
<td>GeneReader DNAseq Targeted Panels v2</td>
<td>Illumina/S5</td>
<td>CLCBio</td>
<td>Mito, SNPs</td>
</tr>
</tbody>
</table>

List not exhaustive – just some common examples

For all workflows please visit: https://strbase.nist.gov/pub_pres/ISFG_NGS_Workflows_Workshop_Sept_9.pdf

Example NGS workflows

For all NGS workflows please visit: http://strbase.nist.gov/pub_pres/ISFG_NGS_Workflows_Workshop_Sept_9.pdf

ForenSeq™ DNA Signature Prep kit - MiSeq FGx

Table 1: ForenSeq DNA Signature Prep Kit - MiSeq FGx

<table>
<thead>
<tr>
<th>Panel</th>
<th>Number # of Loci</th>
<th>10 kb</th>
<th>5 kb</th>
<th>1 kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct STRs</td>
<td>27</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y STRs</td>
<td>24</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X STRs</td>
<td>2</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variability SNPs</td>
<td>64</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenotype SNPs</td>
<td>32</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographic ancestry SNPs</td>
<td>55</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. SNP and STR (short tandem repeat) loci sources include the ForenSeq™ DNA Signature Prep Kit
b. 10 kb, 5 kb, and 1 kb refer to fragment lengths in kbp


Presented at the NIST Forensic DNA typing Continuing Education Day 2020, Gaithersburg, MD
What is the overall goal of library preparation?

• To prepare the PCR products for the sequencer

• Capture a ‘snapshot’ of the PCR products (ratios, abundance)

• We want to avoid
  • Any bias that favors a product based on size, sequence, abundance
  • Uneven yields or representation across samples
  • Inefficient use of the sequencing capability

ForenSeq (Sequenced on MiSeq FGx)

Targeted PCR (tags added) → Enrich Targets (indices added) → Purify libraries

Denature and prepare for loading → Pool libraries → Normalize libraries

Step: Targeted PCR (PCR 1)

Workflow: ForenSeq™ DNA Signature Prep kit

Purpose: Targeted amplification of STRs/SNPs

Target 1 ng of gDNA
18 PCR cycles
5 µl of DNA extract
15 µl total volume
The Answer You Seq: NGS Workflows for Forensic Genetics, Peter Vallone

Presented at the NIST Forensic DNA typing Continuing Education Day 2020, Gaithersburg, MD

---

**Workflow: ForenSeq™ DNA Signature Prep Kit**

- **PCR 1:** Enrich targets (PCR 2)
- **Purpose:** Add sequence indices and P5/P7 sequences to the PCR products

**PCR 1 Products:**
- 15 µL volume
- 15 PCR cycles
- Perform in a different room (not in the original PCR 1 environment)

**Target Region**

<table>
<thead>
<tr>
<th>Seq1 (fPrimer)</th>
<th>i5 index</th>
<th>P5</th>
<th>i7 index</th>
<th>P7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq2 (rPrimer)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Enrich targets, add indices and P5/P7 sequences**

**Seq1 Complement**

**Seq2 Complement**

**Index Adapters**

- 96 unique combination can be created from the eight i5 and twelve i7 indices

---

**PCR 1 products:**
- 15 µL volume
- 15 PCR cycles
- Perform in a different room (not in the original PCR 1 environment)

**Target Region**

<table>
<thead>
<tr>
<th>Seq1 (fPrimer)</th>
<th>i5 index</th>
<th>P5</th>
<th>i7 index</th>
<th>P7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq2 (rPrimer)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Enrich targets, add indices and P5/P7 sequences**

**Seq1 Complement**

**Seq2 Complement**

**Index Adapters**

- 96 unique combination can be created from the eight i5 and twelve i7 indices

---

Presented at the NIST Forensic DNA typing Continuing Education Day 2020, Gaithersburg, MD
Sample Demultiplexing Using Index Sequences

Multiple Samples can be sequenced together... then sorted by index.

Slide courtesy of Meghan Didier (Verogen)

Combinations of i5 (8 nt) and i7 (6 nt) will allow for sample barcoding.

Forward PCR primer

<table>
<thead>
<tr>
<th>5' Index</th>
<th>3' Primer</th>
<th>Target Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>P5 sequence that will bind to the flow cell</td>
<td>Tag that allows for i5 and i5 to be incorporated by PCR and is the sequencing primer binding site</td>
<td></td>
</tr>
</tbody>
</table>

Reverse PCR primer

<table>
<thead>
<tr>
<th>5' Index</th>
<th>3' Primer</th>
<th>Target Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>P7 sequence that will bind to the flow cell</td>
<td>Tag that allows for i7 and i7 to be incorporated by PCR and is the sequencing primer binding site</td>
<td></td>
</tr>
</tbody>
</table>

Workflow: ForenSeq™ DNA Signature Prep kit

Step: Purify libraries

Preparation:
- Purify 45 µL of bead suspension into plate
- Pipette 45 µL of PCR into plate
- Centrifuge 2000 rpm for 2 min
- Let sit for 5 min
- Place on magnetic stand for 2 min (until clear)
- Wash with 200 µL 80% EtOH 2 times
- Add 52.2 µL of resuspension buffer to each well
- Shake 1800 rpm for 2 min
- Place on magnetic stand for 2 min (until clear)
- Recover 50 µL in a fresh plate

Result: purified adapted PCR products

- PCR reaction is mixed with beads
- Adapted PCR libraries bind to the beads
- Excess PCR reagents are washed off
- Purified adapted libraries are recovered
Beads! Beads! Beads!

Solid Phase Reversible Immobilization (SPRI)
Polypropylene core surrounded by a layer of magnetite coated with carboxyl molecules

In the presence of PEG and salt (e.g., 20% PEG and 2.5 M NaCl), the DNA is driven to bind to the negatively charged surface

Wash or clean up (80% EtOH solution)

Resuspend

The ratio of SPRI:DNA can tune the length of DNA fragments bound

Step: Normalize libraries
Workflow: ForenSeq™ DNA Signature Prep kit
Purpose: Ensure that libraries of varying yields are equally represented within the sequencing run

By normalizing the concentration of the libraries, while preserving the content of each library, post-PCR quantification and individual PCR product normalization are not necessary

Mix beads and pure library – 30 min

Different than the previous purification – one strand of the library adapter may be biotinylated (beads would be coated with streptavidin)

Beads bind a fixed amount of library (ideally without bias; locus/adapter size)

2x EtOH wash

Wash with NaOH which denatures the library adapter – leaving single stranded DNA solution (denatured and ready for pooling)

Recover and add to storage buffer

Normalized concentration of ssDNA is ready for pooling

Single stranded DNA is ‘melted’ off the bead
.Binding reagents are washed off

Approx. 2 nM of library post bead normalization

ForenSeq Library Preparation
Bead Normalization

Purified Libraries:
Range of yields

Bead-based Normalization
1. Equal volume of beads added to each well
2. Beads bind equal amount of product per well
3. Excess removed
4. Products eluted off beads

Normalized Libraries:
Equally represented
Sample Kiting
Pool 5 µl of each desired library

Slide courtesy of Meghan Didier (Verogen)
**Streptavidin and Biotin**

DNA can be labeled with biotin moieties. Magnetic beads can be coated with SA. The SA-Biotin complex is stable to organic solvents, denaturants, detergents, temperature, and pH. Very useful for biotech separation applications.

---

**Step: Pool Libraries**

**Workflow: ForenSeq™ DNA Signature Prep Kit**

**Purpose:** Combines equal volumes of normalized library to create a pool of libraries that are sequenced together on the same flow cell.

1. Collect and pool normalized libraries
2. Collect across the plate in an 8-strip tube – then pooled into a single tube

---

**MiSeq FGx Sequencing**

- Preloaded single use reagent cartridge
- Positive consumables tracking
- Auto flow cell positioning
- Walkaway automation

---

**Slide courtesy of Meghan Didier (Verogen)**
Step: Sequencing
Workflow: ForenSeq™ DNA Signature Prep kit
Purpose: Sequence the PCR products

Note: we will not be covering how to set up a specific instrument, loading the system, operational software, etc. This is covered in training materials, software ‘wizards’.

(Fluorescent) Sequencing by synthesis
- The library you created is hybridized to a flow cell
- Individual strands create ‘clusters’ through bridge amplification
- Sequencing proceeds one base per cycle
- Each A, G, C, T has a unique fluorescent dye attached
- Four images of the flow cell per cycle allows for the assignment of sequence at each cluster

Summary
- ForenSeq uses PCR tagging to incorporate library adapters
- Bead-based normalization is used to ensure that each library is of a similar concentration prior to sequencing
- Care should be taken with PCR 2 and i5/i7 indices to avoid contamination
- Sequencing by synthesis – all four bases are incorporated and read per cycle
Acknowledgments

- Verogen
  - Meghan Didier
  - Melissa Kotkin
- NIST
  - Kevin Kiesler
  - Sarah Riman
  - Becky Steffen
  - Katherine Gettings
  - Lisa Borsuk
- NIST
  - Kevin Kiesler
  - Sarah Riman
  - Becky Steffen
  - Katherine Gettings
  - Lisa Borsuk