Strengths and Limitations of NGS for Forensic DNA Analysis

Douglas R. Storts, PhD
Head of Research
# NGS versus Capillary Electrophoresis

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CE</strong></td>
<td><strong>NGS</strong></td>
</tr>
<tr>
<td>- Established technology</td>
<td>- Genotyping based on length and sequence</td>
</tr>
<tr>
<td>- Accepted in court</td>
<td>- Greater multiplex capability</td>
</tr>
<tr>
<td>- Relatively easy workflow</td>
<td>- High dynamic range</td>
</tr>
<tr>
<td>- Complex mixture analysis</td>
<td>- Potential improvement to mixture interpretation</td>
</tr>
<tr>
<td>- Genotyping based on length only</td>
<td>- Smaller amplicons (degraded DNA)</td>
</tr>
<tr>
<td>- Limited multiplex capability</td>
<td>- High cost per sample</td>
</tr>
<tr>
<td>- Complex mixture analysis</td>
<td>- Large amount of data</td>
</tr>
<tr>
<td>- Genotyping based on length only</td>
<td>- Pooling of samples needed to reduce cost per sample</td>
</tr>
<tr>
<td>- No guidelines available yet</td>
<td>- More complex workflow</td>
</tr>
<tr>
<td>- More complex workflow</td>
<td>- Time to result</td>
</tr>
</tbody>
</table>
PowerPlex® Fusion 6C System

23 autosomal STRs ($\text{PI} = 2.3 \times 10^{-32}$), 3 Y-STRs, amel

*A = Amelogenin*
PowerPlex® Fusion 6C Allelic Ladder
PowerSeq™ Systems
Smaller Amplicons

- Amplicon size range is 129-284bp (129-303bp) in the PowerSeq™ Systems
  - Used existing primers where possible to minimize concordance issues
- In contrast, the 27-plex PowerPlex® Fusion 6C System:
General Workflow

1. **Extract & Quantify DNA**
2. **Targeted Amplification**
3. **Library Preparation**
4. **Clonal Amplification/Cluster Generation**
5. **Sequence**
6. **Data Analysis**
7. **PowerSeq™ Systems**

**Kit Components**
- 5X Primer Pair Mix
- 5X Master Mix
- 2800M Control DNA
- Water
Sample Types

**Extracted DNA**
- Extract DNA from sample
- Add up to 15µl of sample into reaction

**FTA® punches**
- Punch FTA® into plate
- Add reaction mix

**NonFTA Punches**
- Punch card into plate, Incubate with PunchSolution™
- Add reaction mix to dried punch

**Swabs**
- SwabSolution™ incubation of swab
- Add 2µl of extract into reaction mix

Cycling time ~85 minutes for all applications
PowerSeq™ Protocol

PCR Setup and Thermal Cycling
1. Thaw all PowerSeq™ components just prior to use.
2. Vortex the components thoroughly for 5 seconds. Centrifuge tubes briefly.
3. Determine the number of reactions including positive and negative controls. Add 1 or 2 reactions to this number.
4. Prepare the PCR amplification mix by combining the components as shown below.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume Per Reaction</th>
<th>Number of Reactions</th>
<th>Final Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water, Amplification Grade</td>
<td>To a final volume of 25μl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PowerSeq™ 5X Master Mix</td>
<td>5.0μl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PowerSeq™ 5X Primer Pair Mix</td>
<td>5.0μl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Template DNA (0.25–0.5ng)</td>
<td>Up to 15μl</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total volume</strong></td>
<td><strong>25μl</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The template DNA will be added at Step 6.
5. Vortex the PCR amplification mix for 5 seconds, then add PCR amplification mix to each reaction well.
6. Add the template DNA.
7. For the positive amplification control, vortex the 2800M Control DNA, then dilute an aliquot to 0.5ng in the desired template DNA volume. Add 0.5ng of diluted DNA to a reaction well containing PCR amplification mix.
8. Seal the plate with strip caps. Briefly centrifuge the plate if desired.
9. Place the plate in the GeneAmp® PCR System 9700 thermal cycler.
10. Program the thermal cycler using the conditions shown in Figure 1. We recommend using 30 cycles. Be sure to select Max mode as the ramp speed.
11. Place the plate in the thermal cycler, run the program and remove the plate from the thermal cycler.

Amplification steps are virtually identical to existing CE-based systems.

Figure 1. The thermal cycling protocol for the GeneAmp® PCR System 9700.
Prototype PowerSeq™ Systems

- Current prototype kits
  - PowerSeq™ Auto (18 and 24-plex)
  - PowerSeq™ Mito (control region)
  - PowerSeq™ Y
  - PowerSeq™ Auto/Mito
  - PowerSeq™ Auto/Y
  - PowerSeq™ Auto/Mito/Y
Data Interpretation

- SoftGenetics
- Battelle
- STRait Razor
- TSSV
- Others
High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing

Xiangpei Zeng, Jonathan L. King, Monika Stoljarova, David H. Warshauer, Bobby L. LaRue, Antti Sajantila, Jaynish Patel, Douglas R. Storts, Bruce Budowle

ABSTRACT

STR typing in forensic genetics has been performed traditionally using capillary electrophoresis (CE). However, CE-based method has some limitations: a small number of STR loci can be used; stutter products, dye artifacts and low level alleles. Massively parallel sequencing (MPS) has been considered a viable technology in recent years allowing high-throughput coverage at a relatively affordable price. Some of the CE-based limitations may be overcome with the application of MPS. In this study, a prototype multiplex STR System (Promega) was amplified and prepared using the TruSeq DNA LT Sample Preparation Kit (Illumina) in 24 samples. Results showed that the MinElute PCR Purification Kit (Qiagen) was a better size selection method compared with recommended diluted bead mixtures. The library input sensitivity study showed that a wide range of amplicon product (6–200 ng) could be used for library preparation without apparent differences in the STR profile. PCR sensitivity study indicated that 62 pg may be minimum input amount for generating complete profiles. Reliability study results on 24 different individuals showed that high depth of coverage (DoC) and balanced heterozygote allele coverage ratios (ACRs) could be obtained with 250 pg of input DNA, and 62 pg could generate complete profiles. These studies indicate that this STR multiplexes system and the Illumina MiSeq can generate reliable STR profiles at a sensitivity level that competes with current widely used CE-based method.
Depth of Coverage
500pg Genomic DNA

Fig. 1. A histogram portrayal of the depth of coverage by locus of one sample (no. 12) with 500 pg of input DNA.

Zeng, et al. 2015. FSI: Genetics 16:38-47
Depth of Coverage
62pg Genomic DNA

Fig. 2. A histogram result of depth of coverage by locus of one sample (no. 12) with 62 pg of input DNA.

Zeng, et al. 2015. FSI: Genetics 16:38-47
# PowerSeq™ Auto System

## Sequence Diversity (18-plex)

<table>
<thead>
<tr>
<th>Population</th>
<th>Samples</th>
<th>Total Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>102</td>
<td>204</td>
</tr>
<tr>
<td>Himalayan</td>
<td>97</td>
<td>194</td>
</tr>
<tr>
<td>Pygmee</td>
<td>99</td>
<td>198</td>
</tr>
<tr>
<td>Total</td>
<td>298</td>
<td>596</td>
</tr>
</tbody>
</table>

## Average Stutter-percentage (± StdDev)

<table>
<thead>
<tr>
<th>Locus</th>
<th>n-1_v1 ±</th>
<th>n-1_v2 ±</th>
<th>n+1_v1 ±</th>
<th>n+1_v2 ±</th>
<th>n-2_v1 ±</th>
<th>n-2_v2 ±</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSF1P0</td>
<td>0.063 ± 0.019</td>
<td>±</td>
<td>0.011 ± 0.007</td>
<td>±</td>
<td>0.005 ± 0.004</td>
<td>±</td>
</tr>
<tr>
<td>D2S1338</td>
<td>0.088 ± 0.024</td>
<td>0.013 ± 0.006</td>
<td>0.012 ± 0.010</td>
<td>0.010 ± 0.010</td>
<td>±</td>
<td>±</td>
</tr>
<tr>
<td>D3S1358</td>
<td>0.091 ± 0.019</td>
<td>±</td>
<td>0.009 ± 0.008</td>
<td>±</td>
<td>0.008 ± 0.005</td>
<td>±</td>
</tr>
<tr>
<td>D5S818</td>
<td>0.069 ± 0.017</td>
<td>±</td>
<td>0.011 ± 0.007</td>
<td>±</td>
<td>0.006 ± 0.004</td>
<td>±</td>
</tr>
<tr>
<td>D7S820</td>
<td>0.052 ± 0.020</td>
<td>±</td>
<td>0.008 ± 0.007</td>
<td>±</td>
<td>0.007 ± 0.005</td>
<td>±</td>
</tr>
<tr>
<td>D8S1179</td>
<td>0.068 ± 0.015</td>
<td>±</td>
<td>0.009 ± 0.007</td>
<td>±</td>
<td>0.006 ± 0.004</td>
<td>±</td>
</tr>
<tr>
<td>D13S317</td>
<td>0.056 ± 0.021</td>
<td>±</td>
<td>0.009 ± 0.007</td>
<td>±</td>
<td>0.006 ± 0.005</td>
<td>±</td>
</tr>
<tr>
<td>D16S539</td>
<td>0.059 ± 0.021</td>
<td>0.005 ± 0.001</td>
<td>0.009 ± 0.005</td>
<td>±</td>
<td>0.007 ± 0.008</td>
<td>±</td>
</tr>
<tr>
<td>D18S51</td>
<td>0.090 ± 0.025</td>
<td>±</td>
<td>0.010 ± 0.009</td>
<td>±</td>
<td>0.009 ± 0.006</td>
<td>±</td>
</tr>
<tr>
<td>D19S433</td>
<td>0.082 ± 0.026</td>
<td>±</td>
<td>0.007 ± 0.005</td>
<td>±</td>
<td>0.008 ± 0.005</td>
<td>±</td>
</tr>
<tr>
<td>D21S11</td>
<td>0.066 ± 0.013</td>
<td>0.012 ± 0.013</td>
<td>0.009 ± 0.004</td>
<td>0.010 ± 0.004</td>
<td>0.008 ± 0.007</td>
<td>0.013 ± 0.004</td>
</tr>
<tr>
<td>FGA</td>
<td>0.079 ± 0.021</td>
<td>±</td>
<td>0.010 ± 0.014</td>
<td>±</td>
<td>0.009 ± 0.006</td>
<td>±</td>
</tr>
<tr>
<td>PentaD</td>
<td>0.016 ± 0.010</td>
<td>±</td>
<td>0.008 ± 0.008</td>
<td>±</td>
<td>0.012 ± 0.037</td>
<td>±</td>
</tr>
<tr>
<td>PentaE</td>
<td>0.038 ± 0.022</td>
<td>±</td>
<td>0.008 ± 0.009</td>
<td>±</td>
<td>0.007 ± 0.006</td>
<td>±</td>
</tr>
<tr>
<td>TH01</td>
<td>0.025 ± 0.012</td>
<td>0.004 ± 0.002</td>
<td>0.007 ± 0.007</td>
<td>±</td>
<td>0.009 ± 0.014</td>
<td>±</td>
</tr>
<tr>
<td>TPOX</td>
<td>0.029 ± 0.013</td>
<td>±</td>
<td>0.005 ± 0.004</td>
<td>±</td>
<td>0.006 ± 0.006</td>
<td>±</td>
</tr>
<tr>
<td>vWA</td>
<td>0.080 ± 0.025</td>
<td>0.011 ± 0.010</td>
<td>0.014 ± 0.024</td>
<td>0.005 ± 0.002</td>
<td>0.008 ± 0.006</td>
<td>0.012 ± 0.006</td>
</tr>
</tbody>
</table>

Data courtesy of Peter de Knijff
PowerSeq™ Auto System
Sequence Diversity (18-plex)

Classical STR alleles  NGS-SNP alleles  NGS-STR alleles

Data courtesy of Peter de Knijff (n=298)
PowerSeq™ Auto - Expanded STRs to 24plex

Added autosomal STRs to generate multiplex equivalent to PowerPlex® Fusion

PowerSeq™ Auto Balance
PowerSeq™ Auto System
Sequence Diversity (24-plex)

188 samples (Caucasian, Afr American, Hispanic)
24 loci sequenced (MiSeq)

Data courtesy of the Applied Genetics Group at NIST
Expanding Beyond Autosomal STRs
Sequencing of Mitochondrial Control Region

PowerSeq™ Mito

Histogram representation of read count data for 2800M DNA library sequenced using MiSeq® v2 chemistry, 1X 260 cycles.
Combined Auto/Mito Multiplex

Demonstrate ability to combine STR primers from 18plex with primers for mitochondrial control region into single 5X mix and run in single amplification reaction.
A Single Multiplex For Multiple Applications

- Single multiplex containing autosomal loci (23-24), Y STR loci (23) and the mtDNA control region
- Equivalent to PowerPlex® Fusion, PowerPlex® Y23 and mitochondrial control region in a single amplification
PowerSeq™ Auto/Y System

Aligned reads per locus using 500pg 2800M Control DNA (n=3)

To control for sample loading on MiSeq, aligned reads per locus normalized to average of total aligned reads.
PowerSeq™ Auto/Mito/Y System
Are “We” Ready for Prime Time?
Relevant Sample Types

- Collaboration with Bruce Budowle
  - Manuscript in preparation demonstrating performance of the 24-plex autosomal STR kit with semen, saliva, blood, and bone samples (single source and mixtures)
  - Data is comparable to CE-based systems
    - Able to generate genotypes from 19:1 mixtures
    - Able to generate full profiles from 67pg DNA
- Collaboration with Peter de Knijff
  - More details to come
- Internal efforts
PowerPlex® Y23 System

Fig. 9. Amplification of a male/female DNA mixture sample containing 125 pg of male DNA and 3000 ng of female DNA. Amplification reactions were performed using a GeneAmp® 9700, and 1 μL of each sample was electrophoresed on a 3130 series instrument using a 3 kV, 5-s injection.

24,000:1 female/male mixture
PowerSeq™ Auto/Mito/Y System
Female:Male Mixtures

Mixtures contained 400ng of female DNA
PowerSeq™ Auto/Mito/Y System
Female:Male Mixtures

Same graph as before, but only showing the Y-STRs

- 8:1
- 24:1
- 72:1
- 216:1
PowerSeq™ Auto/Mito/Y System
Female:Male Mixtures

Total Aligned Reads

- AMEL
- CSF1PO
- D10S1248
- D12S391
- D13S317
- D18S51
- D18S53
- D19S433
- D1S1656
- D21S11
- D22S1045
- D2S1338
- D2S441
- D3S1358
- D5S818
- D7S820
- D8S1179
- D8S1179
- D9S278
- FGA
- FGA
- PENTAD
- PENTAE
- TH01
- TH01
- VWA
- VWA
- DYS19
- DYS385
- DYS389I
- DYS389II
- DYS390
- DYS391
- DYS392
- DYS393
- DYS437
- DYS438
- DYS439
- DYS448
- DYS456
- DYS458
- DYS481
- DYS483
- DYS533
- DYS549
- DYS570
- DYS576
- DYS635
- DYS643
- GATAH4

- Total Aligned Reads

- 0:1
- 8:1
- 24:1
- 72:1
- 216:1
PowerSeq™ Auto/Mito/Y System
Female:Male Mixtures

Same graph as before, but only showing the Y-STRs
PowerSeq™ Auto/Mito/Y System
Female:Male Mixtures

<table>
<thead>
<tr>
<th>Total Aligned Reads</th>
<th>800:1</th>
<th>1600:1</th>
<th>3200:1</th>
<th>6400:1</th>
<th>12800:1</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYS19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS385</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS389I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS390</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS392</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS393</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS437</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS438</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS439</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS448</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS449</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS456</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS458</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS481</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS533</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS549</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS570</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS576</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS635</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS643</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATAH4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary of Y-STR Data

Real-estate limitation on the MiSeq?
Future Focus

- Streamline the protocol to be more in alignment with workflow in a forensic laboratory
- Generate additional data with forensically relevant samples
- Stress test the systems
- Generate additional population data
Are We Ready for Prime Time?

- Need community consensus on dealing with sequence variants
- Need community consensus on types of markers and regions interrogated
  - Some countries forbid use of ancestral markers
  - Is SNP data a burden if you generate an STR profile?
    - If you generate the data, you probably need to report the result
- Need streamlined workflow
- Need to reduce costs
- Need more feedback from the community
Forum for Community Feedback (U.S.)

- Plan to participate in upcoming NIJ grant study (19 month duration) to define requirements for NGS
- Project coordinated by Battelle with a variety of participating laboratories (federal, state, and local)
  - Armed Forces DNA Identification Laboratory, Dover Air Force Base, DE
  - Bureau of Alcohol, Tobacco, Firearms and Explosives, Washington, DC
  - California Department of Justice, Sacramento, CA
  - Harris County Institute of Forensic Sciences, Houston, TX
  - National Institute of Standards and Technology, Gaithersburg, MD
  - Pennsylvania State University, State College, PA
  - Philadelphia Police Department, Philadelphia, PA
Acknowledgments

• Promega
  • Jaynish Patel
  • Spencer Hermanson

• Battelle
  • Seth Faith (now at NCSU)
  • Esley Heizer, Jr.
  • Nancy McMillian
  • Richard Guerrieri
  • Others

• Leiden Univ. Med. Ctr.
  • Peter de Knijff
  • Kristiann van der Gaag

• Thirska Kraaijenbrink
  • Rick van Leeuwen

• NIST
  • Peter Vallone
  • Katherine Gettings

• UNTHSC
  • Bruce Budowle
  • Xiangpei Zeng
  • Jonathan King
  • Monika Stoljarova
  • David Warshauer
  • Bobby LaRue
Thank you