SNP and STR Multiplexes for NGS

Katherine B. Gettings, Ph.D.
Research Biologist, Applied Genetics Group

Workshop: Considerations for Implementing NGS Technologies Into a Forensic Laboratory
68th Annual AAFS Meeting
February 23, 2016
Las Vegas, NV

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

Updated Slides:
http://www.slideshare.net/NIST_AppliedGeneticsGroup

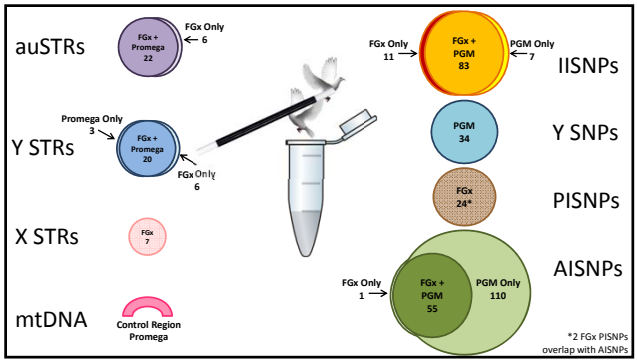
NIST National Institute of Standards and Technology U.S. Department of Commerce

Updated Slides:
http://www.slideshare.net/NIST_AppliedGeneticsGroup

Let's Make a Deal!

estimates do not include instruments, labor or consumables

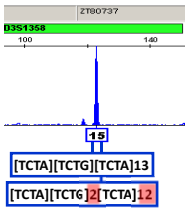
Assay	Number of Samples in Run	Estimated Cost per Sample
Life Tech AmpliSeq SNP Panel (PGM)	6	$174
	30	$122
	59	$115
Illumina ForenSeq (FGx)	8	$215
	32	$84
	96	$55
Promega PowerSeq Auto/Y/mito	8	$249
	32	$110
	96	$68



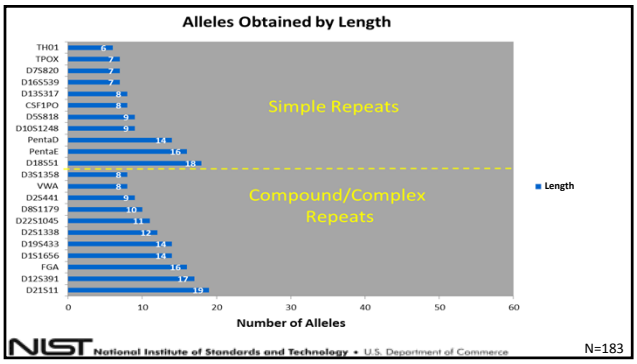
NGS of Forensic STR Loci

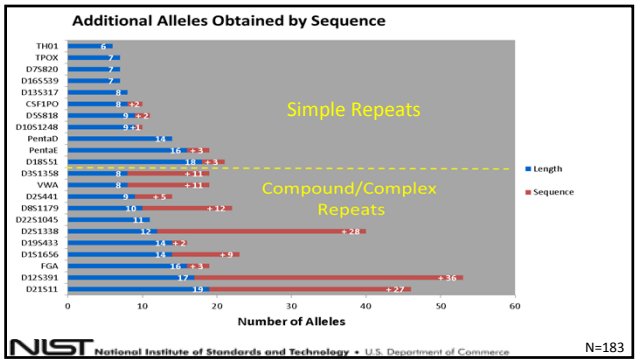
Information

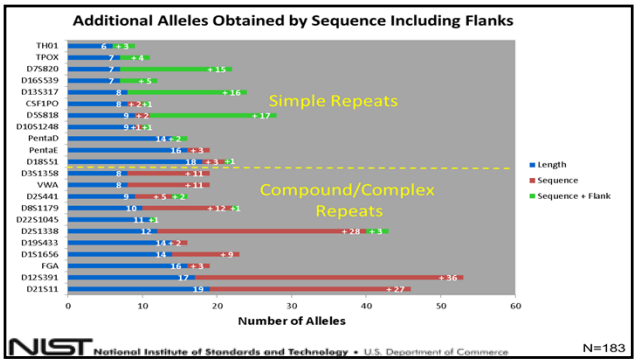
Additional Alleles by Sequence at STR Loci

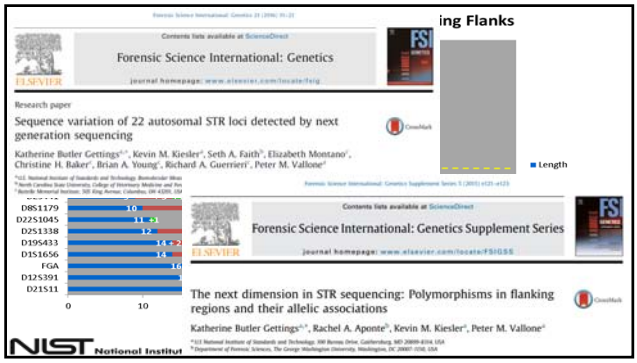


Sequence-Based Heterozygote or Isoallele: A locus that appears homozygous by length-based measurements (CE), but is heterozygous by sequence

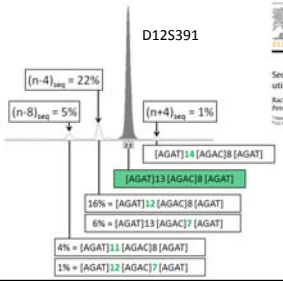






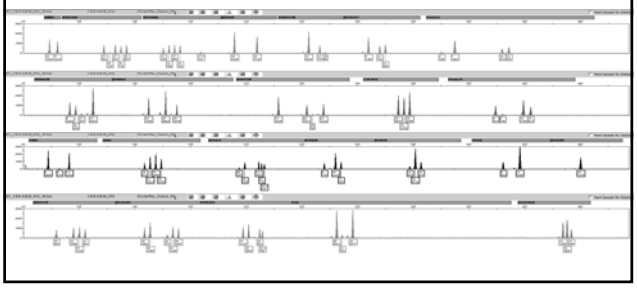


How NGS "Helps" with Stutter



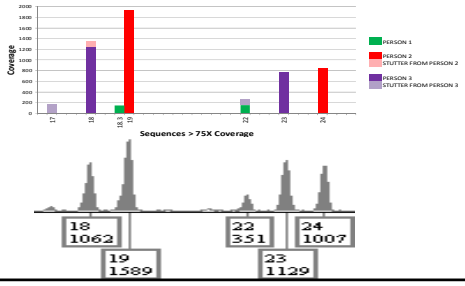
AAFS Criminalistics Session
B180 - Friday Morning
Aponte *et al.*
Sequence-Based Analysis of Stutter at STR Loci: Implementation and Utilization

How NGS Helps in Mixtures



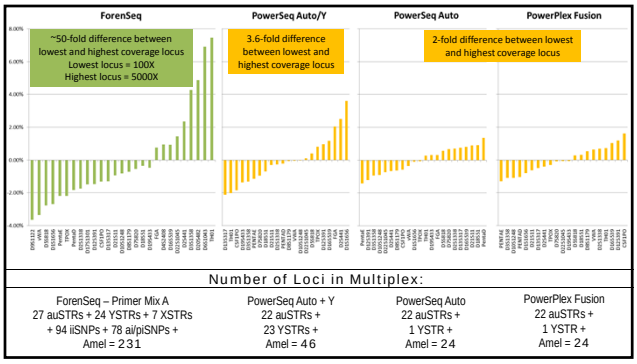
D12S391 1:9:9 Mixture

1 additional allele
2 alleles
help with stutter

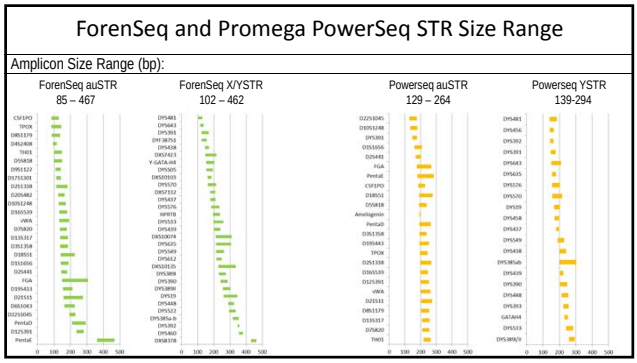


NGS of Forensic STR Loci

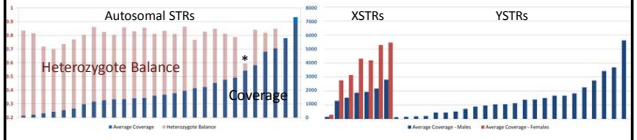
Assay Performance



ForenSeq and Promega PowerSeq STR Size Range



ForenSeq STRs



32 sample run

27 autosomal STRs + Amelogenin

Left side y-axis = heterozygote balance
- ranges from 0.60 to 0.89

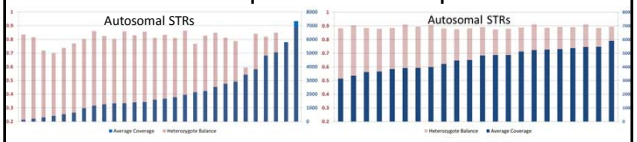
Right side y-axis = average STR coverage
- ranges from 133X to 7332X

7 XSTRs - average coverage ranges from
- 140X to 2807X for males and
- 299X to 5463X for females

24 YSTRs - average coverage ranges from
- 118X to 5628X

* Interpret the locus D2251045 with caution. More imbalance in read counts may be seen between alleles of a heterozygote than observed at other loci. Consider multilocus genotype when determining the presence of a DNA mixture.

ForenSeq and PowerSeq STRs



32 sample run

27 autosomal STRs + Amelogenin

Left side y-axis = heterozygote balance
- ranges from 0.60 to 0.89

Right side y-axis = average STR coverage
- ranges from 133X to 7332X

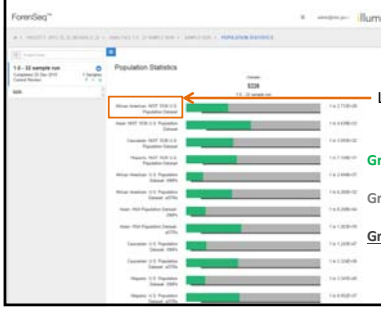
96 sample run

22 autosomal STRs + 1 YSTR + Amelogenin

Left side y-axis = heterozygote balance
- ranges from 0.87 to 0.91

Right side y-axis = average STR coverage
- ranges from 3100X to 5900X

ForenSeq STR and IISNP Statistics



Length-based allele frequencies

Green Bar = sample result

Gray Bar = scaled to range in population

Gray Line = source attribution threshold

NGS of STRs - Conclusions

- Two assays are available for sequencing forensic STR loci
- STR sequencing will increase allelic diversity, improving differentiation among individuals in a mixture
  - Locus specific gains in repeat regions and flanking regions
  - Extent of gain is difficult to quantify
- Characterize "peak height ratios", interlocus balance and stutter by NGS (assay and locus specific)
- Sequence-based allele frequency databases

NGS of Forensic SNP Loci

Information

SNP Information

- IISNP-Individual Identification SNP
- AISNP-Ancestry Informative SNP
- PISNP-Phenotype Informative SNP

ScienceDirect

Research article

Report on ISFG SNP Panel Discussion

SNP Information

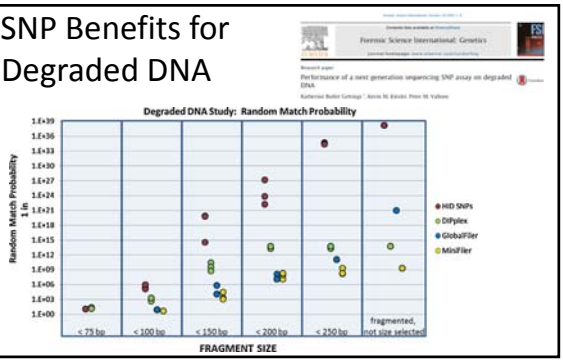
- IISNP-Individual Identification
- PGM Identity SNP Panel and ForenSeq both contain:
  - Kidd 45 and SNPforID52
  - With occasional exceptions

Forensic Science International: Genetics

Research Article

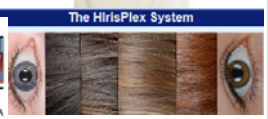
A multiplex assay with 52 single nucleotide polymorphisms for human identification

IISNP Benefits for Degraded DNA



SNP Information

- PISNP-Phenotype
- ForenSeq contains all 24 HIrisplex SNPs
- Only NGS assay with phenotype SNPs

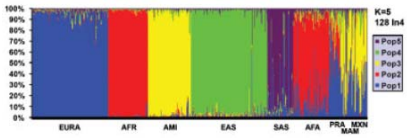


SNP Information

- AISNP - Ancestry Information
- PGM Ancestry SNP Panel contains
  - Seldin 128 (with some exceptions)
  - Kidd 55
- ForenSeq contains Kidd 55

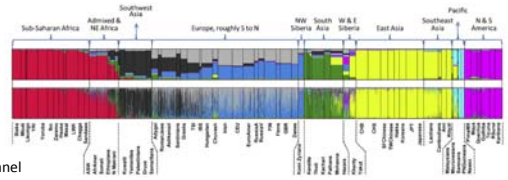
Kosoy 2008 "Seldin 128"

Present in PGM Ancestry Panel



Kidd 2014 "Kidd 55"

Present in ForenSeq and PGM Ancestry Panel



NGS of Forensic SNP Loci

Assay Performance

ForenSeq SNPs

172 SNPs - Identity, Ancestry & Phenotype

Contains (with a few exceptions):

- Kidd 55 for Ancestry
- Kidd 45 + SNPforID52 for Identity
- HIrisplex 24

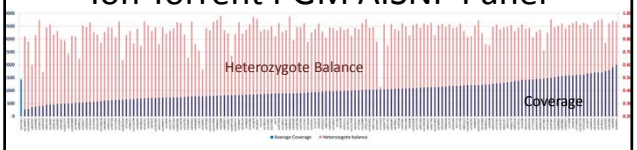
Left side y-axis = average SNP coverage

- ranges from 23X to 3567X (>150 fold)

Right side y-axis = average het balance

- ranges from 0.42 to 0.94

Ion Torrent PGM AISNP Panel



165 SNPs - Ancestry

Contains Kidd 55 and Seldin 128

(with exceptions)

Left side y-axis = average SNP coverage

- ranges from 264X to 2000X (7.6 fold)

Right side y-axis = average het balance

- ranges from 0.43 to 0.98

Ion Torrent PGM AISNP Panel

– PGM plugin uses ALFRED population data

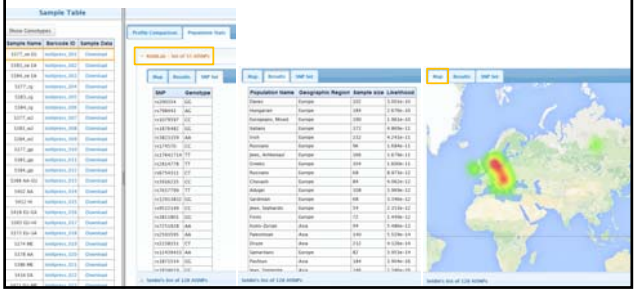
- Evaluates expected occurrence of profile in each ALFRED population
- Potential for very specific ancestry prediction

ALFRED

The ALlele FREquency Database

ALFRED is a resource of gene frequency data on human populations supported by the U. S. National Science Foundation.

Ion Torrent PGM AISNP Panel

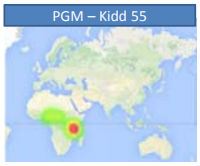


ForenSeq AISNP + PISNP Panel

ForenSeq software uses Hirisplex Model for Phenotype Prediction

And 1000 Genome data for Continental level ancestry prediction

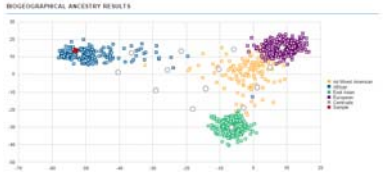
PGM – Kidd 55



PGM – Seldin 128

African American Individual

ForenSeq – Kidd 55



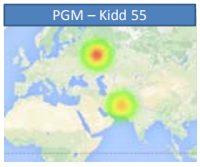
PGM – Kidd 55

PGM – Seldin 128

East Asian Individual

ForenSeq – Kidd 55

PGM – Kidd 55



PGM – Seldin 128

Half European – Half East Asian Individual

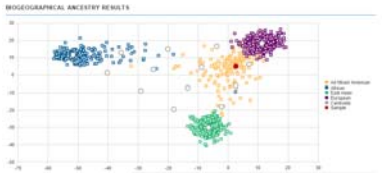
ForenSeq – Kidd 55

PGM – Kidd 55

PGM – Seldin 128

Hispanic Individual

ForenSeq – Kidd 55



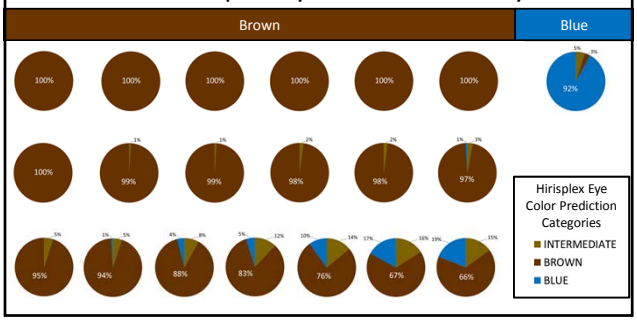
PGM – Kidd 55

Hispanic Individual

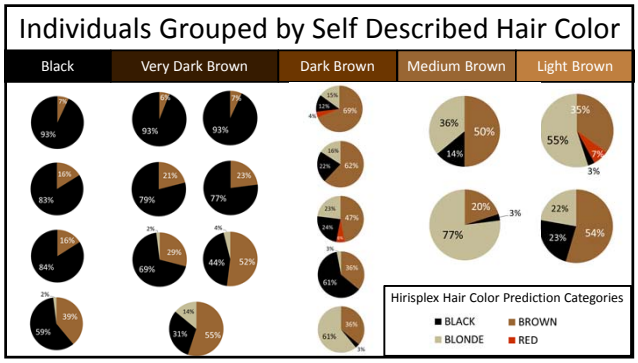
Equally likely to be Central Asian, Native American or European

Over a billion times more likely European / Central Asian / Native American than East Asian / African

Individuals Grouped by Self Described Eye Color



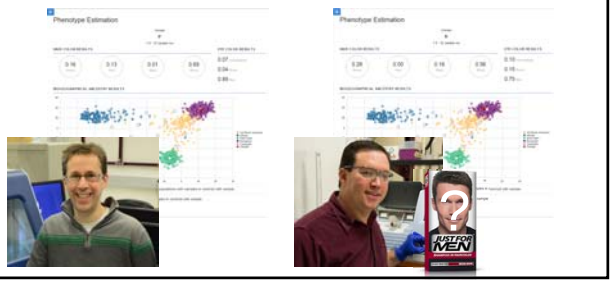
Individuals Grouped by Self Described Hair Color



NGS of SNPs - Conclusions

- More guidance is needed for ancestry and phenotype prediction interpretation
- Identity panels may be useful in degraded samples
- Combining across SNP panels and with STR data may be useful
  - LD must be evaluated

Acknowledgements



Acknowledgements

NIST
Pete Vallone
Kevin Kiesler
Lisa Borsuk
Nate Olson
Becky Steffen
Margaret Kline
Mike Coble
Dave Duewer

GWU
Daniele Podini
Rachel Aponte

illumina
Promega
life technologies

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. **Funding FBI: DNA as a Biometric**

Contact Information
katherine.gettings@nist.gov