AAATCTCTCTAGGATTTTAGCTATAGCTAGCTTAGCTTAGCTAGATATAGCTTAGATCGAGATATATATAGGCGCGCTATTATAGCGCGCGCTAGAGATATATTTATATATTATATAGGAGACATATTAAAGGGACCCATACAGTACGATC
GATCGATATAGAGAGATAGCTAGCTAGCTAGTCGATCCAATAGATATAGATTAGATATACACAGATCTTATAGCGGATCGATCGATGGCTTGGTTAGCTAGACCATAGATCTAATAGGATAGATAGATTAGATAGAGATAGGATAGAGG
ATAGAGATTAGAAAAGAGGATATAGGAGAGAGAGGAGAGATATAGATAGAGATTAGAGATAGAGATTATTATTAGAGAAACCATATAGATCGGGTTTAGGATAGGATAGGATAGGAGGAGAGATTAGAGAGAGATTATATATATATATA
TATAGCGCGCGATCGGCTTTTTCGAGATAGCTAGAGAAATAGTCGATAGATATAGATAGATAGCTCTCGTGTTCGATTCTACAAACAAAACACCCAAAAACATAGCTTAGCTAGATATAGCTTAGATCGAGATATATATAGGCGCGCTAT
TATAGCGCGCGCTAGAGATATATTTATATATTATATAGGAGACATATTAAAGGGACCCATACAGTACGATCGATCGATATAGAGAGATAGCTAGCTAGCTAGTCGATCCAATAGATATAGATTAGATATACACAGATCTTATAGCGGATC
GATCGATGGCTTGGTTAGCTAGACCATAGATCTAATAGGATAGATAGATTAGATAGAGAAATCTCTCTAGGATTTTAGCTATAGCTAGCTTAGCTTAGCTAGATATAGCTTAGATCGAGATATATATAGGCGCGCTATTATAGCGCGCGC
TAGAGATATATTTATATATTATATAGGAGACATATTAAAGGGACCCATACAGTACGATCGATCGATATAGAGAGATAGCTAGCTAGCTAGTCGATCCAATAGATATAGATTAGATATACACAGATCTTATAGCGGATCGATCGATGGCTT
GGTTAGCTAGACCATAGATCTAATAGGATAGATAGATTAGATAGAGATAGGATAGAGGATAGAGATTAGAAAAGAGGATATAGGAGAGAGAGGAGAGATATAGATAGAGATTAGAGATAGAGATTATTATTAGAGAAACCATATAGATC
GGGTTTAGGATAGGATAGGATAGGAGGAGAGATTAGAGAGAGATTATATATATATATATATAGCGCGCGATCGGCTTTTTCGAGATAGCTAGAGAAATAGTCGATAGATATAGATAGATAGCTCTCGTGTTCGATTCTACAAACAAAAC
ACCCAAAAACATAGCTTAGCTAGATATAGCTTAGATCGAGATATATATAGGCGCGCTATTATAGCGCGCGCTAGAGATATATTTCGATGGATCGAAAAAATACCCCCC

Moving Implementation Mountains: Experiencing the Forensic Laboratory NGS Workflow Workshop

ISHI INTERNATIONAL SYMPOSIUM ON HUMAN IDENTIFICATION

MINNEAPOLIS, MN • SEPT. 26–29, 2016

Promega PowerSeq™

- Multiplexed, single tube amplification
- 0.1- 0.5ng input DNA
- Average size ~200bps

22-plex Autosomal STRs

- Loci from PowerPlex® Fusion
- Full CODIS extended panel
- SE33

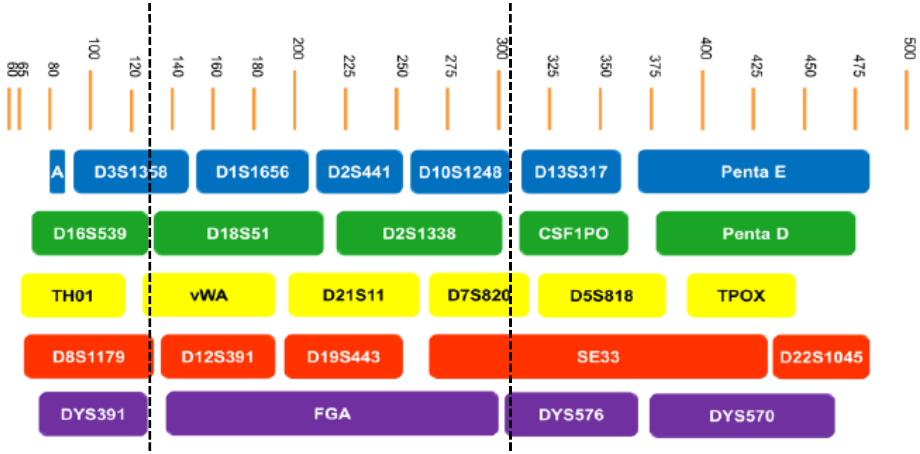
23-plex Y-STRs

- Loci from PowerPlex® Y23

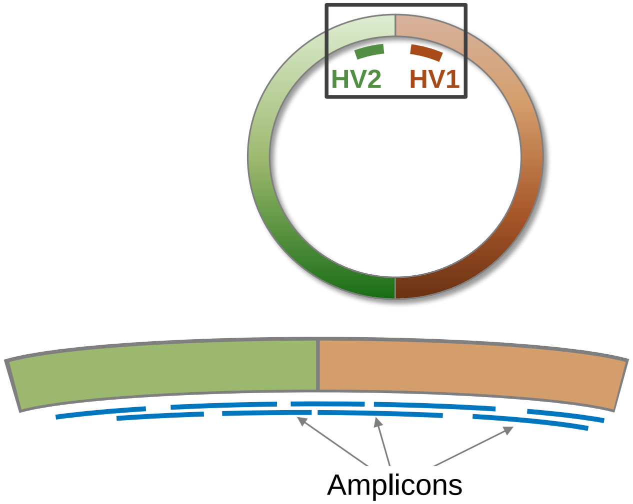
Amelogenin

Mito

- mtDNA d-loop (10 amplicons)

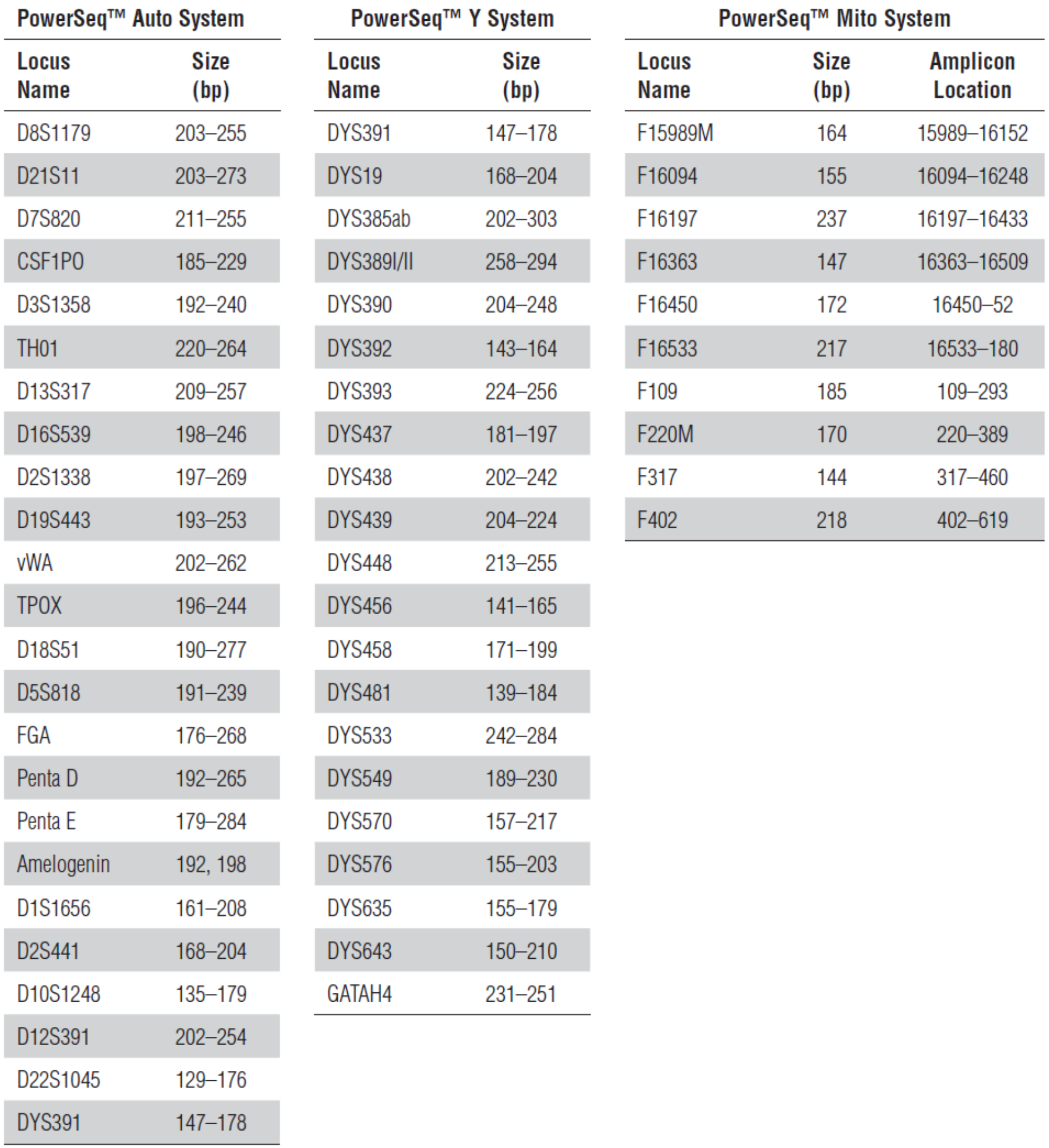


Range for NGS STR PCR



Prototype PowerSeq™ Systems

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



PowerSeq™ Auto System

Locus Name	Size (bp)
D8S1179	203–255
D21S11	203–273
D7S820	211–255
CSF1PO	185–229
D3S1358	192–240
TH01	220–264
D13S317	209–257
D16S539	198–246
D2S1338	197–269
D19S443	193–253
vWA	202–262
TPOX	196–244
D18S51	190–277
D5S818	191–239
FGA	176–268
Penta D	192–265
Penta E	179–284
Amelogenin	192, 198
D1S1656	161–208
D2S441	168–204
D10S1248	135–179
D12S391	202–254
D22S1045	129–176
DYS391	147–178

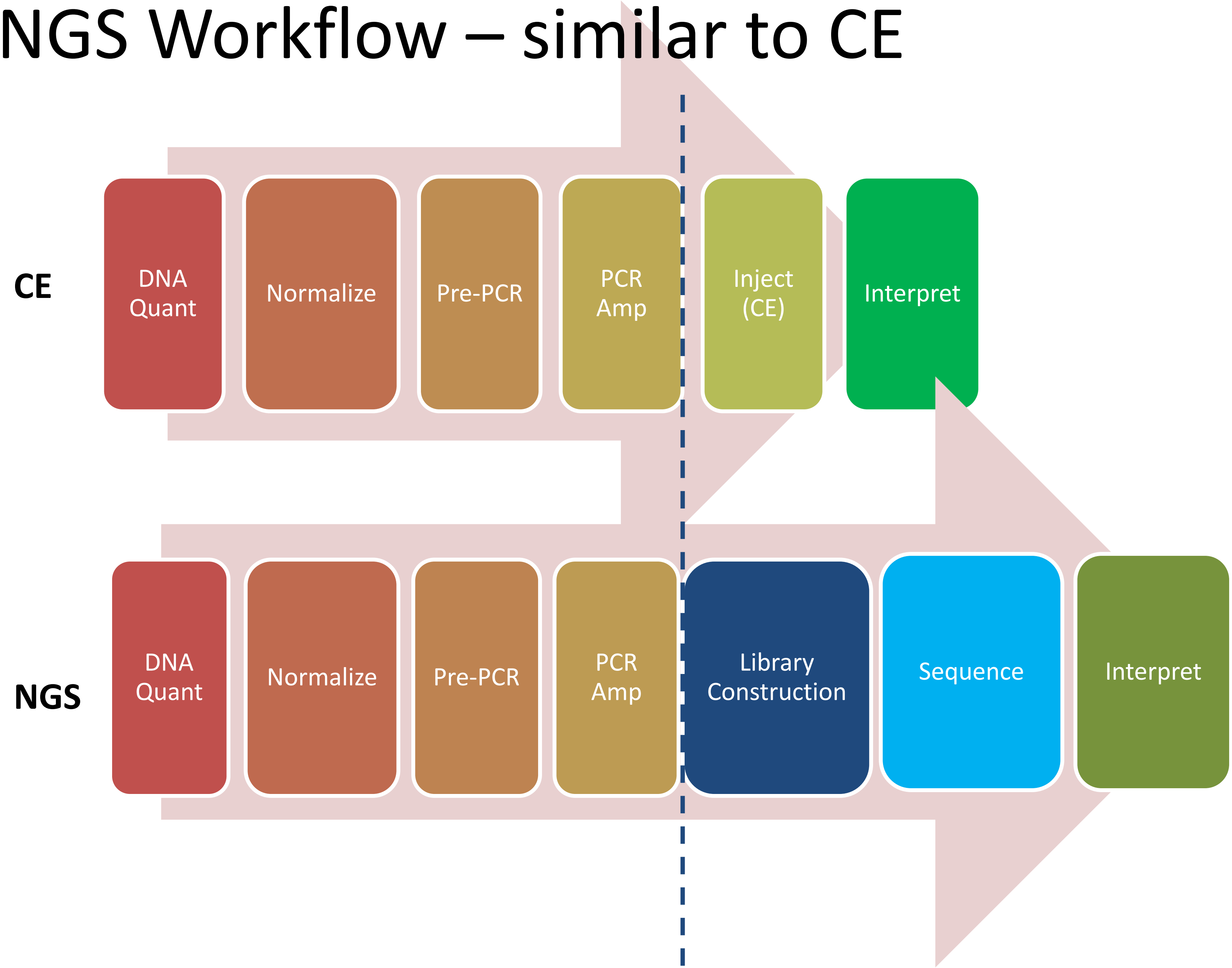
PowerSeq™ Y System

Locus Name	Size (bp)
DYS391	147–178
DYS19	168–204
DYS385ab	202–303
DYS389I/II	258–294
DYS390	204–248
DYS392	143–164
DYS393	224–256
DYS437	181–197
DYS438	202–242
DYS439	204–224
DYS448	213–255
DYS456	141–165
DYS458	171–199
DYS481	139–184
DYS533	242–284
DYS549	189–230
DYS570	157–217
DYS576	155–203
DYS635	155–179
DYS643	150–210
GATAH4	231–251

PowerSeq™ Mito System

Locus Name	Size (bp)	Amplicon Location
F15989M	164	15989–16152
F16094	155	16094–16248
F16197	237	16197–16433
F16363	147	16363–16509
F16450	172	16450–52
F16533	217	16533–180
F109	185	109–293
F220M	170	220–389
F317	144	317–460
F402	218	402–619

NGS Workflow – similar to CE



CE

DNA Quant

Normalize

Pre-PCR

PCR Amp

Inject (CE)

Interpret

NGS

DNA Quant

Normalize

Pre-PCR

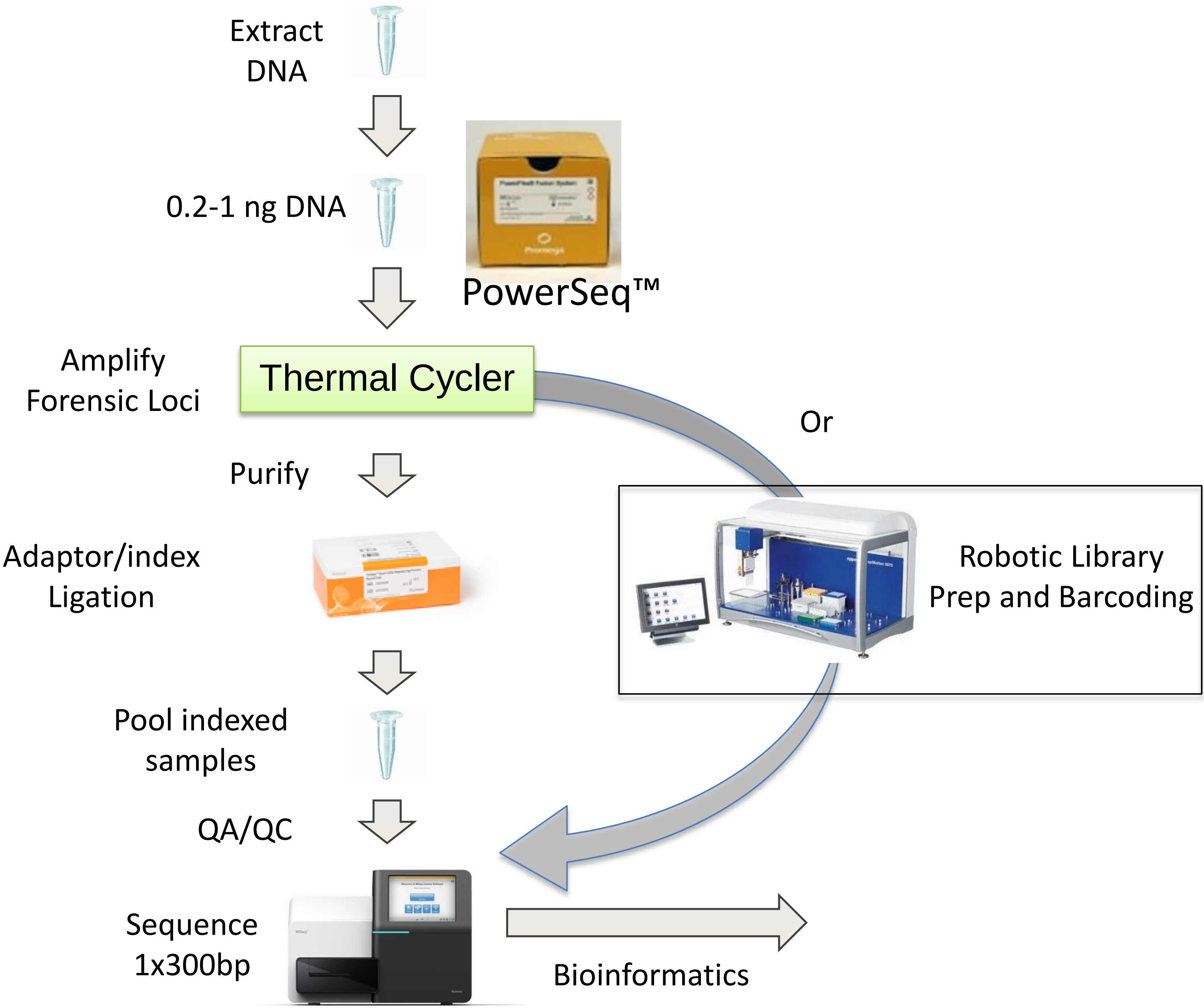
PCR Amp

Library Construction

Sequence

Interpret

PowerSeq™ prep for NGS



Extract DNA

0.2-1 ng DNA

PowerSeq™

Amplify Forensic Loci

Thermal Cycler

Or

Purify

Adaptor/index Ligation

Robotic Library Prep and Barcoding

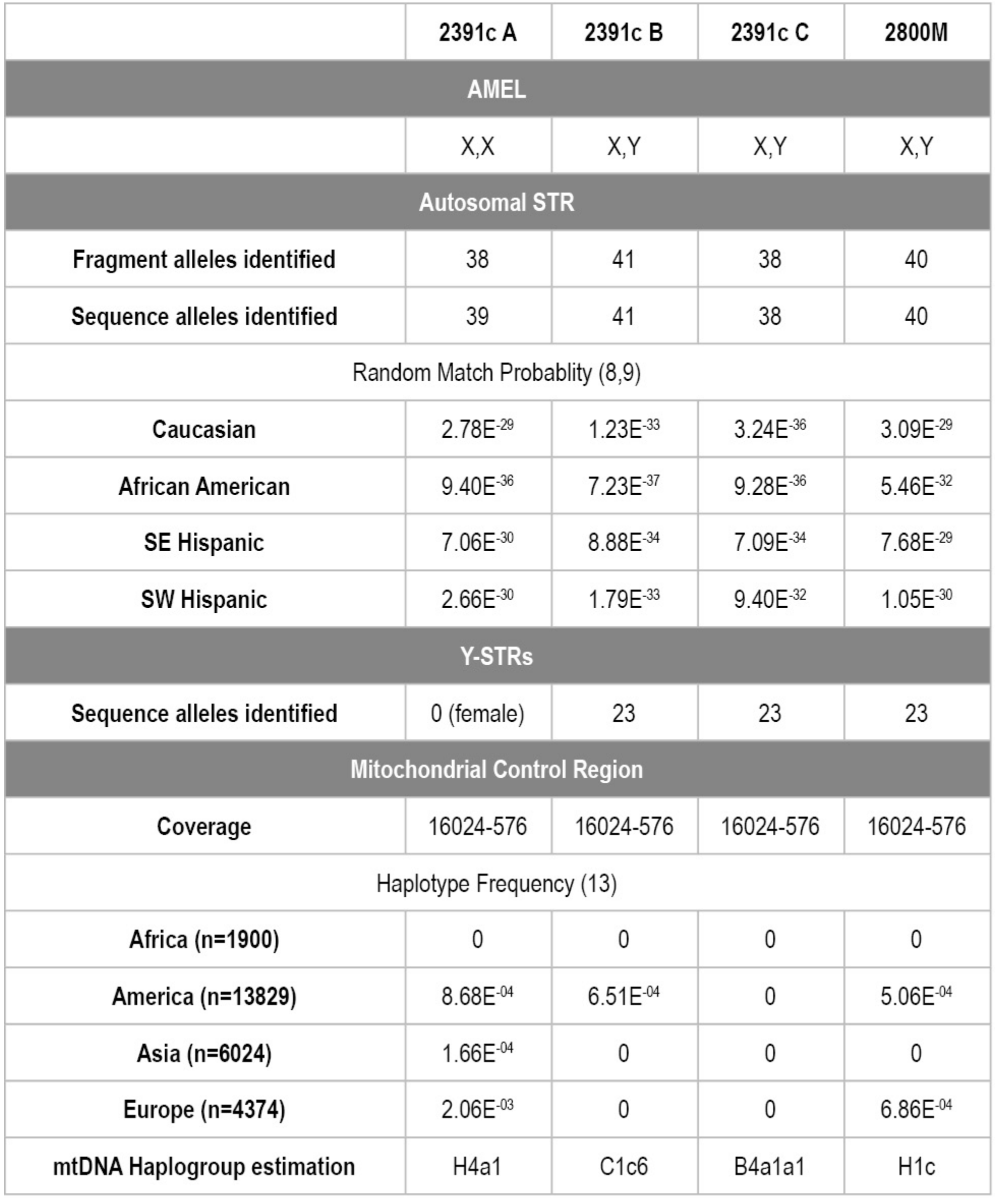
Pool indexed samples

QA/QC

Sequence 1x300bp

Bioinformatics

Summary of PowerSeq™ Auto/mito/Y Output



	2391c A	2391c B	2391c C	2800M
AMEL				
	X,X	X,Y	X,Y	X,Y
Autosomal STR				
Fragment alleles identified	38	41	38	40
Sequence alleles identified	39	41	38	40
Random Match Probablity (8,9)				
Caucasian	$2.78\mathrm{E}^{-29}$	$1.23\mathrm{E}^{-33}$	$3.24\mathrm{E}^{-36}$	$3.09\mathrm{E}^{-29}$
African American	$9.40\mathrm{E}^{-36}$	$7.23\mathrm{E}^{-37}$	$9.28\mathrm{E}^{-36}$	$5.46\mathrm{E}^{-32}$
SE Hispanic	$7.06\mathrm{E}^{-30}$	$8.88\mathrm{E}^{-34}$	$7.09\mathrm{E}^{-34}$	$7.68\mathrm{E}^{-29}$
SW Hispanic	$2.66\mathrm{E}^{-30}$	$1.79\mathrm{E}^{-33}$	$9.40\mathrm{E}^{-32}$	$1.05\mathrm{E}^{-30}$
Y-STRs				
Sequence alleles identified	0 (female)	23	23	23
Mitochondrial Control Region				
Coverage	16024-576	16024-576	16024-576	16024-576
Haplotype Frequency (13)				
Africa (n=1900)	0	0	0	0
America (n=13829)	$8.68\mathrm{E}^{-04}$	$6.51\mathrm{E}^{-04}$	0	$5.06\mathrm{E}^{-04}$
Asia (n=6024)	$1.66\mathrm{E}^{-04}$	0	0	0
Europe (n=4374)	$2.06\mathrm{E}^{-03}$	0	0	$6.86\mathrm{E}^{-04}$
mtDNA Haplogroup estimation	H4a1	C1c6	B4a1a1	H1c

Faith and Scheible, Profiles in DNA (2016)

SIMULATION OF THE PROMEGA POWERSEQ

BIOINFORMATICS EXERCISE

Raw sequence data = Big Data!

Illumina MiSeq

FASTQs

```
@M01020:27:000000000-A3U7D:1:1102:4440:16773 1:N:0:1
AAAGAGCTTACTAAAATGCAACTGGACAATCAGAAAGAGATTGCCGAGATGCAAAATGAGACTCAAAAAGA
GATTGCTGGCATTCAGTCGGCGACTTCACGCCAGAATACGAAAGACCAGGTATATGCACAAAATGAGATGCT
TGCTTATCAACAGAAGGAGTCTACTGCTCGCGTTGCGTCTATTATGGAAAACACCAATCTTTCCAAGCAACAG
CAGGTTTCCGAGCTTATGCGCCAAATGCTTACTCAAGCTCAAACGGCTGGTCAGTCTTTTACCAATGACCAAA
TCAAAGAAATGAACCGAAGGGATACGGCTGGGGTTGACTTAGTTAGACCGGTAGAGCGGGT
+
BBBBAFFDFFFFGBGBGBGFGGHHFFHHFGHHGHGHFHFHHHGGHG?E?FHBAGHGHHHHHHHHHHFHEFE
CBGGFBG5BEHHH5DEDFGGGGGGGGHDBC0FF>FHHD3//?/<BG0FH/FF2FHEGF1?B00<F11DHB11@@
FGDGB1GFFH0CFCG.F0F11<<DD1ECCDGCD-.-
..;0CGCG000:::.;EEBGFFBF0;9BEGBF.9A/F?;000C?99;FFFFA.ADBFFB////BF//9//BFFF/9.A>@BAB//.///
;//9;///9//9B/;.;//B/B/A//;///9----;..../.;9--..-.99/9;B///;9////----.9//;B-9

@M01020:27:000000000-A3U7D:1:1102:7041:167731:N:0:1
ACCCATGTTCCCACTGGCCTGTGGGTCCCCCCATAGATCGTAAGCCCAGGAGGAAGGGCTGTGTTTCAGGGC
TGTGATCACTAGCACCCAGAACCGTCGACTGGCACAGAACAGGCACTTAGGGAACCCTCACTGAATGAATGA
ATGAATGAATGAATGAATGAATGTTTGGGCAAATAAACGCTGACAAGGACAGAAGGGCCTAGATCGGAAG
AGCACACGCCTGAACTCCAGTCACGCCACTATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAAACGAACGC
AGCACGCTCACAAGACCACGGTCCACGTGTGAGTACGTAGAAGAGTGAACTGAGG
+
BCCCGGGGGGGGGGHHHHHHGHHHGGHHHHGGGHHHHHHHHHGGHHHHEGFGFGGHHGGGGGHHHHF
HHGHHDHGHHFHHHHHHHHHHHHGGHEFHFEGG@@GGGHHHHBGGHHHHGCFFHHHGFHHGGGGHHG
HGHFHHGHHHEHHHHGHFHHGGGDDGFHHHHHHH0DFCEGGHGHHGHHHGGGDCFBFFGGFFGGG0AA
EDEFBBFFEGGE.../9A/AD-..AB09B:0;0000;AA...0/00;9...//0;-.-.:0:000BF?0:B.
```

.... 25 Million Reads

Who wants to be a bioinformatician?

```
@M02019:16:000000000-AH1PK:1:1101:12985:1684 1:N:0:1
GTGGGTCCCCCCATAGATCGTAAGCCCAGGAGGAAGGGCTGTGTTTCAG
GGCTGTGATCACTAGCACCCAGAACCGTCGACTGGCACAGAACAGGCAC
TTAGGGAACCCTCACTGAATGAATGAATGAATGAATGAATGAATGAATGT
TTGGGCAAATAAACGCTGACAAGGACAGAAGGGCCG
+
33>ABBBFFFFFGGGGGGGGGGGGGGHHHHGHHHHHHHHHHHFHHH
FHHHHHHHHHHHHHHHHHHHHHHHHGHHGHHGGGHHHHHHHGHH
HHHGHHHHHHHHHHHHGGGEGGGGHHGHHGGHHHHHHHHHHHGH
HGHHHHHHHHHGHGGHHHHHGHHHHGHHHHHGFFHHGGFBCDC--
9::.:
```

Who wants to be a bioinformatician?

```
@M02019:16:000000000-AH1PK:1:1101:12985:1684 1:N:0:1
GTGGGTCCCCCCATAGATCGTAAGCCCAGGAGGAAGGGCTGTGTTTCAG
GGCTGTGATCACTAGCACCCAGAACCGTCGACTGGCACAGAACAGGCAC
TTAGGGAACCCTCACTGAATGAATGAATGAATGAATGAATGAATGAATGT
TTGGGCAAATAAACGCTGACAAGGACAGAAGGGCCG
+
33>ABBBFFFFFGGGGGGGGGGGGGGHHHHGHHHHHHHHHHHFHHH
FHHHHHHHHHHHHHHHHHHHHHHHHGHHGHHGGGHHHHHHHGHH
HHHGHHHHHHHHHHHHGGGEGGGGHHGHHGGHHHHHHHHHHHGH
HGHHHHHHHHHGHGGHHHHHGHHHHGHHHHHGFFHHGGFBCDC--
9::.:
```

Hint: [AATG]x

Don’t sweat!

We’ll make the computers do the hard work

= Bioinformatics

Up the ranks in bioinformatics

Layer 1 – Using web to analyze biological data

Layer 2 – Ability to install and run new programs

Layer 3 – Writing own scripts for analysis in PERL, python or R

Layer 4 – High level coding in C/C++/Java for implementing existing algorithms or modifying existing codes for new functionality

Layer 5 – Thinking mathematically, developing own algorithms and implementing in C/C++/Java

http://www.homolog.us/blogs/blog/2011/07/22/a-beginners-guide-to-bioinformatics-part-i/

www.genomicidlab.com/ishi2016