Strengths and Limitations of NGS for Forensic DNA Analysis

Douglas R. Storts, PhD

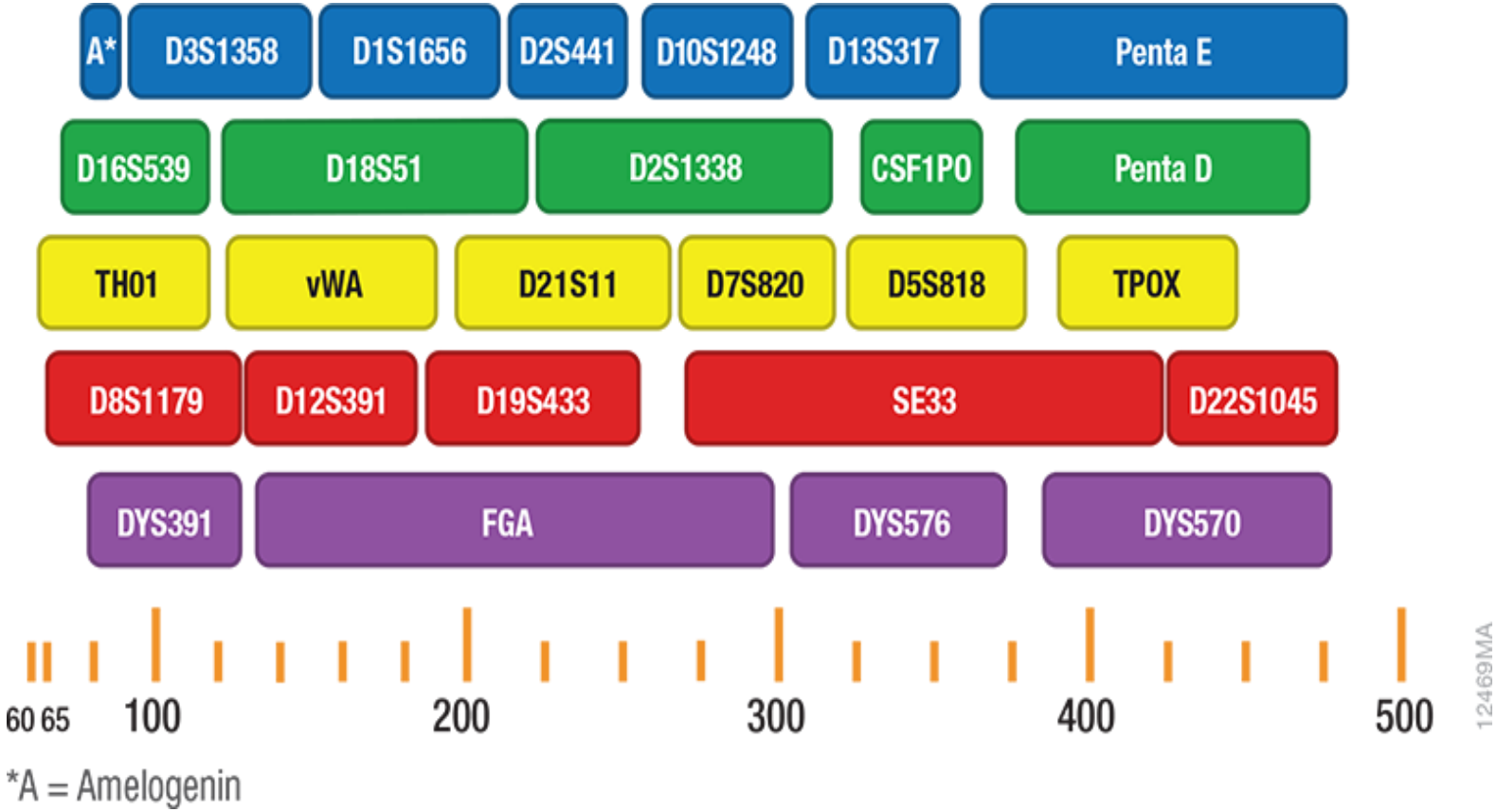
Head of Research

NGS versus Capillary Electrophoresis

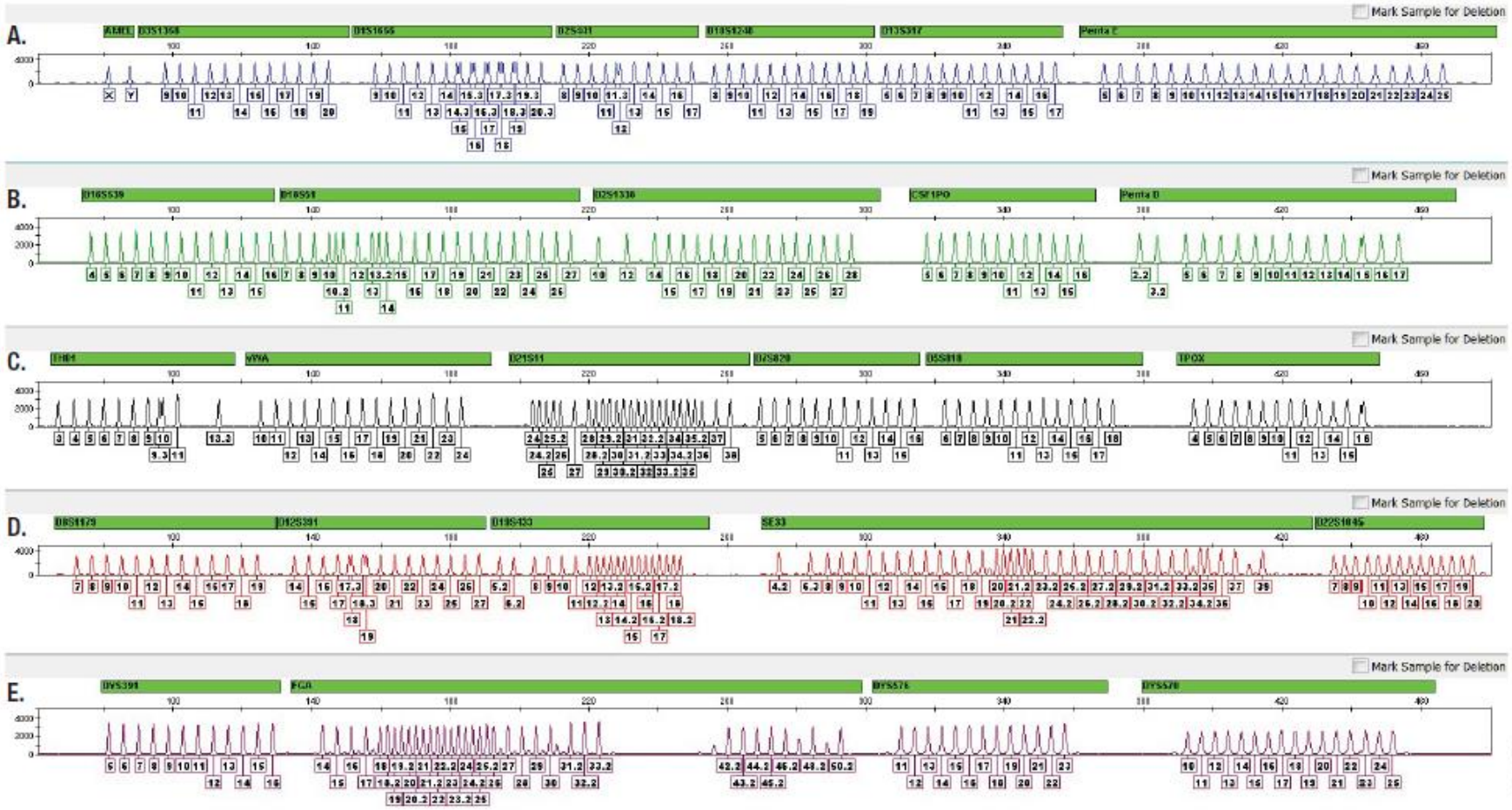
	Advantages	Disadvantages
CE	• Established technology • Accepted in court • Relatively easy workflow	• Limited multiplex capability • Complex mixture analysis • Genotyping based on length only
NGS	• Genotyping based on length and sequence • Greater multiplex capability • High dynamic range • Potential improvement to mixture interpretation • Smaller amplicons (degraded DNA)	• High cost per sample • Large amount of data • Pooling of samples needed to reduce cost per sample • No guidelines available yet • More complex workflow • Time to result

PowerPlex® Fusion 6C System

23 autosomal STRs (PI = 2.3 x 10^{-32}), 3 Y-STRs, amel



PowerPlex® Fusion 6C Allelic Ladder

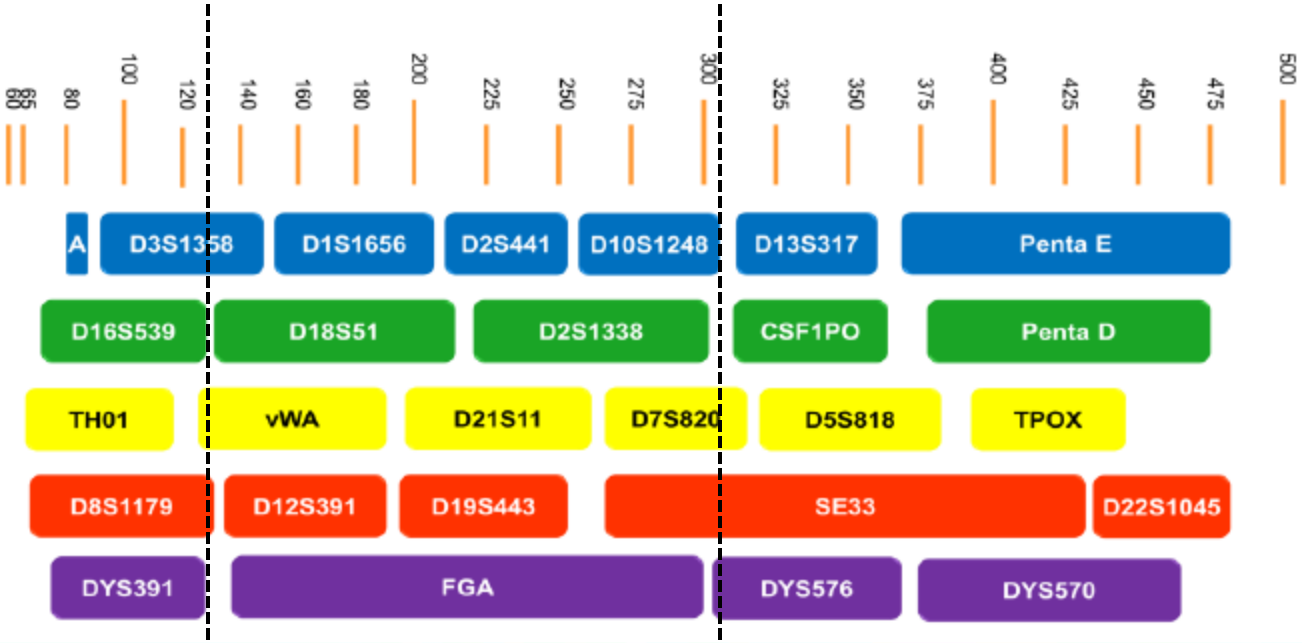


PowerSeq™ Systems

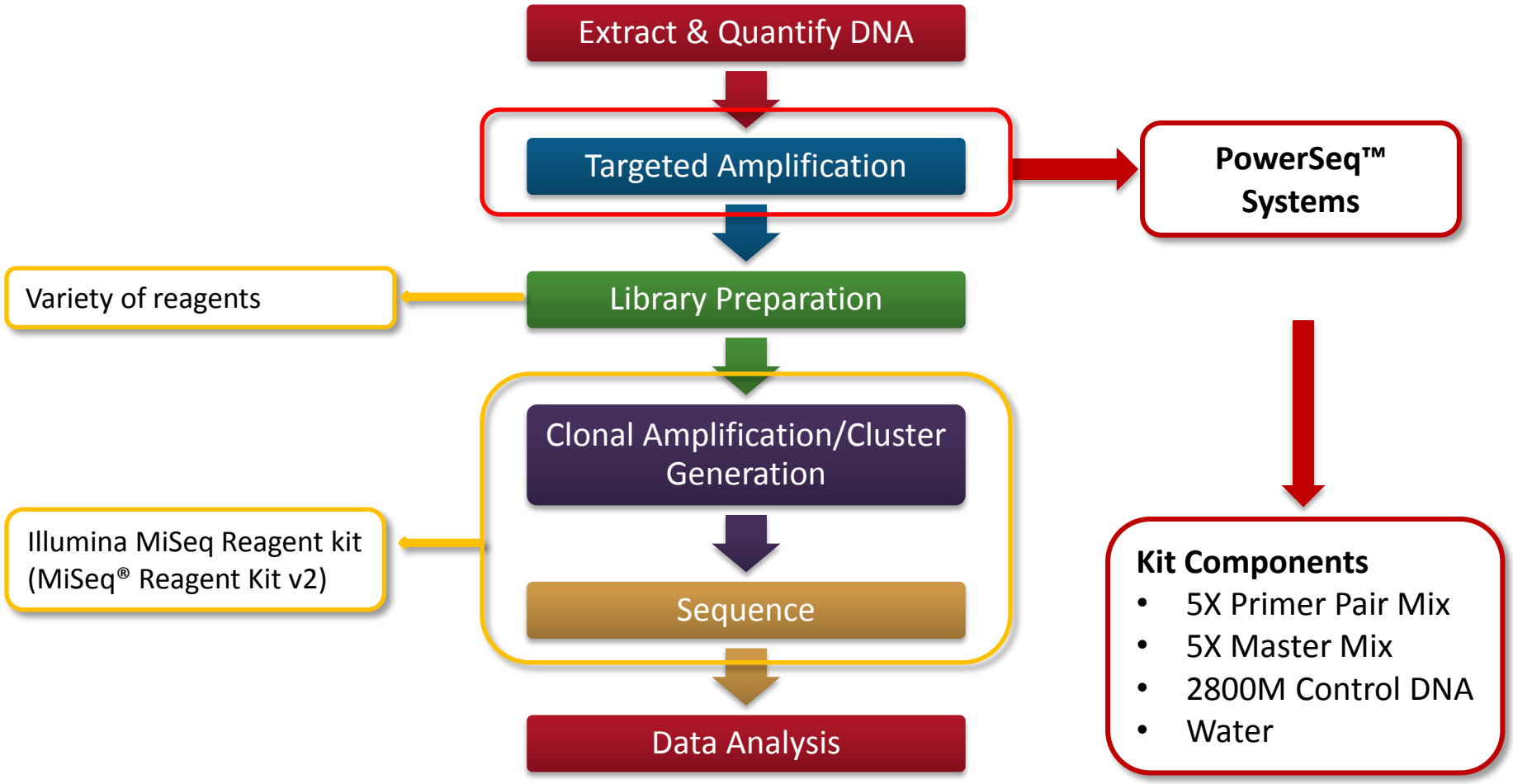
Promega Corporation

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



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.

Component	Volume Per Reaction	Number of Reactions	Final Volume
Water, Amplification Grade	To a final volume of 25µl		
PowerSeq™ 5X Master Mix	5.0µl		
PowerSeq™ 5X Primer Pair Mix	5.0µl		
Template DNA (0.25–0.5ng)[1]	Up to 15µl		
Total volume	**25µl**		

[1]The template DNA will be added at Step 6.

Template DNA (0.25–0.5ng)

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.

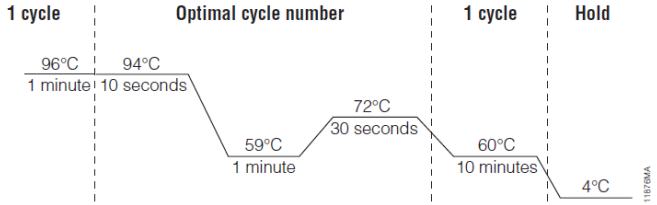


Figure 1. The thermal cycling protocol for the GeneAmp® PCR System 9700.

Amplification steps are virtually identical to existing CE-based systems

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

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

Data Interpretation

- SoftGenetics
- Battelle
- STRait Razor
- TSSV
- Others

Recent Manuscript

Forensic Science International: Genetics 16 (2015) 38–47

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing

Xiangpei Zeng [a], Jonathan L. King [a], Monika Stoljarova [a], David H. Warshauer [a], Bobby L. LaRue [a], Antti Sajantila [a,b], Jaynish Patel [c], Douglas R. Storts [c], Bruce Budowle [a,d,*]

[a] Institute of Applied Genetics, Department of Molecular and Medical Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA
[b] Department of Forensic Medicine, Hjelt Institute, University of Helsinki, P.O. Box 40, 00014 Helsinki, Finland
[c] Promega Corporation, 2800 Woods Hollow Road, Madison, WI 53711, USA
[d] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 22 August 2014
Received in revised form 20 October 2014
Accepted 26 November 2014

Keywords:
Short tandem repeat (STR)
Massively parallel sequencing (MPS)
Illumina MiSeq

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 or nearly complete profiles. These studies indicate that this STR multiplex system and the Illumina MiSeq can generate reliable STR profiles at a sensitivity level that competes with current widely used CE-based method.

© 2014 Elsevier Ireland Ltd. All rights reserved.

Depth of Coverage

500pg Genomic DNA

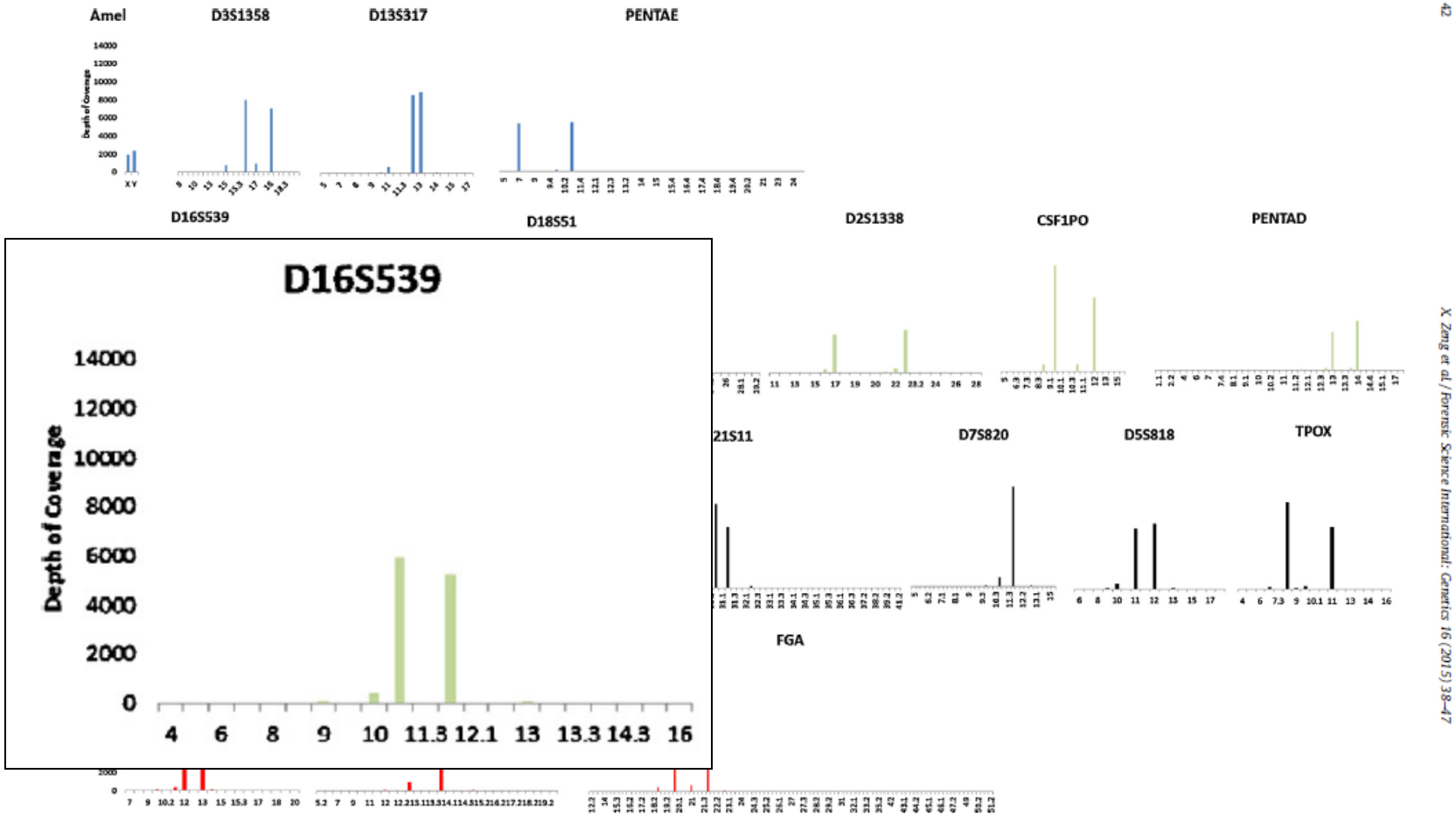


Fig. 1. A histogram portrayal of the depth of coverage by locus of one sample (no. 12) with 500 pg of input DNA.

Zeng, et al. 2015. FSI: Genetics 16:38-47

Depth of Coverage

62pg Genomic DNA

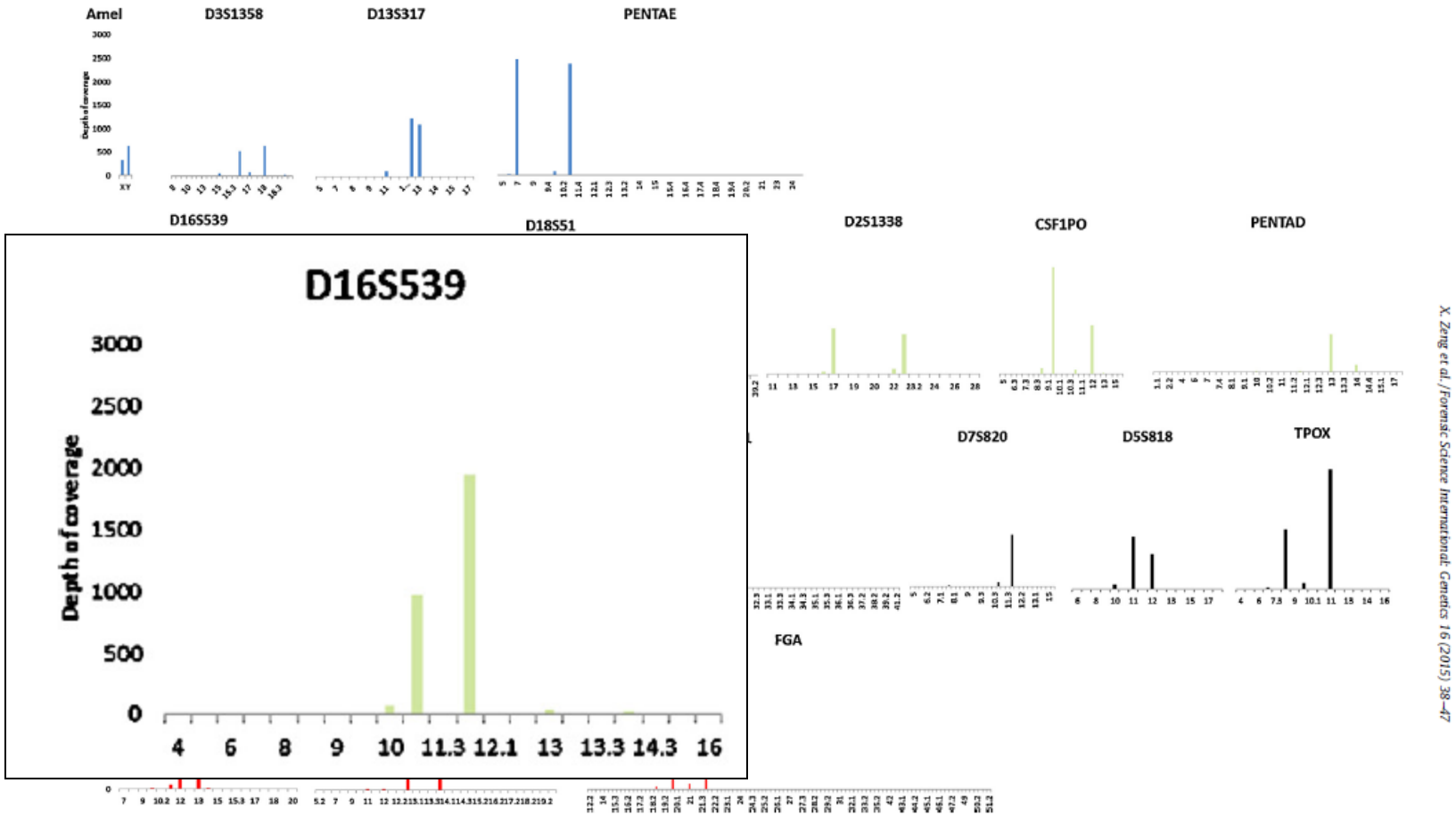


Fig. 2. A histogram result of depth of coverage by locus of one sample (no. 12) with 62 pg of input DNA.

Zeng, et al. 2015. FSI: Genetics 16:38-47

PowerSeq™ Auto System

Sequence Diversity (18-plex)

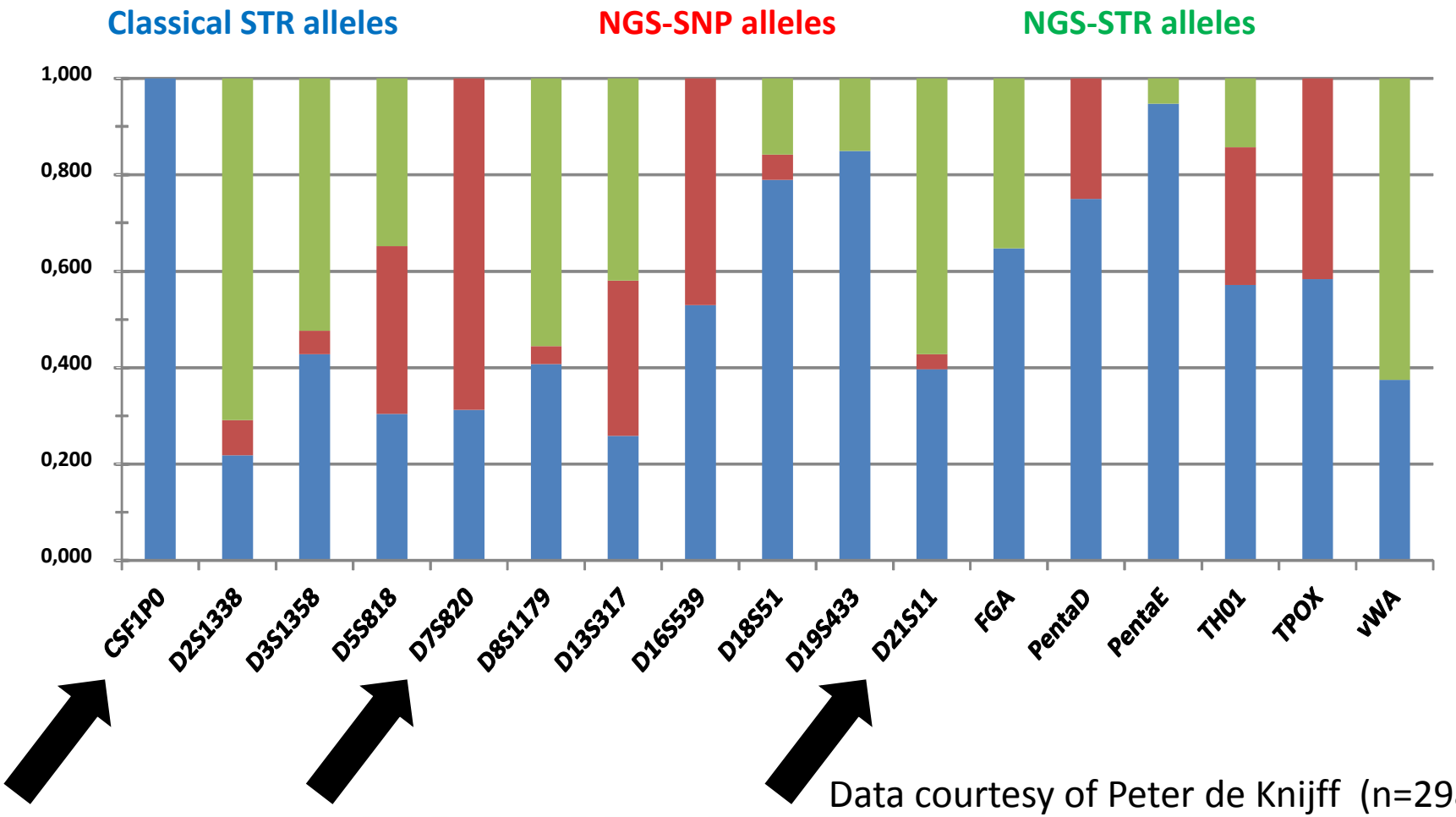
Population	Samples	Total Alleles
Dutch	102	204
Himalayan	97	194
Pygmee	99	198
Total	298	596

Data courtesy of Peter de Knijff

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

PowerSeq™ Auto System

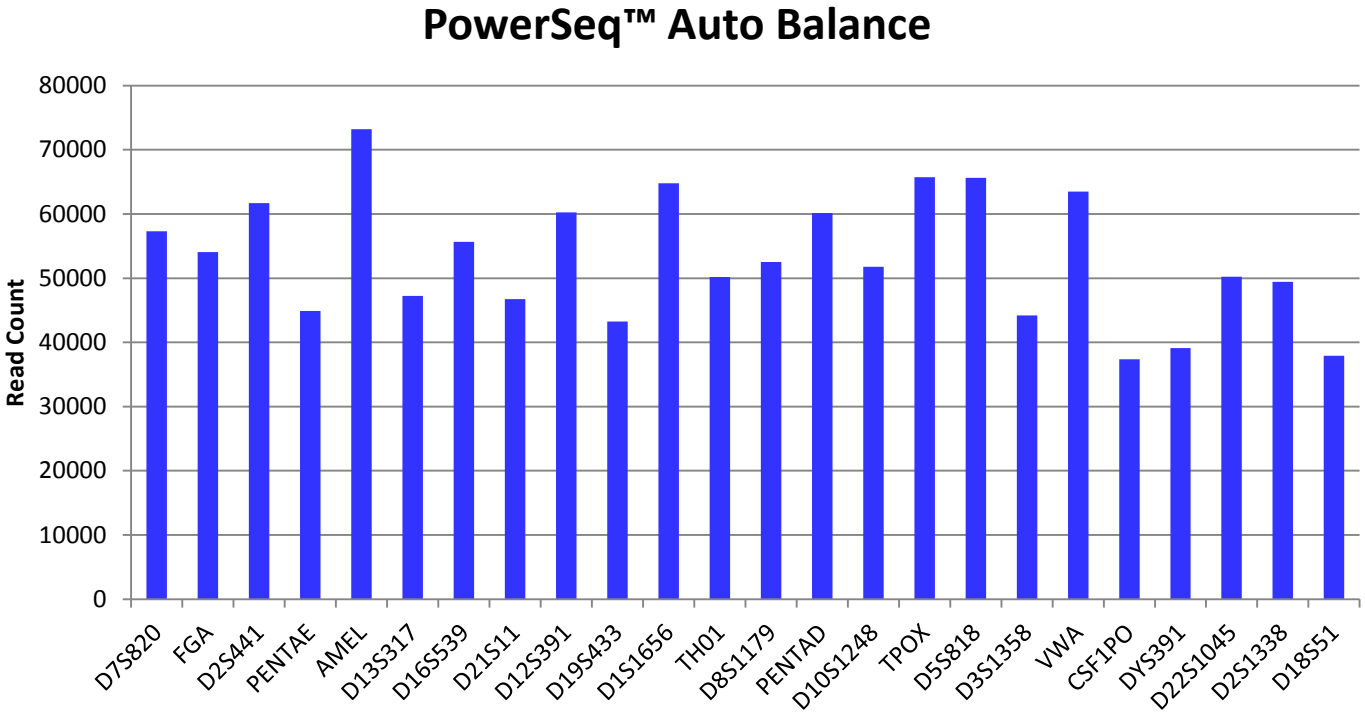
Sequence Diversity (18-plex)



Data courtesy of Peter de Knijff (n=298)

PowerSeq™ Auto - Expanded STRs to 24plex

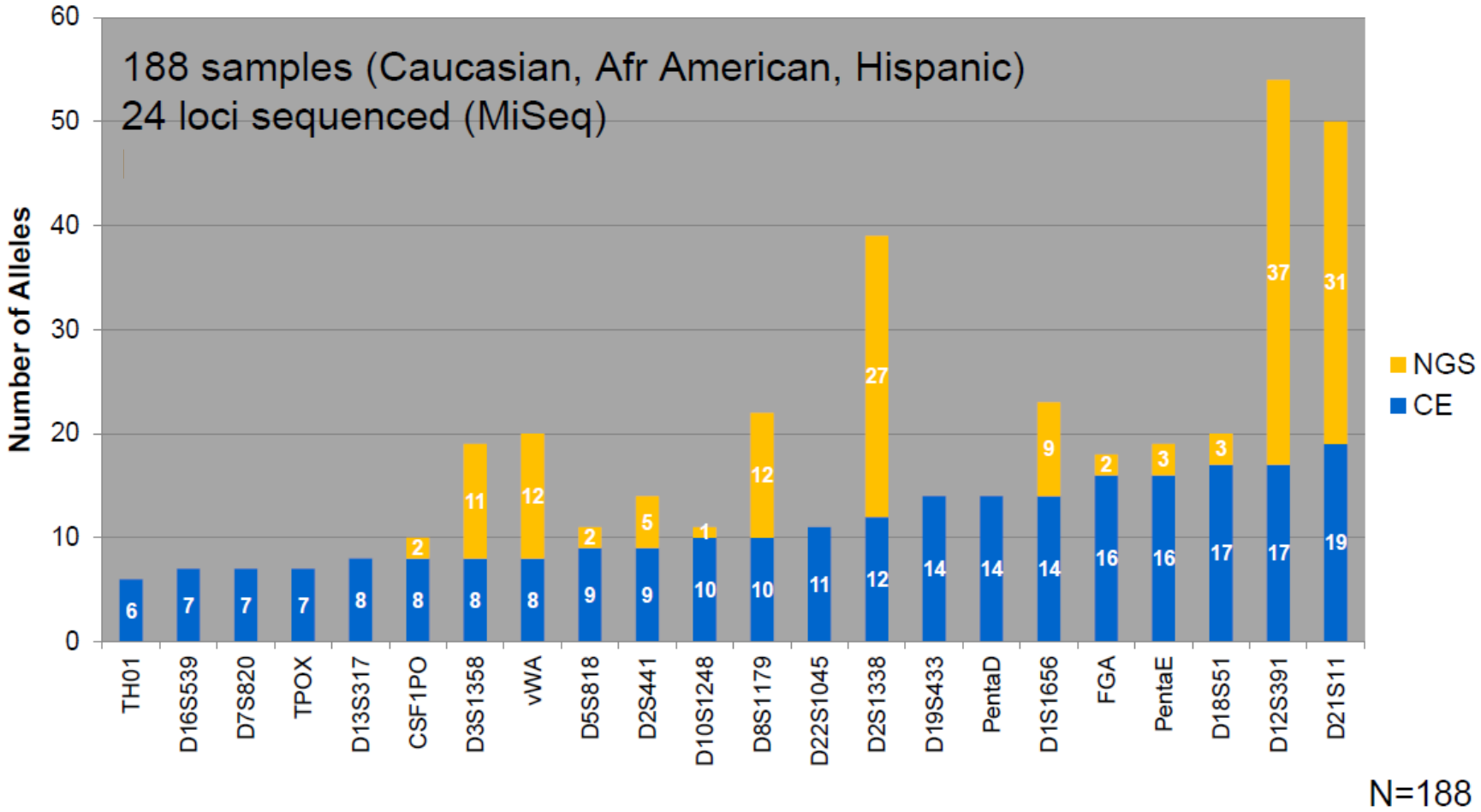
Added autosomal STRs to generate multiplex equivalent to PowerPlex® Fusion



		PowerPlex® Fusion System
		24 loci
ESS/CODIS Loci	D3S1358	•
	D8S1179	•
	D18S51	•
	D21S11	•
	FGA	•
	TH01	•
	vWA	•
CODIS Loci	D5S818	•
	D7S820	•
	D13S317	•
	D16S539	•
	CSF1PO	•
	TPOX	•
ESS Loci	D1S1656	•
	D2S441	•
	D2S1338	•
	D10S1248	•
	D12S391	•
	D19S433	•
	D22S1045	•
Other Loci	Amelogenin	•
	F13A01	
	F13B	
	FESFPS	
	LPL	
	Penta C	
	Penta D	•
	Penta E	•
	SE33	
	D6S1043	
Y-STR Loci	DYS391	•
	DYS389I/II	
	DYS439	
	DYS393	
	DYS390	
	DYS385a/b	
	DYS438	
	DYS437	
	DYS19	
	DYS576	
	DYS448	
	DYS481	
	DYS549	
	DYS533	
	DYS570	
	DYS635	
	DYS392	
	DYS643	
	DYS458	
	DYS456	
	YGATA-H4	

PowerSeq™ Auto System

Sequence Diversity (24-plex)



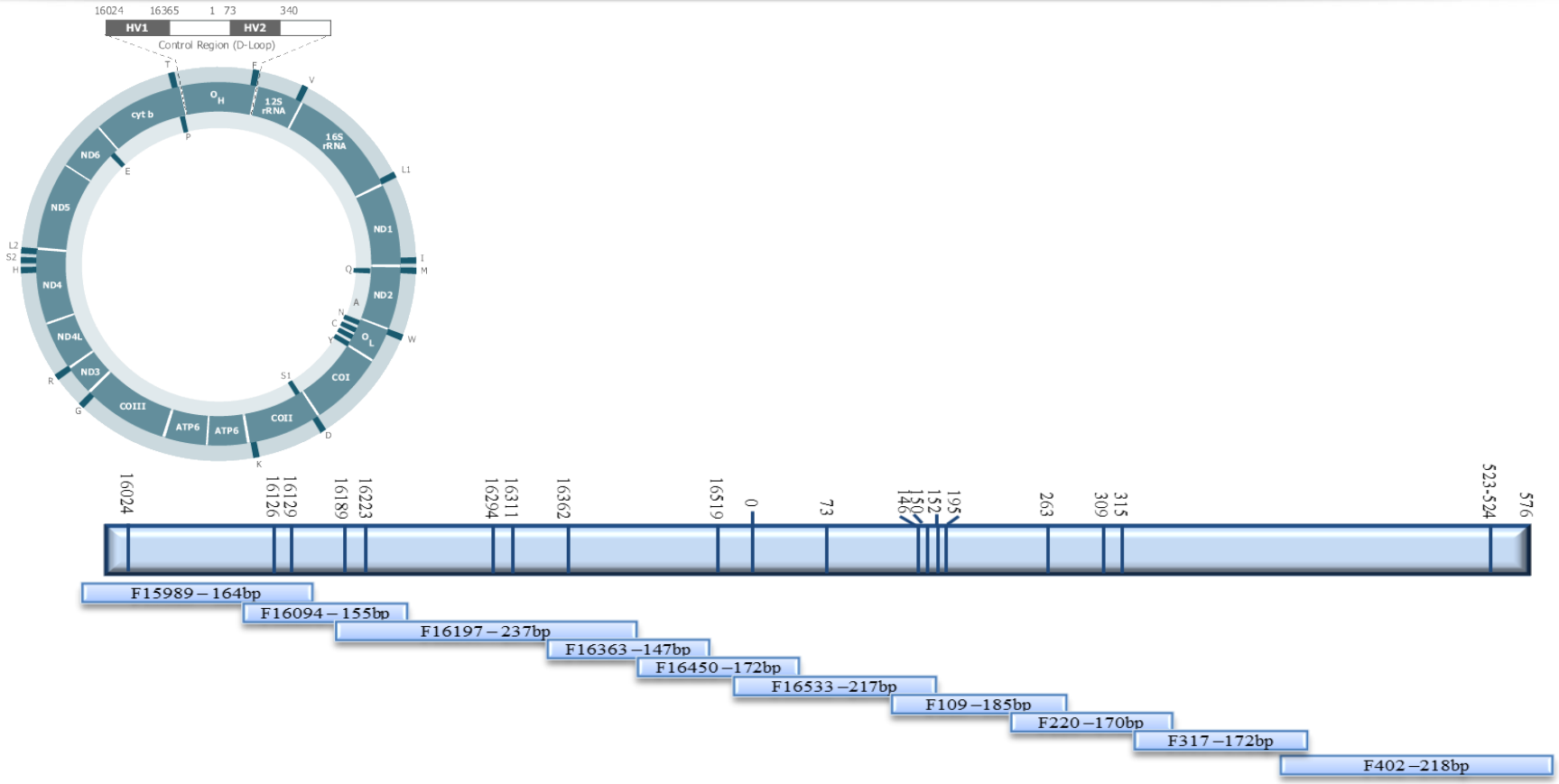
Data courtesy of the Applied Genetics Group at NIST

Expanding Beyond Autosomal STRs

Promega Corporation

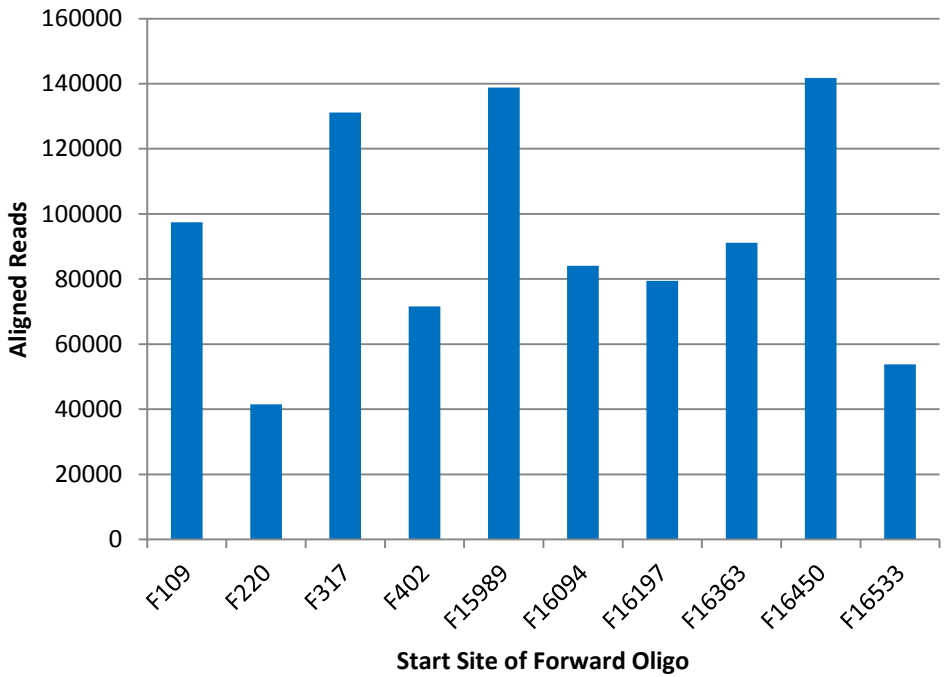
©2013 Promega Corporation. Proprietary Information. Not for further distribution.

Sequencing of Mitochondrial Control Region



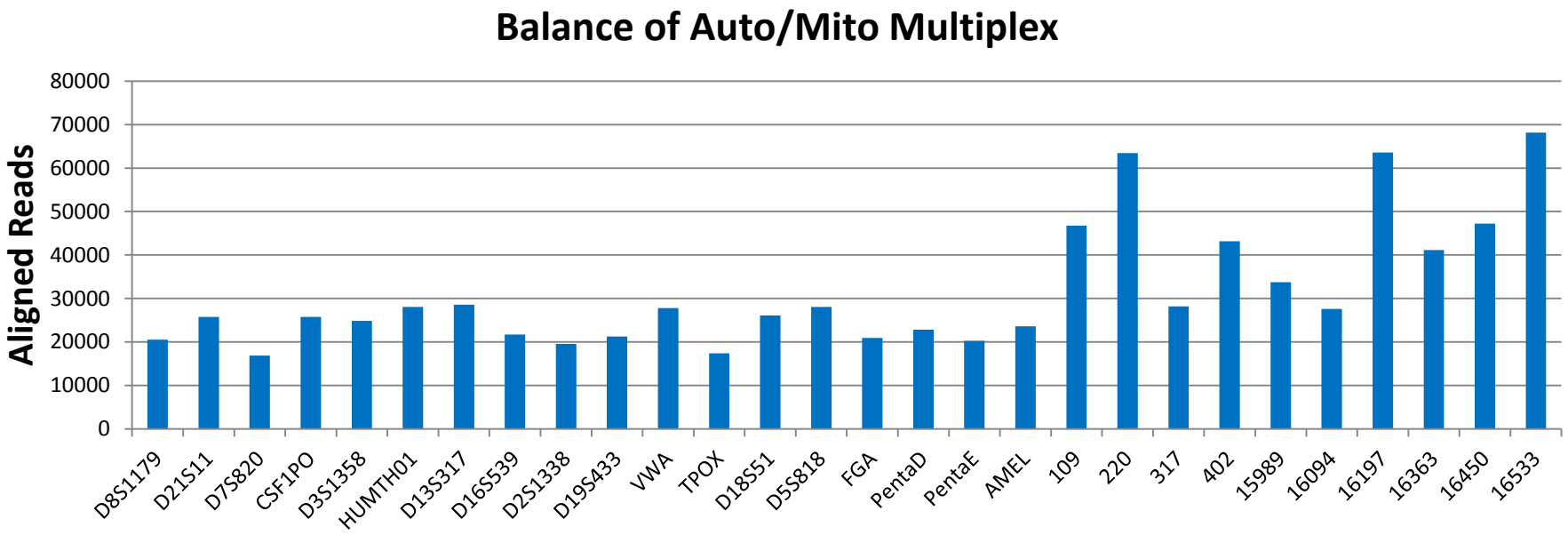
Modification of method described by Eichmann, C. and W. Parson. 2008. 'Mitominis': multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples. Int. J. Legal Med, 122:385-388

PowerSeq™ Mito



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

Combined Auto/Mito Multiplex



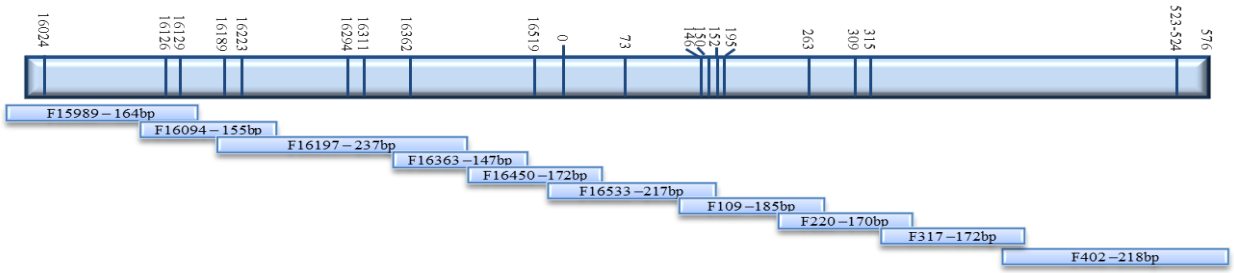
Demonstrate ability to combine STR primers from 18plex with primers for mitochondrial control region into single 5X mix and run in single amplification reaction.

A Single Multiplex For Multiple Applications

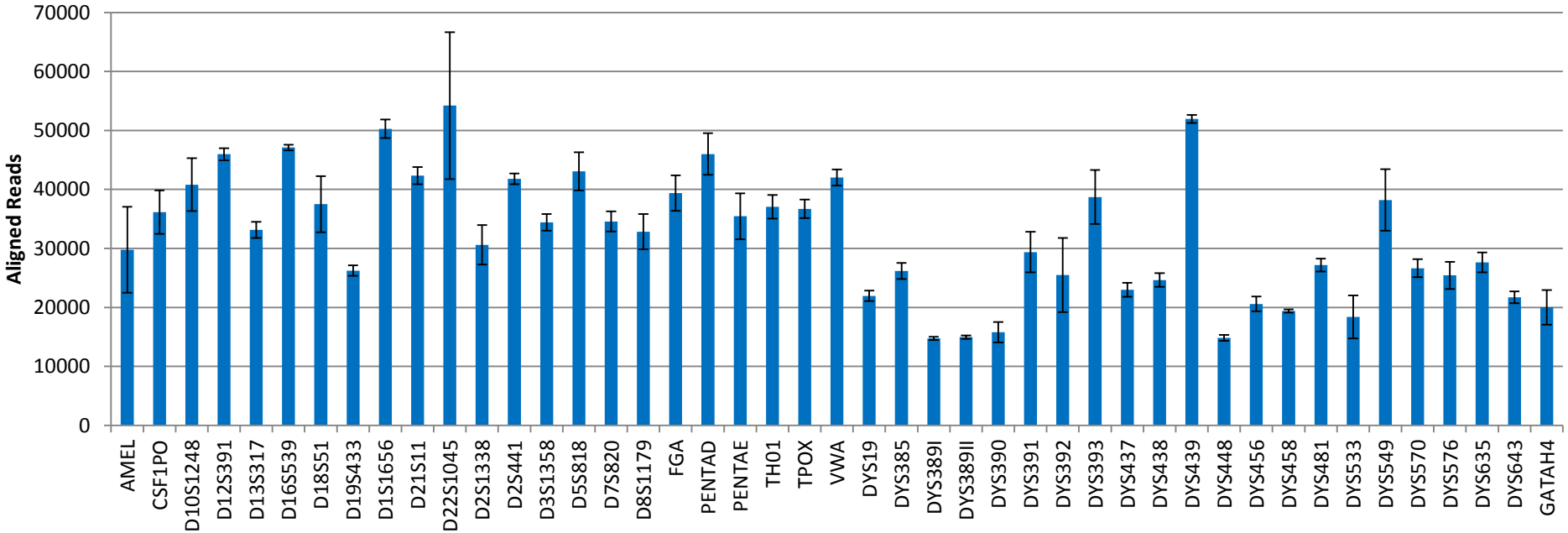
- Single multiplex containing autosomal loci (23-24), Y STR loci (23) and the mtDNA control region
- Equivalent to PowerPlex® Fusion, PowerPlex® Y23 and mitochondrial control region in a single amplification

		PowerPlex® Fusion System	PowerPlex® Y23 System
		24 loci	23 loci
ESS/CODIS Loci	D3S1358	•	
	D8S1179	•	
	D18S51	•	
	D21S11	•	
	FGA	•	
	TH01	•	
	vWA	•	
CODIS Loci	D5S818	•	
	D7S820	•	
	D13S317	•	
	D16S539	•	
	CSF1PO	•	
	TPOX	•	
ESS Loci	D1S1656	•	
	D2S441	•	
	D2S1338	•	
	D10S1248	•	
	D12S391	•	
	D19S433	•	
	D22S1045	•	
Other Loci	Amelogenin	•	
	F13A01		
	F13B		
	FESFPS		
	LPL		
	Penta C		
	Penta D	•	
	Penta E	•	
	SE33		
	D6S1043		
Y-STR Loci	DYS391	•	•
	DYS389I/II		•
	DYS439		•
	DYS393		•
	DYS390		•
	DYS385a/b		•
	DYS438		•
	DYS437		•
	DYS19		•
	DYS576		•
	DYS448		•
	DYS481		•
	DYS549		•
	DYS533		•
	DYS570		•
	DYS635		•
	DYS392		•
	DYS643		•
	DYS458		•
	DYS456		•
	YGATA-H4		•

PowerSeq™ Mito System



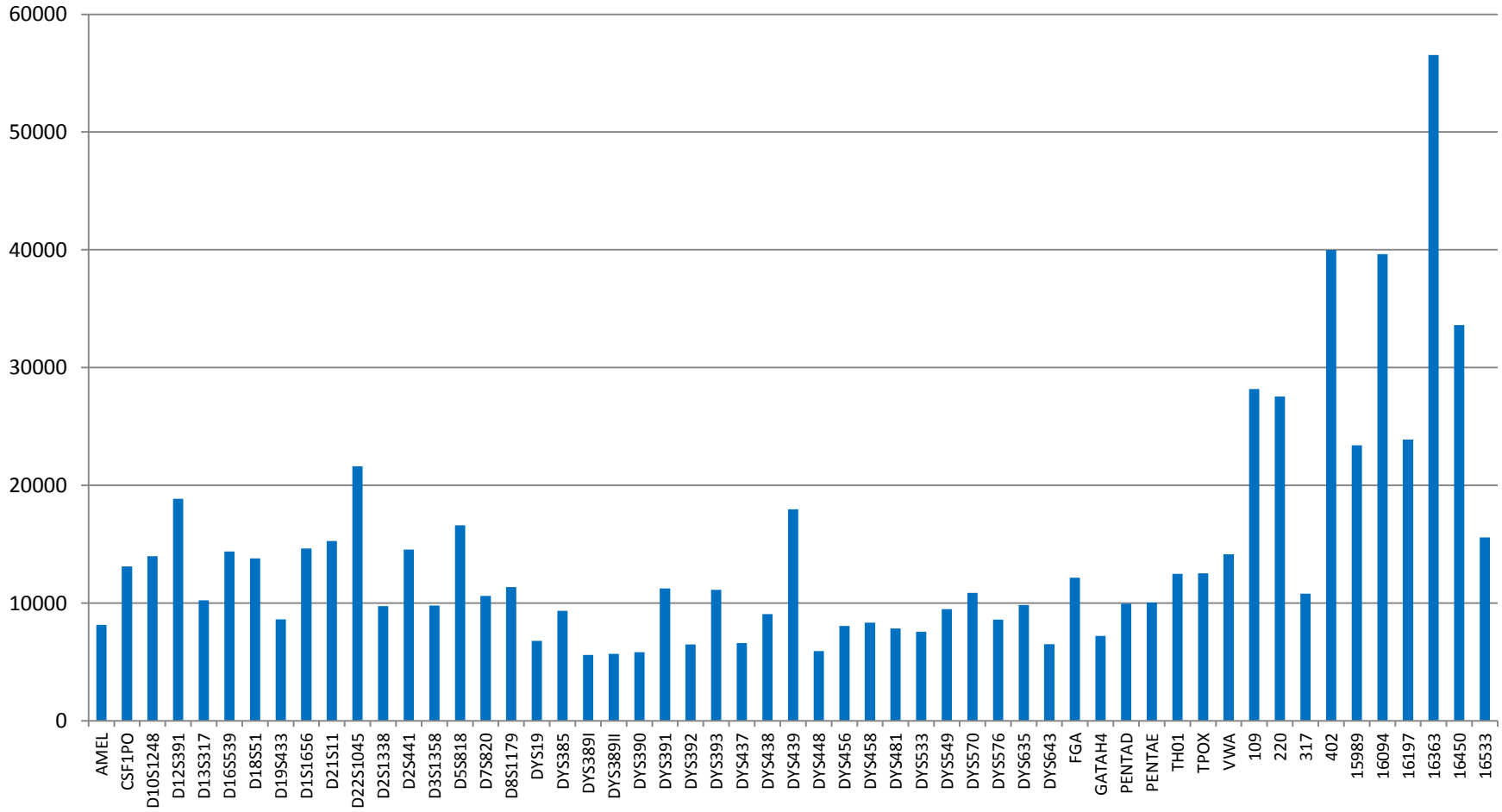
PowerSeq™ Auto/Y System



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?

©2013 Promega Corporation. Proprietary Information. Not for further distribution.

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

©2013 Promega Corporation. Proprietary Information. Not for further distribution.

PowerPlex® Y23 System

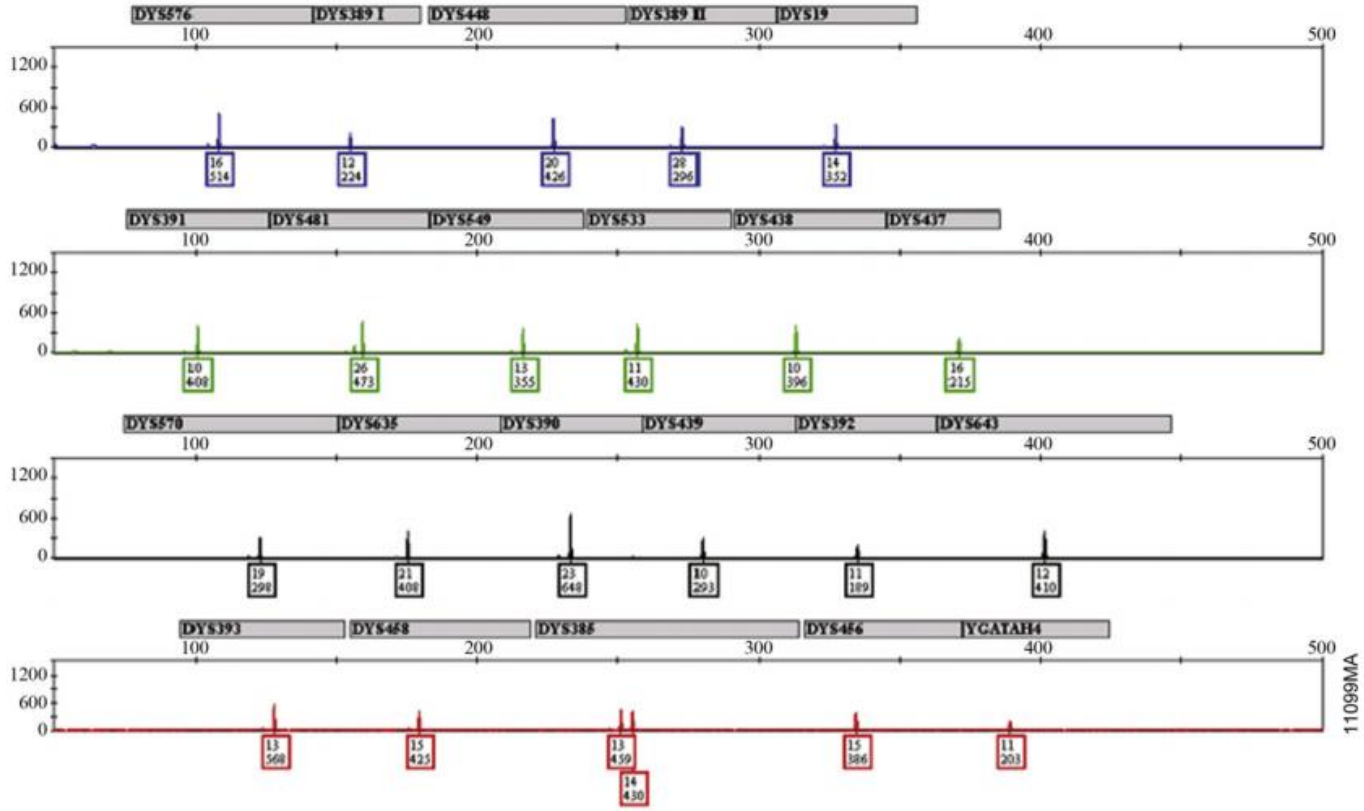
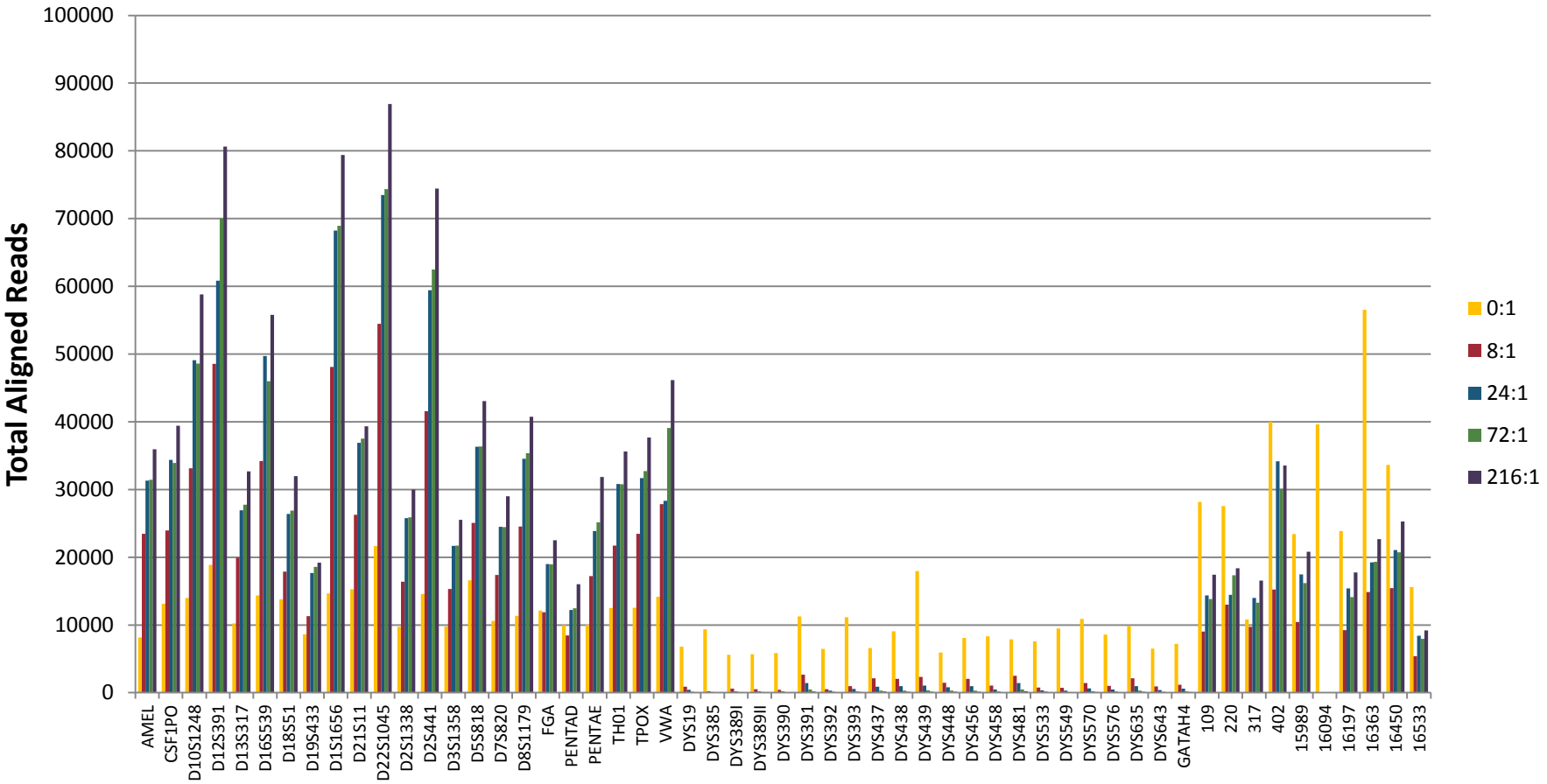


Fig. 9. Amplification of a male/female DNA mixture sample containing 125 pg of male DNA and 3000 ng of female DNA. Amplification reactions were performed using a GeneAmp® 9700, and 1 μl of each sample was electrophoresed on a 3130 series instrument using a 3 kV, 5-s injection.

24,000:1 female/male mixture

PowerSeq™ Auto/Mito/Y System

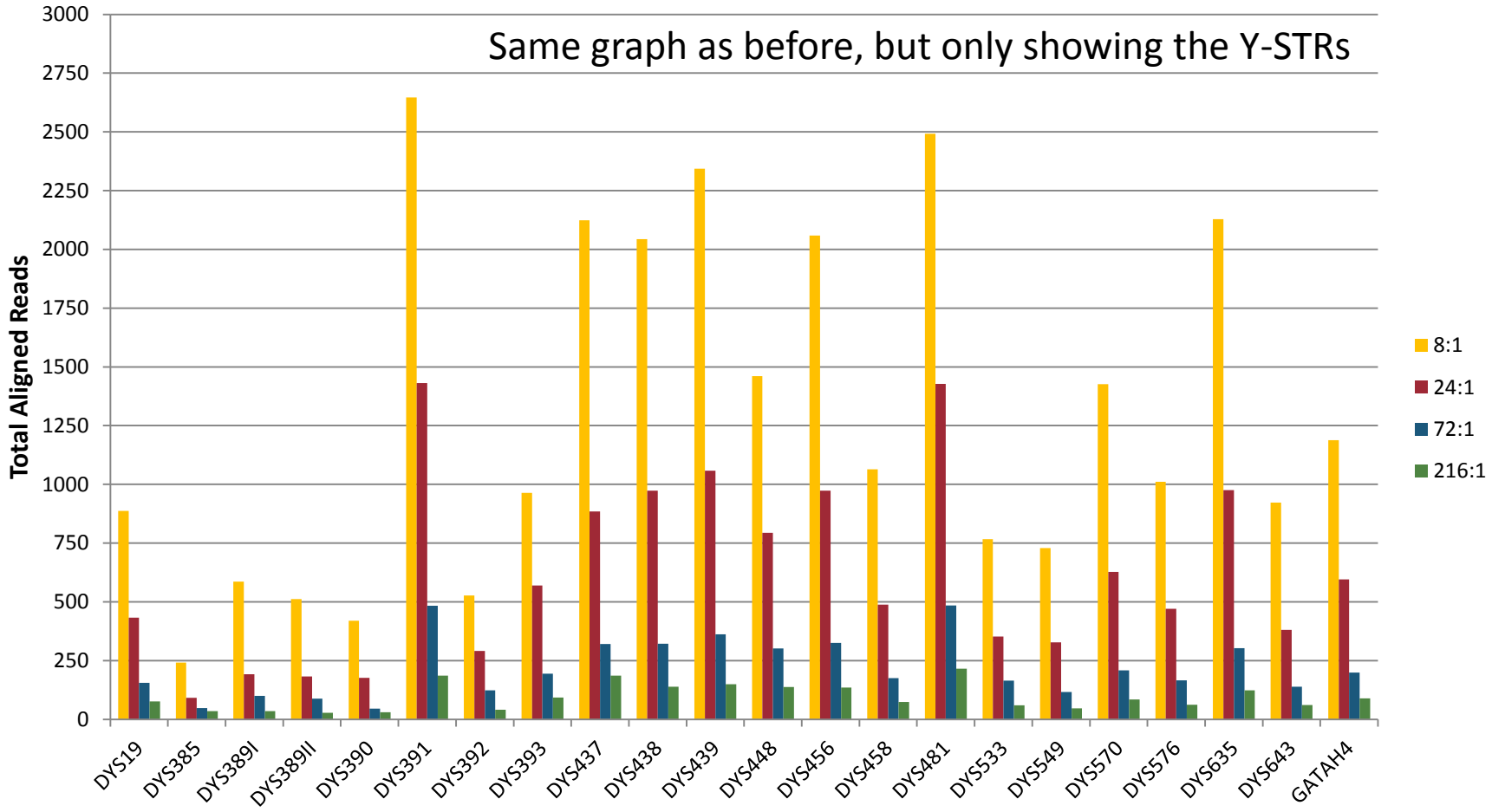
Female:Male Mixtures



Mixtures contained 400ng of female DNA

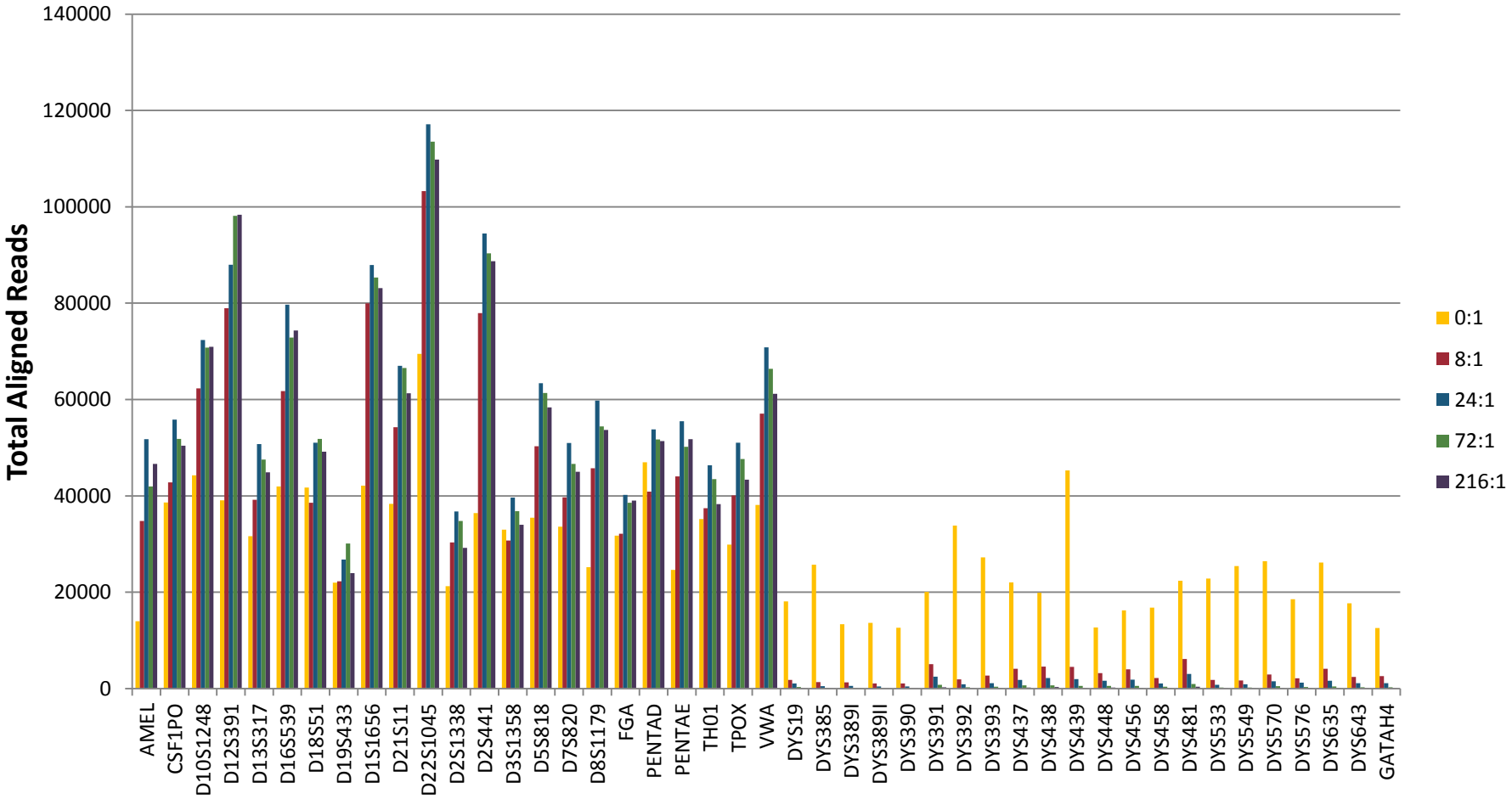
PowerSeq™ Auto/Mito/Y System

Female:Male Mixtures



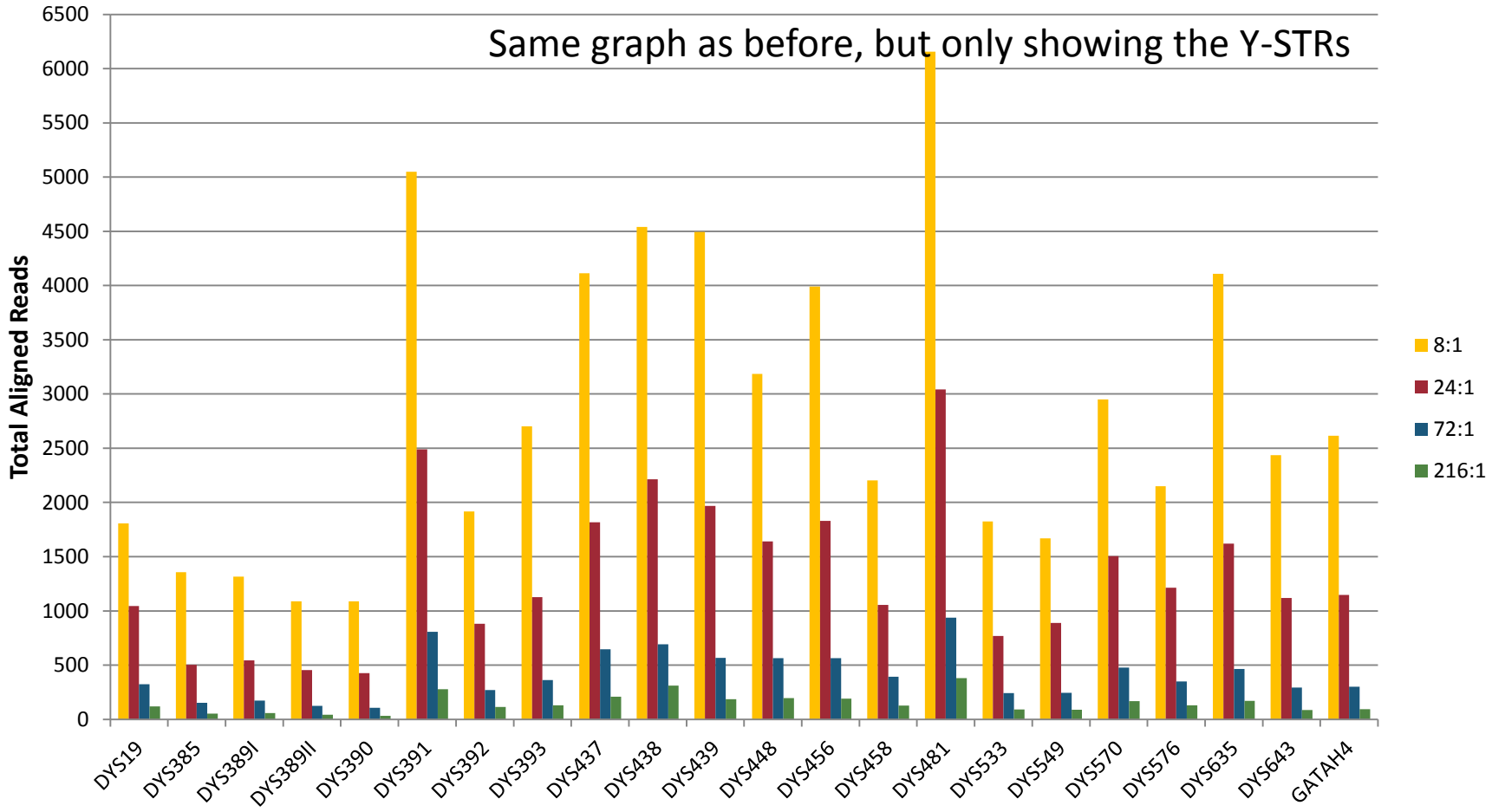
PowerSeq™ Auto/~~Mito~~/Y System

Female:Male Mixtures



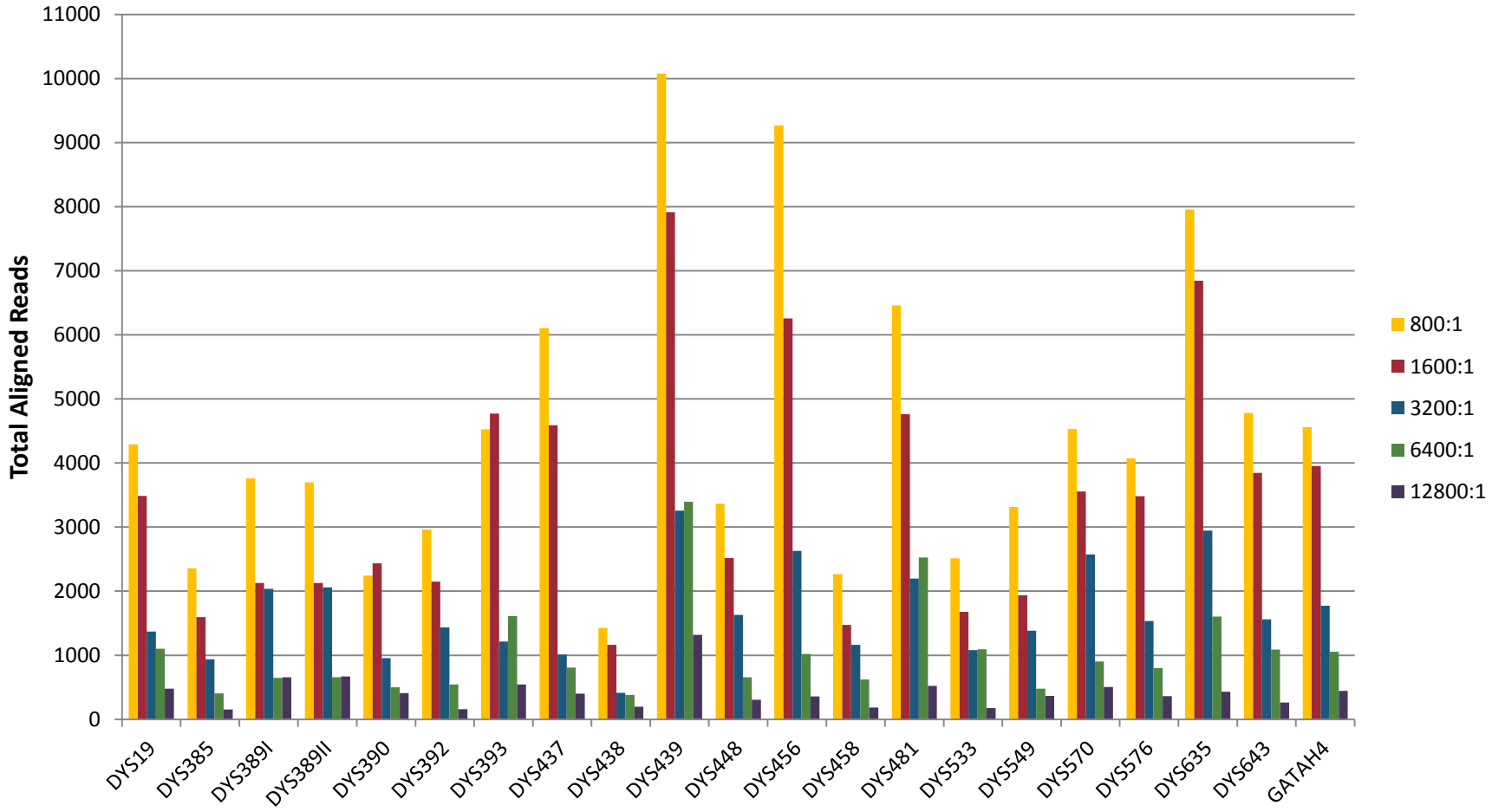
PowerSeq™ Auto/~~Mito~~/Y System

Female:Male Mixtures

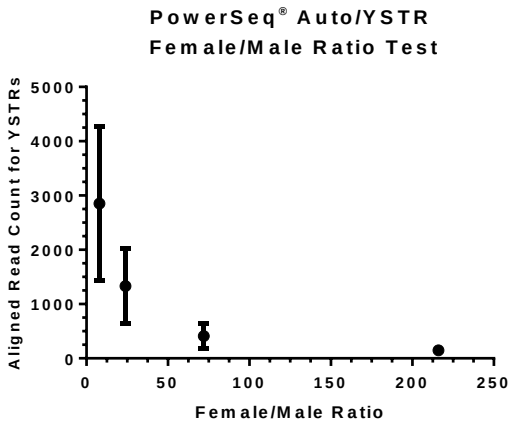


PowerSeq™ ~~Auto/Mito~~/Y System

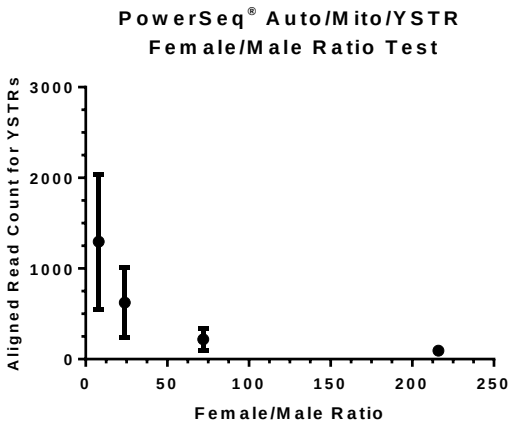
Female:Male Mixtures



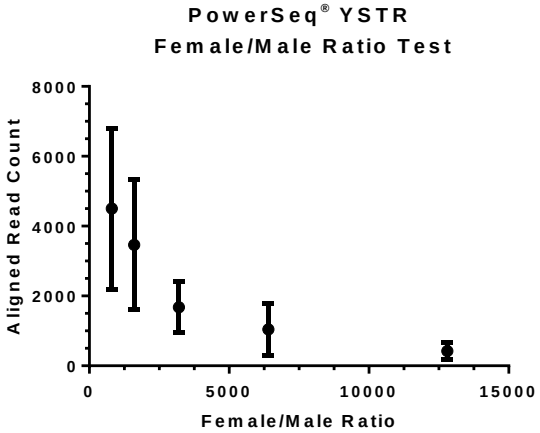
Summary of Y-STR Data



Ratio	Mean	SD
8	2853	1419
24	1332	692
72	412	224
216	148	89



Ratio	Mean	SD
8	1297	745
24	623	392
72	219	127
216	95	56



Ratio	Mean	SD
800	4501	2311
1600	3460	1866
3200	1674	728
6400	1043	737
12800	424	256

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
  - Thirsa Kraaijenbrink
  - Rick van Leeuwen
- NIST
  - Peter Vallone
  - Katherine Gettings
- UNTHSC
  - Bruce Budowle
  - Xiangpei Zeng
  - Jonathan King
  - Monika Stoljarova
  - David Warshauer
  - Bobby LaRue

©2013 Promega Corporation. Proprietary Information. Not for further distribution.

Thank you