Initial Assessment of the Precision ID Globalfiler Mixture ID Panel on the Ion Torrent S5XL DNA Sequencer and Converge V2.0 Software

Kevin M. Kiesler[1], Carolyn R. Steffen[1], Michael D. Coble[1], Matt N. Gabriel[2], Ravi Gupta[2], and Peter M. Vallone[1]

[1]U.S. National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8314, USA
[2]Thermo Fisher Scientific, Human Identification Division, South San Francisco, CA, USA

National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce
Email: Kevin.Kiesler@nist.gov

P01-29
Abstract # 170251

The U.S. National Institute of Standards and Technology (NIST) has recently participated in an assessment of the Precision ID Globalfiler early access Mixture ID Panel (v1.0), which consists of primers for amplification of over 100 forensically relevant loci in the human genome. Markers included in the panel can be divided into four classes: short tandem repeats (STR) (29 autosomal STRs and 1 Y STR), single nucleotide polymorphisms (SNPs) (42 autosomal [1, 2] and 2 Y chromosome), insertion/deletion polymorphisms (Indels) (Amelogenin and Y Indel rs2032678), and microhaplotype blocks (MH) (36 clusters of 2 to 4 SNPs [3]). Several challenging sample types were sequenced and analyzed, including: artificially degraded DNA, multiple-contributor mixtures, mixtures of related individuals, and low DNA input samples. Performance characteristics (e.g. allele coverage ratio, interlocus balance, stutter, drop in/out, benefits and challenges of sequencing versus fragment size analysis) of the Mixture ID Panel relative to state-of-the-art capillary electrophoresis (CE) methods will be presented. Additionally, capabilities of the accompanying Converge v2.0 (beta release) analysis software will be discussed.

Specific Aim: Assess performance of the Precision ID Globalfiler Mixture ID Panel performance relative to state-of-the-art capillary electrophoresis, where possible.

STR Markers

INTRODUCTION: Compatible with traditional CE technology and databases, STR markers are well characterized for their allele frequencies in large population sets, have established interpretation methodology, and known stutter artifact characteristics. Here we examine sequencing technology's benefit to mixture deconvolution by virtue of the additional information contained in the sequence of an allele beyond what CE length-based methods provide.

Four *simulated* mixture evidence samples and scenarios generated for STR evaluation:

- Mix1 - Two person mixture of non-degraded DNA and degraded DNA. "Environmentally degraded evidence at a crime scene contaminated by a crime scene investigator."
- Mix2 - Two person mixture. "Swab of bite mark on neck; good samaritan attacked at bus stop by irrational individual threatening a young woman."
- Mix3 - Three person mixture. "Vaginal swab with suspect, victim, and unknown contributor."
- Mix4 - Four person mixture. "DNA swab of handgun found in vehicle, weapon matched by ballistics to murder scene."

METHODS:
- 1 ng total DNA input to Mixture ID and Globalfiler kits
- Mixture ID run on Ion Torrent S5xl, Globalfiler run on ABI 3500xl
- Data analysis with Converge 2.0 (Ion Torrent) or GeneMapper (CE)
- Mixture Analysis with EuroForMix [4], fully continuous model, runs in "R"
- Comparison of Likelihood Ratio (LR) and deconvolved mixture component ratios
  - LR = Likelihood|Hp/Likelihood|Hd
  - Sequence-based (NGS) or size based (CE) allele frequencies were used [5]
  - STR markers present in both kits (n = 20) used for calculations & comparison

RESULTS:
- Isoalleles for up to five loci were observed in all mixtures (see Table 1)
  - Isoalleles = alleles indistinguishable by size, but resolved by sequencing
- Calculated LR were greater for Mixture ID STRs
  - **Exception**: Mix1, Globalfiler detected three additional markers of degraded component

DISCUSSION:
- The increase in LR for NGS data (versus CE) was primarily due to:
  - More allelic diversity in mixture data
  - More allelic uniqueness in frequency table
- Mixture deconvolution yields similar estimates of component ratios
  - CE and NGS methods perform similarly overall
- EuroForMix was not designed for NGS data
  - Proof of concept only

Figure 1: Screenshot of the EuroForMix GUI. (A) One loads in allele frequencies, evidence and reference data. (B) Model parameters are selected. (C) Likelihoods, LR ratio, and contributor ratios are output.

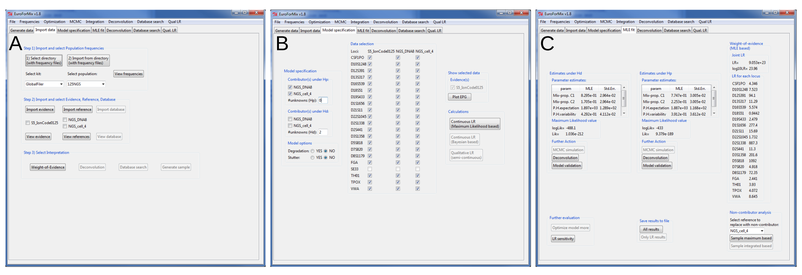


Table 1: Isoalleles detected in mixtures by sequencing.

Mixture	Added # of Alleles	Isoallele Detected at STR Locus
Mix1	1	D8S1179
Mix2	3	D12S391, D3S1358, and D8S1179
Mix3	5	D2S1338, D2S441, D8S1179, D12S391, and D21S11
Mix4	4	D2S1338, D3S1358, D12S391, and D21S11

Table 2: Likelihood ratios (Log_{10}) calculated for each of the four mixtures analyzed with either the Mixture ID panel on the Ion Torrent S5 or Globalfiler kit on the ABI 3500 xl. Twenty STR markers present in both the Mixture ID panel and Globalfiler kit were used for calculations. LR_{log10} is greater for Mixture ID data except in Mix1 because in that case CE detected three additional alleles from the degraded component. In all other mixtures, equal numbers of alleles were detected, with the exception of isoalleles.

Mixture	Mixture ID on Ion Torrent S5 Likelihood Ratio (Log10)	GlobalFiler on ABI3500xl Likelihood Ratio (Log10)
Mix1	4.63	17.15
Mix2	23.96	21.81
Mix3	29.16	19.40
Mix4	100.64	67.22

Table 3: Mixture component ratios calculated for each of the four mixtures analyzed with either the Mixture ID panel on the Ion Torrent S5 or Globalfiler kit on the ABI 3500 xl. Both platforms yielded similar estimates of mixture ratios.

Mixture Component	Mixture ID on Ion Torrent S5 Mixture Proportion	Globalfiler on ABI3500xl Mixture Proportion
Mix1		
Component 1	94.5 % ± 1.7 %	92.4 % ± 1.4 %
Component 2	5.5 % ± 1.7 %	7.7 % ± 1.4 %
Mix2		
Component 1	77.5 % ± 3.0 %	81.7 % ± 1.6 %
Component 2	22.5 % ± 3.0 %	18.3 % ± 1.6 %
Mix 3		
Component 1	61.5 % ± 6.1 %	59.8 % ± 6.0 %
Component 2	31.2 % ± 1.6 %	31.2 % ± 1.9 %
Component 3	7.3 % ± 4.8 %	9.1 % ± 4.4 %
Mix4		
Component 1	21.3 % ± 2.6 %	20.7 % ± 2.2 %
Component 2	57.3 % ± 2.8 %	55.8 % ± 2.7 %
Component 3	12.2 % ± 1.3 %	12.9 % ± 1.2 %
Component 4	9.2 % ± 2.6 %	10.5 % ± 2.5 %

Microhaplotype Markers

INTRODUCTION: A novel marker type in the forensic setting, microhaplotypes (MH) are blocks of two, three, or four SNPs found in a single stretch of less than 200 bp of DNA. They are found on all human chromosomes, are multi-allelic within and between populations, and are useful for individual identification and ancestry prediction. These markers lack the stutter artifact of STRs, however, they also lack the allelic diversity of STR marker systems.

METHODS:
- DNA mixtures of two, three, four, five, and eight contributors at varied ratios prepared
  - Examined with Mixture ID panel on Ion Torrent S5xl
- Data analyzed with Converge 2.0
- Manually deconvoluted some MH loci with no allelic overlap for mixtures 1-4 at left

RESULTS:
- Microhaplotypes generally predict minimum # of contributors (see Table 5)
  - Mostly accurate for 2, 3, and 4 component mixtures; occasional artifacts
  - Higher numbers of contributors and family trios were underestimated
- Deconvolved microhaplotype data matches STR ratio estimates (see Table 4)

DISCUSSION:
- Practical upper limit of contributor estimate using current microhaplotype content
  - Four contributors maximum
- Microhaplotypes can be used for additional deconvolution functionality
- At least three markers should be used to indicate the presence of a mixture
  - Occasional artifacts could cause misclassification (likely early software issue)

Figure 2: Converge 2.0 output showing microhaplotype data for a **three person mixture**. Some alleles are shared between samples. Therefore, not all markers indicate that this is a three-person mixture.

Locus	Genotype	Allele count	Min contributors	IGV
mh13KK-217	AGCG, TGCG, AACG, AATG, AGCA	5	3	
mh11KK-180	ACTC, GCCG, GCTC, AATC, ACCC	5	3	
mh21KK-320	GGCG, GATA, GACA, AACG, AACA	5	3	
mh05KK-170	TAAG, CGGG, CGAA, CAAA, CAAG, TAAA	6	3	
mh13KK-218	TCTT, CTTT, CTTC, TTCT, TTCC, CTCT	6	3	
mh05KK-062	AC, AA, TA	3	2	
mh15KK-104	TAA, TCG, CAG	3	2	
mh01KK-106	TAGG, CGAG, CAGA, CAGG	4	2	
mh15KK-067	GC, GT, TT, TC	4	2	
mh11KK-187	GCGG, CCCA, GTGG	3	2	

Figure 3: Converge 2.0 output showing graphical depiction of NGS reads for 20 (of 36) microhaplotype loci for the three person mixture in Figure 2.

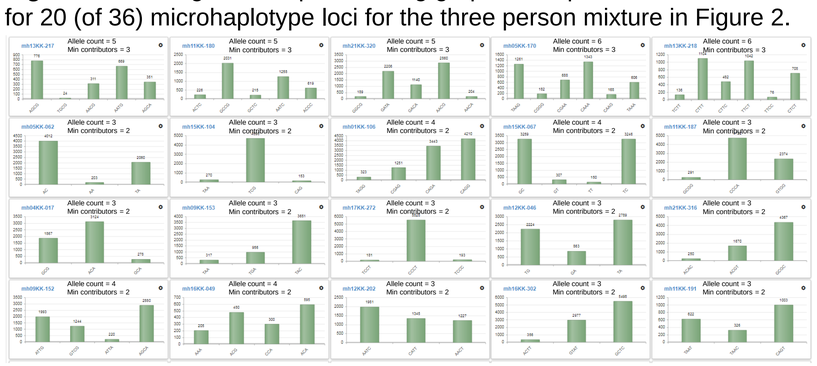


Table 4: Mixture component ratios for microhaplotype markers compared with STR markers. Estimates are similar for both marker types. The microhaplotype data did not detect the degraded component in Mix 1.

Mixture Component	Mixture ID STRs on Ion Torrent S5 Mixture Proportion	Mixture ID Microhaplotypes on S5 Mixture Proportion
Mix1		
Component 1	94.5 % ± 1.7 %	100 %
Component 2	5.5 % ± 1.7 %	Did not detect degraded minor component
Mix2		
Component 1	77.5 % ± 3.0 %	83.7 % ± 3.8 %
Component 2	22.5 % ± 3.0 %	16.3 % ± 1.6 %
Mix 3		
Component 1	61.5 % ± 6.1 %	61.3 % ± 2.0 %
Component 2	31.2 % ± 1.6 %	33.1 % ± 2.1 %
Component 3	7.3 % ± 4.8 %	5.7 % ± 2.1 %
Mix4		
Component 1	21.3 % ± 2.6 %	32.9 % ± 3.5 %
Component 2	57.3 % ± 2.8 %	53.8 % ± 3.4 %
Component 3	12.2 % ± 1.3 %	8.2 % ± 3.0 %
Component 4	9.2 % ± 2.6 %	5.0 % ± 1.4 %

Table 5: Predicted minimum number of contributors by microhaplotype markers in the Mixture ID panel for 21 DNA mixtures designed to test the capabilities of the panel.

Components in Mixture (n)	Microhaplotype Prediction (Minimum # Contributors)	Description
2	2	1:1 ratio, one degraded 200 bp
2	2 *	1:1 ratio
2	1 *	1:1 ratio, one degraded 300 bp
2	2	1:1 ratio
2	2	25:1 ratio, 10 ng DNA input
2	2	50:1 ratio, 10 ng DNA input
2	2 *	25:1 ratio, 1 ng DNA input
2	2	1:1 ratio
2	2	4:1 ratio
2	2 *	9:1 ratio
3	2	1:1:1 ratio, family trio
3	2	6:3:1 ratio, family trio
3	3	1:1:1 ratio
3	3	6:3:1 ratio
3	3	7:2:1 ratio
4	3 *	1:1:1:1 ratio
4	3 *	3:3:2:2 ratio
4	3	6:2:1:1 ratio
5	3	6:5:3:2:1 ratio, family trio plus two unrelated individuals
5	4	6:1:3:3:6 ratio
8	3 *	3:1:1:1:1:1:1:1 ratio

* A single additional marker plus the number listed supports an additional contributor

SNP Markers

INTRODUCTION: Single nucleotide polymorphisms have been in use in forensics for some time in the context of identity matching as well as ancestry and phenotype prediction. The adoption of sequencing technologies is enabling SNP multiplex assays which rival the discrimination power of STR systems. These markers lack stutter artifacts and perform well with degraded DNA due to their smaller amplicon size.

METHODS:
- DNA sheared using Covaris S2 sonicator
- Sheared DNA fractionated using Blue Pippin instrument
  - Fractions contain < 100 bp, < 200 bp, or < 300 bp DNA fragments
- DNA quantity estimated by Quantifiler Trio
- 1 ng total DNA input to Mixture ID on Ion Torrent S5xl
- Data analysis in Converge 2.0
- Random match probabilities (RMP) were calculated
  - Mixture ID SNPs, Mixture ID STRs, GlobalFiler STRs
  - Only STRs present in both Mixture ID and GlobalFiler kits used for calculations

RESULTS (SNPs):
- Partial profile for <100 bp fraction
  - 8 SNPs not called due to insufficient coverage
  - 3 SNPs flagged for strand imbalance, but genotype call was made
- Full and accurate profiles generated for <200 bp and <300 bp fractions
- Discordant allele calls noted in <100 bp fraction
  - Allelic dropout (n = 4) where a heterozygote was called homozygous
  - Incorrect genotype (n = 1) supported by 78 reads, no warning flags

DISCUSSION:
- SNP markers performed well with highly degraded DNA
  - Outperforms STRs (by RMP) with extremely degraded DNA < 100 bp
  - Full and accurate profiles from < 200 bp, < 300 bp template
- Mixture ID STRs rivaled or outperformed SNPs with moderately degraded DNA
  - STR amplicons redesigned for sequencing; smaller, no need for size separation
  - Higher discrimination power using sequence based alleles

Table 6: **Random match probabilities** for Mixture ID panel SNPs, GlobalFiler STRs, and MixtureID panel STRs. No Mixture ID panel STRs produced useful signal with the most highly degraded DNA sample. Shortening of the Mixture ID STRs' amplicon size for Ion Torrent sequencing allows improved performance over CE with degraded DNA.

Degraded DNA Fraction	Mixture ID SNP on Ion Torrent S5	Mixture ID STR on Ion Torrent S5	Globalfiler STR on ABI 3500xl
100 bp	2.5×10^{-13}	0	1.2×10^{-1}
200 bp	8.5×10^{-19}	1.5×10^{-19}	4.3×10^{-11}
300 bp	8.5×10^{-19}	**2.5×10^{-35}**	1.68×10^{-19}

Figure 4: Heatmap of SNP and STR marker performance with degraded DNA fractions.

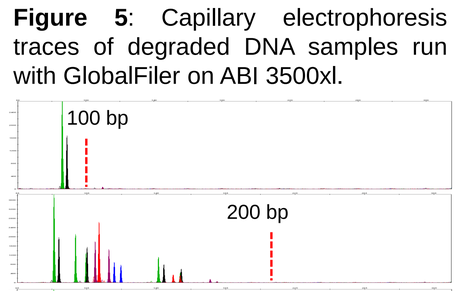
SNP Locus	Degraded DNA Fraction 100 bp	200 bp	300 bp		Mixture ID on Ion Torrent S5 100 bp	200 bp	300 bp	GlobalFiler on ABI 3500xl 100 bp	200 bp	300 bp
rs2292972	CT	CT	CT	AMEL	X	X, Y	X, Y		X, Y	X, Y
rs964681	TT	TT	TT	CSF1PO		10, 12	10, 12			
rs251934	AA	GG	GG	D10S1248			14, 16		14, 16	14, 16
rs9905977	GG	GG	GG	D12S391		22, 22	22, 22			22, 22
rs876724	CT	CT	CT	D13S317		11, 13	11, 13		11	11, 13
rs2831700		GG	GG	D16S539		11, 11	11, 11			11, 11
rs1335873	AA	AA	AA	D18S51			13, 16			
rs221956	CT	CT	CT	D19S433			14, 16.2		14, 16.2	14, 16.2
rs4847034	AA	AA	AA	D1S1656		13, 14	13, 14		13, 14	13, 14
rs6955448	CC	CT	CT	D21S11			30, 31.2			30, 31.2
rs1463729	CT	CT	CT	D22S1045			16, 16		16, 16	16, 16
rs1979255	GG	GG	GG	D2S1338			22, 24			
rs1058083		AG	AG	D2S441		11	11, 15	11	11, 15	11, 15
rs560681		AG	AG	D3S1358		15, 16	15, 16		15, 16	15, 16
rs735155	CC	CC	CC	D5S818		11, 12	11, 12		11, 12	11, 12
rs722098		AA	AA	D7S820		11, 12	11, 12			
rs2111980	CC	CC	CC	D8S1179			13, 16		13, 16	13, 16
rs4288409		CC	CC	DYS391		10	10			
rs2076848	TT	TT	TT	FGA		20	20, 23			20
rs17250845	C	C	C	SE33	-	-	-			
rs2032678	T	T	T	TH01		9, 9.3	9, 9.3			9, 9.3
rs17842518	G	G	G	TPOX		8, 8	8, 8			
rs1454361		TT	TT	Yindel	-	-	-	2	2	2
rs10488710	GG	CG	CG	vWA			16, 18		16	16, 18
rs987640		TT	TT	D12ATA63		15, 17	15, 17	-	-	-
rs1493232	AA	AA	AA	D14S1434		13, 14	13, 14	-	-	-
rs9951171	AA	AA	AA	D1S1677		13, 14	13, 14	-	-	-
rs13218440	GG	GG	GG	D2S1776		10, 12	10, 12	-	-	-
rs1028528	AG	AG	AG	D3S4529		13	13, 15	-	-	-
rs2830795	AA	AG	AG	D4S2408			8, 9	-	-	-
rs1360288	CC	CC	CC	D5S2800		14, 18	14, 18	-	-	-
rs719366	AG	AG	AG	D6S1043		11	11, 14	-	-	-
rs3780962	AG	AG	AG	D6S474		14	14, 16	-	-	-
rs737681	CT	CT	CT							
rs10773760	AA	AA	AA							
rs338882	AG	AG	AG							
rs1109037	AG	AG	AG							
rs717302	AG	AG	AG							
rs4364205		GG	GG							
rs214955	CC	CC	CC							
rs914165	GG	GG	GG							
rs2016276	CT	CT	CT							
rs1015250	CC	CG	CG							
rs1413212	CT	CT	CT							

Key

XX	Alleles detected
XX	Allele dropout
	Locus dropout
XX	Incorrect result

Dancing bear credit: www.Telegraph.co.uk

Figure 5: Capillary electrophoresis traces of degraded DNA samples run with GlobalFiler on ABI 3500xl.



Funding: The FBI Biometrics Center of Excellence (BCOE): Forensic DNA Typing as a Biometric tool and NIST Special Programs Office: Forensic DNA.

Disclaimer: Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. 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 endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

A copy of this presentation is available at:
http://strbase.nist.gov/NISTpub.htm#Presentations

Overall conclusions:

- **Improved performance was noted with Mixture ID STR markers when log likelihood ratios were compared with CE based calculations.**
- **Mixture ratio estimates using STR data were similar between CE based measurements and Ion Torrent sequencing.**
- **Improvements in Mixture ID STR performance noted with degraded DNA due to relaxed constraints on STR amplicon size.**
- **Microhaplotypes added functionality in estimating number of contributors in a mixture, with some limitations in multi-contributor samples. These markers accurately represented contributor ratios compared with STRs when deconvolving mixtures.**
- **SNPs offer potential utility for samples with extremely degraded DNA.**
- **Converge software functioned well as an integrated data analysis platform. However, third party software was required to deconvolute STR mixture data.**

References:
1. Sanchez J, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison C, Musgrave-Brown E, Salas A, Syndercrombe-Court D, Schneider P, Carracedo A, and Morling N: A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis. 2006; 27 page 1713 – page 1724.
2. Pakstis A, Speed W, Fang R, Hyland F, Furtado M, Kidd J, and Kidd K: SNPs for a universal individual identification panel. Human Genetics. 2010; 127: page 315 – page 324.
3. Kidd K, Pakstis A, Speed W, Lagace R, Chang J, Wooton S, Haigh E, and Kidd J: Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. Forensic Science International: Genetics. 2014; 12: page 215 – page 224
4. Ø. Bleka, G. Storvik, P. Gill; EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts; Forensic Sci. Int. Genet., 21:35-44, 2016.
5. Hill, C.R., Duewer, D.L., Kline, M.C., Coble, M.D., Butler, J.M. (2013) U.S. population data for 29 autosomal STR loci. Forensic Sci. Int. Genet. 7: e82-e83.