APPLICATION INFORMATION

DNA Fragment Analysis

ANALYSIS OF 26 WEBER PANEL SIMPLE TANDEM REPEAT POLYMORPHISMS USING THE CEQ™ 2000 FRAGMENT ANALYSIS SYSTEM

Mark Dobbs, Keith Roby, Scott Boyer, Doni Clark, and Graham Threadgill
Beckman Coulter, Inc.

Introduction

Short tandem repeats (STRs) have become a boon to molecular geneticists who localize genes in mammalian genomes. Ideal STR polymorphisms amplify well under standard conditions, have a low amplification background, are easy to score, and are highly informative (Dubovsky *et al.,* 1995). Dr. James Weber of the Marshfield Medical Research Foundation has coordinated an effort to identify and map a large number of ideal human STRs for the genetics community, fulfilling the goal of establishing a physical framework of heritable polymorphic markers that can be used to map any gene of interest (Sheffield *et al.,* 1995). The latest set of markers has an average density of one marker per 10 cM, or a genetic recombination spacing of 10% (~ 10 million bases). Despite the large distance, the gap is readily closed using additional genetic markers and a variety of physical techniques, including chromosome walking.

In this paper, we examine a subset of the Weber panel version 9 using the CEQ™ 2000 Fragment Analysis system from Beckman Coulter. The complete panel consists of nearly 390 dinucleotide, trinucleotide, and tetranucleotide repeat polymorphisms spread throughout the human genome. Twenty six STR loci in three sets of 8 or 10 members each were examined on the CEQ 2000 using the Beckman Coulter WellRed dye-labeled primers. Each set was separated and analyzed on an individual capillary. The results clearly demonstrate that the CEQ 2000 is well suited for the analysis of STRs.

Materials and Methods

Primer sets were obtained from Research Genetics, Inc. (Huntsville, Alabama). Primer sequences were obtained from the Marshfield Medical Research site (http://www.marshmed.org/genetics/). In all cases, the forward primers were dye-labeled as indicated in the fragment lists in Figures 1, 5, 6, and 7. The reverse primers were unlabeled. The genomic DNAs were CEPH Utah pedigree 1331 repository DNA samples NA 07057A or NA 06990 obtained from the Coriell Institute for Medical Research (Camden, NJ). Labeled primers were tested for quality by dilution to 30 fM in deionized formamide, and separated on the CEQ 2000. Primer pairs were quantitated using a DU® 7500 UV Spectrophotometer from Beckman Coulter (Fullerton, CA). Genomic DNA was quantitated on a FLUOstar-P Microplate Fluorometer (BMG Lab Technologies, Offenburg, Germany) using Pico Green* (P-7581; Molecular Probes, Eugene, Oregon). Lambda DNA was used as a quantitative standard for this assay.

Individual amplification reactions were mixed as follows:

10x AmpliTaq Gold PCR Buffer	1.0 μL
5X dNTPs (1.25 mM)	0.8 μL
Primer Mix (5.0 μM each)	0.5 μL
$MgCl_2$ (25 mM)	0.6 μL
DNA (20 ng/μL)	1.0 μL
Enzyme (5U/μL)	0.05 μL
Water	6.05 μL
Total Reaction Vol.	10.0 μL

* *Pico Green is a registered trademark of Moleculer Probes, Inc.*

Now sold through SCIEX Separations
www.sciex.com/ce

SCIEX

BECKMAN COULTER
Capillary Electrophoresis

Cycling conditions were: 95°C – 10 minutes, 1 cycle to activate the enzyme (AmpliTaq Gold, Perkin Elmer, Branchburg, NJ), 95°C – 1 minute, 57°C – 1 minute, 72°C – 1 minute, 40 cycles (MJ PTC 200 Thermocycler, MJ Research, Inc., Waltham, MA). The individual samples were then run on the CEQ™ 2000 to determine the dilutions needed to balance the signal intensities of the size standards and amplification products. Two different approaches to preparing the mixtures were equally successful. In the case of set 3, the amplification products were mixed in calculated ratios before ethanol precipitation. In the case of sets 21 and 31, each amplification product was precipitated, quantitated, and then mixed into deionized formamide to create the sample pool. Forty µL of each mixture, which included 0.5 µL CEQ DNA size standard – 400 (Beckman Coulter P/N 608098), was aliquoted into each sample well. Samples were run on a CEQ 2000 under the Frag-3 method:

Capillary Temperature	50°C
Denature Temperature and Time	90°C for 120 seconds
Injection Voltage and Time	2.0 kV for 30 seconds
Separation Voltage and Time	6.0 kV for 35 minutes

The total cycle time for each row of eight samples, which included denaturation, injection, separation, data analysis, and capillary replenishment was approximately 45 minutes.

Results and Discussion

All of the loci selected for these experiments amplify well using the three novel Beckman Coulter dyes (Figures 1-7). The scale of the reactions and amplification conditions are no different from the conditions used to amplify the same loci with other labeled primers *(e.g,.* ftp://ftp.resgen.com/pub/mappairs/humanset/mappairs_protocol.txt). The peak patterns observed for the loci range from simple to complex. Locus D6S474 demonstrates a simple pattern with approximately 90% of the amplification product present in the true length form, and 10% present in the +A form (Figure 4, 152.30 and 168.38 bases). Each allele from locus D11S2000 displays 3 peaks (Figure 4, 216.73 and 222.82 bases). In the CEQ 2000 Fragment Analysis software, the user may arbitrarily assign any of the visualized peaks as the peak to use in allele identification using the specifications set up for each locus tag (e.g., Figure 8). After visually inspecting the allele patterns and evaluating the level of +A and amount of stutter, the user creates an allele list by calculating slope and offset of the linear relationship between two observed fragment sizes *(Apparent size)* and two expected sizes *(True size)*. Alleles are identified based on their closeness to the apparent sizes. For all of the loci in the studied sets of STRs where more than two alleles were represented by the two individuals examined (20/26), a plot of the observed size versus the expected size was a straight line. In practice, when nothing is known about the alleles or their expected sizes, the first observed fragment may be matched up with any similarly sized true allele in the list. As a rule, because large peaks contain more data points and are more likely to be called accurately, it was best to use the largest peak of a set to identify the alleles in the allele list of a locus tag. The other peaks within the range of a locus tag will be sized by the software and annotated in the *Comment* field of the fragment list (Figure 7). Samples of *Allele Identification Criteria* and an *Allele List* are shown in Figure 8.

The primer sets used in these experiments were suggested by the Marshfield Medical Research Foundation. The number of independent loci examined on individual capillaries was maximized by co-electrophoresis of non-overlapping allele amplification products. Adequate spacing of the loci within a set prevented stutter peaks from one locus from migrating in the range of a neighboring locus tagged with the same color. Nevertheless, if suitable loci are available, it may be possible to increase the number of amplification products that are analyzed simultaneously, further improving the throughput of the system.

In the present experiments, individual PCR reactions or mixtures of independent PCR reactions were precipitated and resuspended in formamide before loading on to the CEQ 2000. It was necessary to precipitate the pooled products to increase the fraction of anions in the injection mixture that were present as DNA molecules. Chloride, phosphate, unincorporated nucleotides and primers decrease the local potential field in the sample, leading to a decrease in the amount of amplified DNA injected. Longer DNA fragments and the size standards also compete with each other for injection. However, from our experience with DNA sequencing, we know that it is possible to inject hundreds of different DNA fragments in large amounts when the injection mixture is relatively free of non-DNA ions.

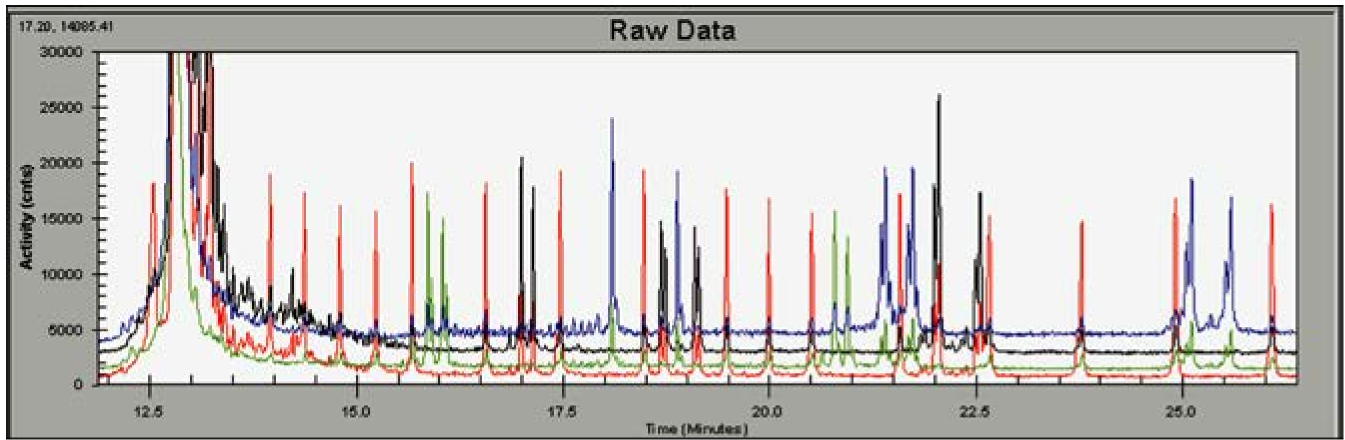
Most of the primer pairs tested on the CEQ 2000 produced labeled DNA fragments in the sizes

expected using slab gel systems (Table 1). The expected sizes were determined by other investigators using a variety of sizing methods, but the amplification products were not sequenced. In DNA NA 07057, one fragment from locus D4S1644 was one full tetranucleotide repeat longer than expected. This was the only example where the relative spacing between alleles in heterozygotes differed from previously reported data. Fragments from loci D7S817, D17S1293, D2S1360, and marker GATA193A07 were exceptional, at approximately 5, 8, 9 and 15 bases larger than expected. It is possible that some DNA fragments in the CEQ™ 2000 separation environment (matrix, denaturant) have slightly different mobilities than in other systems. The CEQ 2000 has yielded highly reproducible fragment sizing results (see Beckman Coulter technical publication A-1876A, entitled *Highly Precise DNA Sizing on the CEQ 2000 Fragment Analysis System*). Therefore, cloning and sequencing of the PCR products may be required to resolve the differences reported between observed and expected sizes. When combining data from different platforms, a few alleles from each locus should be sized on each system to resolve large differences in apparent fragment length.

The CEQ 2000's eight-capillary array operates in a 96-well plate format, a platform utilized by a variety of liquid handling robots, including the Biomek® 2000 and the Biomek FX Automated Laboratory Workstations from Beckman Coulter. At the level of analysis demonstrated here, it would be possible to genotype an individual for the entire Weber panel in 48 separations, or one half of a 96-well sample plate, in 4.5 hours of run time. The CEQ 2000 is thus well suited for gene mapping studies where a large amount of data can be rapidly generated with minimal user intervention.

Table 1. Expected versus Observed Allele Sizes for Weber Panel Loci Studied

Locus	Marker	Allele Size Range	1331-01 (NA 07057)				1331-02 (NA 06990)			
			Allele 1		Allele 2		Allele 1		Allele 2	
			Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed
D6S474	GATA-31	151-167	167.0	168.0	167.0	168.0	167.0	168.4	151.0	152.3
D11S2000	GATA-28D01	199-235	219.0	220.6	209.0	210.4	221.0	222.8	215.0	216.7
D17S1293	GGAA-7D11	262-290	270.0	278.7	266.0	274.6	282.0	291.5	274.0	283.3
D19S591	GATA-44F10	96-112	108.0	108.2	100.0	100.1	108.0	108.7	104.0	104.7
D18S858	ATA-23G05	193-211-	196.0	195.6	196.0	195.6	208.0	208.2	205.0	205.2
D15S642	ATA-41E04	121-139-	133.0	135.1	127.0	129.0	130.0	132.4	127.0	129.4
D18S1364	GATA-7E12	164-208	172.0	172.1	172.0	172.1	172.0	172.5	164.0	164.4
D5S2488	ATA-20G07	230-245	245.0	246.4	236.0	237.8	236.0	237.8	227.0	228.7
D12S2070	ATA-25F09	86-104-	98.0	101.8	86.0	88.9	89.0	92.6	89.0	92.6
D2S1360	GATA-11H10	136-176	144.0	151.3	140.0	147.1	176.0	184.0	160.0	167.7
D6S2427	GGAA-15B08	197-229	217.0	214.5	213.0	210.7	217.0	214.3	201.0	199.2
Unknown	GATA-193A07	339-375	375.0	390.4	347.0	361.5	371.0	386.1	367.0	382.0
D2S2944	GATA-30E06	108-136	120.0	122.4	116.0	118.4	128.0	130.1	120.0	121.9
D7S817	GATA-13G11	157-177-	169.0	175.0	169.0	175.0	169.0	174.7	161.0	166.5
D1S1589	ATA-4E02	199-220	205.0	205.5	202.0	202.5	214.0	214.5	205.0	205.4
D11S1981	GATA-48E02	134-178-	174.0	177.6	170.0	173.6	158.0	161.2	150.0	153.1
D7S820	GATA-3F01	204-240	224.0	227.7	216.0	219.6	220.0	223.4	216.0	219.5
D14S742	GATA-74E02	395-415	403.0	405.4	403.0	405.4	407.0	409.4	403.0	405.3
D4S1647	GATA-2F11	132-156	140.0	142.3	132.0	134.3	132.0	134.2	132.0	134.2
D10S1208	ATA-5A04	179-200-	197.0	198.4	194.0	195.3	194.0	195.4	179.0	180.5
D6S2434	ATA-50C05	224-236	227.0	229.2	224.0	226.1	236.0	238.4	233.0	235.3
D6S1040	GATA-23F08	257-285	281.0	283.6	277.0	278.6	281.0	282.7	273.0	274.5
D9S1122	GATA-89A11	190-210	202.0	202.1	190.0	189.9	198.0	198.0	194.0	194.0
D8S592	GATA-6B02	150-162	158.0	160.3	154.0	156.2	154.0	156.1	150.0	152.1
D4S1644	GATA-11E09	186-206-	202.0	207.0	198.0	198.9	202.0	202.9	198.0	198.8
D3S4542	GATA-148E04	236-260	256.0	258.3	244.0	246.1	256.0	258.3	248.0	250.2



Std	Peak #	Dye	Locus	Allele ID	Estimated Size
*	2	D1			69.79
*	3	D1			80.11
*	4	D1			90.33
*	5	D1			100.38
	6	D3	D19S591	7	104.69
	7	D3	D19S591	8	108.71
*	8	D1			120.05
	9	D2	D15S642	4	129.35
	10	D2	D15S642	5	132.43
*	11	D1			139.34
	12	D4	D6S474	2	152.30
*	13	D1			160.32
	14	D2	D18S1364	2	164.38
	15	D4	D6S474	6	168.38
	16	D2	D18S1364	4	172.49
*	17	D1			180.08
*	18	D1			190.03
*	19	D1			199.99
	20	D3	D18S858	6	205.21
	21	D3	D18S858	7	208.17
	22	D4	D11S2000	5	216.73
*	23	D1			220.05
	24	D4	D11S2000	6.2	222.82
	25	D2	D5S2488	1	228.70
	26	D2	D5S2488	4	237.81
*	27	D1			239.77
*	28	D1			259.84
*	29	D1			279.94
	30	D4	D17S1293	8	283.25
	31	D4	D17S1293	10	291.52
*	32	D1			300.09
*	33	D1			320.09
*	34	D1			340.03

Figure 1. *Plot of run with eight polymorphic Weber markers from DNA NA 06990 electrophoresed in the same capillary (Weber panel set 3) and fragment list with called alleles only.*

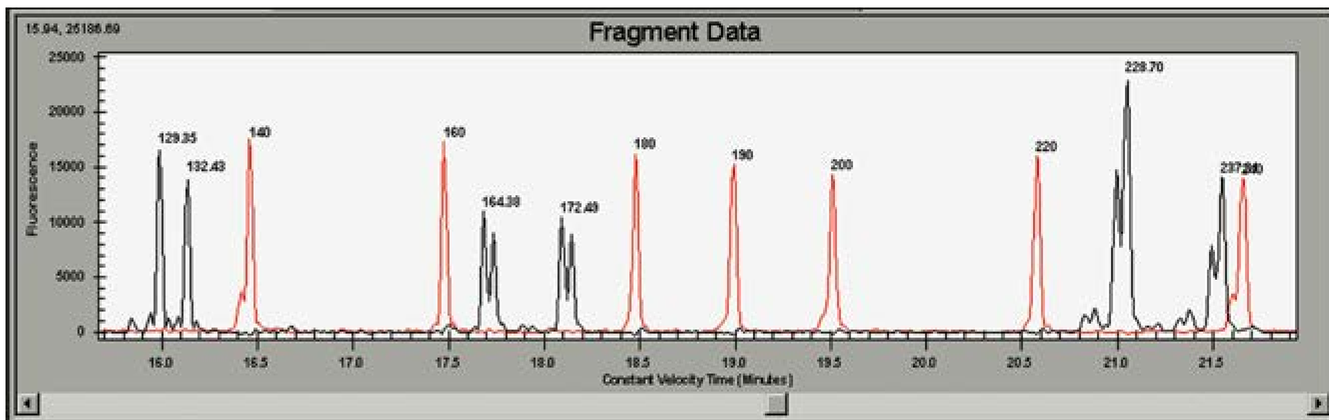


Figure 2. *Same run as Figure 1, displaying only the size standards and the amplification products labeled with dye D2.*

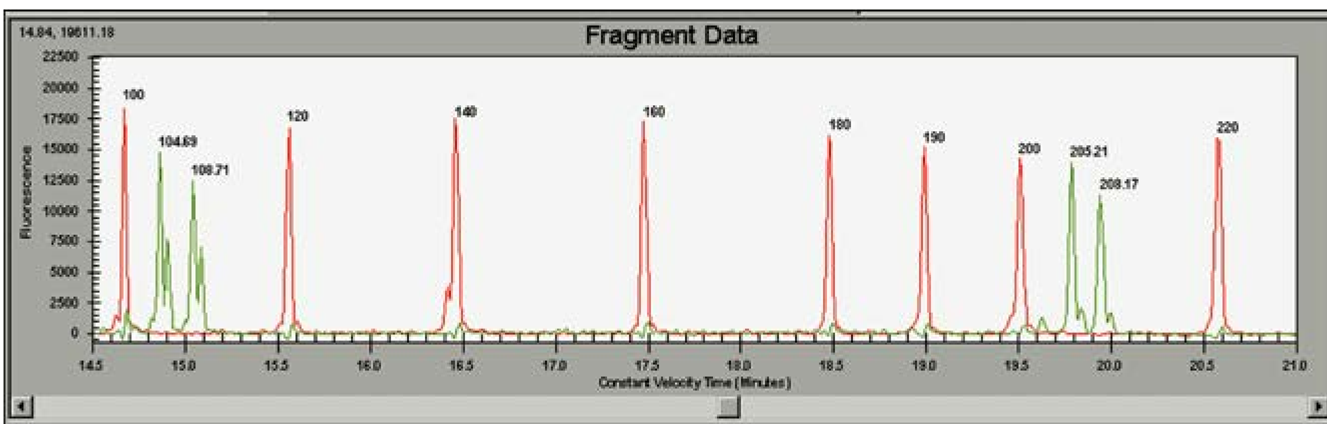


Figure 3. *Same run as Figure 1, displaying only the size standards and the amplification products labeled with dye D3.*

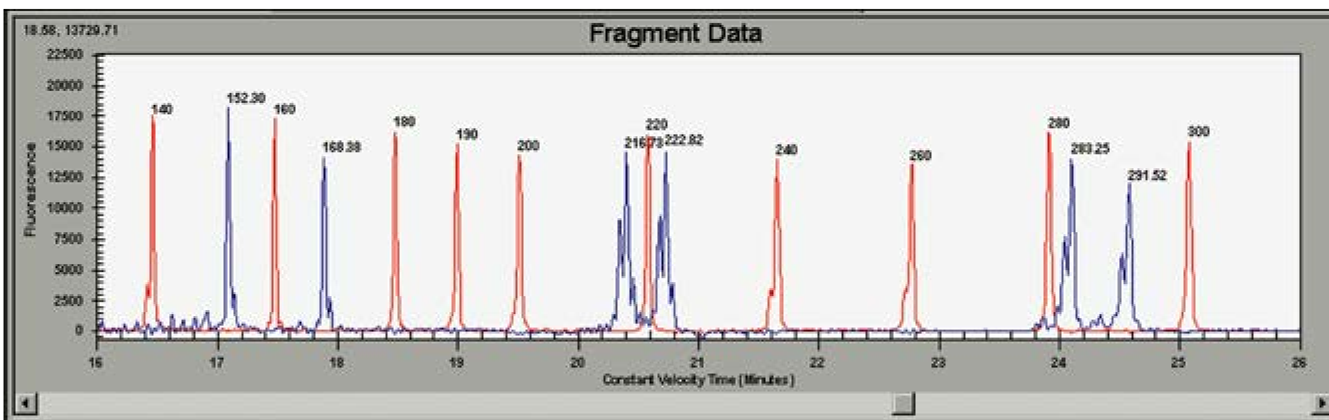
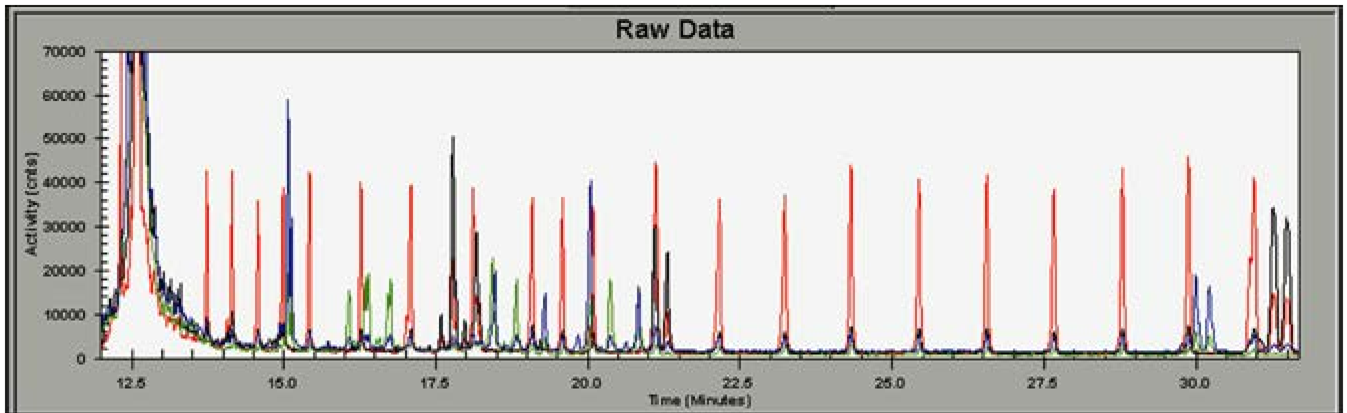
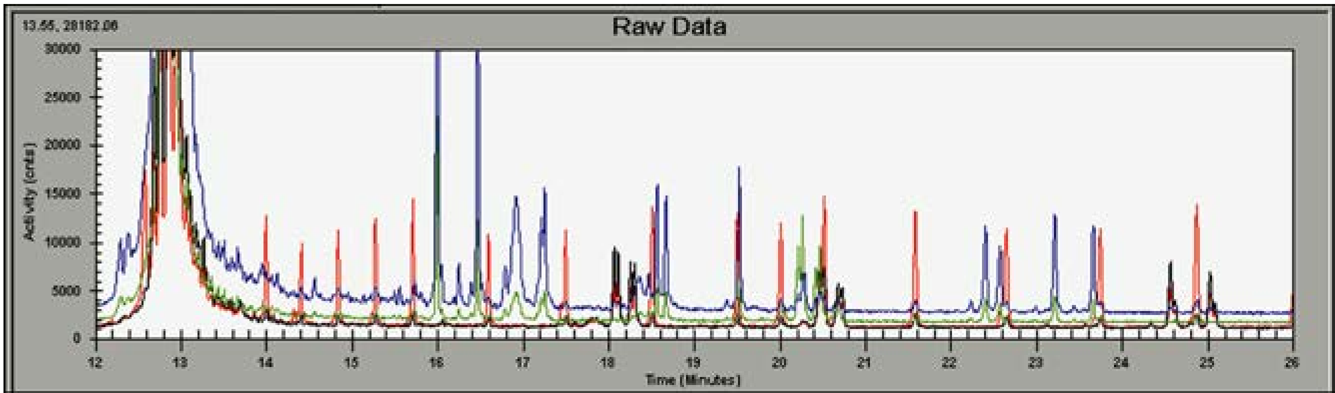


Figure 4. *Same run as Figure 1, displaying only the size standards and the amplification products labeled with dye D4.*



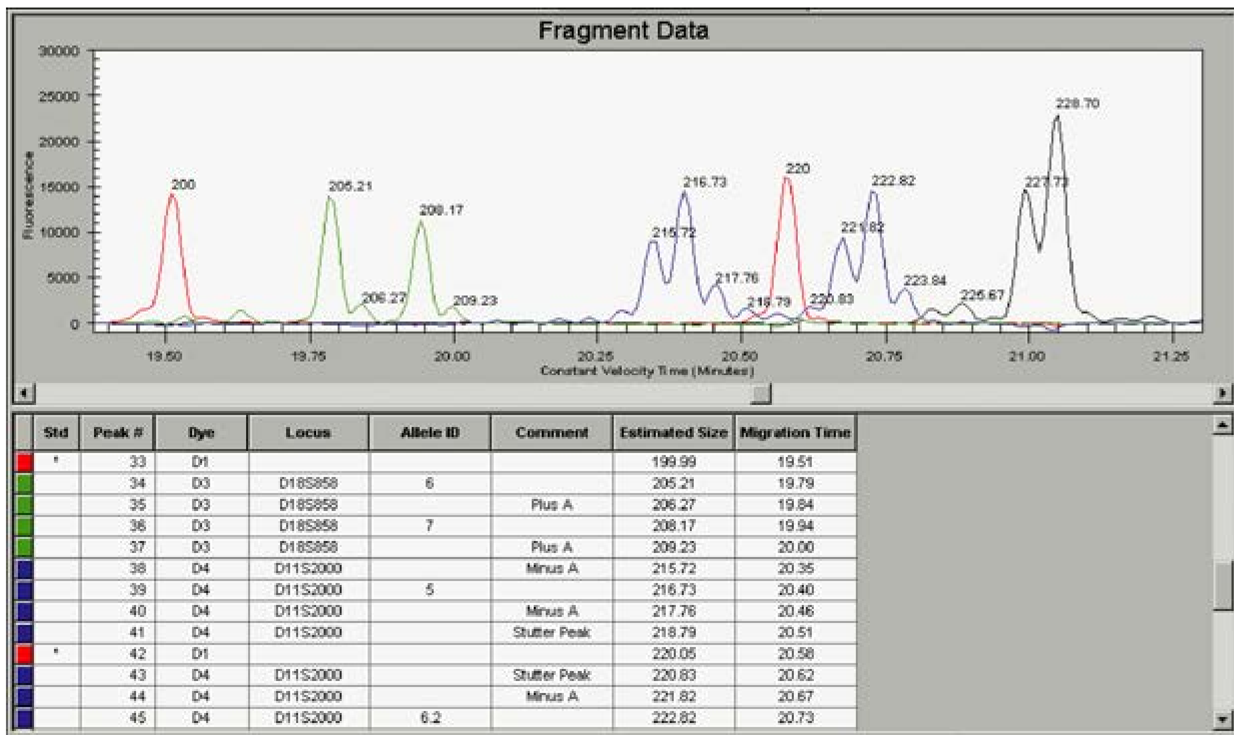
	Std	Peak #	Dye	Locus	Allele ID	Estimated Size
		5	D4	D12S2070	2	92.37
	*	6	D1			100.81
	*	7	D1			120.47
		8	D3	UnknownGATA30E06	4	122.02
		9	D3	UnknownGATA30E06	6	130.22
	*	10	D1			138.60
		11	D2	D11S1981	5	153.30
	*	12	D1			160.31
		13	D2	D11S1981	7	161.48
		14	D3	D7S817	2	166.77
		15	D4	D2S1360	7	167.73
		16	D3	D7S817	4	174.98
	*	17	D1			180.15
		18	D4	D2S1360	11	184.41
	*	19	D1			190.14
		20	D4	UnknownGGAA15B08	2	199.03
	*	21	D1			200.15
		22	D3	D1S1589	3	205.68
		23	D4	UnknownGGAA15B08	6	214.42
		24	D3	D1S1589	6	214.68
		25	D2	D7S820	4	219.61
	*	26	D1			220.21
		27	D2	D7S820	5	223.61
	*	28	D1			239.62
	*	29	D1			259.68
	*	30	D1			279.81
	*	31	D1			300.07
	*	32	D1			320.26
	*	33	D1			340.00
	*	34	D1			360.36
	*	35	D1			379.95
		36	D4	UnknownGATA193A07	8	382.09
		37	D4	UnknownGATA193A07	9	386.20
	*	38	D1			399.78
		39	D2	D14S742	3	405.57
		40	D2	D14S742	4	409.66
		41	D1			410.70

Figure 5. *Plot of run with 10 polymorphic Weber markers from DNA NA 06990 electrophoresed in the same capillary (Weber panel set 21) and fragment list with called alleles.*



	Std	Peak #	Dye	Locus	Allele ID	Estimated Size
	*	3	D1			80.14
	*	4	D1			90.34
	*	5	D1			100.37
	*	6	D1			120.18
		7	D4	D4S1647	2	134.17
	*	8	D1			139.30
		9	D2	D8S592	3	152.13
		10	D2	D8S592	4	156.07
		11	D4	D4S1647	8	157.47
	*	12	D1			160.17
	*	13	D1			180.25
		14	D4	D10S1208	2	180.51
	*	15	D1			189.96
		16	D3	D9S1122	3	193.94
		17	D4	D10S1208	7	195.42
		18	D3	D9S1122	4	198.00
		19	D2	D4S1644	5	198.84
	*	20	D1			199.89
		21	D2	D4S1644	6	202.90
	*	22	D1			220.05
		23	D4	D6S2434	5	235.25
		24	D4	D6S2434	6	238.37
	*	25	D1			239.80
		26	D4	D3S4542	5	250.16
		27	D4	D3S4542	7	258.32
	*	28	D1			259.86
		29	D2	D6S1040	6	274.54
	*	30	D1			279.93
		31	D2	D6S1040	8	282.69
	*	32	D1			300.11
	*	33	D1			320.13
	*	34	D1			340.03
	*	35	D1			360.18

Figure 6. *Plot of run with eight polymorphic Weber markers from DNA NA 06990 electrophoresed in the same capillary (Weber panel set 31) and fragment list with called alleles.*



	Std	Peak #	Dye	Locus	Allele ID	Comment	Estimated Size	Migration Time
	*	33	D1				199.99	19.51
		34	D3	D18S858	6		205.21	19.79
		35	D3	D18S858		Plus A	206.27	19.84
		36	D3	D18S858	7		208.17	19.94
		37	D3	D18S858		Plus A	209.23	20.00
		38	D4	D11S2000		Minus A	215.72	20.35
		39	D4	D11S2000	5		216.73	20.40
		40	D4	D11S2000		Minus A	217.76	20.46
		41	D4	D11S2000		Stutter Peak	218.79	20.51
	*	42	D1				220.05	20.58
		43	D4	D11S2000		Stutter Peak	220.83	20.62
		44	D4	D11S2000		Minus A	221.82	20.67
		45	D4	D11S2000	6.2		222.82	20.73

Figure 7. *Segment of set 3 data with non-allele peaks within a locus range identified in the Comments field.*

Locus Tag Name: D11S200/GATA28D01

Locus:	D 11 S2000
GenBank Accession:	
Primer Set Name:	D 11 S2000
Primer Dye:	D4
Primer Labeled:	Forward
Forward Sequence:	AGTA CAAA AGCC TATT TAGT CAGG
Reverse Sequence:	TTTG AAGATCTG TGAA ATGT GC
Fragment Size Range Min Size:	195
Fragment Size Range Max Size:	239
Repeat Unit Length:	4
Repeat Unit Sequence:	GATA
No. of Repeats in Shortest Allele:	1
Shortest Allele Size:	195
Apparent size includes + A:	Yes

Allele ID Criteria

Allele Confidence Interval:	0.50
Search for stutter:	Yes
Stutter Detection Window Width:	2 repeats
Detect stutter shorter than allele:	Yes
Detect stutter longer than allele:	Yes
Maximum relative stutter peak height:	20%
Detect spurious peaks:	Yes
Maximum height for spurious peaks:	50%
Detect +/- A:	Yes
+ A is dominant:	Yes

Allele List

True Size (nt)	Apparent Size (nt)	Number of Repeats	ID
199	200.49	1	1
203	204.55	2	2
207	208.61	3	3
211	212.67	4	4
215	216.73	5	5
219	220.79	6	6
221	222.82	6.2	6.2
223	224.85	7	7
227	228.91	8	8
231	232.97	9	9
235	237.03	10	10

Figure 8 . *Sample Locus Tag Information. Each field listed is available for data entry by the user.*

All trademarks are the property of their respective owners.

For Research Use Only. Not for use in diagnostic procedures.

© 2014 AB SCIEX. SCIEX is part of AB Sciex. The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners. AB SCIEX™ is being used under license.

www.sciex.com/ce

View SCIEX products at www.sciex.com
Find your local office at www.sciex.com/offices

AB SCIEX Headquarters
500 Old Connecticut Path | Framingham, MA 01701 USA
Phone 508-383-7700
www.absciex.com