Chromosome-specific panels of tri- and tetranucleotide microsatellite markers for multiplex fluorescent detection and automated genotyping: evaluation of their utility in pathology and forensics.

A K Lindqvist, P K Magnusson, J Balciuniene, et al.

Genome Res. 1996 6: 1170-1176
Access the most recent version at doi:10.1101/gr.6.12.1170

RESEARCH

Chromosome-specific Panels of Tri- and Tetranucleotide Microsatellite Markers for Multiplex Fluorescent Detection and Automated Genotyping: Evaluation of Their Utility in Pathology and Forensics

Anna-Karin B. Lindqvist,[1] Patrik K.E. Magnusson,[1] Jorune Balciuniene,[1] Claes Wadelius,[2] Eva Lindholm,[1] Marta E. Alarcón-Riquelme,[1] and Ulf B. Gyllensten[1,3]

[1]Department of Medical Genetics, Biomedical Center, and [2]Department of Clinical Genetics, Uppsala University, Uppsala, Sweden

A set of 391 microsatellite markers (Weber set 6), 85% of which consist of tri- and tetranucleotide repeat markers, were used to design chromosome-specific panels that allowed for a high degree of multiplexing with respect to the fragment size range and fluorophore (FAM, HEX, TET). This marker set has an average coverage of 10.5 cM, with the largest gap being 28.1 cM. The markers were divided into 49 panels, with a maximum degree of multiplexing of 15 markers per panel. The utility of the markers for analysis of DNA from blood, hair, and formalin-fixed archival tissue biopsies was evaluated with respect to amplification efficiency, product yield, and degree of preferential amplification of the shorter allele in heterozygotes. The amplification efficiency was inversely related to repeat length and amplicon length. Based on the analysis of DNA from formalin-fixed biopsies, 51 markers suitable for loss of heterozygosity (LOH) studies were identified. The utility of the marker set for genome scanning, LOH, and forensic analyses is discussed.

The use of markers labeled with fluorophores in genetic linkage analysis has become widespread as a result of a number of advantages, as compared with methods employing radioactivity. These advantages include the lack of need for safety precautions associated with the use of radioactive labels (Schwengel et al.1994) and the possibility of electronic data collection, including the use of special software for fragment size determination and calling of allele and genotype (Ziegle et al. 1992; Reed et al. 1994).

Most of the microsatellite marker sets developed to date for genome scanning are based on highly polymorphic dinucleotide $(CA)_n$ repeats (Weber and May 1989). Indeed, at present >5000 markers of this type have been mapped genetically in the human genome (Dib et al. 1996). However, dinucleotide markers are less suitable for large-scale automated analysis because (1) the fragment resulting from nonspecific addition of a nucleotide may overlap with the allele-peak resulting in ambiguous allele and genotype calling, and (2) strand slippage introduces stutter-bands that contribute to erroneous genotype calling. These artifacts introduced by the PCR pose problems to the software used for automated genotyping. Some of these difficulties may be overcome by modifying the amplification cycle (Smith et al. 1995). Preferentially, these problems can be minimized by use of markers based on tri- and tetranucleotides instead of dinucleotides. One such large set (Weber set 6) has been described recently by Sheffield et al. (1995) at the Cooperative Human Linkage Center (CHLC).

In the present study, we have used the Weber set 6 and designed chromosome-specific panels for fluorescence detection on the ABI 373 and 377 instruments, using three fluorophores (Reed et al. 1994). To determine the efficiency and suitability of individual markers, we have tested these panels for use in genome scanning projects

[3]Corresponding author.
E-MAIL ulf.gyllensten@medgen.uu.se; FAX 46-18-510792.

6:1170–1176 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

based on high molecular weight genomic DNA derived from peripheral blood, as well as their utility in the analysis of forensic material (hair) and pathology material (formalin-fixed biopsies). The latter type of evaluation is of importance for at least two reasons. In linkage studies it is not uncommon that only archival autopsy material, such as formalin-fixed tissue, is available from certain key individuals. Furthermore, the comparison of tumor and normal tissue for loss of heterozygosity (LOH) across the genome is becoming increasingly important for the identification of cancer susceptibility loci (Merlo et al. 1991).

RESULTS

Characteristics of the Marker Set

The complete set of 391 markers consist of 286 tetranucleotides, 48 trinucleotides, and 57 dinucleotides (Table 1). This set was designed to cover the human genome with a mean spacing of 10.52 cM. The largest spacing is a gap of 28.1 cM on chromosome 15. The average heterozygozity for the 391 markers has been estimated to 76%. Table 1 gives a summary of types of repeats on each chromosome, the fluorophores assigned to each marker, the number of markers in each panel, as well as the number of panels for each chromosome. A complete table with the results for the individual markers will be made available from the authors.

Performance of the Marker Primers

Table 2 shows the quality assignment used for each primer pair. The efficiency and quality of the primers was determined by analyzing the sample on the ABI 373 sequencer, employing the

Table 1. Distribution of Markers of Each Chromosome, Repeat Type, Fluorophore, and Number of Markers in each Chromosome-specific Marker Panel

Chromosome	Repeat type			Flurophore			No. panels	No. of markers in each panel				Total no. of markers
	tet	tri	di	FAM	HEX	TET		panel 1	panel 2	panel 3	panel 4	
1	23	5	1	11	10	8	3	15	10	4		29
2	20	2	2	9	9	6	2	14	10			24
3	16	3	3	8	7	7	2	13	9			22
4	16	2	2	7	7	6	2	12	8			20
5	17	2	2	8	7	6	2	12	9			21
6	16	5	2	8	8	7	2	15	8			23
7	13	0	1	6	5	3	2	9	5			14
8	14	3	1	6	6	6	2	12	6			18
9	11	1	4	6	6	4	2	11	5			16
10	11	5	3	7	6	6	2	12	7			19
11	11	1	1	5	4	4	2	9	5			14
12	15	3	0	8	7	3	2	11	7			18
13	9	1	2	6	3	3	2	9	3			12
14	13	2	0	7	4	4	2	13	2			15
15	6	2	3	4	3	4	2	9	2			11
16	8	1	3	5	3	4	2	10	2			12
17	11	0	5	6	5	5	2	11	5			16
18	7	3	3	4	5	4	2	9	4			13
19	7	0	3	4	3	3	2	6	4			10
20	6	1	3	4	3	3	1	10				10
21	5	1	1	3	2	1	1	7				7
22	4	0	2	3	1	1	1	6				6
X	9	0	7	6	4	6	2	10	6			16
Y	4	0	0	2	1	1	1	4				4
Mixed	14	4	3	7	7	7	4	9	6	3	3	21
Total	286	48	57	150	127	114						391

Table 2. Criteria for Assigning Markers to Quality Group

Quality group	Peak height (fluorescent units)
6	> 4000
5	1000–4000
4	500–1000
3	200–500
2	100–200
1	50–100

height of the peaks as an indicator for the amount of PCR product and a threshold of 50 fluorescent units for all fluorophores. Background level and the presence of unspecific peaks were also taken into consideration. By use of a single PCR composition and the same thermocycle parameters, we found that 323 of the 391 markers tested (83%) showed sufficient quality for routine analysis from blood genomic DNA samples (Table 3). The markers that did not fulfill these criteria are listed in Table 4. When considering the repeat type (di-, tri-, or tetranucleotides), primers amplifying trinucleotide repeats were found to have the lowest success rate (Table 4). In addition, primers labeled with HEX showed a lower efficiency, as a result of a weaker signal, than primers labeled with the other fluorophores. The degree of multiplexing for the panels varies from 2 to 15 markers per panel (Table 1) (Fig 1).

Analysis of DNA of Single Hairs

Of the 391 markers, 147 were further tested for their ability to amplify DNA from a single hair. The 147 primers chosen were among those found to have a quality score of 4–6. Of these, a total of 125 markers were able to efficiently amplify from single hair DNA. (The distribution among chromosomes of these markers is shown in Table 5.) In all cases, the alleles obtained from the hair sample corresponded to those of the peripheral blood genomic DNA of the same donor. This indicates that this subset of markers shows enough robustness to be useful for analyses when only limited biological material is available, such as in anthropological or forensic analyses. With the protocol used, the digest from a single hair was found to be sufficient for the PCR analysis of 20 independent microsatellite markers. Further experiments have shown, however, that at least 2–3 times that number of markers can be analyzed

Table 4. Characteristics of Markers That Did Not Amplify Using the Single Amplification

Type/color	No. of markers[a]	Percent[b]
Tetra	20	7
Tri	6	12.5
Di	3	5.3
Total	29	
FAM	10	6.7
HEX	15	11.8
TET	4	3.5
Total	29	

[a]Total of 29 markers; see Table 3.
[b]Percent of the total number of markers of that type/color.

Table 3. Results of Quality Evaluation of the Marker Set

Peak quality[a]	Peak height (fluorescent units)	No. of markers	Percent
5–6	1000–>4000	131	34
4	500–1000	116	30
3	200–500	76	19
2	100–200	25	7
1	50–100	10	3
No amplification		29	7

[a]Quality groups; see Table 2.

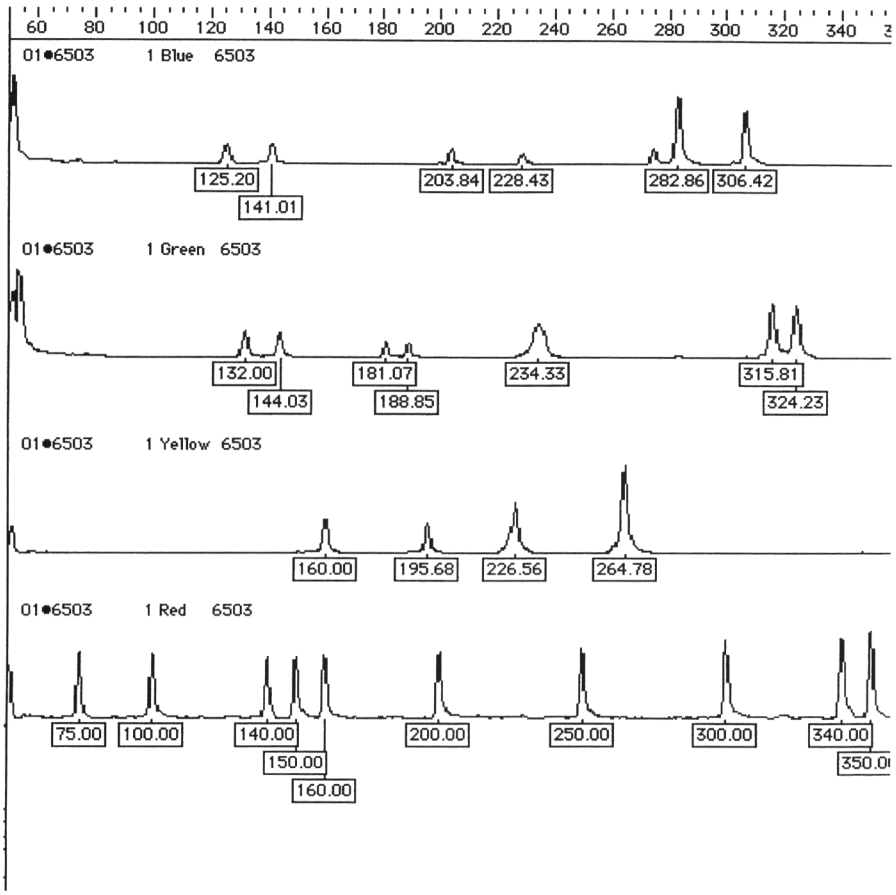

Figure 1 Electropherograms demonstrating the degree of multiplexing of panel 4, chromosome 2. The electropherograms from the *top*: markers labeled with FAM, TET, HEX, size standard GS-350 TAMRA.

from a single hair by reducing the starting amount of template DNA in the PCR (P. Magnusson, unpubl.).

Analysis of Archival Biopsy DNA

The 125 markers that successfully amplified single hair DNA were further tested on DNA prepared from a formalin-fixed biopsy. Of these, 51 markers were found to successfully amplify from this material (Table 5). These markers constitute a suitable set for the localization of LOH.

Relative Efficiency of Different Types of Repeats

First, a comparison was made of the relative efficiency of different types of repeat markers to amplify from the various materials. Among the 147 markers tested on hair DNA, 100% of the dinucleotide repeats, 97% of the tetranucleotide repeats, but only 53% of the trinucleotide repeats amplified DNA from a single hair. Further, of the 125 markers found to amplify from hair DNA, 63% of the dinucleotide repeats, 40% of the tetranucleotide repeats, and 30% of the trinucleotide repeats successfully amplified biopsy DNA. Second, the relationship between amplification success and length of the amplified fragment was investigated. The amplification efficiency of markers with fragments >200 bp decreased markedly on biopsy material as compared with blood-derived or hair-derived DNA (Fig. 2). Finally, we analyzed the relative efficiency of amplification of the two alleles in heterozygous samples. This was done by calculating the ratio between allele peak heights PH1/PH2 (the peak height of the shorter allele divided by the peak height of the longer allele). The mean value for blood DNA and hair DNA were close to unity with a moderate variation (Fig. 3). However, when amplifying from biopsy DNA, there was a preferential amplification of the shorter allele as shown by the mean PH1/PH2 value of 1.41 for biopsy DNA. The corresponding value for hair DNA was 1.10 and for blood-derived DNA was 1.09 (Fig. 3).

DISCUSSION

In the present paper, we describe the use of a set of microsatellite markers for genome scanning, based mainly on tri- and tetranucleotide repeats, which by optimal allocation of fluorophore allows a high degree of multiplexing in the gel analysis. The main advantage of this marker set over those based on dinucleotide repeats is the simple and robust allele and genotype calling. This is attributable to the ease by which alleles can be distinguished and the minimal PCR artifact peaks. By use of one PCR condition, 85% of the markers were successfully amplified. Most

Table 5. Number of Markers That Successfully Amplified DNA Extracted from Hair and Biopsy

Chromosome	Hair	Biopsy
1	10	3
2	14	7
3	11	6
4	5	2
5	13	3
6	7	1
7	5	1
8	2	1
9	7	2
10	6	1
11	3	1
12	9	4
13	3	1
14	5	1
15	4	4
16	2	1
17	4	3
18	2	2
19	5	5
20	1	1
21	1	0
22	n.d.[a]	n.d.
X	6	1
Y	n.d.	n.d.
Mixed	n.d.	n.d.
Total	125	51

[a]n.d. = no data.

likely, many of the markers that were not amplified can be made to work by modifying the buffer and/or ramp conditions. Tetranucleotide repeats were most useful for this study because we encountered fewer problems in calling allele sizes with the Genotyper software with these as compared with dinucleotide repeats.

To enable a high throughput of the genotyping, we divided the markers into chromosome-specific panels that could be analyzed simultaneously on a gel. In the design of the panels, we combined markers with a minimum difference in fragment size of 10 bp. In the panel with the highest degree of multiplexing, PCR amplification products from 15 markers were successfully pooled and analyzed without difficulty in distinguishing the alleles from each other. The large number of markers that were amplified successfully from blood-derived DNA with a wide coverage of the genetic map makes this marker set useful for human linkage projects. The marker set is presently used in a genome scanning project.

The second aim of the study was to evaluate the potential of using this marker set for analysis of biological materials other than peripheral blood DNA. In the present study, we used DNA extracted from roots of hair and from formalin-fixed biopsies. It is important to point out that the DNA extracting methods, as well as the quality of the DNA obtained, differ from that of peripheral blood. This is why we believe it was essential to test the marker set not only on DNA of low concentration but on these authentic materials. Three different conclusions emerged from our study. First, a significant fraction of the markers could be amplified successfully from single-hair preparations. Genotyping of hair and blood DNA from the same donor gave totally concordant results. Although the preparations were done on freshly obtained hair with roots, the results indicate that a significant number of the markers are also suitable for the analysis of hairs that had been collected some time ago. Based on the proper choice of markers, genetic studies of a relatively large number of microsatellite markers may be possible from individual hairs. For instance, single hairs have been used previously for anthropological studies, analyzing the mitochondrial genome (Vigilant et al. 1989). Our results show that similar studies, based on a set of 10–20 microsatellites, should be possible to conduct on single hairs. In forensic casework, single hairs constitute the most frequent type of mate-

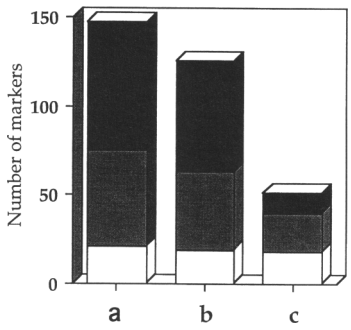

Figure 2 Length distribution among markers successfully amplified from different DNA sources. (a) 147 markers with the best performance on blood DNA were chosen and tested on (b) single hair DNA and (c) DNA from formalin-fixed and paraffin-embedded biopsy. Only markers working on hair DNA were tested with biopsy DNA as template. (Open bars) Markers 80–150 bp; (shaded bars) 150–200 bp; (solid bars) >200 bp.

rial evidence. Although only a fraction of these hairs contain roots, our results indicate that these should be possible to amplify for a set of markers.

A subset of the markers yielded successful amplifications from the biopsy material. Although these markers are too few to allow a complete genome scan, their coverage is sufficient to scan for large deletions on most chromosome arms. Thus, this set of markers may be useful for analysis of genetic diseases when only archival material is available from key individuals in the study. Caution should be exercised, however, when using the marker set on formalin-fixed biopsy material to identify regions displaying LOH. Our results shows that false LOH may occur with the preferential amplification of the shorter allele, in particular, when comparisons are made between fresh, nontumorous material and formalin-fixed tumor biopsies. The magnitude of this problem remains to be determined.

The panels presented here are now being used by a consortium including four different laboratories in Sweden, Finland, and Germany and are available to the labs in the Nordic countries at the Nordic Consortium Primer Resource Center at the Department of Clinical Genetics, Uppsala. The same marker sets also have been made for the ALF and ALF-express systems, to allow the seamless transfer of data between labs using different genotyping systems.

METHODS

Design of the Marker Panels

The primer sequences, marker designation, and chromosomal location, as well as other information for the 391 markers included in the Weber set 6, was obtained through the anonymous ftp (ftp.chlc.org) (Sheffield et al. 1995). Markers located on the same chromosome were divided into panels. Each panel included markers with a difference in fragment size of at least 10 bp and was considered nonoverlapping, and markers within each such group were assigned one of the ABI fluorophores 6-FAM (blue), TET (green), and HEX (yellow) (Applied Biosystems Inc., Foster City, CA). A few panels contained markers from several chromosomes, as they would not fit in the chromosome-specific panels. All oligonucleotides were synthesized, labeled, and cartridge-purified by GenSet, Inc. (France).

DNA Preparation

Peripheral Blood DNA

DNA was prepared from 10 ml of whole blood from one female and one male individual. Red blood cells were lysed in 155 mM NH_4Cl, 10 mM $KHC0_3$, and 0.1 mM ethylenediamine tetraacetate (EDTA). The white blood cell pellet was washed twice in SE-buffer (75 mM NaCl, 25 mM EDTA, 30 mM Tris at pH 8.0) and then vortexed, and 25 µl of proteinase K (10 mg/ml) was added in a final volume of 5 ml SE buffer with 10% SDS. The cells were left overnight at room temperature. The DNA was then extracted by phenol/chloroform precipitated with 3M NaAc and isopropanol, collected by centrifugation, dried, and resuspended in TE-low (10 mM Tris, 0.1 mM EDTA). Each PCR was initiated with 25 ng.

DNA from Single Hair

DNA was prepared from freshly plucked individual hairs (containing a root) of the female blood donor. A single hair was cut about 0.5 cm from the root end and the root-containing part was incubated at 56°C in 100 µl of 50 mM KCl, 10 mM Tris-HCl at pH 8.3, 0.2 mg/ml proteinase K, and 25 mM dithiothreitol (DTT). After 1 hr, an additional 2 µl of proteinase K (10 mg/ml) was added and the mixture incubated for 2 hr at 56°C. The enzyme was inactivated at 95°C for 10 min and the preparation stored at −20°C. For each amplification, 5 µl of the mixture was used.

DNA from Archival Biopsy Samples

A formalin-fixed and paraffin-embedded pancreatic biopsy was kindly provided by Dr. Erik Willander (Department of Pathology, Uppsala University), and DNA was prepared

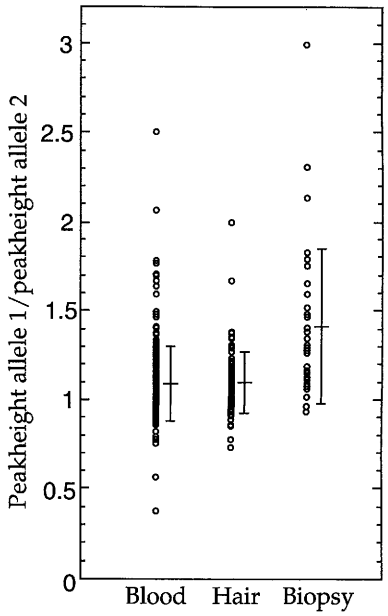

Figure 3 Distribution of peak height ratios for heterozygous markers on blood, hair, and biopsy DNA. Each circle represents the PH1/PH2 value for an individual marker. The error bars show mean and standard deviation.

according to Kösel and Graeber (1994). The biopsy was cut into small grains and divided between 12 Eppendorf tubes. The material in each tube was deparaffinated by adding 1 ml xylene and incubated for 2 hr at room temperature, with an occasional mixing by squeezing the pellet with a plastic stick. The xylene was then aspirated and the material washed for 5 min each with 95% and 70% ethanol. After the material was dried for 20 min at 50°C, 200 µl of a solution containing 0.2 M Tris-HCl at pH 8, 1% SDS, 1 mg/ml proteinase K, and 10 mM EDTA was added. The digestion was left for 3 hr at 50°C followed by a 14-hr incubation at 30°C. The proteinase K was then inactivated and the tubes centrifuged at 14,000 rpm for 5 min. The supernatants were collected into new tubes and the DNA extracted twice with phenol/chloroform. DNA was finally precipitated with 0.1 vol of 3M NaAc at pH 5 and 2.5 vols 95% ethanol and incubated 1 hr at −70°C. The DNA pellets were dissolved in 100 µl Tris–EDTA buffer each. Five microliters of the DNA were used for the PCR amplification.

PCR Conditions and Gel Analysis

All markers were tested under the same conditions. Genomic DNA was prepared from the peripheral blood of one man and one woman. For the PCR amplification, 25 ng of DNA was used in a 15 µl reaction. The PCR contained 50 mM KCl, 10 mM Tris-HCl at pH 8.3, 1.5 mM $MgCl_2$, 2.5 pmole of each primer, 200 mM of each dNTP, and 0.8 units AmpliTaq (Perkin Elmer). The same thermocycle parameters were applied to all markers and all materials: an initial 5 min denaturation at 95°C followed by 15 sec at 95°C, 30 sec at 55°C, and 30 sec at 72°C for 20 cycles. This was followed by 10 cycles of 15 sec at 89°C, 30 sec at 55°C, and 30 sec at 72°C, and a final extension step of 10 min at 72°C. All PCRs were performed on the Perkin Elmer 9600 model.

The amplification products from blood and hair DNA were diluted with water as follows: FAM and TET 1:20, HEX 1:10. Products obtained from amplification of biopsy-derived DNA were diluted: FAM 1:15, TET 1:10, and HEX 1:7.5. A mixture of 1.5 µl of the diluted amplified sample, 3 µl formamide, and 0.5 µl of size standard (GS500-TAMRA) was loaded onto a 6% denaturating polyacrylamide gel and the fragments separated on an ABI 373 instrument. Results were analyzed with the GeneScan 672 (version 1.0) and the Genotyper softwares (version 1.1) from Applied Biosystems.

ACKNOWLEDGMENTS

This work was supported by grants from the Beijer Foundation, the Swedish Medical Research Council, the Nordic Council of Ministers, the Swedish Natural Sciences Research Council, the Marcus Borgström Foundation, the Sörensen Foundation, the Tore Nilsons Foundations, and the Börje Dahlin Foundation.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Dib, C., S. Faure, C. Fizames, D. Samson, N. Droust, A. Vignal, P. Millasseau, S. Marc, J. Hazan, E. Seboun, M. Lathrop, G. Gyapay, J. Morisette, and J. Weissenbach. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380:** 152–154.

Kösel, S. and M.B. Graeber. 1994. Use of neuropathological tissue for molecular genetic studies: Parameters affecting DNA extraction and polymerase chain reaction. *Acta Neuropathol.* **88:** 19–25.

Merlo, G.R., C.S Cropp, R. Callahan, and T.I. Takahashi. 1991. Detection of loss of heterozygosity in tumor samples by PCR. *BioTechniques* **11:** 166–168.

Reed, P.W., J.L. Davies, J.B. Copeman, S.T. Bennett, S.M. Palmer, L.E. Pritchard, S.C.L. Gough, Y. Kawaguchi, H.J. Cordell, K.M. Balfour, S.C. Jenkins, E.E. Powell, A. Vighal, and J.A. Todd. 1994. Chromosome-specific microsatellite sets for fluorescence-based, semi-automated genome mapping. *Nature Genet.* **7:** 390–395.

Schwengel, D.A., A.E. Jedlicka, E.J. Nanthakumara, J.L. Weber, and R.C. Levitt. 1994. Comparison of fluorescence-based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. *Genomics* **22:** 46–54.

Sheffield, V.C., J.L. Weber, K.H. Beutow, J.C. Murray, D.A. Even, K. Wiles, J.M. Gastier, J.C. Pulido, C. Yandava, S.L. Sunden, et al. 1995. A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome wide linkage maps. *Hum. Mol. Genet.* **4:** 1837–1844.

Smith, J.R., J.D. Carpten, M.J. Brownstein, S. Ghosh, V.L. Magnusson, D.A. Gilbert, J.M. Trent, and F.S. Collins. 1995. Approach to genotyping errors caused by nontemplated nucleotide addition by Taq DNA polymerase. *Genome Res* **1:** 312–317.

Vigilant, L., R. Pennington, H. Harpending, T.D. Kocher, and A.C. Wilson. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. *Proc. Natl. Acad. Sci.* **86:** 9350–9354.

Weber, J.L. and P.E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44:** 338–396.

Ziegle, J.S., S. Ying, K.P. Corcoran, L. Nie, P.E. Mayrand, L.B. Hoff, L.J. McBride, M.N. Kronick, and S.R. Diehl. 1992. Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics* **14:** 1026–1031.

Received July 25, 1996; accepted in revised form September 18, 1996.