Am. J. Hum. Genet. 45:547–555, 1989

Detection of DNA Sequence Polymorphisms by Enzymatic Amplification and Direct Genomic Sequencing

David W. Yandell and Thaddeus P. Dryja

Department of Ophthalmology, Harvard Medical School, and Massachusetts Eye and Ear Infirmary, Boston

Summary

The discovery of RFLPs and their utilization as genetic markers has revolutionized research in human molecular genetics. However, only a fraction of the DNA sequence polymorphisms in the human genome affect the length of a restriction fragment and hence result in an RFLP. Polymorphisms that are not detected as RFLPs are typically passed over in the screening process though they represent a potentially important source of informative genetic markers. We have used a rapid method for the detection of naturally occurring DNA sequence variations that is based on enzymatic amplification and direct sequencing of genomic DNA. This approach can detect essentially all useful sequence variations within the region screened. We demonstrate the feasibility of the technique by applying it to the human retinoblastoma susceptibility locus. We screened 3,712 bp of genomic DNA from each of nine individuals and found four DNA sequence polymorphisms. At least one of these DNA sequence polymorphisms was informative in each of three families with hereditary retinoblastoma that were not informative with any of the known RFLPs at this locus. We believe that direct sequencing is a reasonable alternative to other methods of screening for DNA sequence polymorphisms and that it represents a step forward for obtaining informative markers at well-characterized loci that have been minimally informative in the past.

Introduction

Nearly 10 years ago, Botstein et al. (1980) first laid a theoretical framework for the use of restriction endonucleases in a systematic strategy for mapping the human genome. Since then, RFLPs have been used widely as genetic markers for studies of inheritance in humans and for the determination of linkage relationships among human genes. As predicted, they have allowed the construction of a meiotic linkage map of the human genome (White et al. 1985; Donis-Keller et al. 1987). RFLPs have also been used as powerful tools in the investigation of human evolution (Wainscoat et al. 1986) and have greatly advanced our understanding of a variety of human genetic diseases (reviewed by Cooper and Clayton [1988]).

RFLP analysis, however, reveals only those sequence variations that generate a detectably different restriction fragment on a Southern blot. DNA sequence polymorphisms too small to be visualized by conventional Southern blotting, including all single-base changes that do not occur within an enzyme-recognition site, are missed in conventional RFLP screening. In the past, these single-base polymorphisms have been largely ignored as potentially useful genetic markers because of the difficulty inherent in their detection. Several elegant approaches have been devised recently to detect these polymorphic bases. In particular, RNAse mismatch and denaturing gradient-gel electrophoresis (DGGE) are powerful new ways to detect DNA heteromorphism (Myers et al. 1987). We describe here a strategy that complements the other available methods of screening for DNA polymorphism. Our approach is based on direct sequencing of genomic DNA, a method that we believe has several advantages over other methods for screening well-characterized loci from which at least some DNA sequence is already known. This approach also yields information that is not available using other techniques.

Received March 7, 1989; revision received May 31, 1989.

Address for correspondence and reprints: Dr. David W. Yandell, Howe Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, MA 02114.

© 1989 by The American Society of Human Genetics. All rights reserved.

0002-9297/89/4504-0008$02.00

We have extensively studied the human retinoblastoma gene, a gene now implicated as a factor in the development of a variety of malignancies in addition to retinoblastoma. The gene has been cloned (Friend et al. 1986; Fung et al. 1987; Lee et al. 1987), and probes from the cloned gene have been used for DNA-based diagnosis of the hereditary form of retinoblastoma (Wiggs et al. 1988). In a minority (10%–20%) of all families, a deletion or rearrangement within the gene leaves a unique restriction fragment that is diagnostic for the mutant retinoblastoma allele (Dryja et al. 1986). In the remaining families, the methodology for DNA-based diagnosis has been similar to that now being used for many heritable diseases, i.e., an RFLP-based analysis using informative RFLPs closely linked to or within the disease locus. Although several useful RFLPs have been identified in the retinoblastoma gene (Bookstein et al. 1988; Wiggs et al. 1988), approximately 10% of all families referred to our laboratory for DNA-based diagnosis are not informative for any of the RFLPs. In addition, it has often been necessary to base carrier risk estimates on a single RFLP. In an effort to provide informative markers for all families, and, whenever possible, to allow risk estimates to be based on several markers spanning this large gene, we have carried out a base-by-base search for DNA polymorphisms within selected regions of the retinoblastoma gene.

Material and Methods

The 200-kb genomic region spanned by the human retinoblastoma gene was isolated in a series of overlapping inserts from 35 distinct recombinant bacteriophage lambda clones (Wiggs et al. 1988). Exon-containing segments were subcloned into Bluescribe® plasmid cloning vectors (Stratagene, Inc.). Sequencing of cloned plasmid inserts was carried out using conventional methods.

Our basic screening strategy to apply the techniques of polymerase chain reaction (PCR) (Mullis and Faloona 1987) and direct DNA sequencing (Wong et al. 1987) to analyze normal allelic variation in a number of different regions within the retinoblastoma locus. On the basis of a partial genomic sequence (McGee et al. 1989), pairs of 20-base oligonucleotide primers were synthesized so that 13 regions, each 320–1,200 bp in size, could be amplified from genomic DNA by PCR. For all regions screened, amplification and sequencing were carried out on DNA from at least nine unrelated individuals, though for many regions 15 or more individuals were analyzed. Amplified DNA sequences were compared with one another and were checked against sequence data from previously cloned plasmid inserts derived from the same region. Bases obscured by technical artifacts or other ambiguities were not tabulated. In most cases, primer pairs for amplification were derived from intron sequences that flanked one of the 27 exons of the gene, so that the PCR-amplified region contained both intron and exon sequences.

For each amplification reaction, 0.2–1.0 μg genomic DNA was prepared in a reaction buffer containing 20 mM Tris (pH 8.4 or pH 8.6), 50 mM KCl, 30 μg/BSA/ml, 400 μM of each dNTP, $MgCl_2$ at concentrations of 3.5–7.5 mM, 10–50 pmol of each oligonucleotide primer, and 0.25–1.0 units of *Taq* DNA polymerase (Perkin-Elmer Cetus). Optimal $MgCl_2$ concentrations and pH of the PCR reactions varied depending on the amplification primers. PCR amplification (30–35 rounds) was carried out following a cycle of 10 s at 94°C (denaturation), 10 s at 42–55°C (annealing), and 30 s at 70–71°C (polymerization) using a programmable thermal cycler (Ericomp Corp., San Diego). Optimal annealing temperatures varied for each primer pair. All times are based on sample temperature rather than on heat-block temperature and do not include ramping time for the heat block. Genomic DNA from 9–20 individuals was amplified for each region and screened for DNA sequence polymorphisms.

Prior to sequencing, all PCR-amplified DNA samples were treated with proteinase-K and extracted with phenol/chloroform. High-molecular-weight DNA was separated from unused dNTPs and oligonucleotide primers by column purification through sepharose CL-68 (Pharmacia). Double-stranded PCR-amplified template (250–400 ng) was combined with 1–2 pmol (P-32) end-labeled sequencing primer and was heat denatured for 3 min at 94°C. In most cases, the sequencing primer used was one of the two amplification primers, though for very large fragments an internal primer was synthesized. The primer-template mixture above was added to a buffer containing $MgCl_2$ (2.5 mM), Tris-HCl pH 7.5 (5 mM), 6 units Sequenase (U.S. Biochemical), and dithiothreitol (3 mM) and was divided into four reaction mixtures, each containing all four deoxynucleotides (32 μM each) and one dideoxynucleotide (5 μM). This mixture was immediately incubated for 5 min at 37–42°C, and polymerization was stopped with a 0.37% EDTA stop buffer. Prior to being loaded on sequencing gels, the samples were heat denatured at 94°C for 2 min. Conventional 0.4-mm thick, 6% polyacrylamide sequencing gels were used, and autoradiography was typically for 12–24 h without an intensifying screen.

The following procedure was used to assay for alleles of the polymorphism we have called RB1.20: PCR amplification from genomic DNA was carried out as described above (see table 1 for primers used). One-tenth of the amplified DNA (approximately 50 ng) was combined with 10 pM of (P-32) end-labeled oligonucleotide primer, and the mixture was heat denatured for 3 min at 94°C. The primer-template mixture was added to a buffer containing $MgCl_2$ (2.5 mM), Tris-HCl pH 7.5 (5 mM), 2 units Sequenase (U.S. Biochemical), dithiothreitol (3 mM), and all four deoxynucleotides (60 μM each). Polymerization was carried out for 20 min at 37°C and was stopped with a 0.37% EDTA stop buffer. The double-stranded product of this reaction was heat denatured and electrophoresed on 6% polyacrylamide gels as described above for sequencing reactions. Autoradiography was for 2–12 h.

Results

We found four DNA sequence polymorphisms within the 3,712 bp of genomic DNA sequence that we screened (table 2; locations shown in fig. 1). All four polymorphisms occur in introns; of these polymorphisms, one is probably a rare variant (found in only 1 of 15 individuals analyzed). We found no evidence for linkage disequilibrium among these markers.

A representative example (RB1.3) of a single-base-pair polymorphism identified by this method is illustrated in figure 2. This polymorphism occurs near exon 3 of the retinoblastoma gene, where either an A or a G is present at the polymorphic site. Neither form of the polymorphic sequence forms the recognition site of a known restriction enzyme, and hence this DNA sequence polymorphism is not detectable as an RFLP. In a total of 82 chromosomes from unrelated individuals, no other base was observed at this site. Figure 2 also illustrates the value of this polymorphism for diagnostic evaluation. It can be seen that the affected father, who has passed the disease to two children, is heterozygous for RB1.3. Both affected children inherited the G allele, while the unaffected child inherited the A allele. Therefore, in this family the G allele from the affected parent is in phase with and diagnostic for the disease-predisposing mutation.

Figure 3 shows our analysis, using RB1.3, of three

Table 1

Polymorphic Sequences and Primer Pairs Used for PCR Amplification

Polymorphic Sequence	Allele Frequency (%)	Amplification Primer Pair	Fragment Size (bp)	Location
RB1.2:				
TAAAATAAGATCTTAAAG	>95	5′-AAGTGTAATGTTTTTCTAAG-3′	431	124 bp from 5′ end of exon 2
TAAAATAAGACTTAAAG	<5	5′-TAGCAGAGGTAAATTTCCTC-3′		
RB1.3:				
CAGAATTCGTTTCCTTTT	73	5′-TTCAAATATATGCCATCAGA-3′	530	43 bp from 3′ end of exon 3
CAGAATTCATTTCCTTTT	27	5′-GCTTACACATGAATAGTGAGAG-3′		
RB1.20:				
GATTT{CTTT(T)}$_{n\,=\,14\text{–}26}$CCTTTT ...	...[a]	5′-AATTAACAAGGTGTGGTGG-3′	550–600	54 bp from 3′ end of exon 20
		5′-CTTGTAATATGCCTCATAAT-3′		
RB1.26:				
ATTTTTTTAATCTGCAGT	85	5′-ATTCAGTGAAGATATCTAAT-3′	683	10 bp from 5′ end of exon 26
ATTTTTTAAATCTGCAGT	15	5′-TAGTTCCTCTTTGTAGTTCT-3′		

NOTE.—Shown are the sequences and locations of the polymorphic sites and their immediate flanking regions. Also shown are the oligonucleotide primers used to amplify these sequences from human genomic DNA. "Fragment size" refers to the PCR-amplified product. Allele frequencies are based on analysis of the following numbers of individuals (of mixed North American descent): (RB1.2), 14; (RB1.3), 41; (RB1.20), 16; and (RB1.26), 27.

[a]Accurate frequencies not yet determined, as numerous alleles were found.

Table 2

DNA Sequence Polymorphisms Detected by Direct Sequencing of 13 Separate PCR-amplified Regions from the Human Retinoblastoma Locus

	No. of Base Pairs Screened	Polymorphisms
Introns	2,072	4
Exons	1,640	0
Total	3,712	4

other retinoblastoma-prone families. In family RB-32, an intragenic deletion in one copy of the RB gene, presumably causing the predisposition to the tumor, was identified by Southern blotting (data not shown). The deletion involves exons 2–17 and includes the region surrounding RB1.3; therefore, carriers of the disease-predisposing allele are genotypically hemizygous at this polymorphic site. On the basis of these results, it can be predicted that the unaffected child in family RB-32 who has the G− genotype is a carrier of the mutation. As the analysis of RB1.3 clearly shows, two members of this extended pedigree are unaffected yet are carriers for the disease predisposition. This feature of reduced penetrance is characteristic of hereditary retinoblastoma and highlights the need for DNA-based genetic counseling in these families. The other unaffected children in pedigree RB-32 are not carriers of the cancer predisposition. Pedigrees RB-36 and RB-50 also demonstrate cosegregation of the polymorphism and the disease.

The most valuable polymorphism we found, designated RB1.20, is shown in figure 4. At this site near the 3′ end of exon 20 in the retinoblastoma gene, a variable number (n = 14–26) of CTTT(T) repeats occurs, leading to PCR fragments that vary in size from 550 to 600 bp. This sequence variation leads to alleles that may differ in length by as little as 1 bp. This polymorphism is not visible on conventional Southern blots, because of the small size differences among alleles. Direct genomic sequencing of this region is difficult, apparently because of secondary DNA structure associated with the region. For these reasons we have assayed the relative sizes of allelic fragments amplified by PCR by using polyacrylamide gels capable of resolving size differences of a single base pair. Because of the large number of alleles (at least 10) seen at this locus, we have not yet established accurate allele frequencies. However, 94% of the unrelated individuals examined so far have been heterozygotes for this polymorphism. This is consistent with our expectation, based on the number of alleles we found, that at least 90% of all matings will be informative for this polymorphism (Asmussen and Clegg 1985). Mendelian segregation was observed in all families that we analyzed.

From figure 4 it can be seen that the band shared by individuals I-2, II-2, and II-4 tags the retinoblastoma-predisposing allele in this family. Individual III-1 was a newborn child when referred to us for diagnosis. On the basis of the data shown, it can be seen that he has not inherited the retinoblastoma-predisposing allele, and a prediction can be made that this child is not a carrier for the disease. This family was not informative

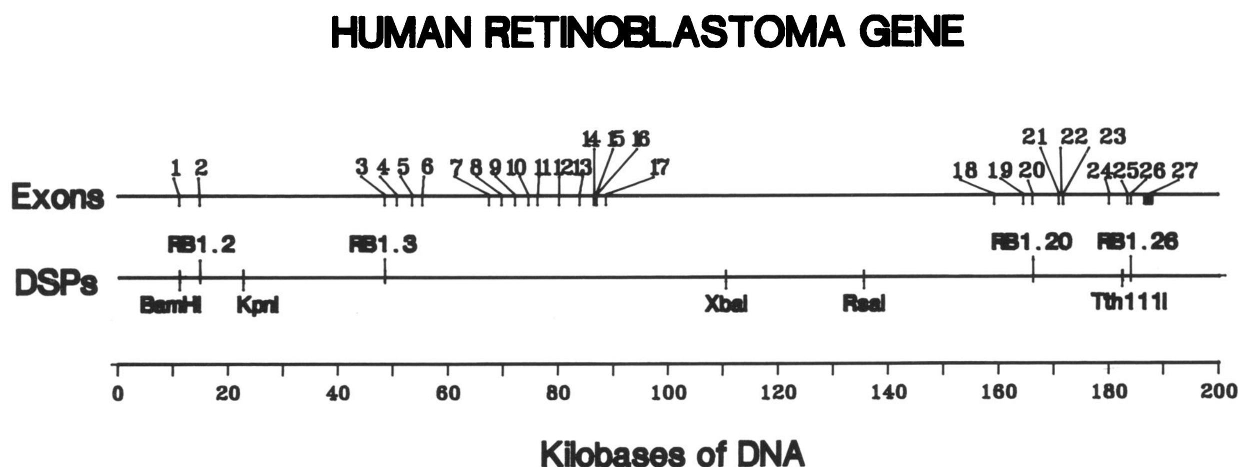

Figure 1 Scale map of the human retinoblastoma gene. The 27 exons make up a 4.7-kb transcript. The figure shows the locations of the DNA sequence polymorphisms (DSPs) identified in this gene. Polymorphisms identified by the name of a restriction enyme are RFLPs; polymorphisms RB1.2, RB1.3, RB1.20, and RB1.26 are not detectable as RFLPs and were found by PCR amplification and direct sequencing.

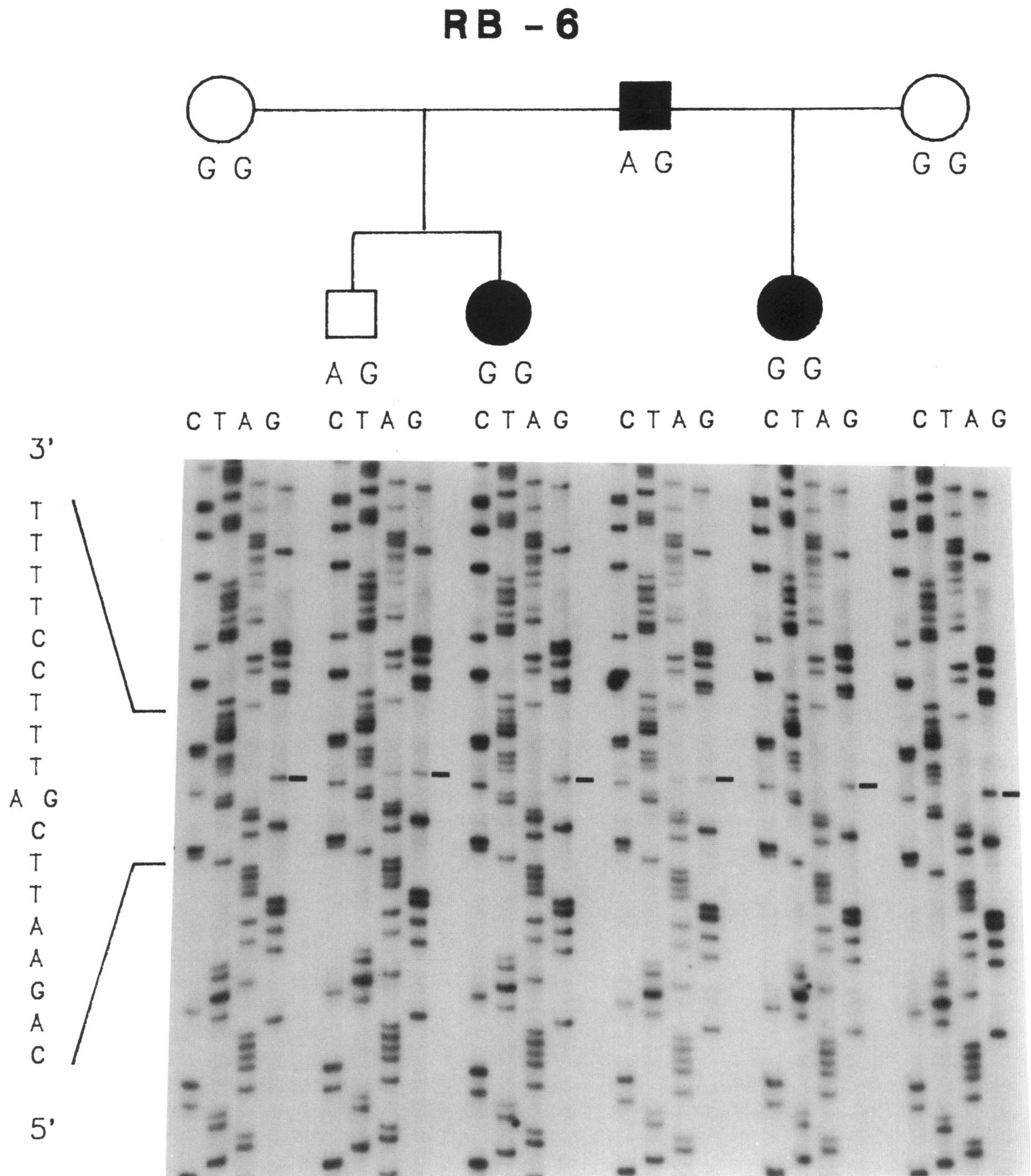

Figure 2 DNA sequence analysis showing inheritance of the polymorphism RB1.3 in a retinoblastoma-prone family. Affected individuals are indicated by filled symbols. Oligonucleotide primers (see table 1) were used to PCR amplify a 530-bp region of the human retinoblastoma gene that includes exon 3. The sequence surrounding the polymorphism is written at the left side of the figure, read 5′ to 3′ from bottom to top, and the polymorphic site in each sample is identified by adjacent tic marks.

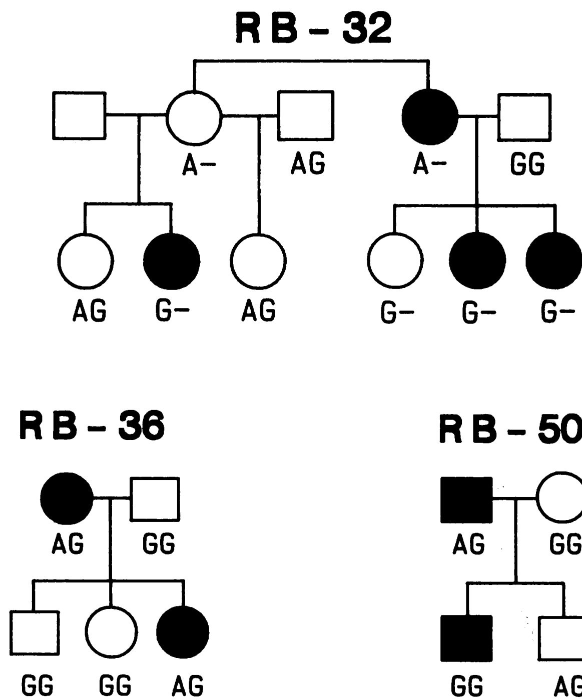

Figure 3 Segregation of the DSP RB1.3 in three families with hereditary retinoblastoma. Alleles are shown beneath the symbol for each person.

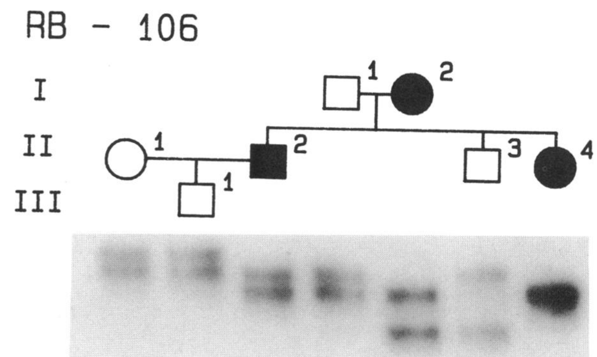

Figure 4 Analysis of the polymorphism RB1.20 in a retinoblastoma-prone family. Bands shown represent variable-sized PCR fragments after labeling of the fragments and separation on a 6% polyacrylamide gel. The fragments vary in size within the range of 550–600 bp.

for any of the known RFLP markers within 13q14. Two additional families we examined (data not shown) that were uninformative for all available RFLP markers were also informative for RB1.20, and thus for the first time they could be offered DNA-based diagnosis and genetic counseling.

Discussion

We have used PCR amplification and direct genomic sequencing to identify several new and very useful DNA sequence polymorphisms in the retinoblastoma gene. None of the markers could have been found by restriction enzyme–based screening. As would be expected, no crossovers were observed between the polymorphic sites and the retinoblastoma-predisposing trait in any informative pedigree with hereditary retinoblastoma. This follows our expectations, since the polymorphisms are within the disease gene. In addition, we have been able to offer DNA-based diagnosis and genetic counseling based on these new polymorphisms to three families that could not be helped with any of the available RFLPs because of a lack of heterozygosity among parents. These results highlight the potential diagnostic value of many as yet undetected DNA sequence polymorphisms.

It is not surprising that additional polymorphic markers can be identified by direct sequencing, as this methodology has been used successfully by us and others to detect disease-causing point mutations and heritable changes in several human genes (Wong et al. 1987; Amselem et al. 1988; Youssoufian et al. 1988; Horowitz et al. 1989). Direct sequencing detects essentially all useful nucleotide variability in the region screened, while RFLP analysis misses what may be a large fraction of single-base polymorphisms. Our previous experience in screening the retinoblastoma gene for RFLPs is a good example. We isolated more than 30 probes from the 200-kb transcriptional unit of the retinoblastoma gene and used those probes to screen DNA from six unrelated individuals digested with 33 different restriction enzymes (Wiggs et al. 1988). On the basis of the restriction enzymes used and the number of nonoverlapping probes screened, we have calculated that this restriction enzyme–based approach surveyed, at most, 5,000 (roughly 2.5%) of the 200,000 bases of the retinoblastoma gene and hence would not have detected more than 2.5% of the existing single-base polymorphisms. This screening required several months of labor and revealed five RFLPs (Wiggs et al. 1988). If this gene is similar to other human genes, the number of polymorphic nucleotides within this 200-kb region is likely to be 600 or more (Nei 1987).

Although the number of known polymorphic DNA markers in the human genome has increased vastly in recent years, we believe that our experience with DNA-based diagnosis using the cloned retinoblastoma gene

is not unusual. Because almost all human genes and their flanking sequences are likely to be polymorphic at the single-base level (Nei 1988), the problem of occasional uninformative matings should be resolvable for most loci. Recent attempts at increasing the number of known informative polymorphic markers in the genome have included systematic strategies for maximizing RFLP detection within a defined region (Aldridge et al. 1984; Feder et al. 1985; Boehm et al. 1987) and large-scale RFLP screening of many hundreds of random genomic clones (Schumm et al. 1988). A very important step forward came with the discovery, by Nakamura et al. (1987), of the relatively common and highly informative length polymorphisms that arise at VNTRs. These discoveries have been particularly valuable for large-scale linkage mapping. However, the RNAse mismatch and denaturing-gradient-gel methods and subsequent improvements in these techniques hold greater potential for improving the PIC value at specific loci that have been minimally informative in the past (Myers et al. 1987; Sheffield et al. 1989). These techniques detect a much higher proportion of single-base changes in a given region than can be detected by RFLP screening, but they do not detect all base changes and hence are less efficient than direct sequencing. A further disadvantage is that while these techniques can efficiently identify the presence of a polymorphism within a 100–1,000-bp fragment, it is often desirable to know the exact sequence surrounding the polymorphism, and these techniques require a second round of characterization (sequencing) before the polymorphism is optimally useful.

The data we present here may also be used to estimate, from a new perspective, the level of nucleotide diversity (heterozygosity at the nucleotide level) in the genome. Previous estimates at other loci have been based primarily on restriction-enzyme screening and are subject to a bias. This bias arises because the sequences recognized by restriction enzymes do not necessarily reflect a random sampling of all 4- or 6-base combinations possible (Bishop et al. 1983; Drmanac et al. 1986), and it is likely that a substantially higher level of polymorphism occurs at CpG pairs than elsewhere (Barker and White 1982; Youssoufian et al. 1988). This is evidenced by the relatively high proportion of RFLPs revealed by such enzymes as *Msp*I (CCGG) and *Taq*I (TCGA) (Feder et al. 1985; Devor 1988). Direct sequencing is not subject to this bias. From our results (table 2), it can be calculated (Nei 1975) that nucleotide diversity within the region we screened is approximately 0.04%. If the intron sequences are considered alone, this estimate increases to 0.07%. These estimates are below the predictions of others, which range from 0.15% to 0.65% or more (Jeffreys 1979; Cooper and Schmidtke 1984; Cooper et al. 1985), and they may reflect the absence from our methods of the bias described above. However, an analogous calculation (Hudson 1982), based instead on the results of our initial RFLP screening of this locus, yields a range of 0.044%–0.087% and is quite consistent with our estimate based on direct sequencing. One explanation for our results is that the human retinoblastoma gene is intrinsically less polymorphic than other regions of the genome, for which the average nucleotide diversity is approximately 0.3% (Nei 1988). Although mutations in the retinoblastoma gene are known to be early events in the formation of several types of cancer (Friend et al. 1987; Harbour et al. 1988; Lee et al. 1988; T'Ang et al. 1988; Weichselbaum et al. 1988), it is unclear why neutral polymorphism at this locus may have been selected against in human evolution (if this is indeed the case).

In summary, we believe that direct sequencing is a viable method for the detection of useful DNA polymorphisms in well-characterized loci. This approach is a valuable complement to RFLP, DGGE, and RNAse mismatch methods, since it detects essentially all polymorphic sites within the region screened and hence will maximize the density of informative markers within the region screened. The sequencing protocol we present utilizes double-stranded PCR-amplified template and requires minimal purification of the PCR-amplified DNA; hence, this approach is very rapid for screening areas where some genomic sequence is already known. Because the only requirement for utilization of these polymorphic markers is knowledge of a unique set of amplification primer sequences and of the polymorphism itself, publication of a polymorphism immediately makes it available to others. Hence, problems and delays associated with the physical transfer of plasmid DNAs between laboratories are avoided, and the costs of maintaining plasmid repositories ultimately will be reduced. Finally, because this technique reveals the sequence surrounding the polymorphism, rapid analysis of these polymorphic markers can be carried out on a large scale with the use of allele-specific oligonucleotide probes for direct hybridization to amplified DNA (Saiki et al. 1986).

Acknowledgments

We are grateful to Terri L. McGee, Tracey A. Campbell,

Siri H. Dayton, and Joyce Rapaport for technical assistance. This work was supported by National Eye Institute grants EY07573 and EY05321, American Cancer Society grant JFRA-211, an award from the Alcon Research Institute, and gifts to the Taylor R. Smith Laboratory of the Massachusetts Eye and Ear Infirmary. T.P.D. is a Research to Prevent Blindness–Dolly Green Scholar.

References

Aldridge J, Kunkel L, Bruns G, Tantravahi U, Lalande M, Brewster T, Moreau E, et al (1984) A strategy to reveal high frequency RFLPs along the human X-chromosome. Am J Hum Genet 36:546–564

Amselem S, Nunes V, Vidaud M, Estivill X, Wong C, d'Auriol L, Vidaud D, et al (1988) Determination of the spectrum of β-thalassemia genes in Spain by use of dot-blot analysis of amplified β-globin DNA. Am J Hum Genet 43: 95–100

Asmussen MA, Clegg MT (1985) Multiallelic restriction fragment polymorphisms in genetic counseling: population genetic considerations. Hum Hered 35:129–142

Barker D, White R (1982) More base pair change polymorphisms at sites containing CpG. Cytogenet Cell Genet 32:253–257

Bishop DT, Williamson JA, Skolnick MH (1983) A model for restriction fragment length distributions. Am J Hum Genet 35:795–815

Boehm TLJ, Werle A, Drahovsky D (1987) A rapid and efficient screening method for DNA restriction fragment length polymorphisms. Clin Genet 32:349–354

Bookstein R, Lee EYHP, To H, Young LJ, Sery TW, Hayes RC, Friedmann T, et al (1988) Human retinoblastoma susceptibility gene: genomic organization and analysis of heterozygous intragenic deletion mutants. Proc Natl Acad Sci USA 85:2210–2214

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Cooper DN, Clayton JF (1988) DNA polymorphisms and the study of disease associations. Hum Genet 78:299–312

Cooper DN, Schmidtke J (1984) DNA restriction fragment length polymorphisms and heterozygosity in the human genome. Hum Genet 66:1–16

Cooper DN, Smith BA, Cooke HJ, Neimann S, Schmidtke J (1985) An estimate of unique DNA sequence heterozygosity in the human genome. Hum Genet 69:201–205

Devor EJ (1988) The relative efficiency of restriction enzymes: an update. Am J Hum Genet 42:179–182

Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, et al. (1987) A genetic linkage map of the human genome. Cell 51:319–337

Drmanac R, Petrovic N, Glisin V, Crkvenjakov R (1986) A calculation of fragment lengths from human DNA with 78 restriction enzymes: an aid for cloning and mapping. Nucleic Acids Res 14:4691–4692

Dryja TP, Rapaport JM, Joyce JM, Petersen RA (1986) Molecular detection of deletions involving band q14 of chromosome 13 in retinoblastoma. Proc Natl Acad Sci USA 83:7391–7393

Feder J, Yen L, Wijsman E, Wang L, Wilkins L, Schroeder J, Spurr N, et al (1985) A systematic approach for detecting high-frequency restriction fragment length polymorphisms using large genomic probes. Am J Hum Genet 37:635–649

Friend SH, Horowitz JM, Gerber MR, Wang XF, Bogenmann E, Li FP, Weinberg RA (1987) Deletions of a DNA sequence in retinoblastomas and mesenchymal tumors: organization of the sequence and its encoded protein. Proc Natl Acad Sci USA 84:9059–9063

Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JA, Albert DA, Dryja TP (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature 323:643–646

Fung YKT, Murphree AL, T'Ang A, Qian J, Hinrichs SH, Benedict WF (1987) Structural evidence for the authenticity of the human retinoblastoma gene. Science 236:1657–1660

Harbour JW, Lai SL, Whang-Peng J, Gazdar AF, Minna JD, Kaye FJ (1988) Abnormalities in structure and expression of the human retinoblastoma gene in SCLC. Science 241:353–357

Horowitz J, Yandell D, Park SH, Canning S, Whyte P, Buchkowich K, Harlow E, et al (1989) Point mutational inactivation of the retinoblastoma antioncogene. Science 243: 937–940

Hudson RR (1982) Estimating genetic variability with restriction endonucleases. Genetics 100:711–719

Jeffreys AJ (1979) DNA sequence variants in the *gamma*-G, *gamma*-A, *delta*-, and *beta*-globin genes of man. Cell 18:1–10

Lee EYHP, To H, Shrew JY, Bookstein R, Scully P, Lee WH (1988) Inactivation of the retinoblastoma susceptibility gene in human breast cancers. Science 241:218–221

Lee WH, Bookstein R, Hong F, Young LJ, Shew JY, Lee EYHP (1987) Human retinoblastoma susceptibility gene: cloning, identification, and sequence. Science 235:1394–1399

McGee TL, Yandell DW, Dryja TP (1989) Structure and partial genomic sequence of the human retinoblastoma susceptibility gene. Gene 80:119–128

Mullis KB, Faloona F (1987) Specific synthesis of DNA *in vitro* via a polymerase catalyzed chain reaction. Methods Enzymol 155:335–351

Myers RM, Maniatis T, Lerman LS (1987) Detection and localization of single base changes by denaturing gradient gel electrophoresis. Methods Enzymol 155:501–527

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T,

Culver M, Martin C, et al (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616–1622

Nei M (1975) Molecular population genetics and evolution. North-Holland, Amsterdam

——— (1987) Molecular evolutionary genetics. Columbia University Press, New York

——— (1988) Relative roles of mutation and selection in the maintenance of genetic variability. Phil Trans R Soc Lond [B] 319:615–629

Saiki RA, Bugawan TL, Horn GT, Mullis KB, Erlich HA (1986) Analysis of enzymatically amplified *beta*-globin and HLA-DQ *alpha* DNA with allele-specific oligonucleotide probes. Nature 324:163–166

Schumm JW, Knowlton RG, Braman JC, Barker DF, Botstein D, Akots G, Brown VA, et al (1988) Identification of more than 500 RFLPs by screening random genomic clones. Am J Hum Genet 42:143–149

Sheffield VC, Cox DR, Lerman LS, Myers RM (1989) Attachment of a 40-base-pair G+C rich sequence (GC clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. Proc Natl Acad Sci USA 86:232–236

T'Ang A, Varley JM, Chakraborty S, Murphree AL, Fung YKT (1988) Structural rearrangement of the retinoblastoma gene in human breast carcinoma. Science 242:263–266

Wainscoat JS, Hill AVS, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, et al (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature 319:491–493

Weichselbaum RR, Beckett M, Diamond A (1988) Some retinoblastomas, osteosarcomas and soft-tissue sarcomas may share a common etiology. Proc Natl Acad Sci USA 85:2106–2109

White RL, Leppert D, Bishop DT, Barker D, Berkowitz J, Brown C, Callahan P, et al (1985) Construction of linkage maps with DNA markers for human chromosomes. Nature 313:101–105

Wiggs JL, Nordenskjold M, Yandell DW, Rapaport J, Grondin V, Janson M, Werelius B, et al (1988) Prediction of the risk of hereditary retinoblastoma using DNA polymorphisms within the retinoblastoma gene. N Engl J Med 318:151–157

Wong C, Dowling CE, Saiki RK, Higuchi RG, Erlich HA, Kazazian HH (1987) Characterization of *beta*-thalassaemia mutations using direct genomic sequencing of amplified single copy DNA. Nature 330:384–386

Youssoufian H, Wong C, Aronis S, Platokoukis H, Kazazian HH Jr, Antonarakis SE (1988) Moderately severe hemophilia A resulting from Glu→Gly substitution in exon 7 of the factor VIII gene. Am J Hum Genet 42:867–871