High-Accuracy DNA Sequence Variation Screening by DHPLC

BioTechniques 29:1084-1092 (November 2000)

Jamie I. Spiegelman, Michael N. Mindrinos and Peter J. Oefner
Stanford Genome Technology Center, Palo Alto, CA, USA

ABSTRACT

*Genetic maps based on biallelic single-nucleotide polymorphisms amenable to microarray-based genotyping have significantly accelerated the mapping of mono- and multigenic traits in model organisms such as *Saccharomyces cerevisiae* and *Arabidopsis thaliana*. This advance needs to be matched by highly accurate, inexpensive and robust methodology for fine-structure mapping of the candidate region(s) and the eventual identification of the causative mutation(s). To establish the usefulness of denaturing high-performance liquid chromatography (DHPLC) for those purposes, we have amplified 476 fragments from two *A. thaliana* ecotypes with an average length of 563 bp covering various candidate regions on chromosomes 1, 2 and 4. Parallel analysis by DHPLC and dye terminator sequencing showed that DHPLC detected 165 out of 166 polymorphic fragments with only four false positives, amounting to a sensitivity, specificity and accuracy of 99.4%, 98.7% and 99%, respectively. It proved beneficial to analyze the fragments not only at the highest but also at the lower temperatures recommended by the algorithm freely available at <http://insertion.stanford.edu/melt.html>.*

INTRODUCTION

Denaturing high-performance liquid chromatography (DHPLC) is a novel technology that compares two or more chromosomes as a mixture of denatured and renatured products ≤ 1000 bp in size (15). The presence of single or multiple mismatches is revealed by the appearance of one or more early eluting peaks in the chromatographic profile representing heteroduplex species. DHPLC has been carried out under both completely and partially denaturing conditions using column temperature to modulate the degree of denaturation (14,15). Applications of DHPLC include the discovery of simple sequence polymorphisms (1,2,18), the mapping of genes (17), the mutational analysis of candidate genes (9,16,19), the targeted screening for induced mutations (12), direct allelic discrimination (14), the analysis of primer extension products (7) and the quantification of gene expression (6).

A number of studies (3–5,8,9,13) have found DHPLC to be superior to single strand conformation analysis (SSCA). Depending on the size and sequence of the PCR products, SSCA detected only 50%–97.5% of the mutations discovered by DHPLC. More generally, DHPLC has been documented to find 95%–100% of mutations identified previously by other methods such as sequencing, SSCA, denaturing gel electrophoresis and protein truncation test (3–5,8,10,13,19). Further, no additional mutations were reported upon resequencing of all or part of the *BRCA1* and *PTEN* exons analyzed prior to that by DHPLC (5,10,19).

The accuracy of DHPLC has been improved substantially by the development of an algorithm that predicts the optimum temperature at which a simple sequence variant in a given melting domain can be detected (8). It is freely available at <http://insertion.stanford.edu/melt.html>. In a blind analysis of 103 known mutations, 99 were detected at the temperature recommended by the algorithm (8). However, the study only assessed the ability of DHPLC to detect single mismatches in 42 different sequence backgrounds with an average size of 308 bp. In reality, one may encounter more than one mutation in a sequence, particularly when the product size is maximized to minimize the number of fragments for screening. To address the concern of whether temperature recommendations made by the algorithm hold for larger fragments containing single or multiple mismatches, we compared a total of 476 PCR products with an average length of 563 bp from the *Arabidopsis thaliana* ecotypes Columbia and Landsberg erecta by both DHPLC and, subsequently, direct sequencing.

MATERIALS AND METHODS

Sequences and Primer Design

The sequences of the bacterial artificial chromosome (BAC) clones listed in Table 1 were downloaded from www.arabidopsis.org or www.ncbi.nlm.nih.gov. The oligonucleotide primers were designed with a uniform melting temperature of 60°C and yielded products with a size range of 464–646 bp.

Table 1. Characterization of Products and Nature of Sequence Variants Detected Between the *A. thaliana* Ecotypes Columbia and Landsberg erecta

Genomic location	Products Analyzed	Average Size (bp)	Polymorphic Products	Products with 1 SNP	Products with 2 SNPs	Products with 3 SNPs	Products with InDels ^d	Products with SNPs + InDels
Chromosome 1 ^a	268	558	48	28	10	9	6	10
Chromosome 2 ^b	137	569	72	20	9	5	3	20
Chromosome 4 ^c	71	569	46	13	2	10	2	19
Total numbers	476	563	166	61	21	24	11	49

^a1BACs T7I23, T25K16, F23A5, T21F11, F5I6, F18B13.
^b2BACs F23H14, F10A8, F2I9, T8O11, T9H9, T9J23, F17A22, T30B22, T8I13.
^cBACs F6N15, F5I10, F6N23, F23K16, T19P19, T5J17.
^dInsertion(s) and/or deletion(s).

PCR Amplification

PCR was performed in 50-μL volumes containing approximately 50 ng genomic DNA from either Columbia or Landsberg erecta, 0.15 μM each primer, 1 U AmpliTaq Gold® DNA Polymerase (Applied Biosystems, Foster City, CA, USA), 250 μM dNTPs (Amersham Pharmacia Biotech, Piscataway, NJ, USA), 10 mM Tris-HCl, pH 8.3, 50 mM KCl and 2.5 mM MgCl₂. Thermal cycling was performed in a GeneAmp® PCR System 9700 (Applied Biosystems). Initial denaturation (of DNA)/activation (of enzyme) at 95°C for 10 min was followed by 16 cycles of denaturation at 95°C for 20 s, primer annealing under touch-down conditions starting at 60°C for 1 min and decreasing by 0.5°C in each consecutive cycle, and primer extension at 72°C for 1 min. We then performed 25 additional cycles at 95°C for 20 s, 52°C for 45 s and 72°C for 1 min. A final extension reaction was performed at 72°C for 7 min, after which the reaction was held at 4°C. Success of amplification was verified by means of agarose gel electrophoresis.

DHPLC

Heteroduplexes were generated by mixing fragments amplified separately from Columbia and Landsberg erecta at roughly equimolar ratio before DHPLC analysis. The mixture was denatured at 95°C for 3 min and allowed to renature

over 30 min by decreasing the temperature from 95°C to 65°C. Each heteroduplex was run on an automated WAVE® DHPLC instrument equipped with a DNasep® column (Transgenomic, San Jose, CA, USA) at the product's recommended temperature(s) as predicted by the melting algorithm available at <http://insertion.stanford.edu/melt.htm>. The mobile phase was 0.1 M triethylammonium acetate (Applied Biosystems), pH 7.0, and 0.1 mM Na₄EDTA (Sigma, St. Louis, MO, USA). The latter reagent was added to chelate metal cations that exert a detrimental effect on separation efficiency (15). DNA fragments were eluted at a flow-rate of 0.9 mL/min with an acetonitrile gradient (J.T. Baker, Phillipsburg, NJ, USA). The start and end points of the gradient were calculated with the WAVE MAKER program (version 3.3.4; Transgenomic), and slight adjustments were made following the first run to ensure a time window of 1.5 min between the void and the product peak. Between runs, the column was flushed for 0.5 min with 50% acetonitrile using the WAVE Accelerator and re-equilibrated for 1.5 min at the starting conditions of the gradient. Chromatograms were examined visually for differences between the Columbia control only and the Columbia Landsberg erecta duplex.

Cycle Sequencing

To remove excess oligonucleotide primers and dNTPs, 10 μL each product

were treated with 1 U each Exonuclease I and Shrimp Alkaline Phosphatase for 30 min at 37°C and 15 min at 80°C. The PCR products were then sequenced with the BigDye™ Deoxy Terminator cycle sequencing kit (Applied Biosystems) according to the manufacturer's instructions. Following solid phase extraction with Sephadex® G-50 (Amersham Pharmacia Biotech), the sequencing reactions were analyzed with a Model 377A sequencer (Applied Biosystems). Sequence analysis was facilitated by the fact that both ecotypes are genetically homogeneous due to inbreeding.

RESULTS AND DISCUSSION

The sensitivity and specificity of DHPLC were determined by the comparative analysis of the Landsberg erecta ecotype to its Columbia allele under temperature conditions predicted by computer simulation of the melting behavior of the DNA segments of interest on chromosomes 1, 2 and 4. The products had an average size of 563 bp and covered a total of approximately 250 kb or about 0.2% of the entire *Arabidopsis* genome. Subsequently, the DHPLC results were confirmed by dye-terminator sequencing. Of the 476 fragments analyzed both by DHPLC and sequencing, 166 were polymorphic (Table 1), 37% (61/166) contained a single-nucleotide polymorphism (SNP) and all but one was detected readily by DHPLC at one or more of the recommended tempera-

Research Report

tures. Hence, the sensitivity of DHPLC for products containing a single base-pair mismatch was 98.4% (Table 2). The single mutation missed was detected successfully at a temperature one degree lower than the recommended one. Thirteen percent of the fragments contained two SNPs, and 14.5% of the samples had three or more SNPs. DHPLC missed none of those, and the same was true for the 11 fragments (6.6%) containing simple insertions and/or deletions (≤ 10 bp) and the 49 fragments (29.5%) containing a combination of SNPs, deletions and insertions. Among them was a fragment with a 185-bp deletion in *Landsberg erecta*, while another product contained a total of 20 SNPs and insertions/deletions.

A previous study (8) had shown that the minimum temperature range over which a mutation could be detected by DHPLC was 4°C. Actually, only 1 out of 32 mutations was not detected over a

Table 2. Sensitivity of DHPLC as a Function of the Number of SNPs per Product and the Analysis Temperature

Product Type	Cumulative Sensitivity (%) ^a		
	1st RTm ^c	1st + 2nd RTm	1st, 2nd + 3rd RTm
1 SNP	91.8	98.4	98.4
2 SNPs	100	100	100
3 SNPs	95.8	95.8	100
InDels	90.9	100	100
Compound ^b	93.4	100	100
Totals	94.0	98.8	99.8
^a Temperature recommendations in 5°C decrements according to melting algorithm provided at http://insertion.stanford.edu/melt.html . ^b Fragments containing both SNPs as well as insertions and/or deletions (InDels). ^c Recommended melting temperature.			

range of at least 6°C. Consequently, if the melting domains in a given product differ by more than 5°C, respectively 10°C, the melting algorithm will recommend one or two additional DHPLC analyses in 5°C decrements. The first recommended temperature always concurs with the temperature at which 50% of the complementary DNA strands of the highest melting domain in a given fragment are denatured. In the present study, the highest temperature recommended led to the detection of 94% of all polymorphic fragments, and sensitivity appeared to be quite uniform irrespective of the number and nature of mismatches contained in a given fragment (Table 2). The high sensitivity attained at this temperature is more striking because 78 of the 166 polymorphic fragments contained melting domains that differed by more than 5°C, while 12 had melting domains that ranged over more than 10°C. Therefore, one might have expected more sequence variants to be missed at the highest recommended temperature, but, as mentioned above, the minimum temperature range over which a mutation can be detected tends to be quite broad, enabling the detection of mutations over a wide range of melting domains. Nevertheless, analyzing those fragments at the additionally recommended temperatures of -5°C and -10°C increased sensitivity to 98.8% and 99.8%, respectively.

The majority, or 120 out of 166 polymorphic fragments, yielded characteristic profiles of two or more discrete peaks (Figure 1B) distinctly different from those of monomorphic segments (Figure 1A). The remaining 46 polymorphic fragments (27.7%) presented only with a fronting shoulder (Figure 1C) or trailing peaks (Figure 2C) that were not observed in the corresponding monomorphic Columbia control. This suggested that in such cases the chosen melting temperature might have already caused the almost complete denaturation of the fragments. This was corroborated by an improved resolution of homo- and heteroduplex species at temperatures lower than the recommended one (compare chromatograms in Figure 1C and Figure 2A). This also allowed the successful detection of the only polymorphic fragment that had been missed at the rec-

ommended temperature (Figure 2, B and C). Such optimization is of particular interest when using DHPLC, not only for the discovery but also for the genotyping of polymorphisms in the course of fine mapping a locus. With improved resolution of homo- from heteroduplex species, it becomes possible to detect a given polymorphism even when it is only present in 1 out of 10 chromosomes as shown previously (10,12). Moreover, the chromatograms in Figure 1 emphasize the importance of including a monomorphic control, not only to recognize sometimes subtle heteroduplex signals but also to reduce the number of false positives. The latter are usually the result of runs of bases or short tandem repeats causing slippage of the polymerase enzyme with subsequent formation of heteroduplexes (15). They may also indicate gene duplication (14). Such false-positive heteroduplex signals can be easily distinguished because they are also observed in the monomorphic control (Figure 1D). However, in four instances, DHPLC of the mixture of Columbia and Landsberg erecta yielded profiles that differed from those of Columbia only, though slightly. Sequencing could not confirm the presence of mismatches. It cannot be excluded entirely that the polymorphic DHPLC profiles actually indicated the presence of mismatches between duplicated regions of the genome that were missed in the sequence analysis. Recently, we have observed that dye-terminator sequencing failed to detect short deletions or insertions if they were only present in one out of five chromosomes, while single-nucleotide substitutions were still detected readily.

Overall, the temperatures recommended by the freely available melting algorithm appear to allow the successful detection of most polymorphic segments of a genome independently of the number and nature of sequence variants. In cases in which a maximum degree of sensitivity is required, smaller temperature decrements may be chosen, even though a very small number of mutations may still go undetected. For instance, a recent study (19) of 183 BRCA1 and BRCA2 mutations reported two mutations that were detected only at a unique temperature. The number of such cases is still too small to

Research Report

Figure 1. Characteristic monomorphic (A and D) and polymorphic (B and C) DHPLC profiles obtained upon the analysis of mixtures of denatured and renatured Columbia and Landsberg erecta products (right column). The left column shows profiles of Columbia fragments only that served as monomorphic controls. Parallel analysis of monomorphic controls facilitates the recognition of subtle changes in chromatographic profiles (row C) indicating the presence of mutation(s) and minimizes false positives that are caused mostly by runs of bases or short variable tandem repeats (row D). Analyses were performed at column oven temperatures of 51°C (A), 46°C (B) and 58°C (C and D), respectively.

Research Report

Figure 2. Effect of temperature optimization on the resolution of homo- and heteroduplex species. Chromatograms in row A show the analyses of the monomorphic control (left) and the mixture of Columbia and Landsberg erecta (right) from Figure 1C at a temperature 4°C lower, which resulted in an improved resolution of the heteroduplex species. Chromatograms in rows B and C show the analysis of the only polymorphic fragment that was missed at the recommended temperature of 56°C (row B), while a clearly polymorphic profile was obtained at 54°C (right chromatogram in row C). The single mismatch in this 534-bp fragment on chromosome 2 (BAC no. F219: forward primer 5'-agaggagagaggaatacacgac-3', reverse primer 5'-atgtggcccggtttattacat-3') was located at nucleotide position 483 and the flanking 20 bp on either side were TTTTAGAAAAATAAACTT[C/T]GATGTATAAAAAGATATGAC.

Research Report

identify a particular sequence motif that may compromise the ability of DHPLC to detect mutations over a wide range of temperatures.

A comparison of the PCR-based sequence data obtained in the present study for Columbia with the published reference sequence revealed 14 discrepancies corresponding to an error rate of 1 in approximately 18 000 bp. The error rates varied slightly among the three chromosomes. It was highest for chromosome 2 (1 in 12 000 bp) and lowest for chromosome 1 (1 in 22 800 bp). This agrees with the reported error rate of 1 mismatch every 12 546 bp that was derived from the independent sequencing of overlapping BAC segments (11).

In conclusion, the high accuracy and low operating cost (approximately 25 cents per run) of DHPLC make it a powerful tool for the identification and scoring of simple sequence polymorphisms in the narrowing of candidate regions identified initially by whole-genome mapping using microarray-based genotyping of biallelic markers (2,20). The high sensitivity of DHPLC also lends reassurance to the eventual screening for the mutations responsible for particular phenotypes. This is especially true in instances of reverse genetics when chemically induced mutations are embedded in thousands of monomorphic bases between the mutant and the wild-type.

ACKNOWLEDGMENTS

This work was funded by the National Institutes of Health grant no. 1R01-HG01932.

REFERENCES

1. Cargill, M., D. Altshuler, J. Ireland, P. Sklar, K. Ardlie, N. Patil, N. Shaw, C.R. Lane et al. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22:231-238.
2. Cho, R.J., M. Mindrinos, D.R. Richards, R.J. Sapolsky, M. Anderson, E. Drenkard, J. Dewdney, L. Reuber et al. 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet.* 23:203-207.
3. Choy, Y.S., S.L. Dabora, F. Hall, V. Ramesh, Y. Niida, D. Franz, J. Kasprzyk-Obara, M.P. Reeve and D.J. Kwiatkowski. 1999. Superiority of denaturing high performance liquid chromatography over single-stranded conformation and conformation-sensitive gel electrophoresis for mutation detection in TSC2. *Ann. Hum. Genet.* 63:383-391.
4. Dobson-Stone, C., R.D. Cox, L. Lonie, L. Southam, M. Fraser, C. Wise, F. Bernier, S. Hodgson et al. 2000. Comparison of fluorescent single-strand conformation polymorphism analysis and denaturing high-performance liquid chromatography for detection of EXT1 and EXT2 mutations in hereditary multiple exostoses. *Eur. J. Hum. Genet.* 8:24-32.
5. Gross, E., N. Arnold, J. Goette, U. Schwarz-Boeger and M. Kiechle. 1999. A comparison of BRCA1 mutation analysis by direct sequencing, SSCP and DHPLC. *Hum. Genet.* 105:72-78.
6. Hayward-Lester, A., P.J. Oefner and P.A. Doris. 1996. Rapid quantification of gene expression by competitive RT-PCR and ion-pair reversed-phase HPLC. *BioTechniques* 20:250-257.
7. Hoogendoorn, B., M.J. Owen, P.J. Oefner, N. Williams, J. Austin and M. O'Donovan. 1999. Genotyping single nucleotide polymorphisms by primer extension and high performance liquid chromatography. *Hum. Genet.* 104:89-93.
8. Jones, A.C., J. Austin, N. Hansen, B. Hoogendoorn, P.J. Oefner, J.P. Cheadle and M.C. O'Donovan. 1999. Optimal temperature selection for mutation detection by denaturing HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis. *Clin. Chem.* 45:1133-1140.
9. Liu, W.O., P.J. Oefner, C. Qian, R.S. Odom and U. Francke. 1998. Denaturing HPLC identified novel FBN1 mutations, polymorphisms, and sequence variants in Marfan syndrome and related connective tissue disorders. *Genet. Test.* 1:237-242.
10. Liu, W., D.I. Smith, K.J. Rechtzigel, S.N. Thibodeau and C.D. James. 1998. Denaturing high performance liquid chromatography (DHPLC) used in the detection of germline and somatic mutations. *Nucleic Acids Res.* 26:1396-1400.
11. Mayer, K., C. Schuller, R. Wambutt, G. Murphy, G. Volckaert, T. Pohl, A. Dusterhoft, W. Stiekema et al. 1999. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* 402:769-777.
12. McCallum, C.M., L. Comai, E.A. Greene and S. Henikoff. 2000. Targeted screening for induced mutations. *Nat. Biotechnol.* 18:455-457.
13. O'Donovan, M.C., P.J. Oefner, S.C. Roberts, J. Austin, B. Hoogendoorn, C. Guy, G. Speight, M. Upadhyaya, S.S. Sommer and P. McGuffin. 1998. Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics* 52:44-49.
14. Oefner, P.J. 2000. Allelic discrimination by denaturing high-performance liquid chromatography. *J. Chromatogr. B* 739:345-355.
15. Oefner, P.J. and P.A. Underhill. 1998. DNA mutation detection using denaturing high-performance liquid chromatography, p. 7.10.1-7.10.12. In N.C. Dracopoli, J.L. Haines, B.R. Korf, D.T. Moir, C.C. Morton, C.E. Seidman, J.G. Seidman and D.R. Smith (Eds.), *Current Protocols in Human Genetics*. Wiley & Sons, New York.
16. Ophoff, R.A., G.M. Terwindt, M.N. Vergouwe, R. van Eijk, P.J. Oefner, S.M.G. Hoffman, J.E. Lamerdin, H.W. Mohrenweiser et al. 1996. Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca^{2+} channel gene CAGNL1A4. *Cell* 87:543-552.
17. Schriml, L.M., R.J. Peterson, B. Gerrard and M. Dean. 2000. Use of denaturing HPLC to map human and murine genes and to validate single-nucleotide polymorphisms. *BioTechniques* 28:740-745.
18. Underhill, P.A., L. Jin, A.A. Lin, S.Q. Mehdi, T. Jenkins, D. Vollrath, R.W. Davis, L.L. Cavalli-Sforza and P.J. Oefner. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* 7:996-1005.
19. Wagner, T., D. Stoppa-Lyonnet, E. Fleischmann, D. Muhr, S. Pagès, T. Sandberg, V. Caux, R. Moeslinger et al. 1999. Denaturing high performance liquid chromatography (DHPLC) detects reliably BRCA1 and BRCA2 mutations. *Genomics* 62:369-376.
20. Winzeler, E.A., D.R. Richards, A.R. Conway, A.L. Goldstein, S. Kalman, M.J. McCullough, J.H. McCusker, D.A. Stevens et al. 1998. Direct allelic variation scanning of the yeast genome. *Science* 281:1194-1197.

Received 12 May 2000; accepted 19 July 2000.

Address correspondence to:

Dr. Peter J. Oefner
Stanford Genome Technology Center
855 California Avenue
Palo Alto, CA 94304, USA
e-mail: oefner@genome.stanford.edu

Eppendorf
4C

Circle Reader Service No. 215