Short communication

Santiago Rodríguez
Xiao-he Chen
Ian N. M. Day

Human Genetics Division,
University of Southampton,
School of Medicine,
Southampton General Hospital,
Southampton, UK

Typing dinucleotide repeat loci using microplate array diagonal gel electrophoresis: Proof of principle

Polymorphic dinucleotide repeat loci ('microsatellite markers') are found in varying abundance throughout the genomes of most organisms. They have been extensively used for genetic studies, but conventional techniques used for their genotyping require sophisticated equipment. Microplate array diagonal gel electrophoresis (MADGE) has previously been extended to economical high-throughput genotyping of trinucleotide and tetranucleotide microsatellite amplicons. However, the capability of this technique to resolve the alleles of dinucleotide repeat loci has not been explored previously. Here we show that a modified microsatellite-MADGE approach can provide sufficient resolution for dinucleotide repeat typing. This enables economical and convenient set up for analysis of single markers in many samples in parallel, suitable, for example, for population association studies.

Keywords: Dinucleotide repeat loci / High-throughput genotyping / Microplate array diagonal gel electrophoresis
DOI 10.1002/elps.200305801

Dinucleotide repeat loci are tandemly repeated tracts of DNA composed of two base pair (bp) long units. Polymorphic dinucleotide repeats have been extensively used for genetic mapping and for population and genetic epidemiological studies because they are typically highly polymorphic and abundant across genomes [1–3]. Both features make microsatellite markers potentially useful for population association studies when they are within genes, or close enough to genes, to tag haplotype blocks [4]. Examples include a dinucleotide repeat in the promoter region of insulin-like growth factor 1 (*IGF1*) [5], two dinucleotide repeats in the plasminogen activator inhibitor 1 (*PAI-1*) gene [6] and the strong linkage disequilibrium of tyrosine hydroxylase (*TH01*) with single nucleotide polymorphism haplotypes in the *IGF2*-insulin (*INS*)-*TH* region [7]. However, large sample sizes (*e.g.* 1000–100 000) are required in such studies in order to have enough power, especially in association studies for complex disease traits [8], for which individual gene contributions are small. Conventional approaches used to type dinucleotide repeats are based on end-labelling of primers and fluorescence-based detection of bands resolved on sequencing gels, and require sophisticated laboratory equipment (*e.g.*, ABI PRISM 377 DNA sequencer machine; Perkin Elmer, Norwalk, CT, USA) and reagents. The genotyping of large samples for single markers using these approaches is thus both expensive and time-consuming. We have developed a utilitarian system (microplate array diagonal gel electrophoresis, MADGE) which is suitable for high-throughput PCR checking, single nucleotide polymorphism (SNP) typing and identification of unknown mutations [9, 10]. It is microplate compatible and requires a minimal hardware investment and minimal reagent costs. This system has recently been adapted for sizing trinucleotide and tetranucleotide microsatellite alleles [11]. Electrophoresis of multiple polyacrylamide gels in a thermostatically controlled tank allows up to 1000 amplification products to be resolved in 90 min. The adaptation of this system to type single dinucleotide repeat loci in many samples in parallel, would enable reduction of costs and increase in throughput, compared with conventional approaches. In this work, we show that with attention to various protocol details, dinucleotide repeats can be genotyped on MADGE gels with sufficient resolution.

The human dinucleotide repeat D11S1760 was amplified by PCR from samples of a DNA bank of unrelated UK Caucasian subjects. This marker has been previously typed by us in other studies [12–14] and was selected for having an expected allele size range from 80 to 100 bp,

Correspondence: Dr. Santiago Rodríguez, Human Genetics Division, University of Southampton, School of Medicine, Duthie Building (MP 808), Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK
E-mail: santi@soton.ac.uk
Fax: +44-0-2380-794264

Abbreviation: MADGE, microplate array diagonal gel electrophoresis

which facilitates resolution of alleles in short tracks compared with longer dinucleotide repeat alleles. The sequences of primers were described by Dib *et al.* [1] as follows: forward primer: 5'-GATCTCAAGTGTTTCCCCAC-3'; reverse primer: 5'-AAACGATGTCTGTCCACTCA-3'. PCR thermal cycling conditions were: 1 cycle at 95°C for 3 min, and then 94°C for 40 s and 60°C for 30 s for 30 cycles, finally 72°C for 2 min. Two fragment sizing markers were used for bracketing the allele size range of D11S1760 in order to ensure accurate sizing without inter-lane variability: one of 104 bp and the other of 74 bp. Both markers were amplification products of the human *IGF1* gene. The forward primer used for both markers was 5'-CCCAAAGCCTCTCATGACACAA-3'. Reverse primers were 5'-GTGCACAATGAGAGGGAGTGAC-3' for the 104 bp marker and 5'-GGGAAGACAGGGGAG TAAAATC-3' for the 74 bp marker. PCR thermal cycling conditions for the 104 bp sizing marker were: 1 cycle at 95°C for 3 min, and then 94°C for 15 s, 63°C for 15 s and 72°C for 15 s, for 35 cycles, finally 72°C for 5 min. PCR thermal cycling conditions for the 74 bp sizing marker were the same as that for the 104 bp marker, except that annealing temperature was 56°C for 30 s. All PCR reactions were performed in 96-well Omniplates (Hybaid, Teddington, Middx., UK), each 10 μL reaction containing 0.4 ng of DNA, 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 0.05% W1, 200 μM each dNTP, 4 pmol each primer (MWG-Biotech UK, Milton Keynes, UK) and 0.1 U of *Taq* polymerase (Gibco, Paisley, UK). $MgCl_2$ concentration was 1.5 mM for all PCRs except that of the 104 bp marker, for which it was 3.0 mM.

Amplification products were resolved on polyacrylamide gels in a modified MADGE format which provides an available track length of 43.5 mm [11]. 1 × Tris-acetate-EDTA (TAE) buffer [15] was used and 150 μL of each of TEMED and 25% ammonium persulphate were added as polymerisation catalysts to 50 mL gel mix. Further modifications were performed in the gel composition, pore size and running conditions used previously in the resolution of trinucleotide and tetranucleotide repeats by microsatellite-MADGE [11]. 7% v/v glycerol was added to the gel in order to increase band definition [16]. The presence of this additive, together with the reduced pore size given by a 9% Duracryl™ matrix, resulted in a larger retarding effect on the electrophoretic migration than that used previously [11]. Increments in voltage (200 V) and in running time (150 min) permitted sufficient resolution of alleles differing two base pairs apart under these electrophoretic conditions (see below). Prior to electrophoresis, the gel was equilibrated in 1 × TAE buffer containing 5 mM KCl for 1 h to equalise local ionic conditions near anode and cathode, which otherwise become imbalanced through differential mobilities of different ions during the run [17]. Loading of samples and set up of gels in the electrophoresis tank containing 1 × TAE buffer pre-equilibrated to a temperature of 40°C, was as previously described [11]. Scanning of the gel stained in Vistra Green™ dye (Amersham Pharmacia, Amersham, UK) and image analysis were as previously described [11] (or using SyBr Green™ or Ethidium Bromide™ by standard UV box and ccd camera). Determination of D11S1760 PCR fragment sizes was performed by the following formula which assumes log-linear relationship of size (bp) and mobility:

$$U = \frac{74}{74^{[(d'74 - d'U)/d'74]}} \quad (1)$$

where U is the unknown fragment size in bp, $d'74$ is the distance between the 74 bp and 104 bp bands, and $d'U$ is the distance between the unknown fragment band and the 104 bp band.

Electrophoresis of 94 D11S1760 PCR products on a MADGE gel is shown in Fig. 1. An enlargement of the amplicons for six unrelated subjects is also depicted. Resolution of bands differing two nucleotides apart (F, Fig. 1) is achieved. Figure 2 shows two sizing markers internal to each track bracketing the band patterns of dinucleotide repeat products for each sample. The markers allow estimation of the allele size for each band and compensate for any variability in relative mobilities between different lanes. Phoretix™ electropherograms for four heterozygous samples with different genotypes are demonstrated in Fig. 3. The two bands corresponding to the genotype 82/80 (bands 3 and 4, Fig. 3), differing by one single dinucleotide repeat, are clearly distinguishable both in the gel and in the electropherogram. This confirms that the electrophoretic conditions used give sufficient resolution to detect differences in size of two nucleotides on a MADGE gel. The distance between the bands 7 and 8 (differing by four nucleotides) is twice the distance between bands 3 and 4, and similar deductions can be made for the distances between bands 1 and 2, and between 5 and 6. This even separation between bands, however, does not ensure that each allele is being accurately sized in different lanes across the gel. In order to know the accuracy of binning when genotyping all the samples present in a MADGE gel, we represented graphically the frequencies of the estimated allele sizes (to the nearest 0.1 bp) of a sample of 94 unrelated individuals. It can be seen (Fig. 4) that there is good clustering of alleles. Clusters fall at near 80 bp, 82 bp, 84 bp, 86 bp, 90 bp, 92 bp, 94 bp and 98 bp, with clear resolution between clusters. Thus, binning to repeat length sizes each differing by 2 bp, is readily achieved. Also in support of the absence of calling errors is the fact that the observed genotypes are in Hardy-Weinberg equilibrium (HWE) (N =

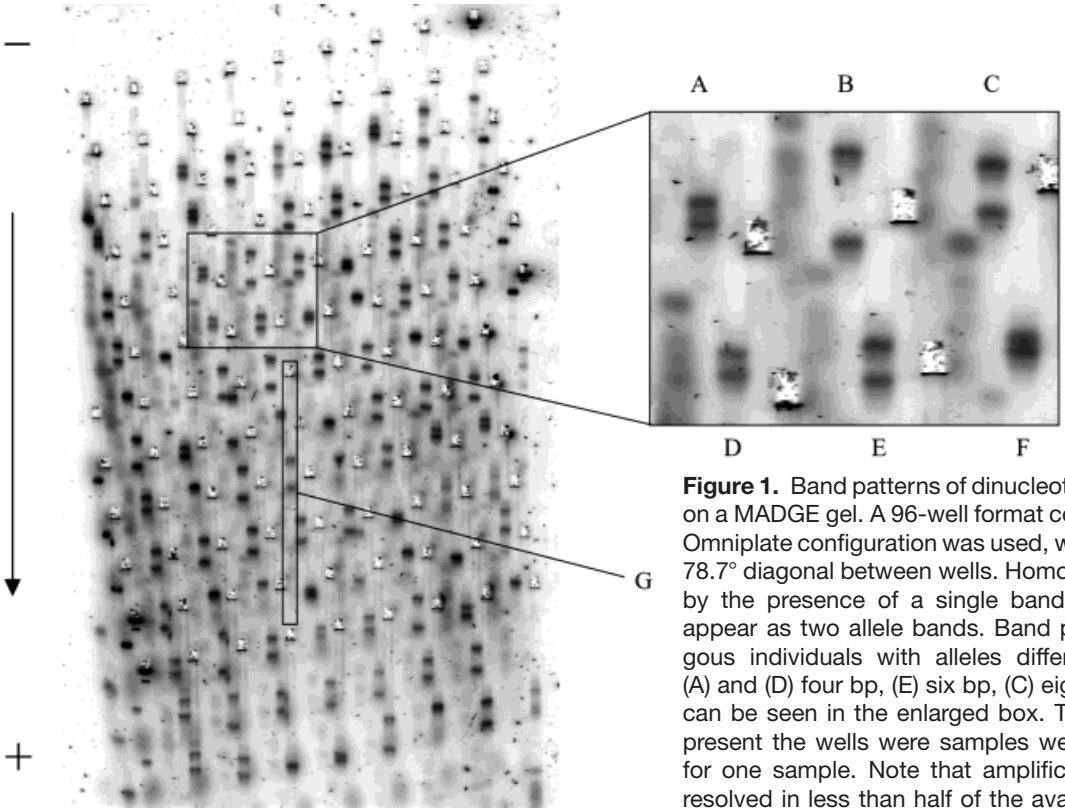

Figure 1. Band patterns of dinucleotide repeat fragments on a MADGE gel. A 96-well format compatible with 8×12 Omniplate configuration was used, with tracks following a 78.7° diagonal between wells. Homozygotes are denoted by the presence of a single band, and heterozygotes appear as two allele bands. Band patterns of heterozygous individuals with alleles differing by (F) two bp, (A) and (D) four bp, (E) six bp, (C) eight bp, and (B) 14 bp can be seen in the enlarged box. The clear squares represent the wells were samples were loaded. (G) Track for one sample. Note that amplification products were resolved in less than half of the available track length.

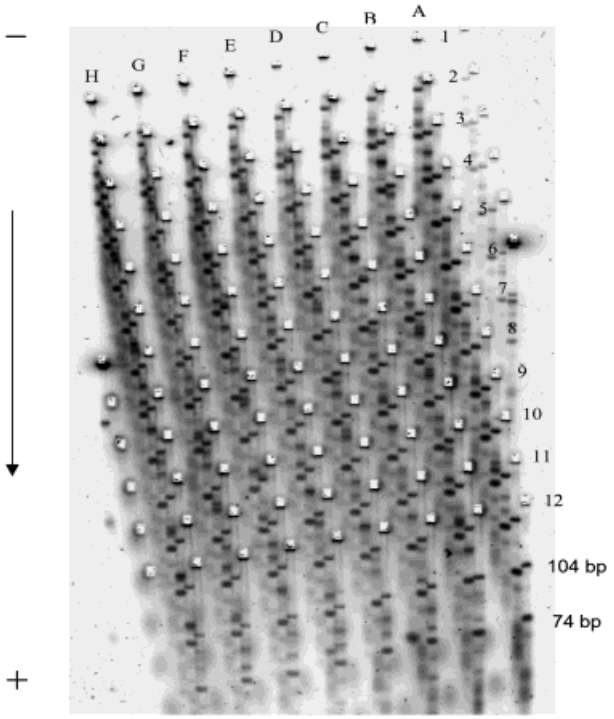

Figure 2. Sizing of dinucleotide repeat alleles on MADGE without inter-lane variability. D11S1760 alleles were bracketed by 104 and 74 bp internal size markers (shown in track A12).

94, χ^2 = 18.5, *df* = 28, *P* = 0.91) whereas typing errors generally give distortions from HWE [18]. In addition, the allele frequency distribution for this sample does not differ significantly (χ^2 = 6.18, *df* = 11, *P* = 0.86) from that previously described [13] in a Galician population sample.

In this work, we have shown that the accurate genotyping of a polymorphic dinucleotide repeat locus on MADGE gels is possible. Allele resolution was achieved in a size range from ~ 70–110 bp. Only 166 of the 5264 polymorphic dinucleotide repeat amplicons described by Dib *et al.* [1] in the human genome are within this size range, using the originally described primers. However, for many loci, improved primer design to give shorter amplicons would be feasible (see below). Furthermore, 128 021 human dinucleotide repeat loci have recently been identified in the human genome sequence using high-performance computing [2]. Many are likely to be polymorphic and may be suitable to tag alleles of the estimated 30 000–300 000 haplotype blocks in the human genome [19]. The allele size range of the PCR amplicons derived from dinucleotide repeats as described by Dib *et al.* [1] is highly dependent on the length of the non-repetitive sequence included. In order to examine this, we have analysed 100 random dinucleotide repeats on human chromosome 1 described by Dib *et al.* (ftp://ftp.genethon.fr/pub/Gmap/Nature-1995/). Alleles for these markers

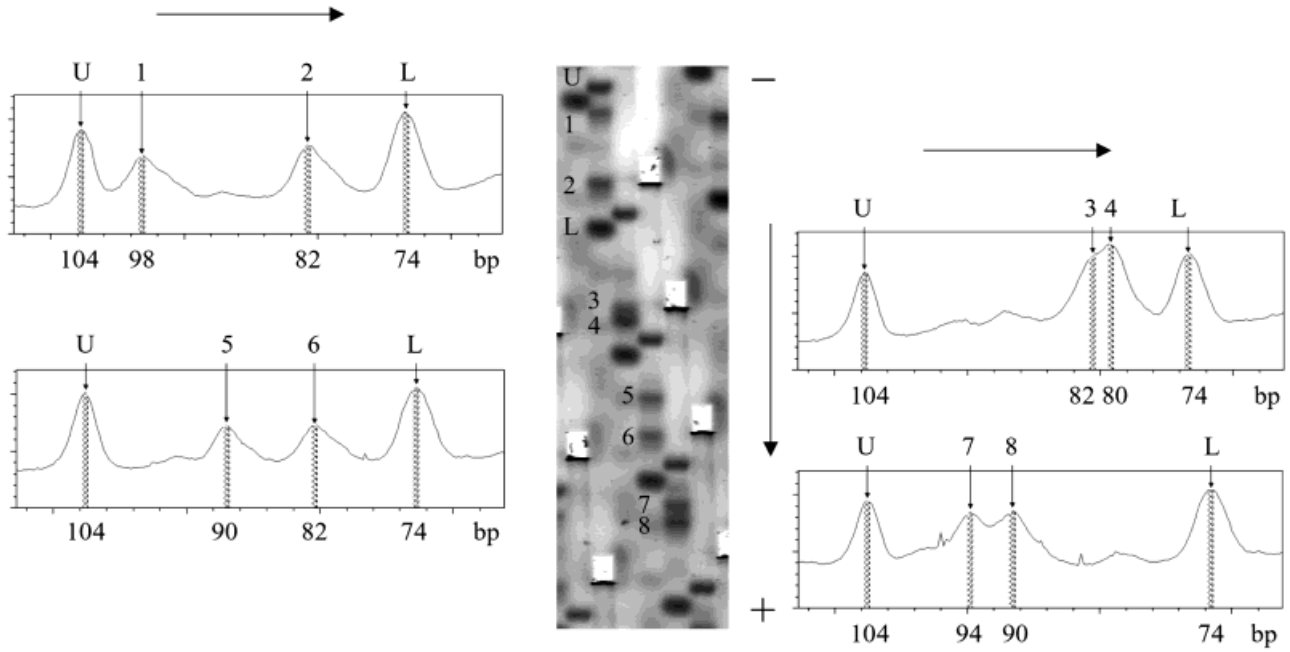

Figure 3. Detail of a MADGE gel used to resolve dinucleotide repeats, with four different tracks, and the Phoretix™ software electropherograms for each track. The numbers denote the different dinucleotide repeat bands detected. Below them is indicated the estimated size (in bp) for each allele. U, upper band in each track (104 bp size marker); L, lower band (74 bp size marker).

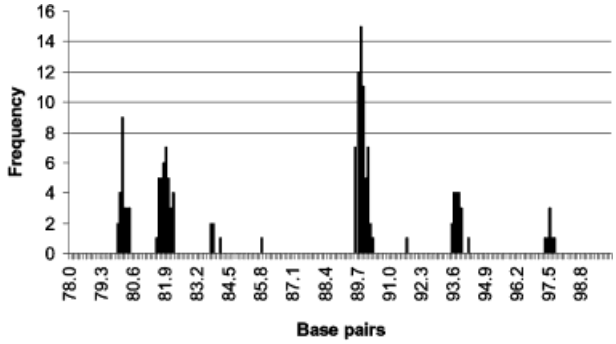

Figure 4. Frequency of estimated allele sizes (in bp) from a random group of nearly 100 individuals. Allele sizes were estimated using the formula given in the text. Eight different bins of alleles are clearly distinguishable, with no overlapping of estimated allele sizes.

range in size from 77 to 322 bp, but the number of AC repeats only range from 10 to 30. Therefore, it can be extrapolated that the length of the repetitive region of almost all the 5264 dinucleotide repeats described by Dib *et al.* [1] is less than 60 bp, and often much shorter. Use of 20 bp primers closer to the repeat would give amplification products that could be resolved on MADGE gels under the conditions presented here. It should be noted that in the resolution of alleles of D11S1760 we have used less that 50% of the available track length. Further adjustments of running conditions would permit use of somewhat longer amplicons, although there might be intrinsic advantages in standarizing amplicon size to ~ 70–110 bp. One worker could run ~ 20 000 tracks per week using this approach (overview in [10]) and convenience of universal PCR, size marker and running conditions, with single amplicon per track, will more likely be preferable than multiplexing within tracks.

MADGE represents an alternative technique for the accurate genotyping of dinucleotide repeats. On average, there may be 1–2 dinucleotide repeats per haplotype block in humans [2, 19] and dinucleotide repeats represent approximately 0.2–2% of all genomic DNA across a wide range of eukaryotic phyla [20]. Traditionally, microsatellite assays have been developed and designed for genomewide and focused linkage studies, but their high density may be suitable to tag haplotypes for association studies. Unlike highly selected, relatively small sample number, genomewide linkage studies, association studies will demand large sample sizes examined for single selected markers. However, the approach described also provides a high throughput and convenient format, with minimal hardware and reagent costs relative to the sophisticated instrumentation and reagent costs commonly inherent in microsatellite analyses. The approach may therefore also lend itself to use in a much wider diver-

sity of laboratories worldwide, in the analysis of dinucleotide repeat loci both in human [1, 2] or any other genome [20].

S.R. was an EC Marie Curie Research Fellow. I.N.M.D. was a Lister Institute Professor. The UK Medical Research Council and the British Heart Foundation are thanked for support.

Received December 9, 2003

References

[1] Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E. *et al.*, *Nature* 1996, *380*, 152–154.

[2] Collins, J. R., Stephens, R. M., Gold, B., Long, B., Dean, M., Burt, S. K., *Genomics* 2003, *82*, 10–19.

[3] Subramanian, S., Mishra, R. K., Singh, L., *Genome Biol.* 2003, *4*, R13.

[4] Cardon, L. R., Abecasis, G. R., *Trends Genet.* 2003, *19*, 135–140.

[5] Polymeropoulos, M. H., Rath, D. S., Xiao, H., Merril, C. R., *Nucleic Acids Res.* 1991, *19*, 5797.

[6] Mansfield, M. W., Stickland, M. H., Carter, A. M., Grant, P. J., *Thromb. Haemost.* 1994, *71*, 731–736.

[7] McGinnis, R. E., Spielman, R. S., *Diabetes* 1995, *44*, 1296–1302.

[8] Risch, N., Merikangas, K., *Science* 1996, *273*, 1516–1517.

[9] Day, I. N., Humphries, S. E., *Anal. Biochem.* 1994, *222*, 389–395.

[10] Gaunt, T. R., Hinks, L. J., Rassoulian, H., Day, I. N., *Nucleic Acids Res.* 2003, *31*, E48.

[11] Chen, X. H., O'Dell, S. D., Day, I. N., *BioTechniques* 2002, *32*, 1080–2, 1084, 1086.

[12] Rodríguez, S., Visedo, G., Zapata, C., *Electrophoresis* 2001, *22*, 2656–2664.

[13] Zapata, C., Rodríguez, S., Visedo, G., Sacristán, F., *Genetics* 2001, *158*, 1235–1251.

[14] Rodríguez, S., Zapata, C., *Mol. Biotechnol.* 2002, *21*, 117–122.

[15] Maniatis, T., Frisch, E. T., Sambrook, J., *Molecular Cloning: A Laboratory Manual*, CSH Laboratory Press, Cold Spring Harbor, New York 1982.

[16] Hu, C. T., O'Shaughnessy, K. M., *Electrophoresis* 2001, *22*, 1063–1068.

[17] Kozulic, B., *Anal. Biochem.* 1994, *216*, 253–261.

[18] Gomes, I., Collins, A., Lonjou, C., Thomas, N. S., Wilkinson, J., Watson, M., Morton, N., *Ann. Hum. Genet.* 1999, *63*, 535–538.

[19] Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., Lander, E. S., *Nat. Genet.* 2001, *29*, 229–232.

[20] Toth, G., Gaspari, Z., Jurka, J., *Genome Res.* 2000, *10*, 967–981.