

Rapid, high fidelity analysis of simple sequence repeats on an electronically active DNA microchip

Ray Radtkey, Lana Feng, M. Muralhidar, Melanie Duhon, David Canter, Deborah DiPierro¹, Sylvia Fallon, Eugene Tu, Kevin McElfresh¹, Michael Nerenberg and Ron Sosnowski*

Nanogen Inc., 10398 Pacific Center Court, San Diego, CA 92121, USA and ¹The Bode Technology Group, Springfield, VA 22150, USA

Received September 9, 1999; Revised January 8, 2000; Accepted January 20, 2000

ABSTRACT

We describe a method for the discrimination of short tandem repeat (STR) alleles based on active microarray hybridization. An essential factor in this method is electronic hybridization of the target DNA, at high stringency, in <5 min. High stringency is critical to avoid slippage of hybrids along repeat tracts at allele-specific test sites in the array. These conditions are attainable only with hybridization kinetics realized by electronic concentration of DNA. A sandwich hybrid is assembled, in which proper base stacking of juxtaposed terminal nucleotides results in a thermodynamically favored complex. The increased stability of this complex relative to non-stacked termini and/or base pair mismatches is used to determine the identification of STR alleles. This method is capable of simultaneous and precise identification of alleles containing different numbers of repeats, as well as mutations within these repeats. Given the throughput capabilities of microarrays our system has the potential to enhance the use of microsatellites in forensic criminology, diagnostics and genetic mapping.

INTRODUCTION

Microarrays present molecular biologists with novel opportunities for high efficiency analysis by allowing miniaturization of many cumbersome techniques and the potential for highly multiplexed testing. Microfabrication of DNA arrays and array substrates is usually accomplished using methods developed by the microchip industry. Currently, the main focus of so-called 'DNA chip' work is on nucleic acid analysis in the fields of genomics, gene discovery and diagnostics (1,2). While allowing massive amounts of data to be accumulated, these methods are limited by the restrictions of passive hybridization kinetics. Because of this, some characterizations of DNA have been impractical by hybridization methods. In fact, the implementation on microarrays of DNA simple sequence repeat analysis, a highly useful tool in human identification, cancer diagnosis and plant breeding, has not been accomplished prior to this report.

In the application of human identification, a standard set of polymorphic loci has been chosen. One criterion for locus selection was a relatively high level of heterozygosity, thereby permitting the generation of statistically unique DNA fingerprints with a minimal number of loci. Typically these microsatellite loci consist of short tandem repeats (STRs) of nucleotides represented in the human population by 4–15 alleles distinguished by differing numbers of repeat units and sometimes include point mutations (3,4). Further, selected STR loci possess conserved, non-repetitive flanking sequences that permit identification among many loci and provide sites to enable specific amplification (5; see <http://ibm4.carb.nist.gov.8800/dna/home.html>).

Amplification techniques allow minute amounts of biological material to be used for genotyping. This has provided a unique and growing opportunity for criminal forensics by enabling the unequivocal identification of individuals from sparse amounts of evidence. The sieving technology now used to assay these simple sequence repeats limits the flexibility of this application to skilled practitioners in well-equipped laboratories.

Slippage hybridization among the repeat units makes accurate allelic discrimination of STR loci unreasonably fastidious by non-active microarray methods (6). This is due to the difficulty of obtaining high fidelity hybridization with both the unique and repeat regions in the proper register (7). Traditional solutions to this problem include protocols that use thermal and/or chemical conditions to promote high stringency hybridization. However, these procedures must be applied simultaneously and uniformly over an entire array, complicating discrimination among hybrids with different sequences. Also, these methods are cumbersome and would require long incubations. We describe here a novel STR oligonucleotide system designed to exploit differences in base stacking energy. Further, we demonstrate the use of high stringency electronic hybridization (8,9) to rapidly and reliably obtain precise STR hybrids. This integrated system allows discrimination of STR alleles at the single repeat unit and microvariant level.

MATERIALS AND METHODS

Oligonucleotides

STR loci sequences and other related information was obtained through the STRbase web site (<http://ibm4.carb.nist.gov.8800/dna/home.html>). Three loci used routinely in human forensics (5) were utilized in the present study: CSFIPO (human *c-fms*

*To whom correspondence should be addressed. Tel: +1 858 410 4600; Fax: +1 858 410 4848; Email: rsosnowski@nanogen.com

Table 1. Oligonucleotides used in the current study

ODN	Sequence
Captures	
TPOX	5'-biotin-GGCACTTAGGGAACCCTCACTG(AATG) ₅₋₁₂ -3'
TH01	5'-biotin-TCCTCCCTTATTTCCCT(CATT) ₄₋₁₀ -3'
CSF1PO half-repeat	5'-biotin-GCCTTCATAGATAGAAGATAGATAGATT(AGAT) _n AG-3', where $n = 6-14$ repeats
CSF1PO full-repeat	5'-biotin-GCCTTCATAGATAGAAGATAGATAGATT(AGAT) _{6,7,13,14} -3'
Reporters	
TPOX	5'-AATGTTTGGGC-BTR-3'
TH01	5'-CATTCACCAT-BTR-3'
TH01prop	5'-propynyl-PO4dCATTACCAT-BTR-3'
TH01-9.3	5'-propynyl-PO4dATCATTCAT-BTR-3'
CSF1PO half-repeat	5'-ATAGGAAGTA-BTR-3'
CSF1PO full-repeat	5'-AGATAGGAAG-BTR-3'
Targets	
TH01	5'-ACACAGACTCCATGGTG(AATG) _{7,11} AGGGAAATAAGGGAGGA-3'
TH01BTR target	5'-ACACAGACTCCATGGTG(AATG) ₁₁ AGGGAAATAAGGGAGGA-BTR-3'
CSF1PO	5'-TGTTCTAAGTACTTCT(ATCT) _{8,15} AATCTATCTATCTTCTATCTATGAAGGC-3'

proto-oncogene for the CSF-1 receptor gene, GenBank accession no. HUMCSF1PO), TH01 (human tyrosine hydroxylase gene, HUMTH01), TPOX (human thyroid peroxidase gene, HUMTPOX). Oligonucleotides (see Table 1) were synthesized by Oligos Etc. (Wilsonville, OR) or Midland Certified Reagent Co. (Midland, TX). All reporter ODNs employed the fluorescent probe Bodipy Texas Red (BTR) (Molecular Probes, Eugene, OR) conjugated to AminoLink (C6 amino linker; Glen Research, Sterling, VA) as described previously (8). Propynyl-modified AminoLink ODNs were synthesized and purified by TriLink (San Diego, CA). The microfabrication of the chip, details of the instrumentation and structure and application of the permeation layer have been described previously (8,10).

Template preparation

Target amplification was carried out as described in the users guide for the monoplex (single STR locus) systems purchased from Promega Inc. (Madison, WI). Genomic DNA used as the target for amplification was obtained from The Bode Technology Group (McClellan, VA). To desalt the STR amplicons, 25 μ l of amplicon solution was diluted to 70 μ l in H₂O and then run through Biospin columns from Bio-Rad, according to the manufacturer's instructions. The sample was then diluted 1:1 in 100 mM histidine to make the final histidine concentration 50 mM. This solution was heated at 95°C for 6 min to denature the double-stranded amplicon and kept at room temperature.

Hybridization

A 1 μ M solution of each capture ODN in 50 mM histidine was electronically loaded to selected sites at room temperature using AC conditions of 0.8 μ A/pad for 25 s with cycles of 38 μ s positive bias and 25 μ s negative bias. Sites were also biased with just histidine to later serve as background controls. Target solutions were electronically hybridized to the whole

array under AC conditions of 1.6 μ A/pad, 19 μ s positive, 12 μ s negative for 3 min. This electronic step was carried out at 32–34°C to prevent slippage hybridization in the repeat region during capture/target hybridization. The array was washed several times with 50/500 buffer (50 mM NaPO₄, pH 7.0, 500 mM NaCl). A solution of 1 μ M reporter (or a mixture of reporters for multiplexing) in 50/500 buffer was added to the array and incubated for 3 min at room temperature. The array was then washed several times with 50 mM NaPO₄, pH 7.0, before an initial fluorescent image was taken. Subsequently the capture/target/reporter sandwich complex was thermally denatured to achieve discrimination between matched and mismatched alleles. During the thermal denaturation process, 50 mM NaPO₄ was used to wash the array three times at each elevated temperature from 30 to 40°C. A fluorescent image of the array was taken after each wash to record the reporter signal left on each site.

RESULTS

Molecular strategy for the STR hybridization assay

Figure 1 depicts the strategy used for discriminating among the different alleles of STR loci by a hybridization method. The key features are the coordinated design of the capture and reporter probes (Fig. 1A) along with addressing of allele-specific capture oligonucleotides (ODNs) to independent sites in an array (Fig. 1B). The system is based on the premise that base stacking stabilizes DNA hybridization (11–15). The design of the capture and reporter ODNs takes advantage of target DNA-directed juxtaposition of the 3'-terminal base of the capture and the 5'-terminal base of the reporter resulting in base stacking between the capture and reporter. This juxtaposition results in increased stability of reporter binding to the complex. If a reporter

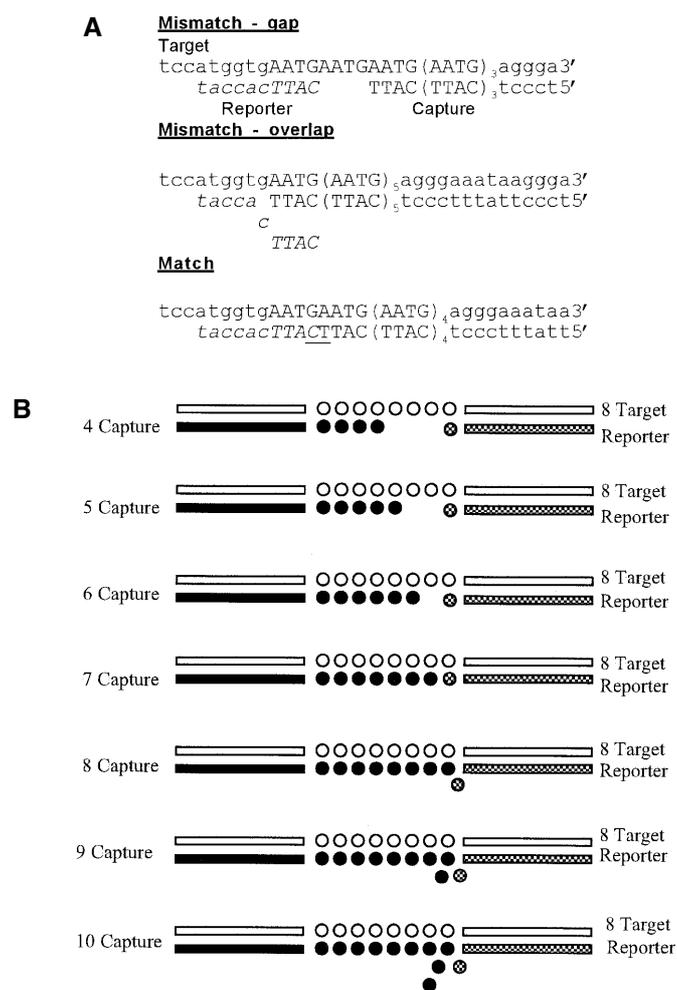


Figure 1. Oligodeoxynucleotide system used to determine STR genotype by hybridization. (A) Three classes of hybridization complexes. Mismatch – gap, in this complex the number of repeats in the target is greater than the repeat number in the capture plus reporter ODNs. Mismatch – overlap, the number of repeats in the target is less than the number of repeats in the reporter plus capture. Match, this is the hybrid complex where the number of repeat units in the capture plus the number of repeats in the reporter equals the number of repeats in the target. Sequences represent the TH01 locus. (B) Diagram depicting an array of possible hybridization complexes for the TH01 locus, homozygous for the 8 repeat allele.

sequence is chosen with less than 12 bases of complementarity to the target, direct hybridization will be very unstable under the conditions described. Only in the presence of base stacking (when the capture and reporter ODNs are in direct apposition) will the hybrid be stabilized. Figure 1A represents the hybridization of a TH01 target sequence containing 6 repeat units at test sites containing complementary capture ODNs with 4, 5 or 6 repeat units. Under the appropriate conditions (discussed below), hybridization of the target to the capture will be nearly equivalent at each site. However, the spatial relationship between the termini of the capture and reporter differs in each case. At the test site with the 4 repeat capture, the resulting capture/target/reporter hybridization complex will contain a gap between the 3'-terminus of the capture and the 5'-terminus of the reporter

(Fig. 1A, Mismatch – gap). Hybridization at the 6 repeat site will result in a complex with a 4 base overlap between capture and reporter ODNs (Fig. 1A, Mismatch – overlap). At the test site with the 5 repeat capture, the hybridization complex will be formed with contiguous capture and reporter termini and base stacking between them (Fig. 1A, Match). Upon application of stringent hybridization conditions, the differential stability of the reporter probe can be ascertained. Determination of the test site with the greatest reporter stability along with knowledge of the number of repeat units present in the capture and reporter at that site allows identification of the STR target allele.

Experiments to test the base stacking model

The role of base stacking in the stability of DNA hybrids has been proposed from the results of several studies (11–15). Two factors shown to affect base stacking energies in nucleic acid sequences are the types of bases involved in the interaction and modification of a base by alkylation (11). We tested the effects of these two variables on stability of the hybridization complexes in our system.

If base stacking is responsible for discrimination among the gap, overlap and match complexes described above, then the relative stability or discrimination ratios should correlate with the reported base stacking energies of different contiguous terminal nucleotides. We used the CSF1PO locus to test this hypothesis. This locus was chosen as we were able to design capture and reporter ODNs to have either T:A or G:A juxtaposed terminal nucleotides. Since base stacking between intrastrand G:A nucleotides is reported to provide greater helix stability than T:A vertically stacked bases (11), a hybridization complex with G:A stacking between capture and reporter contiguous termini should display higher discrimination ratios than the comparable T:A system. The results given in Figure 2A show that this is indeed the case for the CSF1PO locus. The discrimination ratio (DR) was defined as the ratio of matched to mismatched signal following background subtraction from each value. In other experiments, the lengths of reporter ODNs were varied to evaluate how changes in the relative contribution of base stacking to the overall stability of the reporter affected match/mismatch discrimination. Again, these results were consistent with the proposed base stacking mechanism (not shown).

Alkyl modification of bases has also been shown to increase base stacking interactions, presumably by adding more electrons to the base ring structure. Addition of electrons to the ring could contribute to dipoles, π -bonding and dipole-induced dipole mechanisms that have been proposed to be important to vertical base stacking in nucleic acids (11). To further test our hypothesis that base stacking is the major factor contributing to our ability to discriminate among STR alleles, we used reporter ODNs synthesized with propynyl-modified termini at the end contiguous with the capture ODN. The TH01 locus was used for this analysis (Table 1). Discrimination ratios between match and mismatch-gap hybrid complexes were compared for complexes formed with a reporter incorporating a 5'-propynyl-modified cytidine relative to those formed with reporters with an unmodified cytidine 5'-terminus. The results shown in Figure 2B present the relative discrimination ratios obtained using an unmodified reporter versus a propynyl-modified reporter. The discrimination ratios observed with the propynyl-modified reporter are significantly greater than those seen with

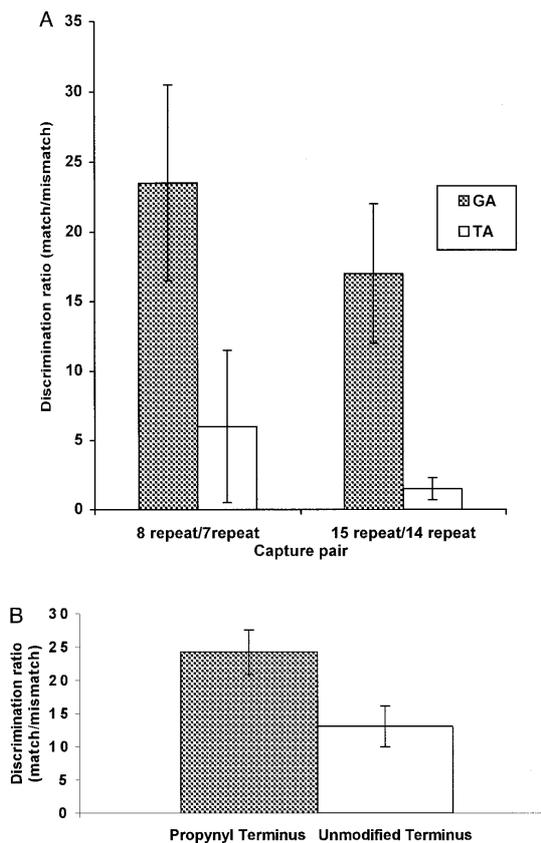


Figure 2. Base stacking and hybridization complex stability. **(A)** Comparisons of discrimination ratios observed using two types of different base stacking pairs in a model CSF system. Capture ODNs with 6, 6.5, 7, 7.5, 13, 13.5, 14 or 14.5 repeat units were electronically addressed to pads on the chip, followed by electronic hybridization of 8 and 15 repeat targets. Two 10 nt reporter molecules were then passively hybridized to the array. One reporter possessed a full repeat at the 5'-end while the second possessed a half repeat (see Table 1). These resulted in either a G:A or T:A stacking interaction at the respective match pads. The results show that the discrimination ratio is greatest for the G:A reporter:capture complex, as predicted by theoretical base stacking calculations. **(B)** Comparison of the discrimination achieved at the TH01 locus between reporter with a natural terminal nucleotide and the same reporter sequence with a propynyl-modified terminus. Capture ODNs with 5, 6, 9 or 10 repeat units were electronically addressed to pads on the chip, followed by hybridization of 7 and 11 repeat targets. The addressed pads were then passively hybridized with reporters containing propynyl-modified termini and subsequently denatured to achieve discrimination. All reporter was then stripped off at 42°C and the standard TH01 reporter was hybridized to the chip. The standard reporter was then denatured to achieve discrimination. The discrimination ratios were calculated by dividing the lowest match signal on each chip by the highest mismatch signal on the chip (average of four chips).

the unmodified reporter. This result fortifies the model of base stacking as a major contributor to hybridization complex stability in our system.

Use of electronic hybridization to achieve precise hybrid complexes

A restrictive difficulty in designing hybridization-based assays to determine the allelic identity of STR loci is obtaining completely synchronous hybridization of repetitive and

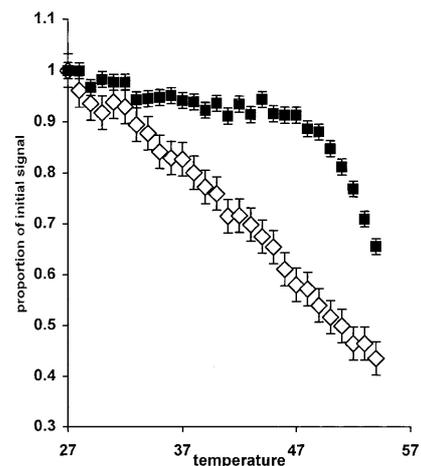


Figure 3. Effect of target hybridization temperature on the dynamics of subsequent denaturation. Fluorescently labeled 11 repeat targets (see Table 1) were hybridized at either 17 (diamonds) or 41°C (squares) followed by passive denaturation using a thermal ramp (see Materials and Methods). Each point represents the average of 20 pads (10 pads from two separate chips). The figure shows that when targets are hybridized at a higher temperature the majority of target molecules remain hybridized until the theoretical melting temperature is approached; when targets are hybridized at a lower temperature the presence of 4 bp repeats leads to misaligned duplexes and heterogeneous melting temperatures.

non-repetitive nucleotides. This is because stable hybridization can occur within the repeat region without alignment of the unique flanking sequences. To obtain precise association of complementary strands, high stringency conditions must be applied during hybridization. Unfortunately, this requires very long hybridization times (3–16 h) to obtain sufficient signal (not shown). We have used the ability of our electronic system to concentrate DNA under high stringency conditions to provide precise STR hybridization in as little as 2 min.

To directly test the fidelity of target hybridization to the capture ODN the following experiment was performed. TH01 capture probes containing unique flanking sequence and four AATG repeat units was addressed to individual test sites (see Table 1). Target ODN (11 repeats plus unique flanking sequence) labeled with BTR was then electronically hybridized under either high (41°C) or low stringency (17°C) conditions, in low conductivity buffer (50 mM histidine). The buffer was changed (50 mM NaPO₄) and fluorescent intensity was monitored throughout an increasing temperature ramp. Figure 3 shows the results of the experiment described above. At test sites where hybridization was done at low stringency, the melting profile describes a shallow slope, presumably reflecting the mixture of T_m values within the heterogeneous hybrid population. Hybridization at the high stringency sites is much more homogeneous, as indicated by a sharp change in fluorescent intensity observed when the melting temperature is reached. These results indicate that hybridization between complementary strands of an STR locus can occur rapidly and accurately when stringency conditions, especially DNA concentration in an electric field, are appropriately controlled.

Alternative reporters for detecting microvariants and attaining redundant analysis of target DNA

Many STR loci contain among their alleles heterogeneity other than varying numbers of repeat units. These generally include variations of one base and may consist of insertions, deletions, transitions or transversions. These variants, especially transitions and transversions, may be difficult to detect by gel electrophoresis methods but should be amenable to hybridization analysis. To demonstrate this we used the TH01 locus which may have a single base deletion in one of the repeat units, most typically in an allele with 9 other intact repeat units, designated TH01-9.3. Four biotinylated captures were addressed to separate electrodes, three of which are described above (TH01 7, 8 and 9 repeats). The fourth has identical flanking sequence but only 3 repeat units. TH01-9.3 homozygote target DNA (K562) was heat denatured in histidine buffer and electronically hybridized as described above. A BTR-labeled reporter specifically designed for the TH01-9.3 microvariant (Table 1) was then passively hybridized, washed and visualized as described above. When hybridized with a TH01-9.3 allele, this reporter allows perfect complementation and vertical base stacking, however, with all other alleles the complementary region is reduced from 10 to 7 bases as well as disrupting base stacking. Figure 4A shows the results of this system when analyzing DNA that is homozygous for TH01-9.3.

STR allele identification can take on a high level of significance when applied to legal issues and therefore it is important to have assays that achieve a high level of accuracy. This can be done with reiterative analysis of the sample either by increasing replicates or by testing with an alternative method that gives confirmatory results. The design of the system described here provides the opportunity for redundant interrogation of the analyte material in a manner that does more than simply increase replicate size.

By altering the reporter, it is possible to interrogate a slightly different sequence position of the target molecule and shift the location of the test site displaying maximum hybridization complex stability. This enables a check for possible sequence-specific anomalies as well as variability among test sites. This is accomplished rather simply by altering the number of repeat units in the reporter. This sequence modification moves the target position being probed by 4 bases and brings in a reporter with a different sequence. The site at which the most stable complex has been formed changes as well. Figure 4B shows that use of an alternative reporter reaffirms the identity of the genetic material tested initially.

Validation with samples typed by an accredited laboratory

Fractions of genomic DNA samples that were STR genotyped by conventional methods in an accredited forensics laboratory (The Bode Technology Group) were received. We amplified these samples with the appropriate primers and subjected them to the analysis described above. Multiplex hybridization analysis in our system of three STR loci for 12 individuals generated results that were 100% concordant with conventional genotyping. Results for four STR loci, hybridized individually, in 25 different samples were also 100% concordant with PAGE analysis (not shown).

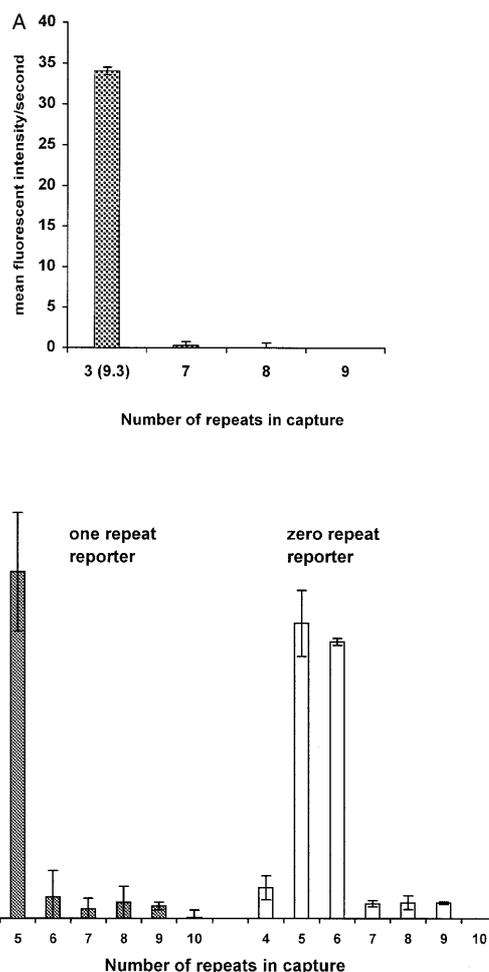


Figure 4. (A) Discrimination of the microvariant locus TH01-9.3. Signal observed (20 pads each from two chips) on four pads after hybridization of TH01-9.3 target and reporter. Because of the nature of the 9.3 microvariant ([AATG]₆ATG[AATG]₃) the positive signal is observed at the pad addressed with the capture ODN containing three repeat units (see Table 1). The 7, 8 and 9 repeat captures correspond to the sites for 8, 9 and 10 repeat targets, respectively. (B) Demonstration of the use of a redundant reporter for the TH01 locus. The set of histograms on the left are the result of the standard TH01 assay using a 1 repeat reporter; the set on the right are results with a 0 repeat reporter. The target used was a PCR product of a heterozygote individual that is heterozygous for TH01 alleles containing 5 and 6 repeats. In the case of the 1 repeat reporter, the match signal appears on captures with $n - 1$ repeats (1 repeat provided by the reporter); in the case of the 0 repeat reporter all repeats in the hybridization are provided by the capture (Table 1).

DISCUSSION

To be useful, a hybridization assay for STR identification must be rapid and accurate. Because of the large number of repeat units present in STR alleles, the only way to obtain precise alignment of complementary strands is by applying high stringency conditions during hybridization. However, this slows the rate of hybridization, resulting in a tedious and inefficient process. We have used the novel method of electronic manipulation of DNA to overcome the significant kinetic problems associated with hybridization-based identification of STR alleles. By using an electric field to concentrate the analyte DNA, the

encounter rate between complementary strands was increased, thereby increasing the hybridization rate. However, since this was done at low salt and high temperature, we were able to select for hybrids with the greatest stability, resulting in a more homogeneous population of hybrids. When subsequently interrogated with a reporter probe, this yields sharp demarcation among different classes of hybridization complexes, allowing accurate determination of allele type. A rapid analysis of STR alleles using passive hybridization techniques (required in non-active microarrays) is currently not possible.

This manuscript describes a method that provides a shift in the fundamental technology used for hybridization of DNA at high stringency. Further, this technology allows many advantages over systems that discriminate among simple sequence repeat alleles by size. A chip-based hybridization assay is highly amenable to miniaturization and easily compatible within a device that could also include DNA purification and amplification units, thereby providing a high degree of portability (16,17). This method also provides significant advantages in speed and convenience by eliminating the need for preparing and running sieving gels. Other features that are unique to this method include highly accurate analysis of microvariants and the ability to perform redundant analysis without requiring additional preparation of the target material.

ACKNOWLEDGEMENT

This work was supported in part by a grant from the National Institute of Justice.

REFERENCES

1. Ramsay, G. (1998) *Nature Biotechnol.*, **16**, 40–44.
2. Phimister, B. (ed.) (1999) *The Chipping Forecast Nature Genetics Supplement*, Vol. 21. Nature America Inc., New York, NY.
3. Weedn, V.W. and Hicks, J.W. (1997) *Natl Inst. Justice J.*, **234**, 16–23.
4. Tautz, D. (1989) *Nucleic Acids Res.*, **17**, 6463–6471.
5. National Research Council (1996) *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington, DC.
6. Hacia, J.G. (1999) *Nature Biotechnol.*, **21** (suppl.), 42–47.
7. Pearson, C.E., Wang, Y.H., Griffith, J.D. and Sinden, R.R. (1988) *Nucleic Acids Res.*, **26**, 816–823.
8. Sosnowski, R.G., Tu, E., Butler, W.F., O'Connell, J.P. and Heller, M.J. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 1119–1123.
9. Edman, C.F., Raymond, D.E., Wu, D.J., Tu, E., Sosnowski, R.G., Butler, W.F., Nerenberg, M. and Heller, M.J. (1997) *Nucleic Acids Res.*, **25**, 4907–4914.
10. Gilles, P.N., Wu, D.J., Foster, C.B., Dillon, P.J. and Chanock, S.J. (1999) *Nature Biotechnol.*, **17**, 365–370.
11. Saenger, W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York, NY, pp. 116–158.
12. Khrapko, K.R., Lysov, Y.P., Khorlyn, A.A., Shick, V.V., Florentiev, V.L. and Mirzabekov, A.D. (1989) *FEBS Lett.*, **256**, 118–122.
13. Pieters, J.M.L., Mans, R.M.W., van den Elst, H., van der Marel, G.A., van Boom, J.H. and Altona, C. (1989) *Nucleic Acids Res.*, **17**, 4551–4565.
14. Yershov, G., Barsky, V., Belgovskiy, A., Kirillov, E., Kreindlin, E., Ivanov, I., Parinov, S., Guschin, D., Drobishev, A., Dubiley, S. and Mirzabekov, A. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 4913–4918.
15. Lane, M.J., Paner, T., Kashin, I., Faldasz, B.D., Li, B., Gallo, F.J. and Benight, A.S. (1997) *Nucleic Acids Res.*, **25**, 611–616.
16. Cheng, J., Sheldon, E., Wu, L., Uribe, A., Gerrue, L., Carrino, J., Heller, M.J. and O'Connell, J.P. (1998) *Nature Biotechnol.*, **16**, 541–546.
17. Swanson, P., Gelbart, R., Atlas, E., Yang, Y., Hay, T.G., Butler, W.F., Ackley, D.E. and Sheldon, E. (2000) *Sensors and Actuators B*, in press.