MASSIVELY PARALLEL DNA TYPING BY CAPILLARY ARRAY ELECTROPHORESIS

NIJ 93-IJ-CX-0010

Project Period: 1/1994 - 12/1995
No-cost Extensions: 1/1996 - 9/1997

Dr. George F. Sensabaugh
Forensic Science Group
School of Public Health
University of California
Berkeley CA 94720-7360

TABLE OF CONTENTS

Note: Most of the literature cited in this report is pre-1995. My entire bibliographic database and my collection of papers relating to capillary electrophoresis, short tandem repeat polymorphisms, and mitochondrial DNA were lost in July 1997; the bibliographic database has not been regenerated as of this writing. The literature cited herein was recovered from a pre-existing database and updates have been added where essential.

EXECUTIVE SUMMARY

The implementation of DNA profile identification databanks requires high throughput DNA typing. The entry of reference profiles, e.g., profiles from convicted offenders, requires facilities with the capacity to perform hundreds of typings per day, day-in and day-out. Regarding the development of profiles from evidence samples, it can be argued that the most cost efficient strategy is to perform batch typing for the initial interrogation of the offender databank; here also, high throughput genetic typing methods would be beneficial.

The overall objective of this project was investigate the use of capillary array electrophoresis (CAE) as an approach to high throughput genetic typing. CAE combines the rapid analysis times of capillary electrophoresis (CE) with the throughput multiplier of running many capillaries in parallel. In a massively parallel CAE mode, 100 or more CE analyses can be conducted simultaneously. The project was undertaken as a collaboration with the Berkeley High Sensitivity DNA Analysis Group headed by R.A. Mathies and A.N. Glazer.

The major project achievements are briefly summarized below.

a. The utility of CAE was demonstrated for the high throughput typing of short tandem repeat (STR) polymorphisms, both singly and in multiplex sets; fifteen STR loci were investigated in the course of this study. Single base resolution has been achieved through the size range of DNA fragments employed in forensic STR analysis (75-350 bp), allowing the detection of off-ladder variants. CAE thus has the resolution and throughput capacity to meet the requirements for forensic databank use.

b. The utility of CAE for high throughput DNA sequence analysis was also demonstrated. This has been demonstrated using mitochondrial DNA (mtDNA) as the target analyte; mtDNA sequence analysis is commonly used in the identification of missing persons and body remains.

c. The use of energy transfer (ET) fluorescent labels, developed by the High Sensitivity DNA Analysis Group, improves both the sensitivity and the quality of analytical output, both for STR typing and for sequence analysis. The quality of the signal with ET labeling is superior to that obtained with single fluorescent labels, the approach currently used in forensic analysis. Moreover, the improved signal quality extends the potential for multicolor detection, making possible more extensive multiplex typing systems.

d. A second generation CAE unit with 4 color detection potential was constructed with shared funding between this project and the High Sensitivity DNA Analysis Group. This unit was designed both for STR typing and for sequence analysis.

e. In-lane allele ladders and in-lane single base step ladders were demonstrated to be reliable

standards for STR typing. Use of the former makes possible typing based on simple comparison of peak position. Use of the latter allows very precise and reproducible estimation of DNA fragment size. Both simplify the detection of off-ladder variants.

f. Cross-compatibility in STR typing between CAE and conventional slab gel methods was demonstrated for the fifteen STR loci, singly and in various multiplex combinations. This work also demonstrated that use of ET labels extends the multiplex capabilities of the slab gel methods.

g. The detection of sample mixtures was investigated using both CAE and conventional slab gel based typing. The detection limit for a mixture was reached when the samples in the mixture were at about 10:1 proportions; mixtures tended to be easier to detect on slab gels subjected to fluorescent scanning than with the CAE system. We established that peak ratio analysis could be used to distinguish the contributors to a mixture in situations involving no shared alleles and in situations involving one shared allele.

h. STR typing was performed on a Japanese population sample of 187 individuals. Analysis of the data indicated a small degree of population structure, consistent with previous studies on this population.

I. New primers were developed for the CSF1PO locus STR to improve typing at this locus in badly degraded samples. The efficacy of these primers was demonstrated.

j. The potential of a new genetic typing approach, Amplified Restriction Fragment Polymorphism (AFLP), was assessed for human genome profiling and found to be relatively undiscriminating.

k. The potential of "CAE on a chip" was demonstrated in preliminary experiments showing that STR typing at the THO1 locus could be achieved with near single base resolution in under 5 minutes. Although the grant terminated before these preliminary experiments on STR typing could be completed, our subsequent studies on a disease marker sequence polymorphism have demonstrated that use of CAE on a chip allows the typing of 96 samples in under 10 minutes. The potential for this high throughput microanalytical technology for DNA typing is thus clearly demonstrated

Each of these areas is described in more detail in the section, Summary of Research Findings, that follows in the body of the report.

Listings of project publications, presentations, and personnel trained are provided in Appendices I - III. Appendix IV consists of copies of the research publications and published abstracts originating from this project.

RESEARCH REPORT

I. RESEARCH RATIONALE

Since the introduction of DNA technology into forensic science in 1985, DNA profiling has become well established as a useful tool in the investigation and prosecution of crimes. Its usefulness has been demonstrated in linking suspects to victims and/or crime scenes, in excluding innocent persons who might otherwise have been suspects, in linking together cases committed by repeat offenders, and in identifying human remains. DNA evidence has also been instrumental in correcting miscarriages of justice; it has been used to exonerate individuals falsely convicted of crimes [1]. More generally, DNA profiling has proven invaluable for establishing the identity of skeletal and other body remains in mass disasters, military repatriations, and investigations of human rights abuses.

The potential value of DNA profile databases for the identification of repeat offenders was recognized early on. A DNA profile database would be used in the same way fingerprint identification databases are currently used: to identify individuals leaving traces at scenes of crimes and to link unsolved crimes together. Many states have legislatively mandated the establishment of databases containing DNA profiles for certain categories of offenders, notably sex offenders and violent criminals. At the national level, the FBI has taken a leading role by defining standards for a combined DNA index system (CODIS) and has established computer networks to connect local jurisdictions to the central databank and to each other. Several states now have databanks in place containing more than a thousand profiles and a few jurisdictions have been able to solve suspectless crimes by searching these databases. The value of offender DNA profile databases has also been demonstrated in the United Kingdom; the U.K. database includes all categories of offenders and has proven very useful in solving property crimes as well as crimes of violence.

The potential of DNA profile databases is constrained at present by two factors. First, the rate of entry of DNA profiles into a database is limited by the throughput potential of the technology used for DNA typing. The DNA typing technology used for CODIS profiling until recently, restriction fragment length polymorphism (RFLP) analysis by Southern blotting, is relatively slow and does not have the capacity to generate DNA profiles on very large numbers of individuals without substantial investments in time and personnel. The magnitude of the typing throughput problem is illustrated by a 1993 survey of 19 states with legislatively mandated databanks; this survey showed that only 12.3% of the more than 141,000 samples collected up to that time had been processed for entry into databases [2]. Although the situation is better now, it is obvious that RFLP analysis as currently done is not a cost effective approach to large scale DNA profiling. The United Kingdom offender profiling project has addressed its throughput needs - more than 150,000 samples processed per year - by dropping RFLP analysis and moving to analysis of polymorphic short tandem repeat (STR) markers using semi-automated typing technologies. The U.S. is following this lead; many forensic laboratories are using STR typing

or are making the transition to STR typing and CODIS expanded in late 1997 to include STR markers in its database index.

The second constraining factor on DNA profile database usefulness is the rate at which DNA profiles can be generated from cases without suspects. The identification of perpetrators in suspectless cases is, after all, the principal rationale for having an offender DNA profile database. The database will be of little use for this purpose if suspectless cases are not worked. However, at present, most crime laboratories are limited by personnel numbers to the analysis of evidence essential to the prosecution of cases, i.e., cases with suspects. It is clear that innovative approaches will be needed to address this problem. One such approach is the clinical process laboratory format. Under this approach, samples from suspectless cases would be sent to a central process laboratory which would utilize automated sample preparation and genetic typing technologies to achieve high throughput. Profiles generated from these samples would be used to interrogate the offender and unsolved crime databases; hits on the database would be referred back to the submitting laboratory for full forensic case work up. A process screening approach such as this would benefit from high throughput genetic analysis systems.

The overall objective of this grant project was to investigate a possible solution to the sample throughput constraint. The principal focus of our investigation was the use of massively parallel capillary array electrophoresis for genetic typing and DNA sequencing. Capillary electrophoresis (CE) has been demonstrated previously to provide rapid, high resolution separations of DNA fragments differing in size; it has been used for the analysis of DNA restriction fragments and polymerase chain reaction (PCR) products [3-7] and for DNA sequencing [8-16]. Typical CE runs are 15-120 minutes in duration, depending on the size range of the DNA fragments to be separated and resolution desired. Among other attractive properties, CE requires very small sample loads and has a loading format that is easily automated [**8-10]. However, simple CE is limited in that a single capillary is run at a time; despite the rapidity of a single CE run, there is no significant gain in overall throughput relative to other methods, e.g., slab gel electrophoresis methods. To gain increased throughput with CE systems, it is necessary either to increase analysis speed by an order of magnitude or to use CE in a massively parallel mode, i.e., run many capillaries at one time.

Capillary array electrophoresis (CAE) follows the massively parallel strategy [14,15,17]. CAE was developed on this campus in the laboratory of R. A. Mathies as part of the Berkeley High Sensitivity DNA Analysis Project; much of the work to be described in this report was a collaboration between our two laboratories. The design of CAE allows many capillaries to be run in parallel array; coupled with the high speed of CE separations, the parallel array allows very high throughput. The capillaries can be individually arranged at the front end to facilitate automated loading. Thus, for example, a 100 capillary instrument operating at one electrophoresis cycle per hour should be capable of completing 800 typing runs per 8 hr day. This single instrument would satisfy the throughput needs of the U.K. database effort, matching in throughput the 25-30 automated sequencers currently in use. This example illustrates the attraction of massively parallel CAE for offender DNA profiling.

The primary objective of the work described in the following sections was the application of CAE to high throughput typing of short tandem repeat (STR) polymorphisms. Four base STR polymorphisms emerged in the early 1990's as the probable markers of choice for the second generation forensic DNA typing. They have been investigated in detail by several groups, notably, Caskey's group at Baylor [18-21] and the Forensic Science Service in Great Britain [22-26], among others. Many hundreds of these markers have now been characterized, primarily for use in high resolution genetic mapping [27,28]. STRs show moderate to high levels of heterozygosity, typically 30-90%, which makes them useful both for genetic mapping and for individualization purposes. The major benefits of STR typing in forensic analysis include:

- STR loci generally have no more than 6-10 alleles. This allows adequate population frequency data to be obtained from relatively small population samples, i.e., 100 or so individuals. Moreover, the alleles are discrete. Hence STR polymorphisms are amenable to traditional genetic analysis, i.e., testing for Hardy-Weinberg equilibrium by enumeration and so on [18-20].
- Amplification of STR alleles yields small products, typically in the 150-250 bp range. This minimizes risk of allelic drop-out due to DNA degradation and/or preferential amplification. This factor contributed significantly to the high success rate of STR typing of degraded remains from mass disasters [e.g., ref.29].
- STR markers can be amplified and analyzed in multiplex; this increases throughput significantly [18-26, 30].
- STRs can be typed by any method that allows single base resolution or near single base resolution, e.g., manual and automated sequencing systems. This allows allelic sizes to be measured on an absolute scale; rare allelic variants and artifacts due to amplification or sample can be readily detected. Moreover, forensic STR analysis can piggyback on the technological advances of the human genome project; any improvement in DNA sequencing technology can be applied to forensic STR analysis.
- An individual is uniquely specified for practical purposes by a genetic type profile consisting of 15-18 high heterozygosity STR loci [31]. Given the number of known STR loci and the relative ease of STR typing, generation of a 15-18 locus profile is a readily achievable goal.

Most forensic DNA labs in the U.S. are either using STRs or are making the transition from RFLP typing to STR typing.

A second objective of our work was the application of CAE to high throughput DNA sequence analysis with a focus on mitochondrial DNA (mtDNA) sequence analysis. It is well established that the D-loop sequence region of mitochondrial DNA is highly variable [32-38]. This sequence region is about 1000 bp in length and contains two hypervariable sequence regions, each about 400 bp in length. The extent of variation is such that most (>95%) unrelated individuals can be distinguished on the basis of their mtDNA type; pairwise sequence comparisons between unrelated individuals show 1-2% sequence differences on average. Unlike the DNA found in the cell nucleus, mtDNA is haploid; that is, each individual possesses only one mtDNA genome, not the two copies of every gene found in the nuclear DNA. mtDNA is inherited maternally and fathers make no contribution to the mtDNA gene pool of their offspring. For this reason, mtDNA analysis has proven useful in missing person identification cases [39-40]; a sequence

match between an unknown and a mother or a sib is strong evidence of a familial relationship. mtDNA analysis has also proven useful with samples containing very little DNA, e.g., hair shafts [42-42].

The mtDNA sequence database now contains sequence data from a wide scattering of populations from all over the world [36]. Nevertheless, extension of this database is needed to provide a better picture of the extent of variation and the rate of mutational change in human populations. DNA sequence analysis, even using automated sequencers, is relatively time consuming due in part to the redundant sequencing and cross-checking that goes into verifying the correctness of a determined sequence; neither manual nor automated sequence analysis have fully overcome problems with error and ambiguity in sequence reading [43-44]. The use of CAE for automated high throughput sequencing is under investigation in several genome research laboratories in addition to the Berkeley High Sensitivity DNA Analysis laboratory.

II. SUMMARY OF RESEARCH FINDINGS

Project research findings are summarized in the sections below. Much of the work is described in publication and when this is the case, the research summaries are abbreviated with reference to the original publication. Work not published in the open literature is summarized in more detail. Project publications are listed in Appendix I and are designated in the following text **I-a**, **I-b**, **I-c**, etc.; project presentations are listed in Appendix II and are designated **II-a**, **II-b**, **II-c**, etc.

A. Use of CAE for STR Typing

Most of our work on STR typing made use of the first generation Berkeley High Sensitivity DNA Analysis Group CAE unit. This CAE unit had an array of 10 capillaries (75 um i.d.) with an effective run length of 25 cm. Fluorescence excitation was provided by an argon ion laser at 488 nm with confocal detection in 2 colors, 510-530 nm (green channel) and > 590 nm (red channel) [17]. The fluorescent signals from the CAE unit are translated to a computer record via an analog to digital converter. Output was stored in digitized form and processed using the programs IPLab, Kalcidagraph, and Canvas. Separations are visualized both as an image of the migrating bands in each capillary and as an electropherogram, a densitometric profile display; the latter allows quantitative comparison of electrophoretic profiles. With appropriate software, genetic types can be called directly; however, development of software compatible with this CAE unit was beyond the scope of this project. During the last two months of the project, experiments were conducted on the second generation 4-color instrument (see section D below); the major focus of these experiments was to test multi-color multiplex combinations.

Experimental details regarding sample preparation, electrophoretic separation, and data analysis have been described in publication [**I-a**, **I-d**, **I-h**] and will not be elaborated here. DNA samples of known STR type were used throughout; types were determined either by typing on slab gels using commercial kits or by measurement against absolute size standards on sequencing gels. DNA samples were amplified by PCR and the amplicons were desalted by flotation dialysis on VCWP membranes prior to electrophoresis; PCR products were loaded on the capillary columns by electrokinetic injection. Initial experiments employed CAE in the matrix 0.8% hydroxyethylcellulose (HEC) under non-denaturing conditions; subsequent experiments employed the same matrix under denaturing conditions.

The two-color detection capacity of the CAE unit was exploited by co-electrophoresing differentially labeled sample and allelic ladder in the same capillary [**I-a**, **I-d**]; this simplifies allele typing calls and facilitates multiplex typing. Energy transfer (ET) dye labeled primers were used for the differential labeling; this approach is discussed in more detail in section C below. Later experiments used a M13 A-track sequence as an absolute sizing standard [**I-h**]; the comparative benefits of the two different standards are discussed in section E below.

Fifteen STR loci were investigated singly and in various multiplex combinations by conventional

slab gel typing and by CAE typing using ET primers. The 15 loci are listed in the table below; of these, 12 are on the set of 13 currently included in CODIS. For several of the loci, primers had to be modified to improve amplification efficiency or to accommodate ET dye placement; these are indicated in the table by *. (The CHLC reference refers to primers from the Cooperative Human Linkage Center database.)

Table 1. LOCI INVESTIGATED FOR CAE ANALYSIS

Locus	Repeat	Het.	Alleles (Size Range)	Primer Reference
THO1	AATG	0.77	8 (179-203)	[45]
TPOX	AATG	0.66	8 (232-248)	[46, 47]
CSF1PO	AGAT	0.74	9 (299-323)	[20, 48]
vWFA31*	AGAT	0.74	11 (139-167)	[22]
FIBRA*	AAAG	0.82	9 (256-284)	[26]
FES/FPS	AAAT	0.70	8 (222-250)	[22]
D8S1179	TCTA	0.78	8 (152-210)	[26]
D18S51	AAAG	0.89	14 (267-319)	[26]
D20S85	AAAG	0.70	6 (123-147)	[26]
D21S11	TCTA/G	0.82	12 (172-264)	[26]
D5S818*	GATA	0.78	6 (139-159)	CHLC
D7S820*	GATA	0.85	7 (226-250)	CHLC
D13S317*	GATA	0.75	7 (175-199)	CHLC
D3S1358	GATA	0.78	7 (97-121)	[49]
D3S1359	GATA	0.89	12 (213-257)	[49]

Overall, the results of our studies establish the feasibility of using CAE for high throughput STR typing in multiplex arrays. With electrophoresis run cycles of 45-60 minutes, and projecting to the 100 capillary array units in development, it should be possible to process more than 800 samples at multiple loci per day; with automated sample loading allowing 24 hour operation, this number increases to 2400 per day. These figures are well within the throughput specifications to sustain databank development.

In addition to the publications noted above, aspects of this work have been presented before a variety of audiences - **II-a**, **II-c**, **II-d**, **II-e**, **II-j**, **II-k**, **II-l**, **II-m**, **II-o**, **II-p**, **II-r**, and **II-w**.

B. Use of CAE for DNA Sequence Analysis

The focus of our sequencing effort was the mtDNA D-loop region; as noted above, this sequence region is highly variable and has found extensive use for the identification of missing persons and body remains. As described in publications **I-e** and **I-I**, ET primers were prepared for the amplification of the D-loop hypervariable region I (HVR-I) and analyses were conducted on the second generation 4-color CAE unit. Sequences were read from the CAE output by a software program and edited by manual review of the CAE electropherograms. The original papers

should be consulted for experimental details.

Test samples were selected from a collection of samples of known mtDNA D-loop sequence; these sequences had been previously determined in this lab by manual sequencing [37, unpublished]. Twelve samples were selected for study based on the nature of the differences in their HVR-I sequences; some differed by but a single base from each other and others contained sequence complexities (e.g., base runs) that pose problems in manual sequencing. Sequences determined by CAE were compared to the known sequences. The raw unedited base calling rate was 68% to 99.6%; most of the difference was due to ambiguous calls. With editing, the sequencing accuracy ranged from 98.9-100%.

This work demonstrated the potential of CAE for high throughput sequence analysis. Using the approach illustrated here, DNA sequences of 450bp or more can be determined in under 3 hours. For a CAE unit with 100 capillaries and automated sample loading, eight sequencing runs per day at 45,000 bp per run could generate 360,000 bp of sequence per day. This work laid the foundation for a small microbial genome sequencing project subsequently undertaken by the Berkeley High Sensitivity DNA Analysis Group. A plan to use the high throughput capabilities of CAE to short cut the sequencing of the human genome has recently been announced [50].

C. Energy Transfer (ET) Fluorescent Labels Improves Signal Sensitivity and Quality

Current high throughput electrophoretic technologies for DNA sequencing, gene mapping, and STR typing employ fluorescence-based detection. A standard labeling strategy is to attach a single fluorescent dye molecule to the PCR primer used to generate the sequencing or STR product. With this strategy, multiple dyes can be used, allowing advantage to be taken of multi-color detection systems. For example, four different fluorescent dyes are routinely used in automated sequencing and in multiplex STR typing on the Applied Biosystems automated sequencer [51]. A limitation of these dyes is that each fluorophore has its own distinctive fluorescence excitation and emission properties; detection systems require compromise excitation wavelengths and software processing to unscramble emission signals.

An alternative dye labeling strategy was developed by the Berkeley High Sensitivity DNA Analysis Group. This strategy exploits energy transfer between two dyes coupled onto the same primer [**I-a**, 52-54]; one dye molecule functions as a fluorescence donor and the other as an acceptor. Incident radiation at a single wavelength (488 nm from an argon ion laser in the CAE units used in this study) is absorbed by the donor which transfers energy to the acceptor; the acceptor then emits at its characteristic fluorescence emission wavelength. By using different acceptor dyes, essentially non-overlapping emission spectra can be obtained with excitation at a single wavelength; in other words, with ET labeling, it is possible to achieve little or no signal overlap between the detection channels. Moreover, the efficiency of fluorescence emission for an ET couple is enhanced relative to the accepter dyes alone. These two factors result in an improvement in the quality of the raw fluorescence output and simplifies the software massage of the signal to obtain a clean readout [52].

The ET couples used for primer labeling in this work along with their excitation and emission wavelengths were as follows: FAM (ex: 488 nm; em: 525 nm), FAM-JOE (ex: 488 nm; em: 555 nm), FAM-TAMRA (ex: 488 nm; em: 580 nm), and FAM-ROX (ex: 488 nm; em: 605 nm). The 4-color CAE unit was designed to detect emissions from these acceptor dyes.

The utility of ET labeled primers was demonstrated both for STR analysis and sequence analysis in publications **I-a**, **I-d**, **I-e**, **I-h**, and **I-I**, and in most of the presentations originating from this project. ET primers for sequencing from M13 cloning vectors are commercially available.

D. Construction of Second Generation CAE with 4-Color Detection Potential

The design of a CAE system with laser excited confocal fluorescence detection was initially described by Huang et al [17]. Unlike most fluorescence detection systems where the fluorescence emission is detected at a 90° angle to the laser excitation beam, this CAE design employs a confocal detection system in which emission is measured 180° to excitation. The confocal detection system utilizes an epi-illumination format; the laser beam is focused on the capillary through a microscope objective and the emitted fluorescence is gathered by the same objective. The ribbon of capillaries is moved back and forth under the focused laser beam on a translation microscope stage so that all capillaries are monitored. The confocal detection system is quite sensitive, with a routine sensitivity limit of about 10 attomoles DNA per band.

Most of the initial work on the project was conducted on this first generation CAE unit, a unit designed for two color detection. With the development of ET labeling, the detection of four or more colors became tractable. To investigate the potential of 4-color analysis for multiplex STR typing and for sequencing, the first 4-color CAE unit was constructed; funding for the construction was split between this grant and the Berkeley High-Sensitivity DNA Analysis Group. The basic design for this unit is shown in fig. 1; see **I-e** and **I-I** for further details. The 4-color unit became operational during the last quarter of the project. When the project was not renewed, work on STR typing had to be shut down and the potential of the unit for multiplex STR analysis could not be further pursued. The unit was subsequently used for some of the sequencing work described in section B above. The DNA Analysis Group subsequently constructed a second 4-color CAE unit and the two units are being used by that group for routine sequencing.

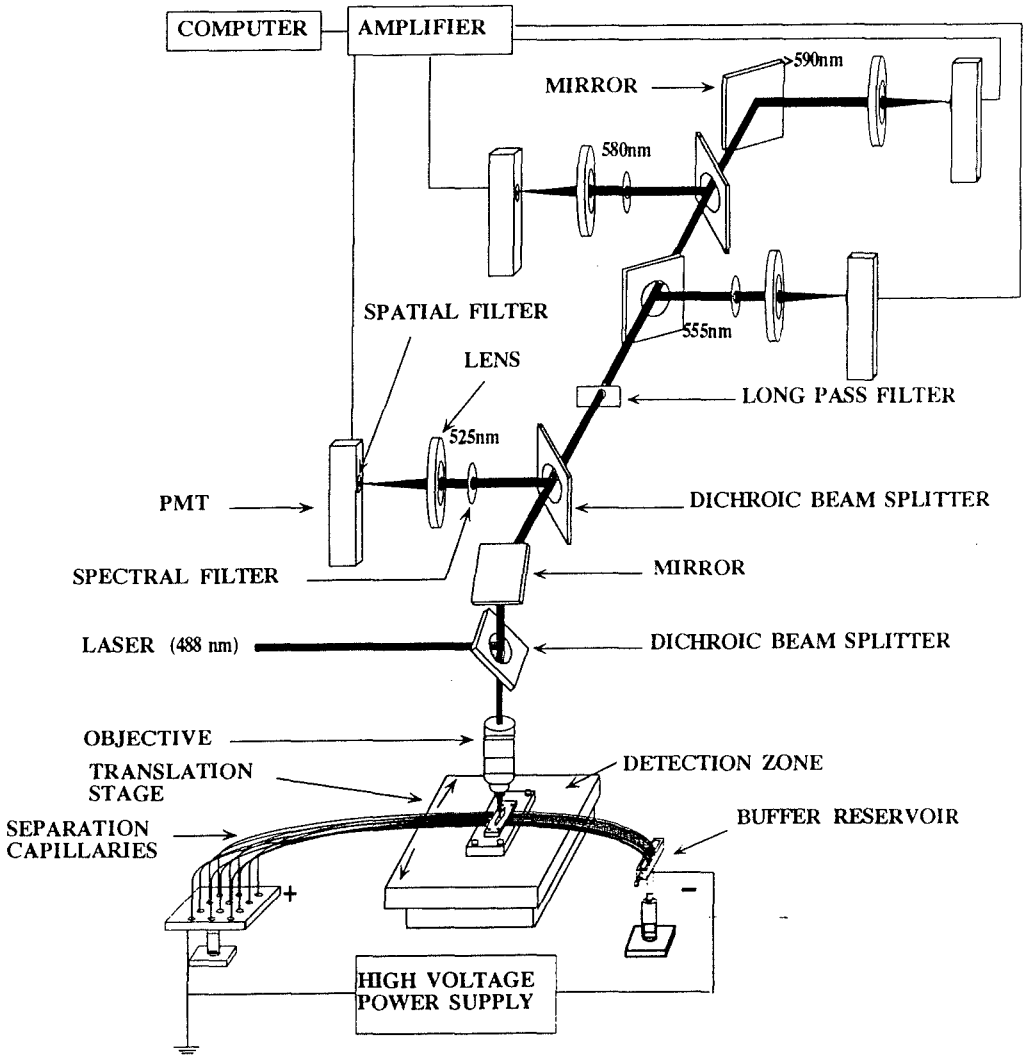


Figure 1. Capillary array electrophoresis unit with 4-color confocal fluorescence detection.

E. Application of In-Lane Allele Ladders and Single Base Sequence Step Ladders

The use of capillary electrophoresis for STR typing requires use of in-lane standards to compensate for electrophoretic variability between capillaries. The typical in-lane standard is a DNA sizing ladder. For example, the in-lane standard recommended for use with the Applied Biosystems 310 CE unit consists of 16 fragments ranging in size from 35 bp to 500 bp. Alleles are identified by measuring their size against the sizing ladder and comparing the determined size to preassigned allele size windows. Although this particular ladder (and, for that matter, this particular instrument) postdate the project period, this approach is typical.

We investigated two alternative approaches to in-lane ladders. The first was a simple allelic ladder; this approach is illustrated in publications **I-a** and **I-d**. For each locus, a mixture of samples was prepared such that all common alleles were represented in the mixture. The allelic ladder was labeled with one ET dye combination and the sample to be typed was labeled with another. Some of the ET labels were found to produce mobility shifts so it was necessary to identify labeling configurations that exhibited little or no shift. This approach would be amenable to simplified allele calling by software based on signal coincidence measurements rather than on fragment size comparisons. An additional advantage of this approach is that off-ladder variants are easily identified. Although this approach limits the number of multiplex sets that can be run at the same time - each multiplex set requires the commitment of two color detection channels, one for the ladder and one for the unknown - the overall throughput potential of CAE is such that this limitation might be offset by the simplicity of the allele calling protocols.

The second approach employed a one base sequencing ladder, specifically, one base sequences from the cloning vector M13. This approach is commonly used for manual di-, tri-, and tetra-nucleotide STR typing on sequencing gels. The approach is illustrated in publications **I-d** and **I-h**. As before, the one base sequencing ladder was labeled with one ET dye combination and the sample to be typed was labeled with another. Allele assignments were made as with a standard sizing ladder as described above, the advantage being that the M13 sizing ladder contains over 50 markers and shows near perfect linearity over the sizing range used for STR typing. Statistical analysis of replicate data showed excellent run-to-run reproducibility. The observed accuracy and precision makes possible straightforward type calling by data processing software.

F. Cross-compatibility for STR Typing Between CAE and Conventional Slab Gels

At the time this project was initiated, STR typing was typically done by slab gel electrophoresis with detection by silver staining, post-staining with fluorescent dyes, fluorescent primer labeling, or, rarely, isotopic labeling. Silver staining represents the low end technology; it is inexpensive but is relatively labor intensive. Fluorescence detection was and still is the high end technology; it requires relatively costly instrumentation in the form of a fluorescence imager or an automated sequencer. It appeared in 1994-95 that forensic laboratories might follow different technological paths to STR typing; it was important to compare the different technologies and verify cross-compatibility in typing.

Four approaches to STR typing were compared: CAE with ET labeled samples, slab gel electrophoresis under denaturing conditions with detection by silver staining, slab gel electrophoresis as above with detection by post-staining using the fluorescent DNA intercalation dye SYBR Green, and slab gel electrophoresis as above with ET labeled samples. For fluorescence detection on the slab gels, we employed a fluorescence imager (Molecular Dynamics FluorImager 575, generously loaned by Molecular Dynamics); this imager had two-color detection potential similar to the CAE unit, i.e., excitation at 488 nm with detection bandpasses centered at 530 nm (red channel) and 610 nm (green channel). Typing on slab gels

entails comparison of the positions of unknown bands to an allelic sizing ladder run in an adjacent or nearby electrophoresis lane. Typing on CAE, in contrast, requires comparison of the unknown to standards in the same capillary. As long as the allelic standards are correctly identified, typing discordances would not be expected with any of these approaches unless electrophoretic artifacts confound interpretation. No confounding artifacts were seen; typings were unambiguous regardless of the electrophoretic system and detection approach used.

ET labeled amplicons were cleanly detected on slab gels with the fluorescence imager. Both red and green labeled samples were detected with comparable sensitivity; as with CAE, there was negligible cross-talk between the detection channels. Fluorescence detection with the CAE system was at least 10x more sensitive than by the fluorescence imager; this is a consequence of the use of confocal optics on the former. FluorImager detection, as with CAE detection, allows relative band densities to be quantitated, a benefit when assessing mixtures (see section G below). The use of the two-color fluorescence imager allows co-electrophoresis of unknowns and standards in the same gel lanes as with CAE although this is not really necessary with slab gel systems; it also makes possible multi-color multiplexes (as illustrated in **I-d**).

The use of SYBR Green post-staining also yielded good results with the fluorescence imager. Detection sensitivity was about the same as with the ET primers. Image quality was perhaps a bit better with the ET primers due to the lack of post-staining artifacts sometimes seen with intercalation dyes such as SYBR Green. With electrophoresis under denaturing conditions, the two DNA strands in a PCR amplicon may migrate with slightly different mobilities, the consequence being that the amplicon is appears as a band doublet. Silver staining was about 3x less sensitive than either ET or SYBR Green staining. As with SYBR Green staining, silver staining reveals both DNA strands.

An important implication of the sensitivity gain with CAE is that fewer amplification cycles can be used when preparing samples for CAE; use of fewer cycles generally improves signal to noise ratios. This was investigated by determining the minimum number of PCR cycles needed to detect product by the four methods. The results were: CAE - less than 20 cycles, ET labeling and SYBR Green staining on gels - 26 cycles, and silver staining - 30 cycles.

In sum, this work demonstrated that STR typing is translatable between four different typing systems, ranging from the low tech silver staining approach to the high tech CAE. Moreover, we showed that ET primers can be used to advantage for STR typing on slab gels with fluorescence imager detection. This work was presented at several meetings [**II-h**, **II-k**, **II-p**] and has been published [**I-b**, **I-d**].

G. Analysis of Sample Mixtures

The interpretation of genetic typing results in samples containing contributions from different individuals has been a problem of historical concern in forensic genetics. Mixtures are readily recognized if typing at a locus shows more than two alleles. Due to the high heterozygosity of

STRs, the chance that two or more donors would contribute no more than two alleles between them across multiple loci is negligibly small. There remains, however, the question of which donor contributed which allele or alleles. With the use of fluorescent labeling, whether with conventional single labels or with ET labels, this question can be addressed by assessing the quantitative proportions of the allelic signals.

We focused on three questions in mixture analysis:

a. What is the detection limit of the lesser contributor in a mixed sample?
b. How well do allele intensity differences reflect contributor proportions?
c. What is the range of allele intensity differences in a heterozygote sample from a single individual? This question addresses the differentiation of contributors where there are two donors, each a homozygote of a different type.

These questions were investigated by measuring allele peak intensities in heterozygote samples from single individuals (question c) and in mixtures of heterozygotes from pairs of individuals giving three allele patterns (one allele shared between the individuals) and four allele patterns (no alleles shared between the individuals). Testing was done using ET labeled THO1 as the model system. Allele peak intensities were determined for samples separated both on conventional slab gels and by CAE. Peak intensities on slab gels were determined by scanning on a FluorImager; peak intensities from the CAE were determined directly from the fluorescence signal output. Integrated peak intensities were used for the analysis.

Questions (a) and (b) are inter-related in that the detected allelic proportions provide answers to both. The three allele mixture sample contained contributions from THO1 types 6,9 and 9,9.3; the four allele mixture was from types 6,9.3 and 7,8. Mixtures in proportions 20:1, 10:1, 5:1, 3:1, 1:1, ...,1:20 were tested. The scanned slab gel images showed the expected alleles in all the tested mixture proportions; in the 20:1 and 1:20 mixtures, the alleles from lesser contributor were just visible and these proportions reflect the practical detection limits with this system. The CAE electropherograms were less sensitive for mixture visualization; mixtures were apparent in the 10:1 and 1:10 samples but not in the 20:1 and 1:20 samples. The CAE unit used for this study was not optimized with regard to signal to noise characteristics; this no doubt accounts for much variability observed on measurements made with this instrument for later analyses performed independently on another project using an instrument optimized for sensitivity detected mixtures beyond 1:20.

The observed detection limits were reflected in the quantitative measurement of the allele peak intensities. The proportions detected off the slab gels by fluorescent imaging were fairly close to the proportions expected for the mixtures throughout the mixture range; for example, in the three allele 1:10 mixture, the observed proportions of the 6, 9, and 9.3 alleles were 1 : 10.4 : 8.6 compared to the expected 1:11:10. With the CAE system, the observed and expected proportions were close over the range 10:1 to 1:10 but for the more extreme 20:1 and 1:20 mixtures, the dominant sample appeared in much greater excess than the expectation.

The question of allelic peak intensities in heterozygotes for a single individual was tested with

THO1 types 7,8 and 6,9.3; these two types were selected to determine if there were differences between alleles adjacent in size and alleles differing substantially in size. The slab gel results showed no differences; the peak ratios for the larger to the smaller allele (that is, 8:7 and 9.3:6) were 0.9 ± 0.1 (range 0.8-1). The CAE results were a little more variable with ratios of 1 ± 0.1 (0.8-1.3) and 1.3 ± 0.3 (1-1.8) respectively. Again, the variability with the CAE unit reflects the non-optimized signal to noise characteristics of the instrument used for this study. Later studies on the optimized instrument noted above showed more precise ratios.

Overall these experiments, although preliminary, indicate that allele proportions are fairly reflective of the relative contributions of the donors to the sample mixture and suggest that allele peak ratio data can be used to distinguish donors to a mixture when the mixtures are 2:1 or more extreme. These findings were presented **[II-l]** and published as an abstract **[I-g]**.

H. STR Typing Study on a Japanese Population

Genetic typing at four STR loci was done on a Japanese population set of 187 individuals consisting of 80 individuals from the metropolitan Tokyo area and 107 individuals from five other geographic regions of Japan, from Hokkaido in the north to Okinawa in the south. The loci typed were CSF1PO, TPO, THO1, and vWFA31. No previously undescribed alleles were noted. At the level of the total population, allele frequencies at the TPO, THO1, and vWFA31 loci were in good agreement with Hardy-Weinberg equilibrium expectations; with the Bonferroni correction for multiple tests, the deviation of CSF1PO was not significant. Pairwise comparison of loci using Fisher's Exact Tests showed no evidence of linkage disequilibrium; a multi-way independence test on all four loci also showed no departure from independence expectations.

Statistical analysis on the allele frequencies in the regional subpopulations showed only two deviations from HW equilibrium expectations at the $p<0.05$ level and none at the $p<0.01$ level. Comparison of the subpopulation frequencies to the total population averages showed only a few instances of significant deviation ($p>0.99$ or $p<0.01$), the preponderance of which being associated with the Okinawa population. Fst analysis showed only small differences between the regional populations, the largest being between Okinawa and the others. The estimated value for Fst in the total entire population was 0.009. These data are comparable to prior Fst analyses based on the classical genetic markers.

These data have been presented **[II-v, II-x]** and published **[I-m]**.

I. Development of New Primers for CSF1PO

CSF1PO is one of the core STR loci used for forensic identification purposes. The PCR primers described by Hammond et al. for CSF1PO typing yield amplicons in the size range 295-327 bp [#]. Fragments in this size range amplify with reduced efficiency in samples containing severely degraded DNA. To extend the usefulness of CSF1PO typing for such samples, the CSF1PO

primers were redesigned to give smaller PCR products(150-182 bp). The reliability of the new primers for CSF1PO typing was demonstrated on known samples and by sequence analysis of the resultant PCR products. These primers are superior to the original primers with regard to electrophoretic resolution and utility for typing of severely degraded DNA.

This work has been presented [**II-v**] and published [**I-j** and **I-l**].

J. Evaluation of Amplified Restriction Fragment Polymorphism (AFLP)

Amplified Restriction Fragment Polymorphism (AFLP) analysis is a technique developed for the identification of strain variation in plant and animals [55]. AFLP employs PCR to amplify a selected set of restriction fragments prepared from total genomic DNA; the result is a complex "fingerprint" type pattern which characterizes the restriction fragment set. By using different combinations of selective primers, distinct and non-overlapping restriction fragment sets can be visualized. Considerable variation has been observed between cultivars within plant species.

We tested a commercial AFLP kit (Life Technologies) for potential usefulness in human differentiation. Genomic DNA samples from 4 individuals were restricted with EcoRI (E) and MseI (M). Restriction fragments were ligated to E and M linkers and E-M fragments were amplified using primers specific for the E and M linker sequences. Twelve sets of E-M fragments were selectively amplified using combinations of E and M primers with different 3 base additions at their 3' ends. Fragments were separated on non-denaturing 5% acrylamide gels, stained with SYBRGreen, and visualized on a FluorImager (Molecular Dynamics). Very little variation was observed between individuals with any of the selective primer combinations tested, suggesting AFLP typing has limited usefulness for human differentiation.

This work was presented [**II-u**] and published as an abstract [**I-k**].

K. Evaluation of Micro-scale "CAE on a Chip"

CAE based genetic typing methodologies provide a stepping stone to a next generation of DNA analysis technologies: miniaturized CAE on a chip [56-60]. The feasibility of STR typing on CAE chips was demonstrated in preliminary experiments using chips described by Woolley and Mathies [60]; see fig 2. This preliminary work has been described in presentations (**II-c**, **II-k**, **II-m**, **II-r**) and publication (**I-d**).

Unfortunately, work with STR typing could not be followed up due to the termination of this grant. However, we have continued to work with CAE chips applied to another genetic typing problem, the high throughput typing of a disease marker gene. These studies have demonstrated that CAE chips can be formatted to allow robotic introduction of sample. Using these chips, we have demonstrated the typing of 96 samples in under 10 minutes [61]. Although this particular application did not require the level of fragment resolution needed for STR analysis, parallel

studies by others in the Berkeley High Sensitivity DNA Analysis Group have demonstrated near single base resolution on chip separations; thus the resolution needed for STR analysis can be achieved.

Single Base Resolution for THO1 Allelic Typing

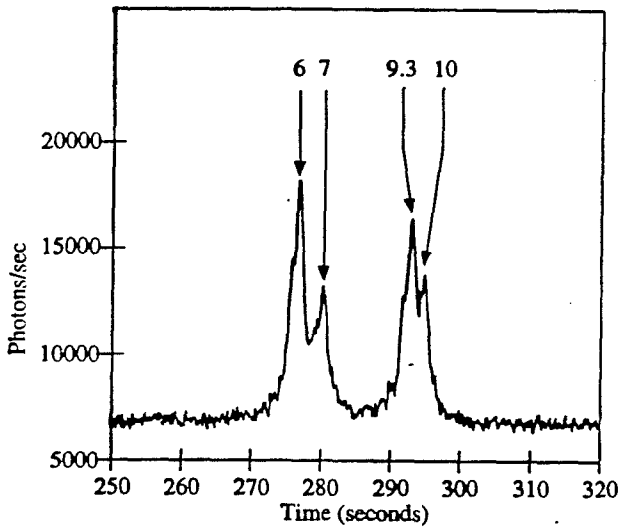


Figure 2. THO1 typing by CE on a chip. Two samples, a 6,9.3 and a 7,10, were mixed to show 4 bp and 1 bp resolution. From **I-d**.

These studies demonstrate that advanced CAE chip technology has great potential for high throughput STR typing. The continued development of CAE chip technology will not only advance throughput, it should also decrease cost per analysis, a significant consideration in making DNA analysis available to forensic laboratories at the broadest level.

SUMMARY

It is recognized that high throughput genetic typing technologies are needed to fully exploit the potential of forensic DNA profiling. The research described in this report demonstrates that massively parallel capillary array electrophoresis possesses the reliability and throughput capability to meet this need.

REFERENCES

1. Connors E; Lundregan T; Miller N; McEwen T. 1996. Convicted By Juries, Exonerated By Science: Case Studies in the Use of DNA Evidence To Establish Innocence After Trial. National Institute of Justice, U.S. Department of Justice, Washington DC.

2. McEwen JE. 1995. Forensic DNA data banking by state crime laboratories. Amer. J. Human Genet. 56:1487-1492.

3. Heiger DN; Cohen AS; Karger BL. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. J. Chromatography 516(1):33-48.

4. Schwartz HE; Ulfelder K; Sunzeri FJ; Busch MP; Brownlee RG. 1991. Analysis of DNA restriction fragments and polymerase chain reaction products towards detection of the AIDS (HIV-1) virus in blood. J. Chromatography 559(1-2):267-283.

5. Ulfelder KJ; Schwartz HE; Hall JM; Sunzeri FJ. 1992. Restriction fragment length polymorphism analysis of ERBB2 oncogene by capillary electrophoresis. Analytical Biochemistry 200(2):260-267.

6. Williams PE; Marino MA; Del Rio SA; Turni LA; Devaney JM. 1994. Analysis of DNA restriction fragments and polymerase chain reaction products by capillary electrophoresis. J. Chromatography A 680:525-540.

7. Marino MA; Turni LA; Del Rio SA; Williams PE. 1994. Molecular size determinations of DNA restriction fragments and polymerase chain reaction products using capillary gel electrophoresis. J. Chromatography A 676:185-189.

8. Cohen AS; Najarian DR; Karger BL. 1990. Separation and analysis of DNA sequence reaction products by capillary gel electrophoresis. J. Chromatography 516(1):49-60.

9. Swerdlow H; Dew-Jager KE; Brady K; Grey R; Dovichi NJ; Gesteland R. 1992. Stability of capillary gels for automated sequencing of DNA. Electrophoresis 13(8):475-483.

10. Swerdlow H; Zhang JZ; Chen DY; Harke HR; Grey R; Wu SL; Dovichi NJ; Fuller C. 1991. Three DNA sequencing methods using capillary gel electrophoresis and laser-induced fluorescence. Analyt. Chem. 63(24):2835-2841.

11. Chen DY; Swerdlow HP; Harke HR; Zhang JZ; Dovichi NJ. 1991. Low-cost, high-sensitivity laser-induced fluorescence detection for DNA sequencing by capillary gel electrophoresis. J. Chromatography 559(1-2):237-246.

12. Swerdlow H; Wu SL; Harke H; Dovichi NJ. 1990. Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. J. Chromatography 516(1):61-67.

13. Swerdlow H; Gesteland R. 1990. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. Nucleic Acids Research 18(6):1415-1419.

14. Huang XC; Quesada MA; Mathies RA. 1992. DNA sequencing using capillary array electrophoresis. Analyt. Chem. 64:2149-2154.

15. Mathies RA; Huang XC. 1992. Capillary array electrophoresis: an approach to high-speed, high throughput DNA sequencing. Nature 359:167-169.

16. Carson S; Cohen AS; Belenkii A; Ruiz-Martinez MC; Berka J; Karger BL. 1993. DNA sequencing by capillary electrophoresis: Use of a two-laser-two-window intensified diode array detection system. Analyt. Chem. 65:3219-3226.

17. Huang XC; Quesada MA; Mathies RA. 1992. Capillary array electrophoresis using laser-excited confocal fluorescence detection. Analyt. Chem. 64:967-972.

18. Edwards A; Civitello A; Hammond HA; Caskey CT. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Amer. J. Human Genet. 49:746-756.

19. Edwards A; Hammond HA; Jin L; Caskey CT; Chakraborty R. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241-253.

20. Hammond HA; Jin L; Zhong Y; Caskey CT; Chakraborty R. 1994. Evaluation of 13 short tandem repeat loci for use in personal identification applications. Amer. J. Human Genet. 55:175-189.

21. Alford RL; Hammond HA; Coto I; Caskey CT. 1994. Rapid and efficient resolution of parentage by amplification of short tandem repeats. Amer. J. Human Genet. 55:190-195.

22. Kimpton CP; Gill P; Walton A; Urquhart A; Millican ES; Adams M. 1993. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Methods Appl. 3:13-22.

23. Kimpton CP; Fisher D; Watson S; Adams M; Urquhart A; Lygo J; Gill P. 1994. Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. Internat. J. Leg. Med. 106:302-311.

24. Urquhart A; Oldroyd NK; Kimpton CP; Gill P. 1995. Highly discriminating heptaplex

short tandem repeat PCR system for forensic identification. BioTechniques 18:1:116-121.

25. Smith RN. 1995. Accurate size comparison of short tandem repeat alleles amplified by PCR. BioTechniques 18:1:122-128.

26. Oldroyd NJ; Urquhart AJ; Kimpton CP; Millican ES; Watson SK; Downes T; Gill PD. 1995. A highly discriminating octoplex short tandem repeat polymerase chain reaction system suitable for human individual identification. Electrophoresis 16:334-337.

27. Beckmann JS; Weber JL. 1992. Survey of human and rat microsatellites. Genomics 12:627-631.

28. Gyapay G; Morissette J; Vignal A; Dib C; Fizames C; Millasseau P; Marc S; Bernardi G; Lathrop M; Weissenbach J. 1994. The 1993-94 Genethon human genetic linkage map. Nature Genetics 7:246-249.

29. Whitaker JP; Clayton TM; Urquhart AJ; Milllican ES; Downes TJ; Kimpton CP; Gill P. 1995. Short tandem repeat typing of bodies from a mass disaster: High success rate and characteristic amplification patterns in highly degraded samples. BioTechniques 18:670-677.

30. Fregeau CJ; Fourney RM. 1993. DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. BioTechniques 15:100-109.

31. National Research Council. 1996. The Evaluation Of Forensic DNA Evidence. National Academy Press, Washington DC

32. Cann RL; Stoneking M; Wilson AC. 1987. Mitochondrial DNA and human evolution. Nature 325:31-36.

33. Vigilant L; Pennington R; Harpending H; Kocher TD; Wilson AC. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86:9350-9354.

34. Hoari S; Hayasaka K. 1990. Interspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Amer. J. Human Genet. 46:828-842.

35. Mountain JL; Hebert JM; Bhattacharyya S; Underhill PA; Ottolenghi C; Gadgil M; Cavalli-Sforza LL. 1995. Demographic history of India and mtDNA-sequence diversity. Amer. J. Human Genet. 56:979-992.

36. Miller KWP; Dawson JL; Hagelberg E. 1996. A concordance of nucleotide substitutions

in the first and second hypervariable segments of the human mtDNA control region. Int. J. Legal. Med. 109:107-113.

37. Melton T; Ginther C; Sensabaugh GF; Soodyall H; Stoneking M. 1997. Extent of heterogeneity on mitochondrial DNA of sub-Saharan African populations. J. Forensic Sci. 42:582-592.

38. Melton T; Wilson M; Batzer M; Stoneking M. 1997. Extent of heterogeneity on mitochondrial DNA of European populations. J. Forensic Sci. 42:437-446.

39. Holland MM; Fisher DL; Mitchell LG; Rodriguez WC; Canik JJ; Merril CR; Weedn VW. 1992. Mitochondrial DNA sequence analysis of human skeletal remains: Identification of remains from the Vietnam War. J. Forensic Sci. 38:542-553.

40. Gill P; Ivanov PL; Kimpton C; Piercy R; Benson N; Tully G; Evett I; Hagelberg E; Sullivan K. 1994. Identification of the remains of the Romanov family by DNA analysis. Nature Genetics 6:130-135.

41. Higuchi RG; vonBeroldingen CH; Sensabaugh GF; Erlich HA. 1988. DNA typing from single hairs. Nature 322:543-546.

42. Wilson MR; Polanskey D; Butler J; DiZinno JA; Replogle J; Budowle B. 1995. Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. BioTechniques 18(4):662-669.

43. Churchill G; Waterman M. 1992. The accuracy of DNA sequences: Estimating sequence quality. Genomics 14:89-98.

44. Koop BF; Rowan L; Chen W-Q; Deshpande P; Lee H; Hood L. 1993. Sequence length and error analysis of sequenase(A) and automated Taq cycle sequencing methods. BioTechniques. 14:442-447.

45. Puers C; Hammond HA; Jin L; Caskey CT; Schumm JW. 1993. Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG]n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Amer. J. Human Genet. 53:953-958.

46. Anker R; Steinbrueck T; Donis-Keller H. 1992. Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. Hum. Molec. Genetics 1(2):137.

47. Huang NE; Schumm J; Budowle B. 1995. Chinese population data on three tetrameric short tandem repeat loci--HUMTHO1, TPOX, and CSF1PO--derived using multiplex PCR and manual typing. Forensic Science Int'l. 71:131-136.

48. Polymeropoulos MH; Rath DS; Xiao H; Merril C. 1991. Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES) Nucleic Acids Res. 19(14):4018.

49. Li H; Schmidt L; Wei MH; Hustad T; Lerman MI; Zbar B; Tory K. 1993. Three tetranucleotide polymorphisms for loci: D3S1352; D3S1358; D3S1359. Hum. Mol. Genet. 2:1327.

50. Marshall E; Pennisi E. 1998. Hubris and the human genome. Science 280:994-995.

51. Hunkapiller T; Kaiser, RJ; Koop BF; Hood L. 1991. Large-scale and automated DNA sequence determination. Science 254:59-67.

52. Ju J; Ruan C; Fuller CW; Glazer AN; Mathies RA. 1995. Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. Proc. Natl. Acad. Sci. USA 92:4347-4351.

53. Ju J; Glazer AN; Mathies RA. 1996. Energy transfer primers: a new fluorescence labeling paradigm for DNA sequencing and analysis. Nature Medicine 2:246-249.

54. Ju J; Kheterpal I; Scherer JR; Ruan C; Fuller CW; Glazer AN; Mathies RA. 1995. Design and synthesis of fluorescence energy transfer dye-labeled primers and their application for DNA sequencing and analysis. Anal. Biochem. 231:131-140.

55. Vos P; Hogers R; Bleeker M; Reijans M; van de Lee T; Freijters A; Pot J; Peleman J Kuiper M; Zabeau M. 1995. AFLP - a new technique for DNA fingerprinting. Nuc. Acids Res. 23:4407-4414.

56. Harrison DJ; Manz A; Fan Z; Lludi H; Widner HM. 1992. Capillary electrophoresis and sample injection systems integrated on a planar glass chip. Analyt. Chem. 64:1926-1932.

57. Harrison DJ; Fluri K; Seiler K; Fan Z; Effenhauser CS; Manz A. 1992. Micromachining a miniaturized capillary electrophoresis-based chemical analysis system on a chip. Science 261:895-897.

58. Effenhauser CS; Paulus A; Manz A; Widner HM. 1994. High-speed separation of antisense oligonucleotides on a micromachined capillary electrophoresis device. Analyt. Chem. 66:2949-2953.

59. Schmalzing D; Koutny L; Adourian A; Belgrader P; Matsudaira P; Ehrlich D. 1997. DNA typing in 30 seconds with a microfabricated device. Proc. Nat. Acad. Sci. 94:10273-10278.

60. Woolley AT; Mathies RA. 1994. Ultra-high-speed DNA fragment separations using

microfabricated capillary array electrophoresis chips. Proc. Natl. Acad. Sci. USA 91:11348-11352.

61. Simpson PC; Roach D; Wooley AT; Thorsen T; Johnston R; Sensabaugh GF; Mathies RA. 1998. High-throughput genetic analysis using microfabricated 96-sample capillary array electrophoresis microplates. Proc. Nat. Acad. Sci. 95:2256-2261.

APPENDIX I: PUBLICATIONS

a. HIGH SPEED, HIGH THROUGHPUT THO1 ALLELIC SIZING USING ENERGY TRANSFER PRIMERS AND CAPILLARY ARRAY ELECTROPHORESIS. Y. Wang, J. Ju, B. Carpenter, J. Atherton, G.F. Sensabaugh, and R. Mathies. Anal. Chem. 67: 1197-1203 (1995)

b. VERTICAL COMPATIBILITY OF SHORT TANDEM REPEAT POLYMORPHISM TYPING USING ENERGY TRANSFER FLUORESCENT PRIMERS. J. Atherton Wallin, B. Carpenter, Y. Wang, J. Ju, R. Mathies, and G.F. Sensabaugh. CACNews, Fall 1995, pp. 16-17 (1995)

c. RAPID DETECTION OF GENETIC TYPES A POLYMORPHIC SHORT TANDEM REPEAT (STR) LOCI USING TWO COLOR CAPILLARY ARRAY ELECTROPHORESIS. Y. Wang, J. Ju, R. Mathies, B. Carpenter, J. Atherton, and G.F. Sensabaugh. Science and Justice 35:240 (1995). (abstract)

d. SHORT TANDEM REPEAT (STR) POLYMORPHISM ANALYSIS USING ENERGY TRANSFER FLUORESCENT PRIMERS. J. Atherton Wallin, J. Ju, Y. Wang, B. Carpenter, R. Mathies, and G.F. Sensabaugh. Proceedings, 6th International Symposium of Human Identification - 1995. Promega Corp. (1996) pp.72-80.

e. DNA SEQUENCING USING FOUR-COLOR CAPILLARY ARRAY ELECTROPHORESIS AND ENERGY TRANSFER PRIMERS. I. Kheterpal, J. Ju, A. Radhakrishnan, G.S. Brandt, C.L. Ginther, S.M. Clark, J.R. Scherer, G.F. Sensabaugh, and R.A. Mathies. Proc. Soc Photo-Opt. Instrum. Engineers 1996, 2680:204-213 (1996).

f. FORENSIC APPLICATIONS OF PCR TECHNOLOGY: THE VIEW FROM THE FIRST DECADE. G.F. Sensabaugh. Proceedings, International Workshop on Forensic DNA Analysis of Evidential Samples. National Research Institute for Police Science, Tokyo, Japan (1996) p.13 (Abstract)

g. SHORT TANDEM REPEAT (STR) POLYMORPHISM ANALYSIS USING ENERGY TRANSFER FLUORESCENT PRIMERS. J. Atherton Wallin, Y. Wang, R. Mathies, and G.F. Sensabaugh. Science and Justice 36:199-200 (1996) (Abstract)

h. HIGH RESOLUTION CAPILLARY ARRAY ELECTROPHORETIC SIZING OF MULTIPLEXED SHORT TANDEM REPEAT LOCI USING ENERGY TRANSFER FLUORESCENT PRIMERS. Y. Wang, J. M. Wallin, J. Ju, G.F. Sensabaugh, and R. Mathies. Electrophoresis 17:1485-1490 (1996).

PUBLICATIONS (Cont'd)

i. DNA SEQUENCING USING A FOUR-COLOR CONFOCAL FLUORESCENCE CAPILLARY ARRAY SCANNER. I. Kheterpal, J.R. Scherer, S.M. Clark, A. Radhakrishnan, J. Ju, C.L. Ginther, G.F. Sensabaugh, and R.A. Mathies. Electrophoresis 17: 1852-1859 (1996).

j. EVALUATION OF NEW PRIMERS FOR CSF1PO. K. Yoshida, K. Sekiguchi, K. Kasai, H. Sato, S. Seta, and G.F. Sensabaugh. Int. J. Legal Med. 110:36-38 (1997).

k. AMPLIFIED RESTRICTION FRAGMENT POLYMORPHISM (AFLP) ANALYSIS OF HUMAN GENOMIC DNA. K. Andera, J. Liu, and G.F. Sensabaugh. Science and Justice 38:62 (1998) (Abstract)

l. CSF1PO POLYMORPHISM: EVALUATION OF NEW PRIMERS AND ALLELE FREQUENCIES IN THE JAPANESE POPULATION. K. Yoshida, K. Sekiguchi, K. Kasai, H. Sato, S. Seta, and G.F. Sensabaugh. Science and Justice 38:62 (1998) (Abstract)

m. GENETIC VARIATION AT 6 STR IN THE JAPANESE POPULATION. K. Yoshida, K. Kasai, H. Sato, C. Brenner, and G.F. Sensabaugh. Advan. Forensic Haemogenetics 7 (1998) In Press.

APPENDIX II: PRESENTATIONS

(Presenter's name underlined)

a. RAPID DETECTION OF GENETIC TYPES AT POLYMORPHIC SHORT TANDEM REPEAT (STR) LOCI USING TWO COLOR CAPILLARY ARRAY ELECTROPHORESIS. Y. Wang, B. Carpenter, J. Atherton, R. Mathies, and G.F. Sensabaugh. Joint Meeting, California Association of Criminalists and The Forensic Science Society, Pasadena CA, October 1994.

b. FORENSIC DNA TECHNOLOGY. G.F. Sensabaugh. DNA Conference, Science in the Courtroom, Univ. of San Francisco School of Law Program in Continuing Legal Education, Nov. 1995.

c. RAPID HIGH THROUGHPUT DETECTION OF GENETIC TYPES AT POLYMORPHIC SHORT TANDEM REPEAT (STR) LOCI USING TWO COLOR CAPILLARY ARRAY ELECTROPHORESIS. Y. Wang, B. Carpenter, J. Atherton, R. Mathies, and G.F. Sensabaugh. Invited presentation, International Forensic Science Symposium, Taipei, Taiwan, November 1994.

d. DNA TECHNOLOGY, FORENSIC SCIENCE, AND CRIMINAL JUSTICE. G.F. Sensabaugh, Rosenblatt Lecturer, The Rosenblatt Memorial Lectures, Northeastern University, Boston MA, February 1995.

e. DNA PROFILING AND DATABANKS. G.F. Sensabaugh, Rosenblatt Lecturer, The Rosenblatt Memorial Lectures, Northeastern University, Boston MA, February 1995.

f. FORENSIC DNA. G.F. Sensabaugh, Invited Seminar, Medical Genetics Research Conference, UCSF, April 1995.

g. FORENSIC DNA TYPING - BEFORE AND AFTER. G.F. Sensabaugh. Invited Presentation to Meeting, California Association of Criminalists, Walnut Creek CA, May 1995.

h. VERTICAL COMPATIBILITY OF STR TYPING USING ENERGY TRANSFER FLUORESCENT PRIMERS. B. Carpenter, J. Atherton, Y. Wang, J. Ju, R. Mathies, and G.F. Sensabaugh. Meeting, California Association of Criminalists, Walnut Creek CA, May 1995.

i. FORENSIC APPLICATION AND ANALYSIS OF DNA. G.F. Sensabaugh, invited lecture, UC Davis, May 1995.

PRESENTATIONS (Cont'd)

j. GENETIC IDENTIFICATION FROM PROTEINS TO DNA. G.F. Sensabaugh, Invited Presentation, Summer Toxicology Meeting, Genetic and Environmental Toxicology Association and Northern California Chapter of the Society of Toxicology, Menlo Park CA, June, 1995.

k. SHORT TANDEM REPEAT (STR) POLYMORPHISM ANALYSIS USING ENERGY TRANSFER FLUORESCENT PRIMERS. J. Atherton Wallin, J. Ju, Y. Wang, B. Carpenter, R. Mathies, and G.F. Sensabaugh. Invited presentation, 6th International Symposium of Human Identification, Scottsdale AZ, September 1995.

l. SHORT TANDEM REPEAT (STR) POLYMORPHISM ANALYSIS USING ENERGY TRANSFER FLUORESCENT PRIMERS. J. Atherton Wallin, Y. Wang, R. Mathies, and G.F. Sensabaugh. Meeting, California Association of Criminalists, San Pedro CA, October, 1995.

m. HIGH THROUGHPUT DNA ANALYSIS USING CAPILLARY ARRAY ELECTROPHORESIS. Y. Wang, J. Ju, B. Carpenter, J. Atherton, R. Mathies, and G.F. Sensabaugh. Invited presentation, Eastern Analytical Symposium, Somerset NJ, November 1995.

n. DNA TECHNOLOGY AND WHAT IT IS DOING TO THE REST OF FORENSIC SCIENCE. G.F. Sensabaugh. Invited presentation, New Jersey Association of Forensic Scientists, Somerset NJ, November 1995.

o. HIGH THROUGHPUT STR TYPING USING CAPILLARY ARRAY ELECTROPHORESIS. G.F. Sensabaugh. Invited seminar, New York Medical Examiners Office, New York City, November 1995.

p. HIGH THROUGHPUT DNA ANALYSIS USING ENERGY TRANSFER FLUORESCENT PRIMERS AND CAPILLARY ARRAY ELECTROPHORESIS. G.F. Sensabaugh. Invited seminar, Institute for Forensic Genetics, Copenhagen, Denmark, November 1995.

q. FORENSIC APPLICATIONS OF PCR TECHNOLOGY: THE VIEW FROM THE FIRST DECADE. G.F. Sensabaugh. Invited presentation, International Workshop on Forensic DNA Analysis of Evidential Samples. National Research Institute for Police Science, Tokyo, Japan, December 1995.

r. MULTIARRAY CAPILLARY ELECTROPHORESIS. G.F. Sensabaugh. Workshop presentation, Advanced DNA Technologies, American Academy of Forensic Sciences, Nashville TN, February 1996.

PRESENTATIONS (Cont'd)

s. CAPILLARY ELECTROPHORESIS. G.F. Sensabaugh. Meeting, Northern California DNA Study Group, Berkeley CA, March 1996.

t. FORENSIC APPLICATION OF GENETIC IDENTIFICATION TECHNOLOGIES. G.F. Sensabaugh. Presentation to California State Senate Select Committee on Genetics and Public Policy, Sacramento CA, May 1996

u. AMPLIFIED RESTRICTION FRAGMENT POLYMORPHISM (AFLP) ANALYSIS OF HUMAN GENOMIC DNA. K. Andera, J. Liu, and G.F. Sensabaugh. Meeting, California Association of Criminalists, San Jose Ca, May 1996.

v. CSF1PO POLYMORPHISM: EVALUATION OF NEW PRIMERS AND ALLELE FREQUENCIES IN THE JAPANESE POPULATION. K. Yoshida, K. Sekiguchi, K. Kasai, H. Sato, S. Seta, and G.F. Sensabaugh. Meeting, California Association of Criminalists, San Jose Ca, May 1996.

w. CAPILLARY ELECTROPHORESIS: FROM ARRAYS TO CHIPS. G.F. Sensabaugh. Workshop presentation, Advanced DNA Technologies: Automation and Applications, American Academy of Forensic Sciences, New York City NY, February 1997. (Presentation not made due to illness.)

x. GENETIC VARIATION AT 6 STR IN THE JAPANESE POPULATION. K. Yoshida, K. Kasai, H. Sato, C. Brenner, and G.F. Sensabaugh. 17th Internat. Congress, Internat. Soc. Forensic Haemogenetics, Oslo, Norway, Sept. 1997 (Poster presentation)

APPENDIX III: PERSONNEL TRAINED

GRADUATE STUDENTS

Brooke Carpenter - STR analysis using ET primers on Fluorimager
Current position: Santa Clara County Criminalistics Laboratory, San Jose CA

Jennifer Griggs - STR analysis on Fluorimager
Current position: Human Identification Group, Roche Molecular Systems, Alameda CA

Jeanette Atherton Wallin - STR analysis using ET primers on Fluorimager and CAE
Current position: Human Identification Group, Applied Biosystems Division, Foster City CA

Stephenie Winter - Preparation of DNA samples
Current position: Los Angeles County Coroner's Office Forensic Laboratory, Los Angeles CA

Todd Thorsen - Training in STR analysis
Current position: Graduate student, California Institute of Technology, Pasadena, CA

POST-DOCTORAL RESEARCHER

Dr. Charles Ginther - mtDNA sequencing
Current position: Independent consultant, Los Angeles CA

VISITING SCIENTIST

Kanako Yoshida - Population studies on STR loci in Japanese
Current position: National Research Institute of Police Science, Tokyo Japan

APPENDIX IV

RESEARCH PUBLICATIONS AND PUBLISHED ABSTRACTS

Anal. Chem. **1995**, *67*, 1197–1203

Rapid Sizing of Short Tandem Repeat Alleles Using Capillary Array Electrophoresis and Energy-Transfer Fluorescent Primers

Yiwen Wang,[†,‡] Jingyue Ju,[†] Brooke A. Carpenter,[§] Jeanette M. Atherton,[§] George F. Sensabaugh,[§] and Richard A. Mathies*,[†]

Department of Chemistry and Forensic Science Group, School of Public Health, University of California, Berkeley, California 94720

Genetic typing of the short tandem repeat (STR) polymorphism HUMTHO1 has been performed using capillary array electrophoresis and energy-transfer fluorescent dye-labeled polymerase chain reaction primers. Target alleles were amplified by use of primers labeled with one fluorescein at the 5′ end and another fluorescein at the position of the 15th (modified) base to produce fragments that fluoresce in the green (λ_{max} = 525 nm). Unknown alleles were electrophoretically separated together with a standard ladder made up of alleles having 6, 7, 8, and 9 four-base pair repeats, each of which was amplified with an energy-transfer primer having a donor fluorescein at the 5′ end and a rhodamine acceptor at the position of the 7th (modified) base to produce standard fragments fluorescing in the red (>590 nm). Separations were performed on arrays of hollow fused-silica capillaries filled with a replaceable sieving matrix consisting of 0.8% hydroxyethyl cellulose plus 1 μM 9-aminoacridine to enhance the resolution. The labeled DNA fragments were excited at 488 nm, and the fluorescence was detected with a two-color confocal fluorescence scanner. Separations are complete in less than 20 min and allow sizing with an average absolute error or accuracy of less than 0.4 base pair and an average standard deviation of ~0.5 base pair with no correction for mobility shift and cross-talk between the fluorescence channels. This work establishes the feasibility of high-speed, high-throughput STR typing of double-stranded DNA fragments using capillary array electrophoresis.

DNA sequences containing di-, tri-, tetra-, and pentanucleotide repeats are often genetically polymorphic.[1–3] Over 2000 of these short tandem repeat (STR) polymorphisms have been mapped on the human genome,[4] and it is estimated that thousands more remain to be discovered.[5] Because of the abundance of this type of polymorphism and the relative ease of STR detection following amplification by the polymerase chain reaction (PCR), STRs have found widespread use as markers in gene mapping studies[4] and are emerging as potential markers for use in testing for paternity and personal identity.[6–9]

The analysis of STR markers in gene mapping and in the development of population polymorphism data banks is typically performed by slab gel electrophoresis. As the number of interrogated STR markers increases and as one desires to have a more complete polymorphism data base, the speed, throughput, and sample handling associated with slab gels becomes limiting. One approach for increasing the speed and throughput of these electrophoretic separations is through the use of capillary array electrophoresis (CAE).[10–12] In this technique, the speed of electrophoretic separations is increased ~10-fold through the use of high electric fields, and the throughput is increased by using large numbers of capillaries in an array. High sensitivity detection of fluorescently labeled DNA fragments is achieved through the use of a laser-excited, confocal-fluorescence scanner.[13] An additional advantage of CAE is that samples are easily loaded in parallel through standard electrokinetic injection techniques. Thus far, CAE has been used for DNA sequencing[10,11] and for DNA fragment sizing,[14,15] but it has not been used to analyze STRs.

In this paper, we explore the development of methods for rapid STR sizing on capillary arrays. One goal was to determine whether accurate sizing can be performed by separating ds-DNA fragments on capillaries containing easily replaceable nondenaturing hydroxyethyl cellulose (HEC) sieving solutions. If successful, this would dramatically increase the speed of separations and simplify the separation procedures by eliminating the need for denaturing columns. We show here that, by using the appropriate separation buffer and labeling methods, sizing of

[†] Department of Chemistry.

[‡] Performed as a part of the Ph.D. research of Y.W. in the Graduate Group in Biophysics.

[§] Forensic Science Group.

(1) Tautz, D. *Nucleic Acids Res.* **1989**, *17*, 6463–6471.
(2) Edwards, A.; Civitello, A.; Hammond, H. A.; Caskey, C. T. *Am. J. Hum. Genet.* **1991**, *49*, 746–756.
(3) Charlesworth, B.; Sniegowski, P.; Stephan, W. *Nature* **1994**, *371*, 215–220.
(4) Gyapay, G.; Morissette, J.; Vignal, A.; Dib, C.; Fizames, C.; Millasseau, P.; Marc, S.; Bernardi, G.; Lathrop, M.; Weissenbach, J. *Nat. Genet.* **1994**, *7*, 246–339.
(5) Beckmann, J. S.; Weber, J. L. *Genomics* **1992**, *12*, 627–631.
(6) Hammond, H. A.; Jin, L.; Zhong, Y.; Caskey, C. T.; Chakraborty, R. *Am. J. Hum. Genet.* **1994**, *55*, 175–189.
(7) Kimpton, C. P.; Gill, P.; Walton, A.; Urquhart, A.; Millican, E. S.; Adams, M. *PCR Methods Appl.* **1993**, *3*, 13–22.
(8) Frégeau, C. J.; Fourney, R. M. *BioTechniques* **1993**, *15*, 100–119.
(9) Alford, R. L.; Hammond, H. A.; Coto, I.; Caskey, C. T. *Am. J. Hum. Genet.* **1994**, *55*, 190–195.
(10) Huang, X. C.; Quesada, M. A.; Mathies, R. A. *Anal. Chem.* **1992**, *64*, 967–972.
(11) Huang, X. C.; Quesada, M. A.; Mathies, R. A. *Anal. Chem.* **1992**, *64*, 2149–2154.
(12) Mathies, R. A.; Huang, X. C. *Nature (London)* **1992**, *359*, 167–169.
(13) Quesada, M. A.; Rye, H. S.; Gingrich, J. C.; Glazer, A. N.; Mathies, R. A. *BioTechniques* **1991**, *10*, 616–625.
(14) Clark, S. M.; Mathies, R. A. *Anal. Biochem.* **1993**, *215*, 163–170.
(15) Zhu, H.; Clark, S. M.; Benson, S. C.; Rye, H. S.; Glazer, A. N.; Mathies, R. A. *Anal. Chem.* **1994**, *66*, 1941–1948.

alleles of the human tyrosine hydroxylase locus THO1 differing by a single base pair can be achieved. A second goal was to develop methods for two-color labeling and detection so that a standard STR ladder could be electrophoresed along with the unknown amplified alleles. We have accomplished this by amplifying the standard THO1 allelic ladder and the unknown with the energy-transfer (ET) dye-labeled primers recently developed by Ju and co-workers.[16] ET primers are valuable because they absorb strongly at a common laser wavelength while also exhibiting intense and nonoverlapping emissions. The THO1 allelic sizing experiments presented here validate the use of energy-transfer dye-labeled primers together with capillary array electrophoresis for rapid and high-throughput STR sizing.

EXPERIMENTAL SECTION

Instrumentation. Capillary array electrophoresis separations were detected with the laser-excited, confocal-fluorescence scanner as previously described by Huang et al.[10,11] Briefly, excitation light at 488 nm from an argon ion laser is reflected by a long-pass dichroic beam splitter, passed through a 32×, NA 0.4 microscope objective, and brought to a 10 μm diameter focus within the 75 μm i.d. capillaries in the capillary array. The fluorescence is collected by the objective, passed back through the first beam splitter to a second dichroic beam splitter that separates the red (λ > 565 nm) and green (λ < 565 nm) detection channels. The emission is then focused on 400 μm diameter confocal pinholes, spectrally filtered by a 590-nm long-pass filter (red channel) or a 20-nm band-pass filter centered at 520 nm (green channel), followed by photomultiplier detection. The output is preamplified, filtered, digitized, and then stored in an IBM PS/2 computer. A computer-controlled stage is used to translate the capillary array past the optical system at 20 mm/s. The fluorescence is sampled unidirectionally at 1500 Hz/channel. The scanner construction and operation have recently been described in detail.[17] Postacquisition image processing was performed with the programs IPLab, KaleidaGraph, and Canvas.

Capillary Electrophoresis. Polyacrylamide-coated, fused-silica capillaries were prepared using a modification of the procedure described by Hjertén.[18] A 2–3 mm wide detection window was produced by burning off the polyimide coating with a hot wire followed by cleaning the external surface with ethanol. The detection window was placed 25 cm from the injection ends of the 75 μm i.d., 350 μm o.d., 50 cm long fused-silica capillaries (Polymicro Technologies, Phoenix, AZ). The inner walls of the capillaries were incubated with 1 N NaOH for 30 min at room temperature, followed by rinsing with deionized water. The capillaries were then treated overnight at room temperature with γ-methacryloxypropyltrimethoxysilane (1:250 dilution with H_2O adjusted to pH 3.5 with acetic acid) to derivatize the walls for polyacrylamide binding. Freshly made 4% *T* acrylamide solution in $^1/_2\times$ TBE buffer (45 mM tris, 45 mM boric acid, 1 mM EDTA, pH 8.3) was filtered with a 0.2-μm syringe filter and degassed under vacuum for 30 min. *N,N,N',N'*-Tetramethylethylenediamine (TEMED; 1 μL) and 10 μL of 10% ammonium persulfate (APS) solution were added to 1 mL of gel solution. The solution was immediately forced into the capillary with a 100-μL syringe. After 30 min, the polyacrylamide solution was flushed out with deionized water and capillaries were filled with buffer consisting of hydroxyethyl cellulose (HEC; M_n = 438 000, Aqualon Co. Hopewell, VA) dissolved in $^1/_2\times$ TBE. The separation buffer was prepared by adding 0.8 g of HEC to 100 mL of $^1/_2\times$ TBE and dissolved by stirring overnight at room temperature.[14] The HEC buffer was degassed under vacuum for 30 min, centrifuged for 20 min on a tabletop centrifuge, and drawn into a 100-μL syringe, and 3 μL was injected into each capillary. Capillaries were prerun at 80 V/cm for 5 min before each experiment. Diluted and deionized PCR samples were injected by inserting the capillary in a 5-μL sample volume held in an Eppendorf tube followed by electrokinetic injection (80 V/cm for 3 s). After injection, the sample tubes were replaced with tubes containing 0.8% HEC plus $^1/_2\times$ TBE buffer. Electrophoresis was performed at 80 V/cm using five-capillary arrays held at ambient temperature (22 °C). The low (80 V/cm) electrophoresis voltage was used to avoid undersampling of the bands with our current detection system, which is limited to 1-Hz scan rates. When the experiments were complete, capillaries were flushed with water and then with methanol, followed by drying. Our coated capillaries could be refilled 20–25 times before the quality of the separations deteriorated. Methods for the further extension of the lifetime of capillary columns have been described.[19]

PCR Amplification of THO1 Loci. DNA was isolated from blood by using standard methods.[20] The human tyrosine hydroxylase locus HUMTHO1, chromosomal location 11p15.5, contains a polymorphic four-base STR sequence (AATG) in intron 1.[21] PCR amplification of this polymorphic region produces allelic fragments designated 5–11, according to the number of AATG repeats; an additional allele designated 9.3 differs from allele 10 by a single base deletion. The primer sequences used for PCR are 5′-ATTCAAAGGGTATCTGGGCTCTGG-3′ (THO1-A) and 5′-GTGGGCTGAAAAGCTCCCGATTAT-3′ (THO1-B).[2] PCR amplifications were performed in 50-μL volumes by using 10 ng of genomic DNA template, 0.5 μM of each primer, 5 units of Taq DNA polymerase, 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl at pH 8.3, and 200 μM dNTPs (final concentrations indicated). The PCR cycle protocol using a Perkin-Elmer Cetus Model 480 was as follows: (1) melting at 95 °C for 5 min; (2) 30 cycles of 95 °C for 1 min, 58 °C for 1 min, and 72 °C for 1 min; (3) 72 °C for 7 min to complete extention. The PCR sample was then dialyzed for 30 min by pipeting 8–10 μL onto a 0.10-μm VCWP membrane filter (Millipore, Bedford, MA), which was floated on deionized water in a beaker held at 4 °C. Dialysis was used to remove salts which can interfere with sample injection. Following dialysis, the samples were diluted with deionized water 100–1000 times (depending on product concentration) before electrokinetic injection. The amplifications with fluorescently labeled primers were also performed as described above. Initially, four sets of fluorescent primers were used for PCR amplification and the mobility shifts of the products were evaluated with capillary electrophoresis. For these mobility shift experiments, 1–2 ng of unlabeled PCR product was reamplified by 20 PCR cycles using the

(16) Ju, J.; Ruan, C.; Fuller, C. W.; Glazer, A. N.; Mathies, R. A. *Proc. Natl. Acad. Sci. U.S.A.*, in press.
(17) Mathies, R. A.; Scherer, J. R.; Quesada, M. A.; Rye, H. S.; Glazer, A. N. *Rev. Sci. Instrum.* **1994**, *65*, 807–812.
(18) Hjertén, S. *J. Chromatogr.* **1985**, *347*, 191–198.
(19) Schmalzing, D.; Piggee, C. A.; Foret, F.; Carrilho, E.; Karger, B. L. *J. Chromatogr.* **1993**, *652*, 149–159.
(20) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular cloning: a laboratory manual*, 2nd ed.; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.
(21) Puers, C.; Hammond, H. A.; Jin, L.; Caskey, C. T.; Schumm, J. W. *Am. J. Hum. Genet.* **1993**, *53*, 953–958.

Chart 1. Structures of the Six PCR Primers Used for the Amplification of the HUMTHO1 Loci[a]

F10F FAM-5'-ATTCAAAGGGT*ATCTGGGCTCTGG-3'
(CH)$_2$(CO)–NH–(CH$_2$)$_6$–NH–**FAM**

F14F FAM-5'-GTGGGCTGAAAAGCT*CCCGATTAT-3'
(CH)$_2$(CO)–NH–(CH$_2$)$_6$–NH–**FAM**

F2R FAM-5'-ATT*CAAAGGGTATCTGGGCTCTGG-3'
(CH)$_2$(CO)–NH–(CH$_2$)$_6$–NH–**ROX**

F6R FAM-5'-GTGGGCT*GAAAAGCTCCCGATTAT-3'
(CH)$_2$(CO)–NH–(CH$_2$)$_6$–NH–**ROX**

THO1-A 5'-ATTCAAAGGGTATCTGGGCTCTGG-3'

THO1-B 5'-GTGGGCTGAAAAGCTCCCGATTAT-3'

[a] The fluorescent primers are labeled with a common fluorescein donor (F) at the 5′ end and either a second fluorescein or a rhodamine (R) acceptor at the indicated locations of a modified T in the sequence. The number of nucleotides between the two fluorophores is indicated in the primer designation

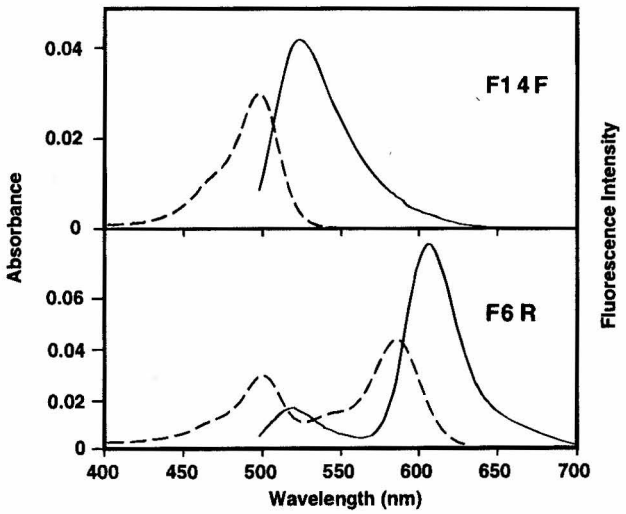


Figure 1. Absorption (- - -) and fluorescence emission (—) spectra of the fluorescently labeled THO1 primers F14F and F6R measured in 1× TBE. F14F exhibits strong absorption at ~488 nm and intense fluorescence emission with a maximum at 525 nm. F6R also exhibits intense absorption at ~488 nm, but the maximum emission is shifted out to ~600 nm.

appropriate fluorescent primers. THO1 types for all samples used in this study were independently determined by analysis on slab gels essentially as described by Puers.[21] Standard reference alleles were determined by sequence analysis (M. Savill and G. Sensabaugh, unpublished).

Design and Synthesis of PCR Primers. Chemicals were purchased from Applied Biosystems (Foster City, CA). Oligodeoxynucleotides were synthesized by the phosphoramidite method on an Applied Biosystems 392 DNA synthesizer. The structures of two unlabeled (THO1-A and THO1-B) and four energy-transfer dye-labeled PCR primers are presented in Chart 1. The nomenclature and procedures for preparing and purifying energy-transfer fluorescent primers have been described by Ju et al.[16] The F10F and F14F primers required for the studies reported here were synthesized with the incorporation of two 5-carboxyfluorescein (FAM) fluorophores at the locations indicated in Chart 1. The F2R and F6R primers were similarly prepared with the incorporation of FAM as a donor and 6-carboxy-X-rhodamine (ROX) as an acceptor.

The energy-transfer dye-labeled primers are advantageous for two-color fragment sizing because the 488-nm exciting light is optimally absorbed by the FAM chromophore in these primers followed by enhanced emission at the FAM wavelength in the case of F10F and F14F or very distinctively Stokes-shifted emission following energy transfer in the case of F2R and F6R. Absorption spectra of the primers were measured on a Perkin-Elmer Lambda 6 UV–visible spectrophotometer and fluorescence emission spectra were taken on a Perkin-Elmer Model MPF 44B spectrofluorimeter. As representative examples, spectra of F14F and the energy-transfer dye-labeled primer F6R are presented in Figure 1. F14F exhibits strong absorption at ~488 nm and intense fluorescence emission with a maximum at 525 nm. F6R also exhibits intense absorption at 488 nm, but because of the fluorescein-to-rhodamine fluorescence energy transfer, the emission maximum is shifted out to ~600 nm. Primers were dissolved in 10 mM Tris-HCl, 1 mM EDTA buffer at a final concentration of 10 pmol/μL for PCR reactions.

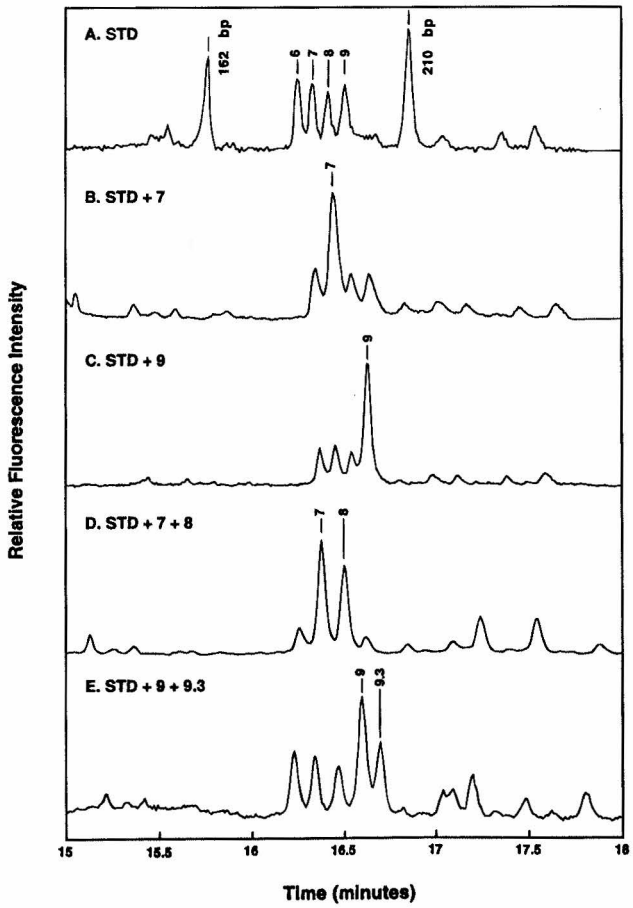


Figure 2. Electropherograms of (A) the THO1 standard ladder consisting of the 6, 7, 8, and 9 alleles which have been mixed and then coinjected with a ΦX-174 RF DNA–*Hinc*II digest. (B) Standard allelic ladder spiked with "unknown" allele 7. (C) Standard allelic ladder spiked with allele 9. (D) Standard allelic ladder spiked with alleles 7 and 8. (E) Standard allelic ladder spiked with alleles 9 and 9.3. These separations were run with 0.8% HEC, ½× TBE, and 1 μM thiazole orange in the running buffer and detected in the green channel. The additional weak peaks appearing at detection times greater than ~17 min are due to heteroduplex formation.

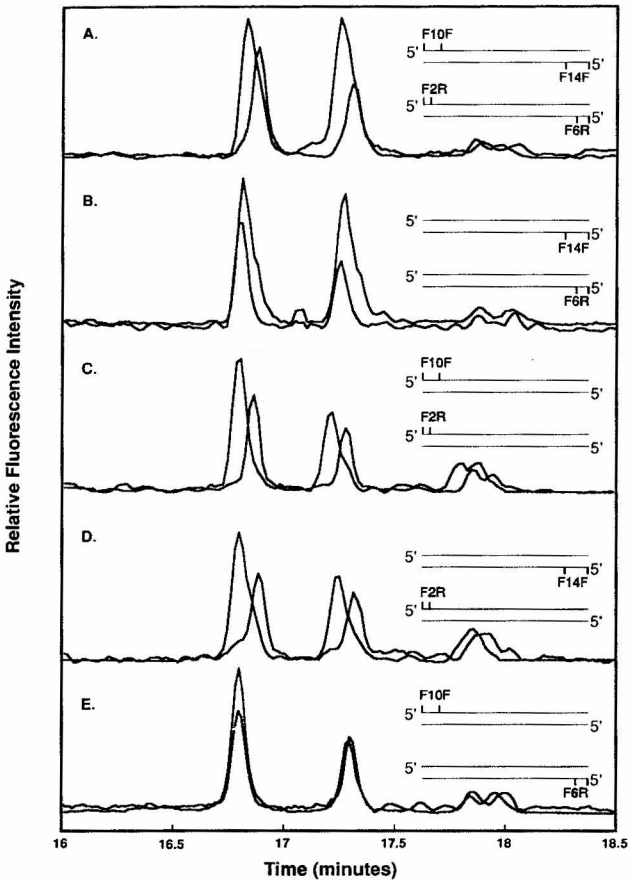


Figure 3. Comparison of the mobility shift using five different methods for attaching energy-transfer-coupled fluorophores to THO1 target alleles 6 and 9.3. The green line indicates the fluorescence intensity in the green channel, and the red line indicates the intensity in the red channel. The structures of the primer sets used are indicated. Electrophoresis was performed using 0.8% HEC, $^1/_2\times$ TBE, and 1 μM 9-AA in the running buffer.

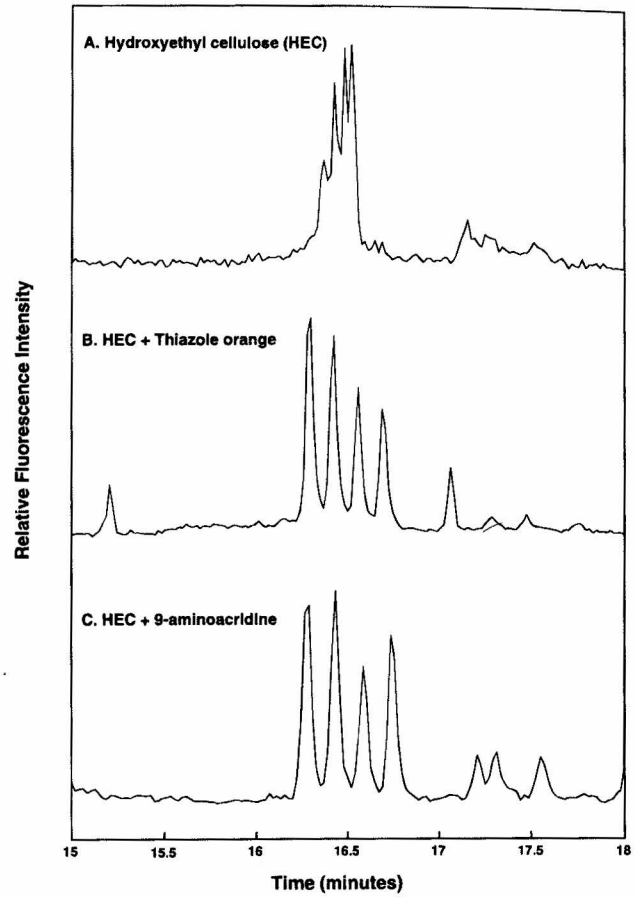


Figure 4. Electropherograms of standard THO1 allelic ladder separations using three different running buffers. (A) 0.8% HEC plus $^1/_2\times$ TBE. (B) 0.8% HEC plus $^1/_2\times$ TBE and 1 μM thiazole orange. The additional peaks at 15.2 and 17.1 min in this run are due to ΦX-174 RF DNA–*Hinc*II DNA. (C) 0.8% HEC plus $^1/_2\times$ TBE and 1 μM 9-AA. The standard ladder was amplified with F6R and detected in the red channel.

RESULTS AND DISCUSSION

One-Color Sizing of THO1 Alleles. Figure 2 presents capillary array electrophoresis separations of the standard THO1 ladder that has been labeled on-column with the intercalating fluorophore, thiazole orange (TO).[22] To form this standard ladder, two individual heterozygote samples containing alleles 6 + 9 and 7 + 8 were amplified and mixed. Trace A presents the electropherogram of the allelic ladder which was injected along with a ΦX-174 RF DNA–*Hinc*II restriction digest (Pharmacia, Piscataway, NJ) as a control. The four alleles which are 4 bp apart are nearly baseline-resolved with a separation time of less than 17 min. In trace B, an "unknown" PCR-amplified sample is added to the allelic ladder and coinjected. The increased intensity of allele 7 identifies the unknown allele. Traces C–E in Figure 2 similarly demonstrate the detection of samples containing alleles 9, 7 + 8, and 9 + 9.3. This work demonstrates the feasibility of producing the THO1 ladder for allelic typing with capillary array electrophoresis using nondenaturing HEC separation matrices. The use of a standard allelic ladder to perform accurate sizing has previously been reported.[2, 21] High-resolution separation of ds-DNA fragments on capillary columns has also been demonstrated

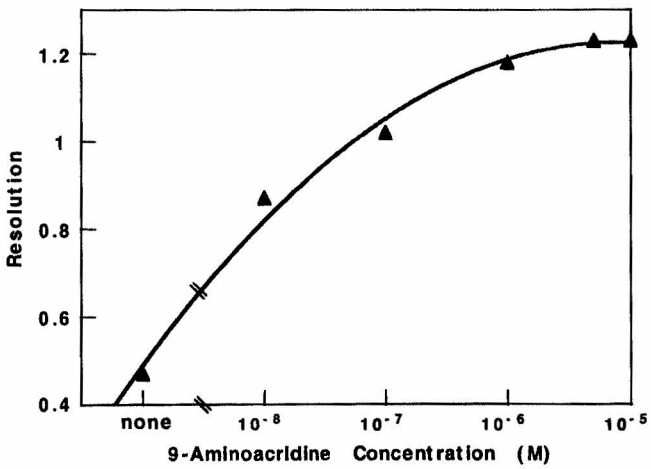


Figure 5. Resolution of the THO1 ladder as a function of 9-aminoacridine concentration. The resolution of last two adjacent peaks (8, 9) in the THO1 allelic ladder was calculated using the equation $R = (T_2 - T_1) \times 1.18/(\text{fwhh}_1 + \text{fwhh}_2)$, where T and fwhh are the peak migration time and full width at half height, respectively. The THO1 ladder was amplified with F6R as described in the text and separated on 0.8% HEC at 80 V/cm.

by McCord and co-workers.[23,24] However, for one-color labeling of the allelic ladder and unknown to work reliably and precisely, it is necessary to have good control over the concentration of the

(22) Rye, H. S.; Quesada, M. A.; Peck, K.; Mathies, R. A.; Glazer, A. N. *Nucleic Acids Res.* **1991**, *19*, 327–333.

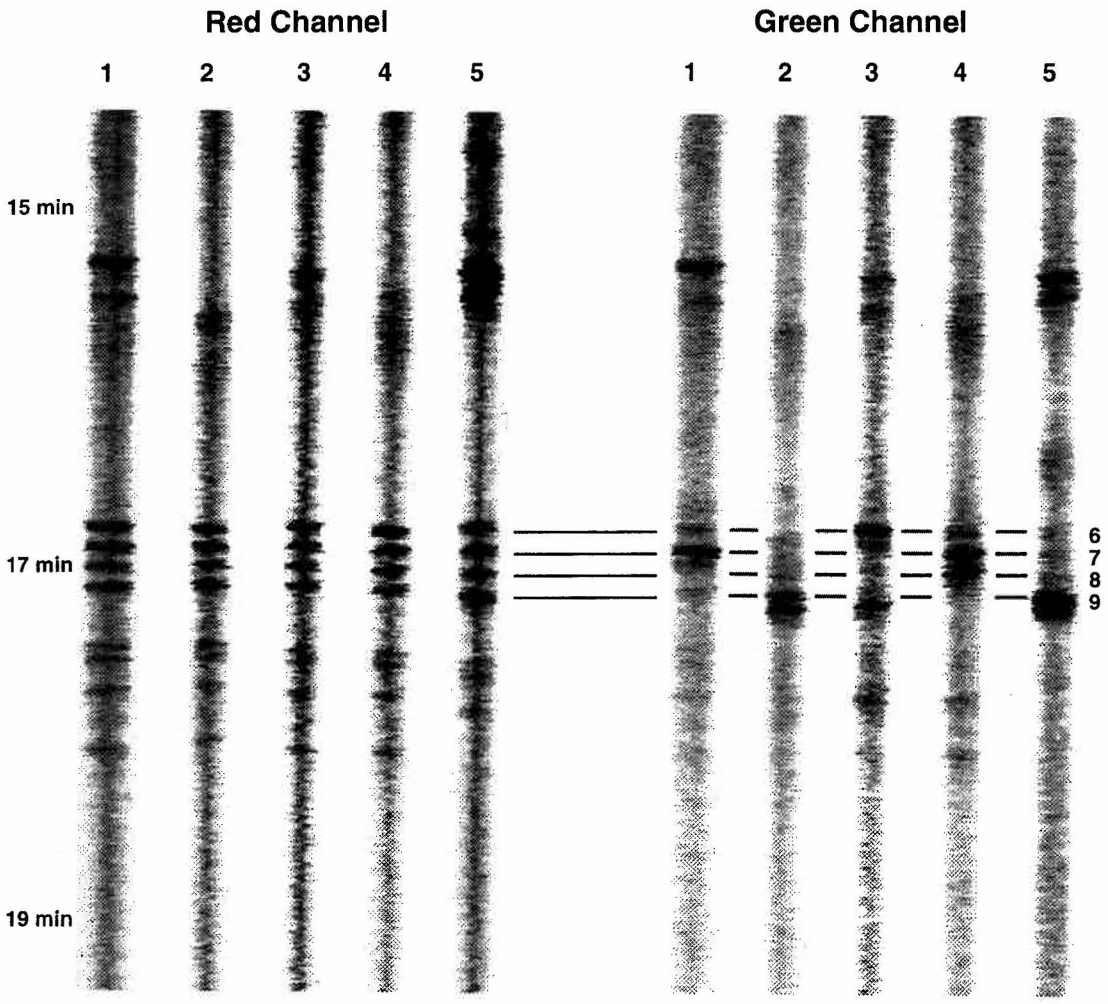


Figure 6. Images of the fluorescence from a five-capillary array separation of THO1 alleles. The left image presents the fluorescence signal as a function of time detected in the red (>590 nm) channel, and the right image presents the fluorescence signal from the green (λ_{max} = 525 nm) channel. The standard THO1 allelic ladder (6 + 7 + 8 + 9) was amplified with the red-emitting ET primer F6R and detected in the red channel; unknown alleles were amplified with the green-emitting primer F14F and detected in the green channel. These images have been adjusted for the 1–2% capillary-to-capillary variance in mobility by shifting the time axes so that the allelic ladder is detected at the same time in all capillaries. These separations were performed with 0.8% HEC and 1 μM 9-AA in the running buffer at 80 V/cm.

DNA in the unknown samples. Alternatively, one could use a nonallelic ladder as the standard but there could be some loss of accuracy in the resulting interpolation.

Evaluation of Energy-Transfer Primer Labeling. For routine sizing experiments, it is desirable to perform two-color detection where the allelic standards are amplified with one fluorescent primer and the unknowns are amplified with a second fluorescent primer having a distinctive emission. Ju and coworkers[16] recently reported the synthesis and use of a new class of fluorescent primers labeled with pairs of dye molecules that are coupled by fluorescence energy transfer. These ET primers have the advantage of providing strong absorption at a common laser excitation wavelength (488 nm). Following the fluorescence energy transfer, the ET primers emit at a Stokes-shifted wavelength determined by the properties of the acceptor. Thus, the fluorescence emission of the ET primers is very intense, and the emission spectra of the different ET dye-labeled primers are distinctively Stokes-shifted. These primers have been shown to provide 2–6 times the signal strength compared to conventional single dye-labeled fluorescent primers in DNA sequencing applications.[16] Furthermore, the mobility shift of DNA fragments generated with ET primers depends on the spacing between the dyes. We therefore performed experiments to evaluate the mobility shift of the amplified fragments for all combinations of singly and doubly labeled targets.

In trace A of Figure 3, both strands of the amplified 6 and 9.3 targets have been labeled. In one case, the F10F primer is extended to form the (+) strand and the F14F primer is extended to form the (−) strand producing the green-emitting fragments. In the second case, the ET dye-labeled primer F2R is extended to produce the (+) strand and F6R is used to extend the (−) strand producing the red-emitting fragments. These fragments were mixed and electrophoresed on HEC-filled capillaries in the presence of 1 μM 9-aminoacridine (9AA, see below). The mobility shift between the green fragments and the red fragments in trace A was found to be ~2 bp. We therefore decided to evaluate amplifying with just one fluorescent primer per ds-DNA fragment to see if a particular combination of labels would reduce the mobility shift. Trace B in Figure 3 shows that amplifying the (−)

(23) McCord, B. R.; McClure, D. L.; Jung, J. M. *J. Chromatogr. A* **1993**, *652*, 75–82.
(24) McCord, B. R.; Jung, J. M.; Holleran, E. A. *J. Liq. Chromatogr.* **1993**, *16*, 1963–1981.

strand with the F14F primer or with the F6R primer generates fragments having almost no mobility shift (≤0.3 bp) between the green- and the red-labeled fragments. The two other labeling methods shown in traces C and D resulted in larger mobility shifts. However, amplifying the (−) strand with F6R and the (+) strand with F10F also produced fragments having almost no mobility shift (trace E). We decided to perform the subsequent experiments using the labeling method illustrated in trace B because these fragments are labeled on the same strand and can thus also be sized under denaturing conditions if necessary.

Resolution Enhancement with 9-Aminoacridine. To achieve satisfactory resolution of the THO1 allelic ladder, we found it necessary to include an intercalating dye in the running buffer. Trace A of Figure 4 shows that poor resolution is obtained when the allelic ladder (amplified with the F6R primer) is run in 0.8% HEC alone. Trace B shows the separation of the same ladder in 0.8% HEC plus 1 μM of the intercalating dye thiazole orange. The resolution of the separation in the presence of TO is dramatically enhanced. Electrophoresis in the presence of the intercalator ethidium bromide has also been shown to improve the electrophoretic resolution of ds-DNA.[25,26] Unfortunately, TO contributes to the signal in the green channel of our two-color detection system, rendering it unsuitable for use in the desired two-color labeling scheme. It is thus necessary to use a nonfluorescent intercalator to improve the resolution. In electrophoretic separations of preformed dimeric dye–DNA complexes, Zhu et al.[15] observed that the addition of the nonfluorescent dye 9-aminoacridine (9-AA) can be used to dramatically improve ds-DNA separations much like TO and ethidium. We therefore decided to evaluate the effect of 9-AA. Trace C of Figure 4 presents a separation of the THO1 allelic ladder when the column is filled with 0.8% HEC and 1 μM 9-AA. This separation is as good as that obtained in the presence of TO. The dependence of the resolution of the THO1 ladder separation on 9-AA concentration is presented in Figure 5. The resolution improves significantly up to 1 μM 9-AA and is only slightly better at 5 μM. 9-AA concentrations above 50 μM were found to quench the fluorescence.

Two-Color THO1 Sizing with Capillary Array Electrophoresis. Figure 6 presents the results of a typical THO1 sizing experiment performed using two-color capillary array electrophoresis. The standard allelic ladder was amplified with F6R and detected in the red channel, while the unknown alleles were amplified with F14F and detected in the green channel. The signal detected as a function of time is presented as two images: the left image is the signal in the red channel and the right image is that detected in the green channel. The alleles appear ~17 min after injection. The signal in the red channel is predominantly from the red-labeled standard ladder, and the expected four band patterns are seen in all capillaries. The unknown, amplified with the green-emitting primer, is detected in the green channel. Allelic bands are identified as the intense green bands coinciding in mobility with allelic ladder bands in the red channel. To illustrate, the right image of Figure 6 reveals intense green bands corresponding to allele 7 in lane 1, alleles 9 + 9.3 in lane 2, alleles 6 + 9.3 in lane 3, alleles 7 + 8 in lane 4, and allele 9 in lane 5. Figure 7 presents electropherograms derived from the image in Figure 6. Very clear discrimination is observed between the green-labeled fragments and the red-labeled fragments with almost no cross-talk between the green and red channels. In some cases, the amplification products of heterozygote samples contain heteroduplex bands that migrate behind the allelic ladder. Additional weak "noise" bands may represent nontemplated base addition[7] known to occur with PCR amplification of THO1 (M. Savill and G. Sensabaugh, unpublished data).

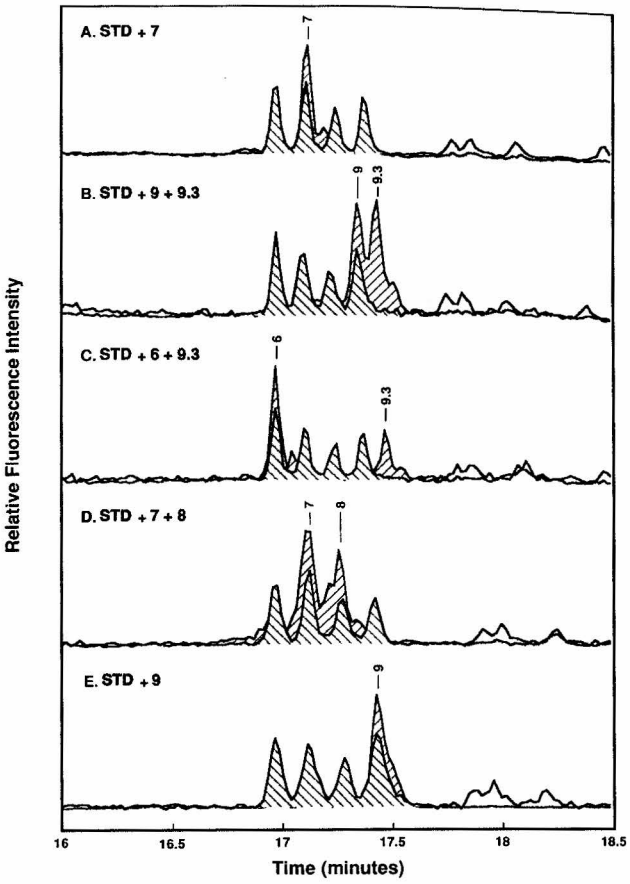


Figure 7. Electropherograms of the THO1 fragment sizing separations presented in Figure 6. The green signal is from the unknown alleles, and the red signal is from the standard THO1 ladder. Traces A–E correspond to lanes 1–5 in Figure 6.

The accuracy and precision of allelic sizing using CAE was tested by performing multiple runs on 11 different samples. These results are summarized in Table 1. Since a linear relationship exists between molecular weight and migration time in the region of interest, the allelic ladder was used with a linear regression analysis to size the unknown fragments. The calculated sizes of unknown alleles are compared to true sizes based on sequence analysis and verified by denaturing polyacrylamide gel electrophoresis. The average absolute difference of the determined allele sizes from the true allele sizes was 0.41, and over 70% of the determined values were within 0.5 bp of the true value. The reproducibility is excellent (relative standard deviation less than 0.4% for each allele) and there was no ambiguity in allele assignments. Two alleles, 9.3 and 10, which differ by a single base pair deletion cannot be electrophoretically resolved when paired but can be correctly assigned when in combination with any other allele. It should, however, be possible to separate these two fragments on columns containing higher concentrations of HEC (J. Bashkin and R. Johnston, personal communication).

(25) Schwartz, H. E.; Ulfelder, K.; Sunzeri, F. J.; Busch, M. P.; Brownlee, R. G. *J. Chromatogr.* **1991**, *559*, 267–283.
(26) Guttman, A.; Cooke, N. *Anal. Chem.* **1991**, *63*, 2038–2042.

Table 1. Statistical Analysis of THO1 Fragment Sizing[a]

allele	length (bp)	no. of detns	mean size[b]	SD[c] (%)
6	183	6	183.1	0.61 (0.33)
7	187	23	187.0	0.41 (0.22)
8	191	15	191.2	0.69 (0.36)
9	195	11	195.4	0.60 (0.31)
9.3	198	8	198.3	0.52 (0.26)
10	199	6	199.0	0.31 (0.15)

[a] Eleven different amplified samples (7, 8), (6, 9.3), (9, 9.3), (6, 9), (7, 8), (8, 9), (7, 9.3), (7, 10), (6, 9.3), (7), and (9) were run 8, 2, 4, 3, 5, 2, 1, 6, 1, 3, and 2 times, respectively. [b] Mean PCR product size as determined by linear regression using the allelic ladder as the sizing standard. [c] Standard deviation (SD) in terms of base pairs for the indicated number of determinations. The percent relative SD is given in parentheses.

CONCLUSIONS

The purpose of this work was to develop methods for high-speed, high-throughput sizing of short tandem repeats using capillary array electrophoresis. This can be broken down into the steps of (1) developing methods for achieving single-base (or near single base) resolution of ds-DNA fragments using replaceable separation matrices and (2) developing methods for multiplex labeling to achieve on-column sizing of unknown STRs in the presence of an allelic standard. We demonstrate here that these objectives can be achieved using ET primers in combination with CAE on HEC capillary supports containing 9-AA. The advantages of the method presented here are that ET primers provide higher sensitivity and better color discrimination than that possible with conventional single dye-labeled fluorescent PCR primers. Second, the HEC solutions used as the sieving media provide high speed (<20 min) separations on columns that can be easily refilled and rerun. Finally, the CAE format provides the ability to run 50 or more separations in parallel.[27] The amplification of multiple STR targets with different fragment lengths in the 100–500-bp size range plus the use of up to four different ET primers[16] should provide an additional increase in the throughput of STR typing.

(27) Bashkin, J. S.; Roach, D.; Rosengaus, E.; Barker, D. L. Human Genome Program, Contractor-Grantee Workshop III, U.S. Department of Energy 1993; p 116.

ACKNOWLEDGMENT

We thank the members of the Berkeley High-Sensitivity DNA Analysis Group for their support, Alexander N. Glazer and Steven M. Clark for valuable discussions, and Alan Taur for expert preparation of the figures. This research was supported by a grant from the Director, Office of Energy Research, Office of Health and Environmental Research of the U.S. Department of Energy under Contract DE-FG-91ER61125 to R.A.M. and by a grant from the National Institute of Justice (93-IJ-EX-0010) to G.F.S. J.J. was supported by a Human Genome Distinguished Postdoctoral Fellowship sponsored by the U.S. Department of Energy, Office of Health and Environment Research, and administrated by the Oak Ridge Institute for Science and Education. Additional support from the Amersham Life Science Corp. is also gratefully acknowledged.

Received for review November 10, 1994. Accepted January 13, 1995.[⊗]

AC941100O

[⊗] Abstract published in *Advance ACS Abstracts*, February 15, 1995.

Vertical Compatibility of Short Tandem Repeat Polymorphic Typing Using Energy Transfer Fluorescent Primers

Jeanette M. Atherton Wallin[1], Brooke A. Carpenter[3], Yiwen Wang[2], Jingyue Ju[2], Richard A. Mathies[2], and George F. Sensabaugh[1]

We have previously demonstrated the application of fluorescent dye-labeled energy-transfer (ET) primers to short tandem repeat polymorphism (STR) sizing using capillary array electrophoresis (CAE).[1] CAE offers rapid high-throughput typing on STRs but may not be readily available to smaller laboratories. In this presentation, we demonstrate the advantages of ET primers in STR typing on conventional slab gel typing systems such as those currently used in many forensic laboratories.

We have employed the STR locus HUMTHO1[2] as our standard since it is the model STR used by many other forensic laboratories. The primer sequences described by Edwards et al.[3] were used, producing PCR fragments of 179-203 base pairs in length. Two ET dye-labeled reverse primers were used. Both have a fluorescein derivative, FAM, covalently bonded to the 5' end which serves as the common fluorescence donor. One reverse primer has another molecule of FAM covalently bonded to a modified thymidine residue in the fifteenth position from the 5' end (the F14F primer). The other primer has a molecule of ROX, a rhodamine derivative, covalently bonded to a modified thymidine residue in the seventh position from the 5' end (the F6R primer). The forward primer is unlabeled. For a more complete explanation of ET labeling and mechanism, refer to Ju et al.[4,1]

The absorption-emission spectrum of F14F shows maximum absorption at approximately 496 nm and maximum emission in the green at 525 nm. F6R shows absorption at 496 nm and Stokes-shifted emission in the red at 605 nm, with negligible residual emission at 525 nm. The distinctly separate emission peaks of F14F and F6R allow two-color detection with a single argon ion laser (488 nm) as an excitation source. The ET primers provide 2-8 times more signal than single FAM and ROX labeled primers when excited with a single laser line at 488 nm, providing more sensitivity for detection.

Our work using ET primers on CAE has been described.[1] Briefly, the two-color CAE instrument used has a standard CE set-up with an argon ion laser, stationary confocal optics, a scan stage allowing use of multiple capillaries, and two photomultiplier tubes allowing simultaneous two-color detection.[5] CAE separations were performed under native conditions using 0.8% of the replaceable sieving matrix hydroxyethylcellulose and 1-2 mM 9-aminoacridine in 0.5x TBE. Separations take approximately 15 minutes using 100 V/cm. Electrophoretic migration of identical F14F and F6R products show nearly the same mobility with an average standard deviation of approximately 0.5 base pairs. For THO1 typing, we created an allelic ladder using F6R. Unknowns were amplified with F14F and sized by comparison with the ladder.

For the present study, we investigated STR analysis on conventional slab gels using three methods of detection: ET dye-labeled PCR products, post-staining with SYBR Green, and silver staining. The first two detection modes entailed the use of a FluorImager 575, kindly loaned to us by Molecular Dynamics Inc.

The ET dye-labeled PCR products are both detectable on the FluorImager, which excites with a 488 nm argon ion laser. F14F is optimally detected using a 530DF30 band pass filter and F6R is optimally detected using a 610EFLP long pass filter; there was no significant crosstalk between the two channels. SYBR Green products were detected using a 530DF30 bandpass filter.

Separations of THO1 PCR products, amplified with ET primers, were performed on a 6% polyacrylamide gel run for 2.25 hours at 40W. Near single base resolution was achieved. Alleles 9.3 and 10 could be distinguished when run in adjacent lanes but could not be resolved in a 9.3/10 heterozygote, unless using 2 color detection.

An advantage of the two color system is that the red F6R ladder and green F14F sample may be placed together in the same lane. In a one color detection system, every few lanes on the gel must contain a ladder. This limits the number of samples that may be loaded on the gel. To increase throughput, a two color system allows every lane to be used for samples. This also serves to provide more accurate sizing of unknown samples since lane-to-lane variation (e.g. smiling) is eliminated.

To compare relative sensitivities of the various detection modes, analyses were performed on 1:3 serial dilution series made with F14F and F6R THO1 PCR products. The CAE could detect F14F PCR products at a 1:2187 dilution with resolution well above baseline; F6R PCR products were detected at a 1:243 dilution. Therefore, F14F is the more sensitive label with CAE detection. With the FluorImager, both F14F and F6R PCR products could be detected at a 1:27 dilution. This same gel was then SYBR Green post-stained. When denaturing gels are stained with SYBR Green, alleles are seen as band doublets due to the staining of both strands. The samples could still be detected down to 1:27 with the SYBR Green staining. Another gel was run with the same PCR product dilutions, and then silver stained. The samples could only be detected down to a 1:9 dilution (Table 1).

Since use of the ET Primers

[1]Forensic Science Group, School of Public Health and [2]Department of Chemistry, University of California, Berkeley, CA 94720, [3]Santa Clara Crime Lab.

provides increased sensitivity, a reduction in the number of PCR cycles was studied. Fewer PCR cycles means less time, a better signal-to-noise ratio, and less manipulation of the sample. We began with the traditional 30 cycles and then tried 28, 26, 24, 22, and 20 cycles. Two different samples, a THO1 7,8 and a THO1 6,9, were amplified with F14F at these various cycles. On the CAE, both samples were seen well above baseline at 20 cycles. The FluorImager could detect samples well above background that had been amplified with as few as 26 cycles; with manipulation of the FluorImager software, there were slight indications of detected product cycled 24 times. This same gel was then SYBR Green stained, and as few as 26 cycles could be detected (Table 2).

Conclusions

This research has shown that ET primers may be used with both CAE and FluorImager detection, thereby demonstrating the vertical compatibility of this technology between detection modes. It was also demonstrated that in using ET primers, the CAE instrument is more sensitive than the FluorImager. The CAE can detect a larger dilution of PCR product and product cycled fewer times. Thus, in cases when only a limited amount of a forensic sample is available, less sample may be used to generate PCR product and less product may be used for STR sizing.

ET primers and SYBR Green staining are both more sensitive than silver staining. The ET primers offer about the same level of sensitivity as SYBR Green staining on the FluorImager instrument. However, there are several advantages in using ET primers with the FluorImager versus staining with SYBR Green. First of all, no post staining is required; this saves time, avoids increased background signal, and eliminates the need for disposal of SYBR Green waste. Second, two-color detection is available at one excitation wavelength, saving space on gels and requiring only one laser line. Thirdly, because only one strand is labeled, each allele is represented by a single band rather than by a doublet as seen with SYBR Green staining and silver staining.

References

1. Wang, Y., J. Ju, B. A. Carpenter, J. M. Atherton, G. F. Sensabaugh, R. A. Mathies. 1995 *Anal. Chem.* 67;1197-1203.
2. Puers, C., H. A. Hammond, L. Jin, C. T. Caskey, J. W. Schumm. 1993 *Am. J. Hum. Genet.* 53;953-958.
3. Edwards, A., A. Civitello, H. A. Hammond, C. T. Caskey. 1991 *Am. J. Hum. Genet.* 49;746-756.
4. Ju, J., R. Chihchuan, C. Fuller, A. N. Glazer, R. A. Mathies. 1995 *Proc. Natl. Acad. Sci.* 92;4347-4351.
5. Huang, X. C., M. A. Quesada, R. A. Mathies. 1992 *Anal. Chem.* 64;967-972.

Acknowledgements

We would like to thank Dick Rubin, Elaine Mansfield, Marylinn Munson, and Jennifer Worley of Molecular Dynamics Inc. for their generous loan of the FluorImager 575 and their assistance in its use.

This article was voted best paper at the Spring 1995 CAC Seminar.
—Ed.

Table 1
RELATIVE DETECTION LIMITS OF ET DYE-LABELED THO1 PCR PRODUCTS

Method	Sensitivity
ET Primers on CAE	
Green (F14F)	1 : 2187
Red (F6R)	1 : 243
ET Primers on FluorImager	
Green (F14F)	1 : 27
Red (F6R)	1 : 27
SYBR Green Stain on FluorImager	1 : 27
Silver Stain	1 : 9

Table 2
DETECTION LIMITS OF THO1 PCR PRODUCTS AS A FUNCTION OF CYCLE NUMBER

Method	Minimum Cycle Number
ET Primers on CAE Green (F14F)	< 20
ET Primers on FluorImager Green (F14F)	<=26
SYBR Green on FluorImager	26

bands for heterozygotes in Hae III RFLP analysis (particularly at locus D10S28). These 'extra bands' usually come in pairs, are reproducible and consistent for a given sample and are always smaller than the primary band(s). These results suggest that this phenomenon is not due to incomplete enzyme digestion (partials), mixtures of samples, or 'extra bands' due to unstripped probes from previous hybridizations.

This phenomenon has been observed for control DNA samples (K562 cell line) and DNA extracted from blood or semen stains. The 'extra bands' produced are not equimolar with the primary bands (usually less hybridization signal is detected for the 'extra bands').

We are evaluating the possibility that these are the result of a relaxation of the Hae III endonuclease specificity for the sequence GGCC to include similar but specific sequences. Purposeful overdigestion (with orders of magnitude greater enzyme concentration than our current protocol) can produce similar 'extra bands' patterns, but because this was not a systematic result for every set of samples, we believe there may be other factors involved.

We have also observed that DNA extracted from reference blood samples stored as liquid in preservatives (as opposed to stored as a stain) produce 'shadow bands' that are predictably detected just below the primary diagnostic band. This band often appears around 150 base pairs below the diagnostic band, again primarily with D10S28. This band is often unresolvable from the diagnostic band, especially in longer autorad exposures. It is worth noting that the size of this difference is approximately the same as the amount of DNA wound around nucleosomes, which are known to be hypersensitive sites.

Rapid detection of genetic types at polymorphic short tandem repeat (STR) loci using two-color capillary array electrophoresis

Y Wang, J Ju and R Mathies, Department of Chemistry, School of Public Health, University of California, Berkeley, CA 94720

B Carpenter, J Atherton, and George Sensabaugh, Forensic Science Group, School of Public Health, University of California, Berkeley, CA 94720

The implementation of DNA profile databanks will require a DNA typing technology and information management system that can accommodate the typing of many hundreds of samples per day. This requirement defines to some extent the markers chosen to establish the database. With this consideration in mind, we have investigated the use of capillary array electrophoresis (CAE) for the high throughout typing of polymorphic short tandem repeat (STR) loci. We demonstrate here typing at the THO1 locus using a two-color labelling system in which the allelic ladder is labelled with one fluorescent dye label and the test sample is labelled with a second. Labelled PCR products are detected using a laser excited confocal fluorescence scanning system. Near baseline resolution is achieved for the 4-bp repeat allelic products in 20 minute runs using capillaries filled with 0·8% (w/v hydroxyethyl-cellulose (BEC), 0·5X TBE pH 9·3, and 1uM 9-aminoacridine. The 9·3 and 10 alleles, which differ by a single base, can be distinguished in isolation but could not be resolved in a 9·3,10 heterozygote. With this excepti genetic typing at the THO1 locus is unambiguous; typi results can be translated by software directly into datab files. A very large number of polymorphic STR loci a known and use of five or more, singly or in multiplex, c provide high discriminatory power.

New non-radioactive, filmless imaging system for gene typing

R Rubin, T Worley, D Hanzel, E Mansfield, Molecu Dynamics

The closing this summer of the Barnwell, South Carol radioactive waste site to California and other states l accelerated interest in non-isotopic DNA detection me ods. The new era of fluorescence imaging offers a versati high throughout, quantitative complement to chemilumin cence for non-radioactive methods.

For human identification analysis, we have recently used Molecular Dynamics FluorImager Scanning System DIS80 typing and STR multiplex analysis in polyacrylam gels and single-copy Southern blot detection of RFLI Additionally, we have used the FluorImager System's 9-v microplates scanning capability for high through quantitation of total DNA in human samples.

For the PCR-based assays, DIS80 and STR, the FluorIma detects either fluorescein labeled or unlabeled P products. The unlabeled assays require just a 10 min post-electrophoresis stain in SYBR Green I (Molec Probes) with no destaining step. In either case, wet gels scanned eliminating any drying, fixing or blotting time.

In RFLP analysis we have detected target DNA in as li as 100 ng (0.05 attomole) genomic DNA using a single lc probe to D2S44 (YNH24) (Promega Corp). We adapte chemiluminescent protocol using alkaline phospha coupled to the probe. The fluorescent substrate, AttoF (JBL Scientific) is substituted for a chemilumines substrate eliminating the use of film.

Typical scans on the FluorImager System require und minutes. Scanning occurs off-line of electrophoresis allo the scanner to be shared by several tabs which proce many gels or blots in parallel. Fluorescence offers a gre dynamic range and more linear response than film use autoradiography or chemiluminescence. Images are digit and become available immediately for quantitative in analysis using Molecular Dynamics ImageQuaNT Fragment Analysis software programs. Both image ana packages completely preserve the integrity of the raw files while allowing optimal on-screen contrast of dark weak spots over the entire four order dynamic range o instrument.

Co-amplification of short tandem repeats (STRs) amelogenin for rapid genetic typing and ge identification

NM Duda and SB Lee, Department of Justice I Laboratory, Berkeley, CA 94710

JW Schumm, C Sprecher and A Lins, Promega Corpor 2800 Woods Hollow Road, Madison, WI 53711

Polymorphic short tandem repeats (STRs) are val genetic markers because they exhibit easily detec

PROCEEDINGS, SIXTH INTERNATIONAL SYMPOSIUM ON HUMAN IDENTIFICATION 1995 (PROMEGA CORPORATION)

Short Tandem Repeat (STR) Polymorphism Analysis Using Energy Transfer Fluorescent Primers

J. Atherton[1], J. Ju[2], Y. Wang[2], B. Carpenter[1], R.A. Mathies[2], and G.F. Sensabaugh[1].
[1] Forensic Science Group, School of Public Health
[2] Chemistry Department, University of California, Berkeley, CA 94720

<> <> <> <> <> <> <> <> <> <> <> <> <> <> <> <> <> <> <>

DNA profiling has been established as a valuable tool in the investigation and prosecution of violent crimes, particularly sex offenses. It has also proven of great use in the identification of skeletal and other body remains in criminal, civil, and human rights cases. The next great challenge is the development and implementation of DNA profile identification databanks that can be used in the same way fingerprint identification databanks are currently used. Many states have legislatively mandated the establishment of databanks containing DNA profiles of certain categories of offenders, notably sex offenders, and the framework of a national system for the U.S. has been established under CODIS (1).

For the full potential of DNA databanks to be realized, two conditions must be met. First, high throughput DNA profiling technology will be required to process the large number of samples submitted for databank entry. Some states have inputs exceeding 1000 databank samples per month to process. Second, casework laboratories will have to begin doing DNA profiling on every case in which there is viable biological evidence. Needless to say, DNA profile databanks are anticipated to be of great value in helping solve suspectless cases, cases which at present are rarely worked up in crime laboratories due to limited personnel and technology.

These two conditions impose constraints on both the genetic markers and analytical technologies to be used in profiling. The markers used must be capable of being typed on multiple platforms–high throughput platforms for databank process laboratories and low cost platforms for routine casework laboratories. The current consensus favors the use of polymorphic short tandem repeat (STR) loci for profiling, in great because this method meets both conditions. Typing throughput–a primary consideration for databank labs–is also of concern for casework labs; the greater the information that can be generated per unit analysis, the greater the efficiency. Typing throughput, however, means little without better approaches to sample preparation. Sample preparation remains a rate limiting step; no matter how rapid the DNA typing technology, the generation of profiles can proceed no faster than the rate of DNA isolation. Advances in sample preparation will benefit throughput in both databank and casework laboratories, particularly if sample preparation can be integrated with typing technologies. Finally, it is likely that today's high throughput analytical systems will be superseded by more advanced technologies in the future. Accordingly, the analytical typing technologies developed to meet present needs should be capable of evolving. Advanced throughput technologies will bring future benefit to casework throughput as well, much as advances in computer chip technology have brought computers into the home.

To meet high-end throughput needs, we are evaluating capillary array electrophoresis (CAE) with laser excited confocal fluorescence detection (2-4). We have previously demonstrated the feasibility of using CAE for the rapid analysis of STR polymorphisms (5,6); this provides a foundation for high throughput DNA profiling based on these markers. We have also investigated the use of an intermediate technology–detection of fluorescently labeled STRs on a fluorescence imager (FluorImager, Molecular Dynamics) (7). In both these studies, we have exploited fluorescence energy transfer (ET) to optimize multi-color detection. The concept of using PCR primers labeled with fluorescent ET dyes has been recently described by Ju *et al.* (8,9) who demonstrated their superior properties for automated DNA sequence analysis. In this paper, we summarize our studies on STR analysis using ET primers.

CAPILLARY ARRAY ELECTROPHORESIS (CAE)

Capillary electrophoresis (CE) is capable of providing rapid, high resolution separations of DNA fragments; it has been used for the analysis of DNA

restriction fragments and polymerase chain reaction (PCR) products (10-18) and for DNA sequencing (2,3,19-25). Typical CE runs are 15-120 minutes in duration, depending on the size range of the DNA fragments to be separated and resolution desired. CE requires very small sample loads and has a loading format that is easily automated (8-10). The presentation by McCord at this symposium describes CE in more detail. Capillary array electrophoresis (CAE) is CE with many capillaries run simultaneously in parallel array (2-4). The design of CAE, coupled with the high speed of CE separations, allows very high throughput. The CAE unit used in this study, a prototype model built in the Mathies laboratory (4), has an array of 10 capillaries (75 um i.d.) with an effective run length of 25 cm. Fluorescence excitation is provided by an argon ion laser at 488 nm with confocal detection at 510-530 nm (green channel) and >590 nm (red channel). Output is stored in digitized form and processed using the programs IPLab, Kaleidagraph, and Canvas. The confocal detection system is quite sensitive, with a routine sensitivity limit of about 10 attomoles DNA per band. A schematic of the laser excited confocal fluorescence CAE system set up for two-color detection is shown in Figure 1.

With CE, it is necessary to co-electrophoresis the test sample with a typing standard. The two-color detection system allows a test sample and typing standard, each labeled with a different fluorescent dye, to be analyzed in the same capillary. The CAE unit illustrated in Figure 1 has also been constructed in a four-color detection mode; the four-color unit can be used both for DNA sequence analysis and STR typing analysis.

ENERGY TRANSFER (ET) PRIMERS

Primers labeled with different fluorescent dyes are routinely used in automated sequencing and in multiplex STR typing on the Applied Biosystems automated sequencer (26-32). A limitation of these dyes is that each fluorophore has its own distinctive fluorescence excitation and emission properties; detection systems require compromise excitation wavelengths and software processing to unscramble emission signals. ET labeling represents a significant advance on this situation (6,8,9). Primers are labeled with two dye molecules, one of which functions as a energy donor and the other as an acceptor; the design of a typical ET primer is shown below.

F6R

FAM-5'-GTGGGCTGAAAAGCTCCCGATTAT-3'
|
$(CH)_2(CO)$-NH-$(CH_2)_6$-NH-**ROX**

The designation F6R indicates that 6 nucleotides intervene between the donor dye FAM and the modified T to which the acceptor dye ROX is attached. Incident radiation at a single wavelength (in our case, 488 nm) is absorbed by the donor which transfers energy to the acceptor; the acceptor then emits at its characteristic fluorescence emission wavelength. Use of different acceptor dyes allows generation of distinctive emission spectra to be obtained from excitation at a single wavelength (Fig. 2). For example, FNF labels have their emission maximum in the green at 525 nm; FNR labels emit maximally in the red at 605 nm. Due to the high efficiency of energy transfer, there is no need to make primer concentration adjustments to compensate for fluorescence intensity differences among the primers as must be done for primers labeled with single fluors; ET primers with different acceptors have roughly comparable fluorescence emission properties (8,9).

STR TYPING USING CAE AND ET PRIMERS

Our initial efforts focused on the STR marker THO1, a tetranucleotide STR; this is one of the best studied STRs and is a consensus marker in many forensic STR typing systems (32,33). Samples of previously determined THO1 type were amplified by PCR using ET primers; PCR products were desalted by flotation dialysis on VCWP membranes and were loaded on the capillary columns by electrokinetic injection.

Initial experiments demonstrated that THO1 alleles could be separated with near baseline resolution on capillary columns containing the non-denaturing separation matrix 0.8% hydroxyethylcellulose (HEC) and the non-fluorescent intercalator 9-aminoacridine; under non-denaturing conditions, the presence of an intercalator is needed to achieve near single base resolution. ET primers differing in the number of nucleotides between donor and acceptor dyes were tested to determine effects on mobility. A pair of primers, F14F and F6R, were found to yield PCR products with coincident mobilities and these primers were used in subsequent experiments involving non-denaturing separations.

Two samples amplified using one ET primer, F6R, were combined to form an allelic ladder containing the 6, 7, 8 and 9 alleles. Test samples were amplified using the other ET primer, F14F. Test samples were co-electrophoresed with the allelic ladder. Typing the test samples was straightforward; the allelic ladder peaks that coincide with the test sample peaks indicate the type. Electrophoresis runs were completed in under 20 min. Statistical analysis of replicate data shows mean sample sizing on each allele to be accurate within 0.4 bp with standard deviations less than 0.7 bp. This accuracy and precision makes possible straightforward type calling by data processing software. This work has been published (6).

Electrophoresis under non-denaturing conditions detects heteroduplex DNA as well as homoduplex DNA (34). The heteroduplex bands migrate outside the region of the homoduplex allelic bands and hence do not interfere with typing. They do, however, potentially intrude on migration zones that might be occupied by other markers in a multiplex set. To avoid heteroduplex formation, we have recently evaluated capillary electrophoresis under denaturing conditions employing a separation matrix containing 2% HEC and 7M urea. Under these conditions, baseline resolution is achieved for the tetranucleotide repeat alleles and alleles differing by a single nucleotide (such as the 9.3 and 10 alleles in THO1) can be distinguished. Mobility shifts due to the primer labeling configuration are insubstantial and typing against an allelic ladder is straightforward.

We are currently working toward multiplex typing by CAE using previously characterized markers, *e.g.*, the markers developed by Promega and the Home Office Forensic Science Service (27-32,33,35-38). In some cases, primers have been modified to improve amplification properties. CAE typing exploiting ET primers has been developed for the loci shown in table 1. Multiplex sets have been constructed using non-overlapping markers; an example using the markers vWFA, TPOX, THO1, and D18S51 is illustrated in Figure 3.

Overall, the results of these studies establish the feasibility of using CAE for STR typing. With electrophoresis cycles of 30 minutes, and projection to a 48 capillary unit, it should be possible to process more than 500 samples per day; if samples are typed with a 4 locus multiplex, this translates to more than 2000 typings per day. These figures are well within the throughput specifications required for databank development. If desired, throughput can be increased using 4-color detection. Two 4-color configurations are possible. First, two multiplexes can be typed in concert against the corresponding allelic ladders, each test sample multiplex and each allelic ladder being detected in a different color. Alternatively, three multiplexes can be typed in concert against a standard sizing ladder; this approach requires allele-calling based on size measurement rather than comparison to the allelic ladder.

The potential for even higher throughput is on the horizon. Preliminary experiments show that STR typing can be done using the CAE on a chip developed by Woolley and Mathies (39); a result is shown in Figure 4. With a run time of under 5 minutes, the CAE on a chip is at least 4 times faster than conventional CAE. This result demonstrates the feasibility of STR typing on a next generation technology.

STR TYPING ON SLAB GELS USING ET PRIMERS

ET labeling offers many advantages for STR typing on conventional slab gels using fluorescence detection (7). This is demonstrated by comparison of the several detection formats that can used with conventional slab gel electrophoresis: silver staining, post-staining with the DNA binding fluorescent dye SYBR Green, single fluorescent dye labeling, and ET labeling. For fluorescence detection, we have employed a fluorescence imager (Molecular Dynamics FluorImager 575) with a two-color detection potential similar to the CAE unit, *i.e.*, excitation at 488 nm with detection using 530 nm bandpass and 610 nm longpass filters. SYBR Green labeled products and FAM labeled products were detected with the 530 nm bandpass filter; FNR labeled products were detected with the 610 nm longpass filter.

Slab gel electrophoresis was performed on 6% polyacrylamide gels (19:1 acrylamide/bis), 32 cm in length, for 2.25 hours at 40 watts. Near single base resolution was achieved for THO1; the 9.3 and 10 alleles could be resolved when run in adjacent lanes or when in different colors in the same lane. Initial studies with THO1 PCR products showed that the primer labeling configuration affected mobility. For example, the F14F labeled primer exhibited an approximate +1 base shift relative to the F6F and

F6R labeled primers; the two F6 primers were nearly identical in mobility. In contrast, use of the THO1 primer singly labeled with fluorescein (Promega) gave an approximate -1 base shift. This consideration is important if test samples and allelic ladder standards are to be run in the same lane.

Electrophoresis under denaturing conditions separates the DNA strands of PCR products. As a consequence, both silver stained and SYBR Green stained alleles are detected as band doublets. In contrast, use of fluorescently labeled primers, whether in the single fluor or ET mode, presents a single band for each allele. This simplifies pattern interpretation, particularly in situations involving PCR artifacts such as non-templated base addition and stutter banding. With regard to detection sensitivity, fluorescent primers and SYBR Green staining are comparable; all are about 3 times more sensitive than silver staining. Use of the Fluorimager for detection allows relative band densities to be quantitated, a benefit when assessing mixtures.

A singular advantage of ET labeling coupled with fluorescence imager detection is the throughput gain offered by two-color detection: it is possible to co-electrophorese two multiplexes, each labeled with a different color, in a single lane. Because there is negligible cross talk between the red and green detection channels, the red and green labeled samples are completely distinguished. FNF and FNR labeled products are detected with comparable sensitivity. The simultaneous typing of seven loci in two multiplexes is shown in Figure 5.

CONCLUSIONS

ET labeling represents a significant advance on fluorescent labeling technology. We have demonstrated here the advantages of using ET labeling in high throughput STR typing by CAE and in conventional slab gel electrophoresis coupled with fluorescence imager detection.

REFERENCES

1. McEwen J.E. (1995) Forensic DNA data banking by state crime laboratories. *Amer. J. Hum. Genet.* 56:1487-1492.

2. Huang X.C., Quesada M.A. and Mathies R.A.(1992) DNA sequencing using capillary array electrophoresis. *Anal. Chem.* 64:2149-2154.

3. Mathies R.A. and Huang X.C. (1992) Capillary array electrophoresis: an approach to high-speed, high throughput DNA sequencing. *Nature* 359:167-169.

4. Huang, X.C., Quesada M.A. and Mathies R.A. (1992) Capillary array electrophoresis using laser-excited confocal fluorescence detection. *Anal. Chem.* 64:967-972.

5. Wang Y., Ju J., Carpenter B.A., Atherton J.M., Sensabaugh G.F. and Mathies R.A. (1994) Rapid detection of genetic types at polymorphic short tandem repeat (STR) loci using two-color capillary array electrophoresis. *California Association of Criminalists Fall Seminar*, Pasadena, CA, 19-22 October, 1994.

6. Wang Y., Ju J., Carpenter B.A., Atherton J.M., Sensabaugh G.F. and Mathies R.A. (1995) Rapid sizing of short tandem repeat alleles using capillary array electrophoresis and energy-transfer fluorescent primers. *Anal. Chem.* 67:1197-1203.

7. Atherton J.M., Carpenter B.A., Atherton J.M., Sensabaugh G.F. and Mathies R.A. (1995) Vertical compatibility of STR typing using energy transfer fluorescent primers. *California Association of Criminalists Spring Seminar*, Walnut Creek, CA, 9-13 May, 1995.

8. Ju J., Ruan C., Fuller C.W., Glazer A.N. and Mathies R.A. (1995) Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc. Natl. Acad. Sci. U.S.A.* 92:4347-4351.

9. Ju J., Kheterpal I., Scherer J.R., Ruan C., Fuller C.W., Glazer A.N. and Mathies R.A. (1995) Design and synthesis of fluorescence energy transfer dye-labeled primers and their application for DNA sequencing and analysis. *Anal. Biochem.* 231:131-140.

10. Heiger, D.N., Cohen, A.S. and Karger, B.L. (1990) Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. *J. Chromat.* 516:33-48.

11. Schwartz H.E., Ulfelder K., Sunzeri F.J., Busch M.P. and Brownlee R.G. (1991) Analysis of DNA restriction fragments and polymerase chain reaction products towards detection of the AIDS (HIV-1) virus in blood. *J. Chromat.* 559:267-283.

12. Ulfelder K.J., Schwartz H.E., Hall J.M. and Sunzeri F.J. (1992) Restriction fragment length polymorphism analysis of ERBB2 oncogene by capillary electrophoresis. *Anal. Biochem.* 200:260-267.

13. Williams P.E., Marino M.A., Del Rio S.A., Turni L.A. and Devaney J.M. (1994) Analysis of DNA restriction fragments and polymerase chain reaction products by capillary electrophoresis. *J. Chromat. A* 680:525-540.

14. Marino M.A., Turni L.A., Del Rio S.A. and Williams P.E. (1994) Molecular size determinations of DNA restriction fragments and polymerase chain reaction products using capillary gel electrophoresis. *J. Chromat. A* 676:185-189.

15. McCord B.R., McClure D.L. and Jung J.M. (1993) Capillary electrophoresis of polymerase chain reaction-amplified DNA using fluorescence detection with an intercalating dye. *J. Chromat. A* **652**:75-82.

16. McCord B.R., Jung J.M. and Holleran E.A. (1993) High resolution capillary electrophoresis of forensic DNA using a non-gel sieving buffer. *J. Liq. Chromat.* **16**:1963-1981.

17. Srinivasan K., Morris S.C., Girard J.E., Kline M.C. and Reeder D.J. (1993) Enhanced detection of PCR products through use of TOTO and YOYO intercalating dyes with laser induced fluorescence-capillary electrophoresis. *Appl. Theor. Electrophoresis* **3**:235-239.

18. Butler J.M., McCord B.R., Jung J.M., Lee J.A., Budowle B. and Allen R.O. (1995) Application of dual internal standards for precise sizing of polymerase chain reaction products using capillary electrophoresis. *Electrophoresis* **16**:(forthcoming).

19. Cohen A.S., Najarian D.R. and Karger B.L. (1990) Separation and analysis of DNA sequence reaction products by capillary gel electrophoresis. *J. Chromat.* **516**:49-60.

20. Swerdlow H., Dew-Jager K.E., Brady K., Grey R., Dovichi N.J. and Gesteland R. (1992) Stability of capillary gels for automated sequencing of DNA. *Electrophoresis* **13**:475-483.

21. Swerdlow H., Zhang J.Z., Chen D.Y., Harke H.R., Grey R., Wu S.L., Dovichi N.J. and Fuller C. (1991) Three DNA sequencing methods using capillary gel electrophoresis and laser-induced fluorescence. *Anal. Chem.* **63**:2835-2841.

22. Chen D.Y., Swerdlow H.P., Harke H.R., Zhang J.Z. and Dovichi N.J. (1991) Low-cost, high-sensitivity laser-induced fluorescence detection for DNA sequencing by capillary gel electrophoresis. *J. Chromat.* **559**:237-246.

23. Swerdlow H., Wu S.L., Harke H. and Dovichi N.J. (1990) Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. *J. Chromat.* **516**:61-67.

24. Swerdlow H. and Gesteland R. (1990) Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucl. Acids Res.* **18**:1415-1419.

25. Carson S., Cohen A.S., Belenkii A., Ruiz-Martinez M.C., Berka J. and Karger B.L. (1993) DNA sequencing by capillary electrophoresis: Use of a two-laser-two-window intensified diode array detection system. *Anal. Chem.* **65**:3219-3226.

26. Hunkapiller T., Kaiser R.J., Koop B.F. and Hood L. (1991) Large-scale and automated DNA sequence determination. *Science* **254**:59-67.

27. Kimpton C.P., Gill P., Walton A., Urquhart A., Millican E.S. and Adams M. (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods Appl.* **3**:13-22.

28. Fregeau C.J. and Fourney R.M. (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *BioTechniques* **15**:100-109.

29. Kimpton C.P., Fisher D., Watson S., Adams M., Urquhart A., Lygo J. and Gill P. (1994) Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. *Intl. J. Leg. Med.* **106**:302-311.

30. Urquhart A., Oldroyd N.K., Kimpton C.P. and Gill P. (1995) Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. *BioTechniques* **18**:116-121.

31. Smith R.N. (1995) Accurate size comparison of short tandem repeat alleles amplified by PCR. *BioTechniques* **18**:122-128.

32. Oldroyd N.J., Urquhart A.J., Kimpton C.P., Millican E.S., Watson S.K., Downes T. and Gill P.D. (1995) A highly discriminating octoplex short tandem repeat polymerase chain reaction system suitable for human individual identification. *Electrophoresis* **16**:334-337.

33. Puers C., Hammond H.A., Jin L., Caskey C.T. and Schumm J.W. (1993) Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG]$_n$ and reassignment of alleles in population analysis by using a locus-specific allelic ladder. *Amer. J. Hum. Genet.* **53**:953-958.

34. Cheng J., Kasuga T., Mitchelson K.R., Lightly E.R.T., Watson N.D., Martin W.J. and Atkinson D. (1994) Polymerase chain reaction heteroduplex polymorphism analysis by entangled solution capillary electrophoresis. *J. Chromat. A* **677**:169-177.

35. Hammond H.A., Jin L., Zhong Y., Caskey C.T. and Chakraborty R. (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Amer. J. Hum. Genet.* **55**:175-189.

36. Anker R., Steinbrueck T. and Donis-Keller H. (1992) Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. *Hum. Molec. Genetics* **1**:137.

37. Huang N.E., Schumm J. and Budowle B. (1995) Chinese population data on three tetrameric short tandem repeat loci-HUMTHO1, TPOX, and CSF1PO-derived using multiplex PCR and manual typing. *Forensic Sci. Intl.* **71**:131-136.

38. Polymeropoulos M.H., Rath D.S., Xiao H. and Merril C. (1991) Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES). *Nucl. Acids Res.* **19**:4018.

39. Woolley A.T. and Mathies R.A. (1994) Ultra-high-speed DNA fragment separations using microfabricated capillary array electrophoresis chips. *Proc. Natl. Acad. Sci. U.S.A.* **91**:11348-11352.

Table 1.

Locus	Repeat	Het.	Alleles (Size Range)	Reference
THO1	AATG	0.77	8 (179-203	(33, 34)
CSF1PO	AGAT	0.74	9 (299-323)	(35)
TPOX	AATG	0.66	8 (232-248)	(36, 37)
vWFTA31	AGAT	0.74	11 (139-167)	(32)
FIBRA	AAAG	0.82	9 (256-284)	(32)
FES/FPS	AAAT	0.70	8 (222-250)	(38)
D6S502	TCTA	0.78	8 (152-210	(32)
D18S51	AAAG	0.89	14 (267-319)	(32)
D20S85	AAAG	0.70	6 (123-147)	(32)
D21S11	TCTA/G	0.82	12 (172-264)	(32)

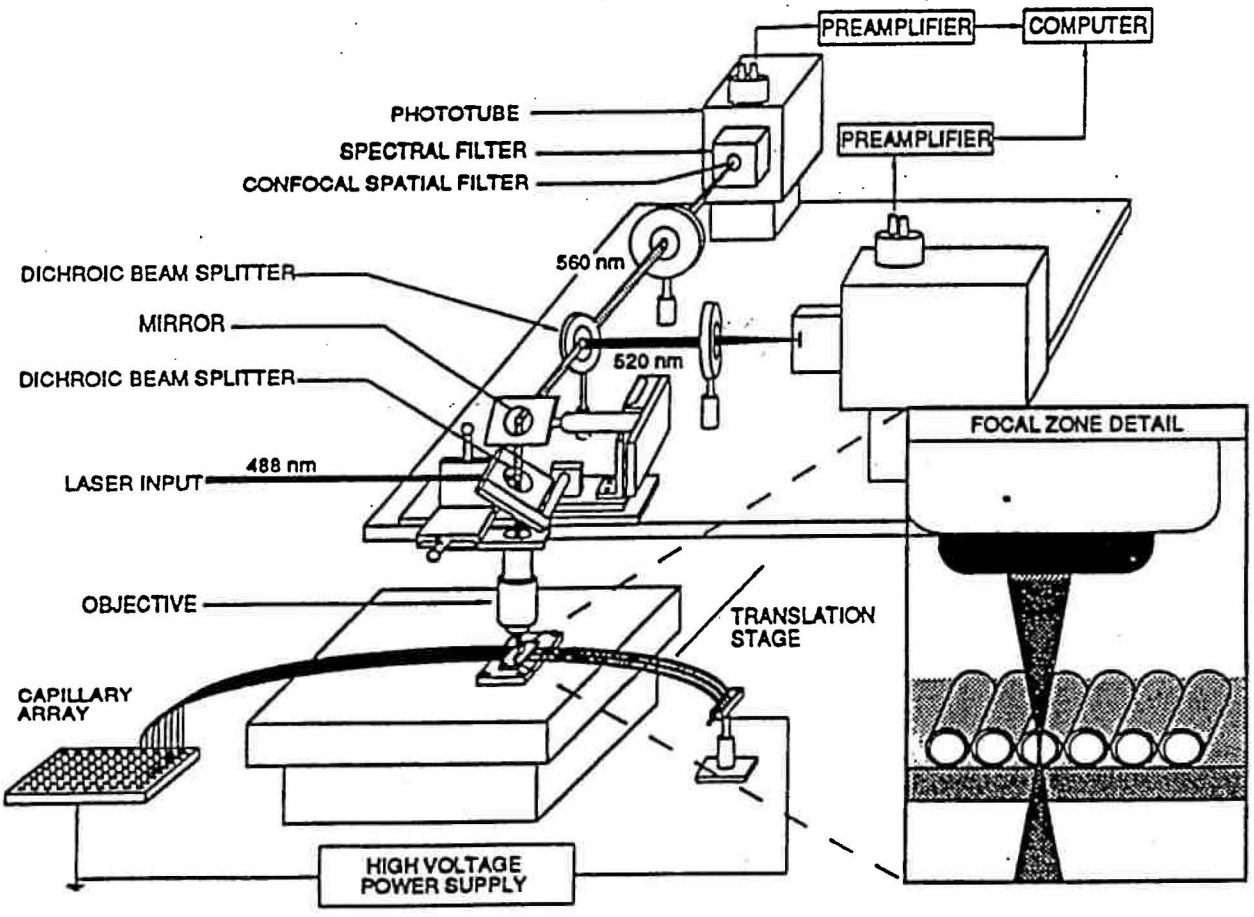


Figure 1. Capillary array electrophoresis unit with two-color confocal fluorescence detection.

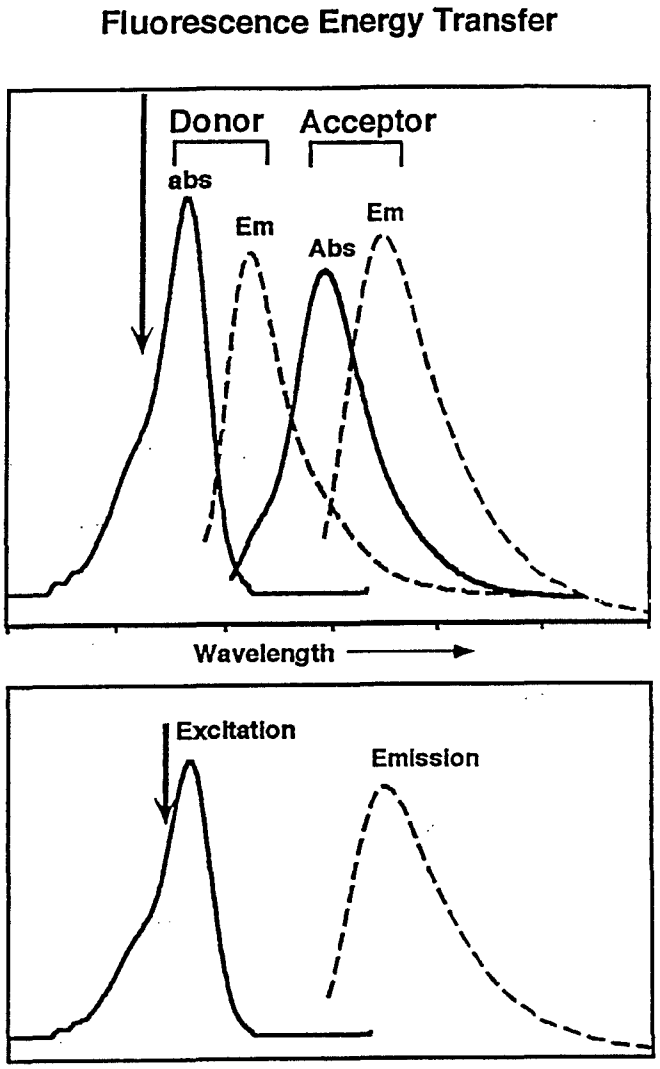


Figure 2. Fluorescence energy transfer. The donor molecule is excited at the wavelength indicated by the arrow (top panel). The excited donor molecule transfers its excitation energy to an acceptor molecule which then emits fluorescence at the acceptors characteristic emission wavelength (bottom panel). Pairing a single donor with different acceptors permits efficient fluorescence emission at different wavelengths from excitation at a single wavelength.

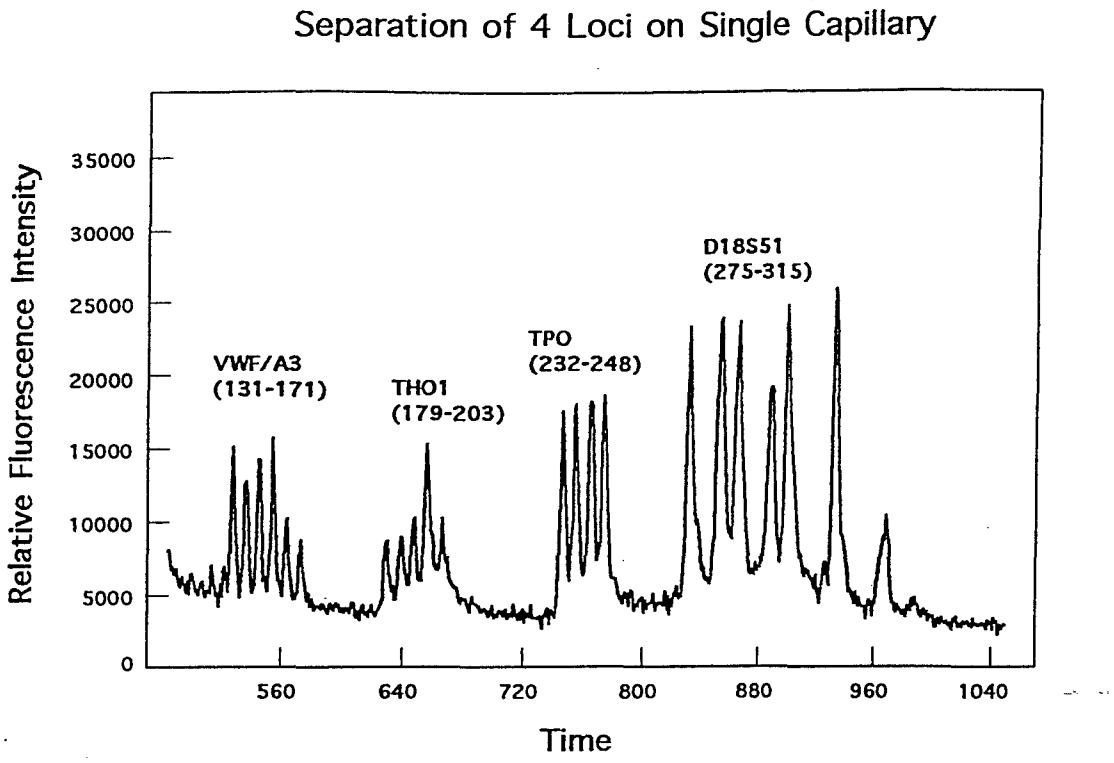


Figure 3. Multiplex typing of four loci on a single capillary run. Allelic ladders for the vWFA, THO1, TPO, and D18S51 loci are shown.

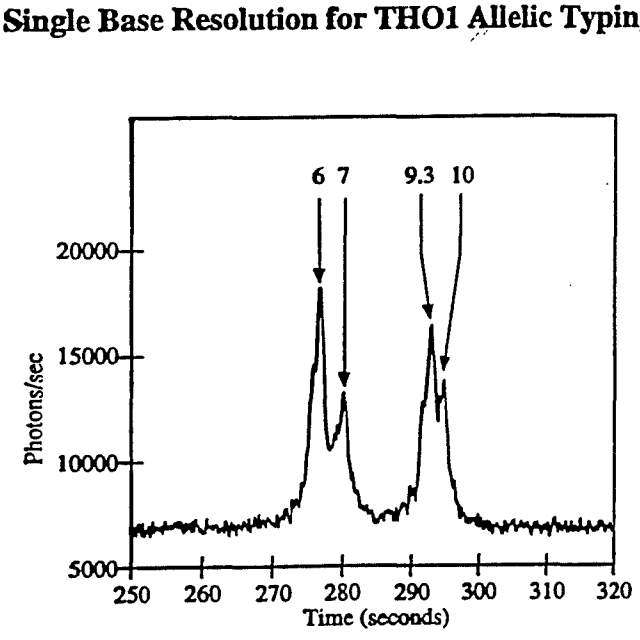


Figure 4. THO1 typing with CAE on a chip. Two samples, a 6,9.3 and a 7,10, were mixed to show 4 bp and 1 bp resolution.

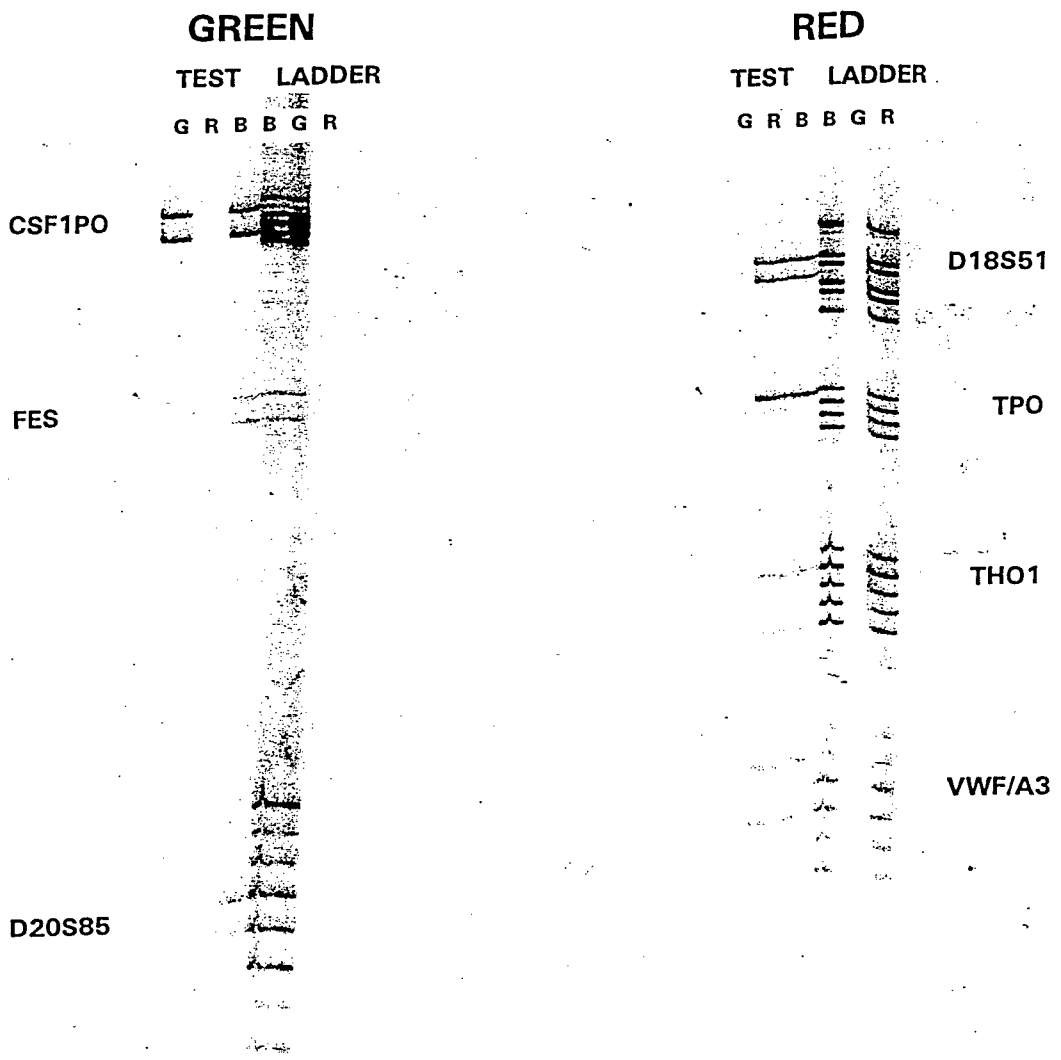


Figure 5. Two-color detection of two multiplexes on a slab gel. The triplex CSF1PO- FES/FPS-D20S85 was labeled using FNF primers and detected in the green channel (530 nm bandpass filter). The quadruplex D18S51-TPO-THO1-vWFA was labeled with FNR primers and detected in the red channel (610 nm longpass filter). Each multiplex was run separately (the G lanes for the triplex and the R lanes for the quadruplex) and combined (the B lanes). The test sample was typed by comparison to the allelic ladder lane. Note the absence of fluorescence crosstalk between the green and red detection channels.

DNA Sequencing Using Four-Color Capillary Array Electrophoresis and Energy Transfer Primers

Indu Kheterpal, Jingyue Ju, Arun Radhakrishnan, Gabriel S. Brandt, Charles L. Ginther[1], Steven M. Clark, James R. Scherer, George F. Sensabaugh[1] and Richard A. Mathies

Department of Chemistry and [1]School of Public Health
University of California, Berkeley, CA

ABSTRACT

A practical capillary array electrophoresis sequencer has been constructed and tested for large scale DNA sequencing. This system employs a scanning confocal optical system to detect DNA fragments electrophoresed on a capillary array. By using the recently developed energy transfer fluorescent primers, sensitive four-color detection is facilitated while exciting with just a single laser wavelength (488 nm) from an argon ion laser. Software for facile reduction of the images to four-color trace files and for automated base-calling have also been developed. This system can detect up to 25 capillaries at a time and has a raw sequencing rate of ~ 6 kilobases/hour. Applications to mitochondrial D-loop DNA sequencing are presented as a demonstration.

Keywords: DNA sequencing, energy transfer primers, capillary electrophoresis, confocal fluorescence detection, mitochondrial DNA

1. INTRODUCTION

The human genome is the ultimate source of genetic information for a variety of biological purposes including disease diagnosis and therapy, identification of individuals, and exploring genetic heritage. Sequencing the human genome is essential for understanding these biological processes. To accomplish this task in a reasonable amount of time, the current sequencing rates of 1-10 million bases per year per genome center have to be increased to at least ~ 1 billion bases per year. Therefore, new techniques for DNA sequencing have to be developed. Capillary electrophoresis (CE) has emerged as a very powerful high speed analytical tool for DNA analysis.[1] In 1992, we developed a two-color confocal capillary array electrophoresis (CAE) scanner and thereby combined the high speed and sensitivity of CE with the ability to run up to 25 samples simultaneously.[2] We have presented the applications of this scanner for sequencing,[3,4] restriction fragment sizing[5] and short tandem repeat typing.[6] A two-color binary coding

method was used to fluorescently label the fragments for sequencing as only two of the available dyes could be effectively excited with one laser. Since then, we have developed energy transfer (ET) dye-labeled primers and demonstrated that they can dramatically enhance the capabilities of four-color sequencing.[7-9] The design, construction and operation of a practical four-color CAE sequencer is presented here and applied to large scale mitochondrial DNA sequencing.

We are using mitochondrial DNA for testing our four-color scanner because it provides a convenient source of a large number of interesting DNA templates. Mitochondrial DNA is a circular, double-stranded molecule consisting of 16,569 base pairs.[10] Its D-loop control region contains great diversity making it useful for evolutionary studies.[11] Since, mtDNA is maternally inherited, it is also used for maternally linked individual identification.[12] African populations show the greatest range of diversity in the mtDNA control region. Ginther et al. (unpublished) have determined D-loop sequences from over a hundred individuals from Sierra Leone and have identified twelve distinctive motifs. We present here representative sequences from ten of these motifs obtained using our four-color CAE scanner. This is the first demonstration of complete DNA sequencing runs using ET primers to sequence on capillary arrays and obtaining the base calls automatically. We are now in the process of completing the construction of a CAE sequencing facility for large scale DNA sequencing projects. This facility will be used for instruction, studying population trends and as a further testing site for new sequencing protocols, dyes and instrumentation.

2. MATERIALS AND METHODS

Four-color capillary array electrophoresis. A schematic diagram of the CAE sequencer is presented in Figure 1. It utilizes an argon ion laser (Spectra-Physics, Model 2020, Mountain View, CA) for excitation (488 nm, 3 mW). The beam is expanded, collimated and focused into the capillary through a microscope objective (20X, 0.5 N.A., Rolyn Optics, Covina, CA). The fluorescence is collected by the objective and passed through a series of dichroic beam splitter assemblies followed by confocal detection. The fluorescence in each wavelength region is focused on a 200-μm confocal pinhole (Melles Griot, Boulder, CO) with a 105 mm acromat lens (Rolyn Optics) and detected by the Hammatsu HC 120-05 photomultipliers (Photonics Research Corp., San Jose, CA). The output is amplified and filtered using SR 640 dual channel low pass filters (Stanford Research Systems, Inc., Sunnyvale, CA), digitized using a 16-bit ADC board (DAS-HRES, Keithley Instruments, Inc. Taunton, MA) and stored in a PC.

The capillary array is placed on a computer controlled translation stage (Model SX6, Parker Hannifin, Rohnert Park, CA). The stage continuously scans the capillary array at 10 mm/sec perpendicular to the

direction of the electrophoresis. The spatial dimension of the acquired image represents the physical image of the capillary and the temporal dimension reports on the concentration of labeled fragments as a function of time. Four such 16-bit image files, corresponding to the four detection channels are recorded for each experiment. Data for each channel is sampled at 1000 Hz and the image resolution is 10 μm/pixel; therefore 10 pixels represent the entire 100-μm inner diameter of the capillary. High frequency noise is reduced by the 500 Hz electronic low-pass filter. An image of the migrating bands is built bidirectionally with periodic 0.5 second sweeps. The transit time for the DNA fragments through the detection zone varies from 10 to 15 seconds and therefore at least 20 points are collected per peak. Data processing was performed post run by automatically summing the 10 pixels across each capillary interior and outputting the four-color trace files in a format readable by the base-calling program.

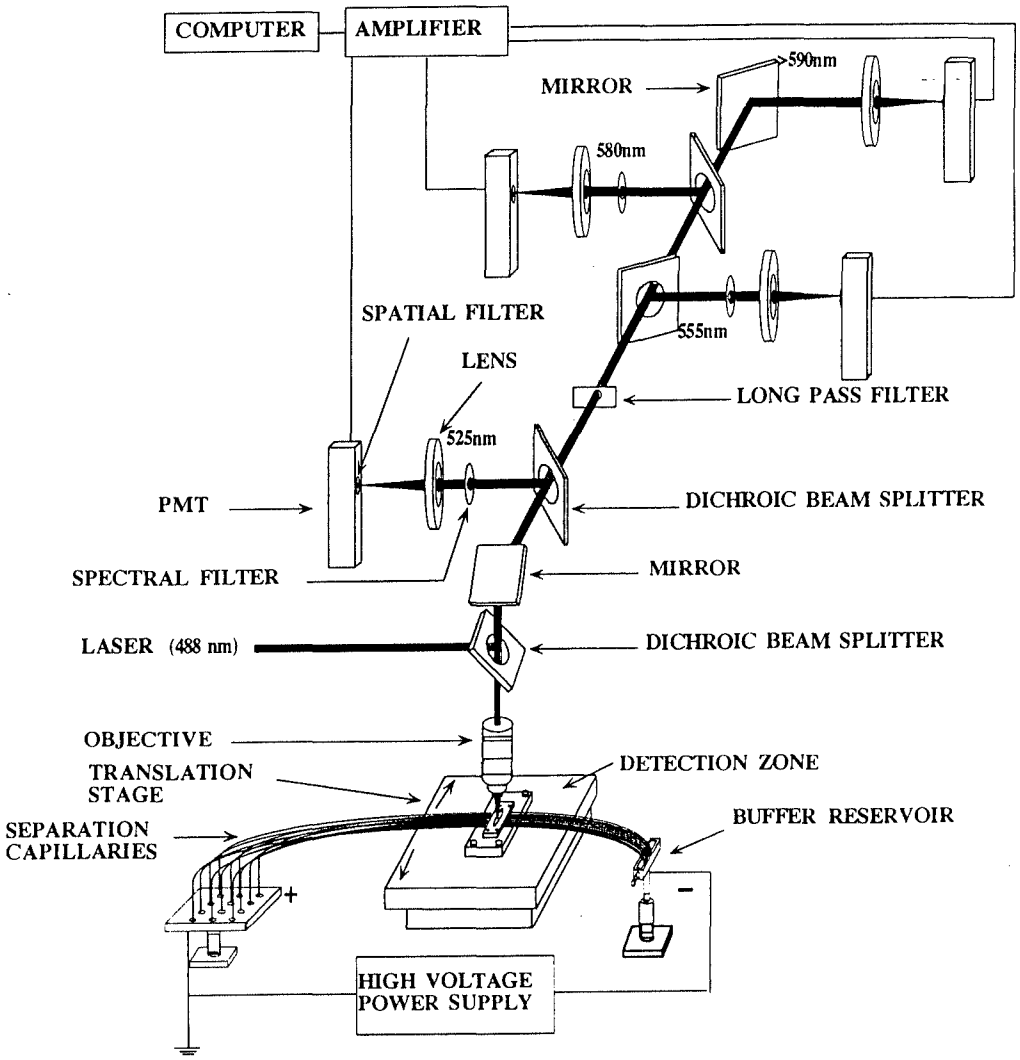


Fig. 1 Schematic of the four-color capillary array electrophoresis DNA sequencer.

Preparation of Capillary Arrays. The fused silica capillaries (100 μm i.d., 200 μm o.d., Polymicro Technologies, Phoenix, AZ) used for sequencing were pre-treated as described before.[13] Briefly, the capillaries were washed with 0.5 M NaOH, 0.1 M HCl and 0.1 M NaOH for an hour each. The capillaries were filled with 50% solution of 3-(trimethoxysilyl)propylmethyacrylate (Aldrich, Milwaukee, WI) in methanol and stored overnight. The separation matrix was prepared according to Best et al.[14] Twenty microliter of 10% APS (ammonium persulfate) and 2 μl of TEMED (tetraethylmethylenediamine) were added to a 5 ml aliquot of the freshly degassed 6% T acrylamide in a 7 M urea, 1X TBE buffer (90 mM Tris, 90 mM boric acid and 2 mM EDTA). Capillaries were filled with this solution and stored overnight at 4^0 C. Gel filled capillaries were pre-electrophoresed for 30-40 minutes while ramping the voltage from 0-10 kV. The capillaries were placed in a holder and located with grooves at 220 μm intervals. The DNA sequencing samples were electrokinetically injected for 20 sec at 170 V/cm and separated at 200 V/cm.

Energy transfer primers. The structures of the fluorescent oligonucleotide primers used for sequencing are shown in Figure 2. The design and synthesis of these energy transfer (ET) primers have been presented in detail by Ju et al.[7-9] These primers contain a common donor dye at the 5'-end and an acceptor dye bound to a modified thymidine residue (T*) eight to nine bases down in the sequence. Since these primers exploit Förster resonance energy transfer, they can all be efficiently excited at 488 nm and they exhibit distinct and strong emission maxima (525, 555, 580 and 605 nm). The common donor used for these primers is 6-carboxy-fluorescein (FAM, F) and the acceptors used are 2',7'-dimethoxy-4',5'-dichloro-6-carboxy-fluorescein (JOE, J),

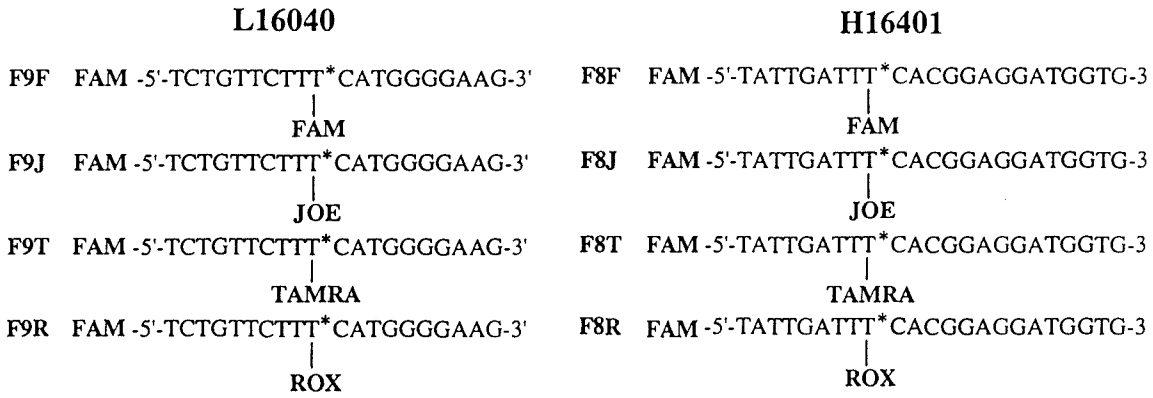


Fig. 2. Structures of the ET primers used for mitochondrial DNA sequencing. The number in the abbreviated name indicates the number of nucleotides between the donor and the acceptor. For the heavy strand primer (H16401) an eight nucleotide spacer was used; the light strand primer (L16040) used a spacer of nine nucleotides.

N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA, T) and 6-carboxy-X-rhodamine (ROX, R) at the 3' end. Specific primers have been synthesized for the light and heavy strands of the hypervariable region I of the D-loop of mitochondrial DNA. The light and heavy strand primers are denoted as L16040 and H16401, respectively.

Preparation of DNA sequencing samples. Mitochondrial DNA was isolated from blood stained samples in the collection of G.F.S.; these were collected in Sierra Leone in 1981. DNA sequencing samples were produced using a Sanger dideoxy terminator method. The reactions containing ddCTP were run with the FNF primer, ddATP terminators were run with the FNR primer, ddGTP terminators were run with the FNT primer, and ddTTP terminators were run with the FNJ primer. For every reaction, 0.4 pmoles of the primer was mixed with 0.1-0.4 μg of the amplified hypervariable region I of mtDNA and 0.5 μL of the ThermoSequenase (Amersham Life Sciences Inc., Cleveland, Ohio) reaction buffer (260 mM Tris/HCl, pH 9.5, 65 mM $MgCl_2$). Three microliter of the dNTP/ddNTP mixture (300:1) was added to each reaction tube. dGTP was replaced with dITP and the ratio of dITP to ddGTP was 3000:1. The tubes were preheated for 1 minute at 95^0 C. ThermoSequenase (1.6 units) was added to each tube followed by thermal cycling (15X). The cycle sequencing temperatures were 95^0 C for 30 seconds, 56^0 C for 20 seconds, and 72^0 C for 2 minutes. The samples were then pooled into stop solution (300 μL ethanol, 40 μL of 4 M ammonium acetate and 10 μL of 50 mM EDTA) and ethanol precipitated. The precipitated DNA was resuspended in 5 μL of 98% formamide in 1 mM EDTA. A 2 μL aliquot of this sample was used for injection. The samples can also be made with only one cycle followed by resuspension in only 2 μL of 98% formamide.

Image Processing. The fluorescence intensity data collected using the sequencer were formatted to be directly read and called by a program written by Anthony J. Berno.[15] This program was modified for operation on a Power Macintosh 8100/80 by Alan B. Greenfield and Jeffrey R. Sachs (Daniel H. Wagner Associates, Sunnyvale, CA). The base-calling program filters the traces to remove high-frequency noise, normalizes the raw data, color separates to remove cross talk due to spectral overlap, performs mobility shift corrections and event filters using second and fourth derivative Fourier filtering. The program uses a graph theoretic approach to assign the base-calls.[15]

3. RESULTS

Figure 3 presents an image of the migrating DNA fragments in an array of 11 capillaries. Four channels are displayed individually as four lanes for every capillary. The capillary images have been background subtracted and color thresholded for display. The sequence starts within half an hour of injection. All the bases in each capillary are well resolved. There is some spectral overlap in the emission spectra of the primers and therefore a band is observed in all four lanes but the highest intensity band represents the terminating

base. The capillaries shown in Figure 3 present the sequences of the light strands from different motifs of the Sierra Leone population. The 16142-16192 section of mtDNA presented here contains differences at eleven different positions from the known Anderson sequence.[10] The position of these mutations is marked in the figure by arrows.

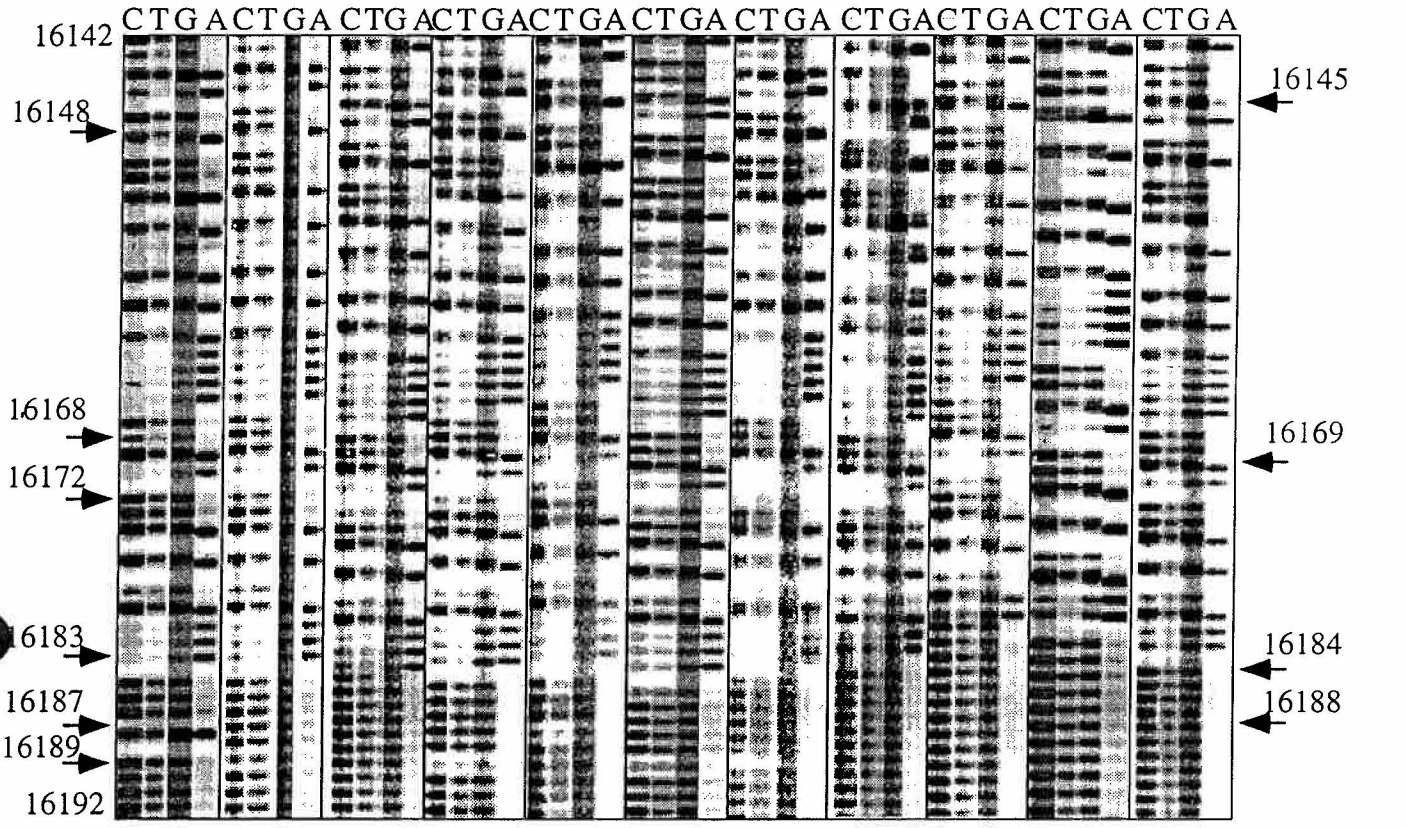
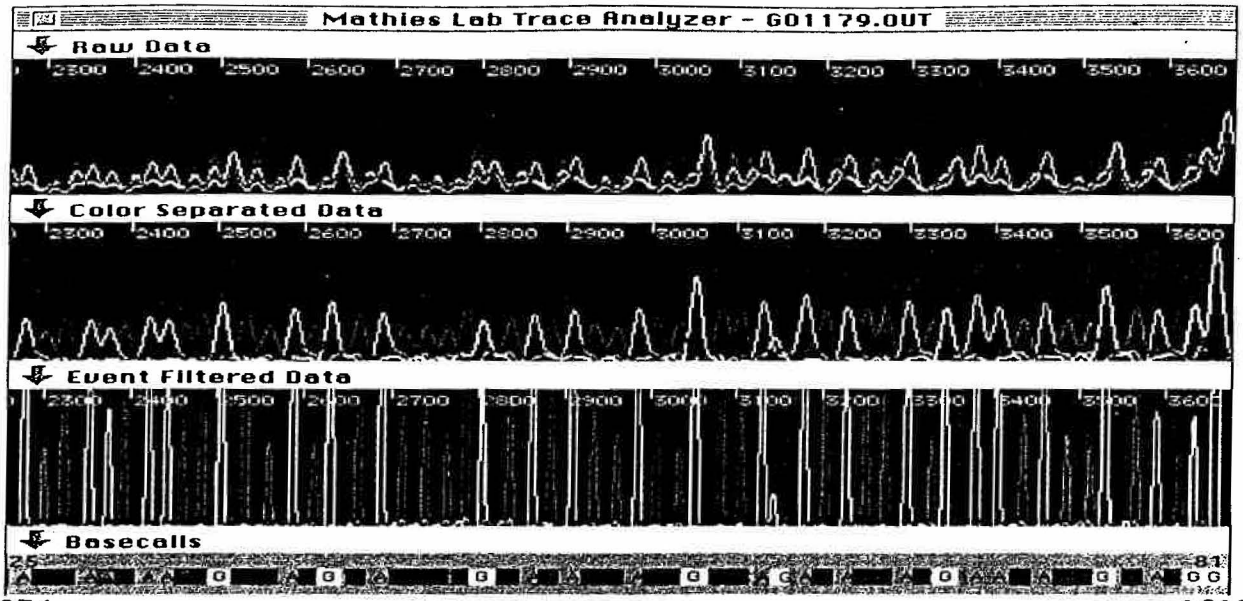


Fig. 3 Image of hypervariable region I mtDNA fragments separated in an eleven capillary array.

The sequencing data is integrated post run across the entire width of each capillary and these fluorescence intensity traces are directly transferred to the base-calling program. Figure 4 presents a screen image illustrating the operation of the base-calling software. The sequence shown is 25-81 bases from the primer. The top panel presents the raw data plotted as fluorescence intensity vs time. The color separated data shown in the second panel is obtained by means of multidimensional analysis. The matrix for such analysis can either be a default matrix or it can be calculated dynamically. The third panel presents the event filtered data obtained by using second and fourth derivative Fourier filtering. The base calls are indicated at the bottom.


Mathies Lab Trace Analyzer - G01179.OUT
Raw Data
Color Separated Data
Event Filtered Data
Basecalls

16074 16130

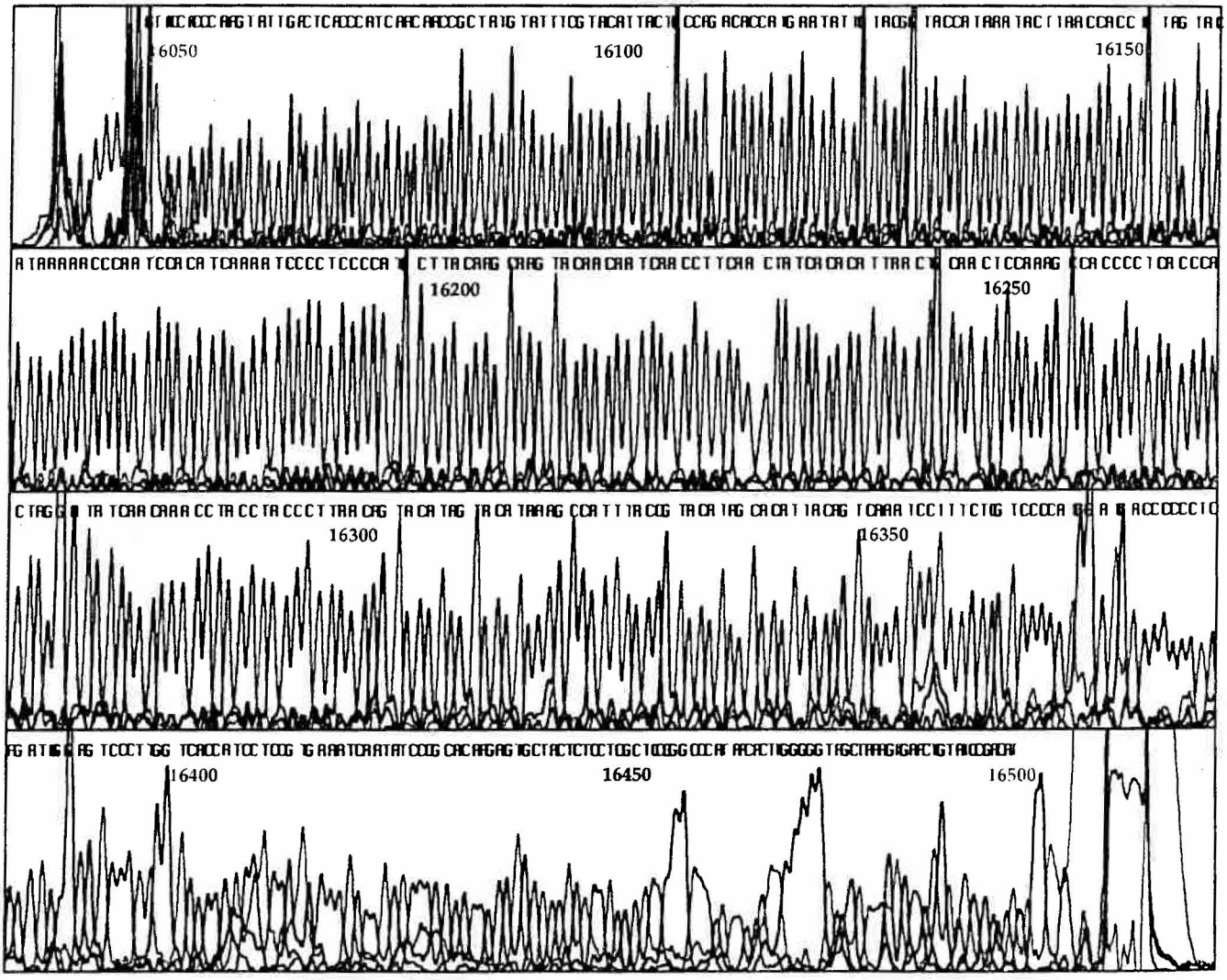

16050
16100
16150
16200
16250
16300
16350
16400
16450
16500

Fig. 4 (Top) Screen image demonstrating the operation of the base calling program.

Fig. 5 (Bottom) Analyzed four-color DNA sequencing data of the L strand of the hypervariable region I of the D-loop using ET primers. The sequence is identified by the color of each peak. Blue (F9F), C; Green (F9J), T; Black (F9T), G; Red (F9R), A. The total electrophoresis time is 122 minutes.

Figure 5 represents a detailed analysis of the sequence obtained in one representative capillary in an eleven capillary array. The data shown has been color separated and the base calls have been assigned automatically. The hypervariable region I in mitochondrial D-loop extends from 16020-16500. The primer sequence for the light strand extends from 16020-16040. The bases can be called starting from 16049 to over 16500. The entire sequencing run was complete in 2 hours. There is only one mistake in the sequence giving it an accuracy of 99.8%.

4. CONCLUSION

We have demonstrated that arrays of capillaries can be run simultaneously and high quality sequence data can be obtained within two hours using our sequencer. We have successfully sequenced the L-strand of the hypervariable region I of the control region of mitochondrial DNA. The construction of a CAE sequencing facility is in process where large scale sequencing will be performed. We plan to have two CAE sequencers with the capability of running 48 capillaries each. Using these sequencers our raw sequencing rate will be 24,000 bases/hour. This facility will also be used as a testing site for the development of new protocols and instrumentation.

5. ACKNOWLEDGEMENTS

We thank Alan B. Greenfield and Jeffrey R. Sachs for modification of the base-calling program and Mark Bennett for technical assistance. This research was supported by the Director, Office of Energy Research, Office of Health and Environmental Research of the U.S. Department of Energy under Contract DE-FG-91ER61125, by the National Institutes of Health (HG01399-01), and by the National Institute of Justice (93-IJ-EX-0010). J.J. was supported by a Human Genome Distinguished Postdoctorol Fellowship sponsored by the U.S. Department of Energy, Office of Health and Environment Research, and administered by the Oak Ridge Institute for Science and Education. We also acknowledge additional support from Amersham Life Sciences Inc.

REFERENCES

(1) Baker, D. R. *Capillary Electrophoresis*, Wiley: New York, 1995.

(2) Mathies, R. A.; Huang, X. C., "Capillary array electrophoresis: An approach to high-speed, high-throughput DNA sequencing," Nature (London), 359, 167-169, 1992.

(3) Huang, X. C.; Quesada, M. A.; Mathies, R. A., "Capillary Array Electrophoresis Using Laser-Excited Confocal Fluorescence Detection," Analytical Chemistry, 64, 967-972, 1992.

(4) Huang, X. C.; Quesada, M. A.; Mathies, R. A., "DNA sequencing using capillary array electrophoresis," Analytical Chemistry, 64, 2149-2154, 1992.

(5) Clark, S.; Mathies, R. A., "High-Speed Parallel Separation of DNA Restriction Fragments Using Capillary Array Electrophoresis," Analytical Biochemistry, 215, 163-170, 1993.

(6) Wang, Y.; Ju, J.; Carpenter, B. A.; Atherton, J. A.; Sensabaugh, G. F.; Mathies, R. A., "Rapid Sizing of Short Tandem Repeat (STR) Alleles Using Capillary Array Electrophoresis and Energy-Transfer Fluorescent Primers," Analytical Chemistry, 67, 1197-1203, 1995.

(7) Ju, J.; Glazer, A. N.; Mathies, R. A., "Fluorescence Energy Transfer Dye-Labeled Primers for DNA Sequencing and Analysis," Proceedings of the National Academy of Sciences USA, 92, 4347-4351, 1995.

(8) Ju, J.; Kheterpal, I.; Scherer, J. R.; Ruan, C.; Fuller, C. W.; Glazer, A. N.; Mathies, R. A., "Design and Synthesis of Fluorescence Energy Transfer Dye-Labeled Primers and Their Application for DNA Sequencing and Analysis," Analytical Biochemistry, 231, 131-140, 1995.

(9) Ju, J.; Glazer, A. N.; Mathies, R. M., "Energy transfer primers: A new fluorescence labeling paradigm for DNA sequencing and analysis," Nature Medicine, 2, 180-182, 1996.

(10) Anderson, S.; Bankier, A. T.; Barrell, B. G.; de Bruijn, M. H. L.; Coulson, A. R.; Drouin, J.; Eperon, I. C.; Nierlich, D. P.; Roe, B. A.; Sanger, F.; Schreier, P. H.; Smith, A. J. H.; Staden, R.; Young, I. G., "Sequence and organization of the human mitochondrial genome," Nature, 290, 457-464, 1981.

(11) Vigilant, L.; Stoneking, M.; Harpending, H.; Hawkes, K.; Wilson, A. C., "African Population and the evolution of human mitochondrial DNA," Science, 253, 1503-1507, 1991.

(12) Gill, P.; Ivanov, P. L.; Kinpton, C.; Piercy, R.; Benson, N.; Tully, G.; Evett, I.; Hagelberg, E.; Sullivan, K., "Identification of the remains of the Romanov family by DNA analysis," Nature Genetics, 6, 130-135, 1994.

(13) Huang, X. C.; Mathies, R. A. "Application of CAE to DNA sequencing" , eds. M. D. Adams, C. Fields, and J. Craig Venter, 17-28, Academic Press: London; San Diego, 1994.

(14) Best, N.; Arriaga, E.; Chen, D. Y.; Dovichi, N. J., "Separation of Fragments up to 570 bases in length by use of 6% T non-cross-linked polyacrylamide for DNA sequencing in capillary electrophoresis," Analytical Chemistry, 66, 4063-4067, 1994.

(15) Berno, A. J., "A Graph Theoretic Approach to the Analysis of DNA Sequencing Data," Genome Research, in press, 1996.

FORENSIC APPLICATIONS OF PCR TECHNOLOGY: THE VIEW FROM THE FIRST DECADE

Sensabaugh, G.F.
Forensic Science Group, School of Public Health
University of California, Berkeley, CA 94720, UNITED STATES

The polymerase chain reaction (PCR) was first described just ten years ago. Its potential application to the analysis of biological evidence was evident from the beginning and it is now used worldwide in forensic laboratories. PCR has extended the horizon of biological evidence analysis in several significant ways.

1. PCR can be used to amplify very small amounts of DNA. This has opened the way to genetic typing of evidence materials such as hair, saliva traces, and bone, items which previously had been difficult or impossible to type.

2. PCR can be used to amplify degraded DNA. This has opened the way to genetic typing of old and degraded evidence samples such as buried skeletal remains and tissue traces from disaster sites.

3. PCR based typing has emerged as the dominant method used in human genome mapping, resulting in the discovery of many thousands of new genetic markers. It is now possible to achieve virtual individualization using PCR based typing.

4. PCR based methods have simplified DNA sequencing, making possible DNA sequence analysis as an approach to identification. The analysis of mtDNA sequences has become an important tool in the identification of human remains.

5. PCR based methods are extensively used for the genetic characterization of animals, plants, and microbes. This opens the way to the use of virtually any biological material as evidence.

6. PCR based genetic typing is susceptible to automation, making possible the high throughput analyses necessary for the establishment of DNA profile databases.

In sum, over the course of the past decade, PCR based methods have greatly extended the potential of forensic biology. It is now in our hands to turn this potential into reality.

California Association of Criminalists 86th Semi-annual Seminar

The Eighty-sixth Semi-annual Seminar of the California Association of Criminalists was held at San Pedro, California in Autumn 1995. Abstracts of the papers presented at the meeting are printed below.

Handling high profile cases: a team approach

Lakshmanan Sathyavagiswaran MD, and Steven Dowell, Los Angeles County Department of Coroner

As Chief Medical Examiner – Coroner of Los Angeles County, Dr Lakshmanan has been involved with many complex cases, several of which were high profile. He shared his experiences and insights of handling those cases. Preparation for courtroom presentation as applicable was discussed. The interaction between criminalists, toxicologists and investigators was discussed. Additionally, a protocol for high profile cases that has been developed was presented.

Forensic casework analysis using STRs, DQα and PM in combination

Robin Cotton PhD, Anjali A Ranadive, Melisa A Weber and Charlotte J Word PhD, Cellmark Diagnostics

Analysis of polymorphic genetic loci using the polymerase chain reaction (PCR) enables results to be obtained from small and/or degraded samples. PCR analysis of multiple polymorphic loci has significantly enhanced the amount of information obtainable from biological evidence. The Amplitype HLA DQα Forensic DNA Amplification and Typing Kit and the Amplitype PM PCR Amplification and Typing Kit allow typing of six polymarker loci using a reverse dot format. The recent availability of reagents to amplify short tandem repeat (STR) loci such as the HUMCSFIPO, HUMTPOX and HUMTHO1 triplex (CTT) using Geneprint STR systems adds additional loci which can be analyzed using the PCR. Although validation studies have been conducted using the CTT triplex, limited data was available on the efficiency of amplification of the CTT triplex in comparison to the DQα and PM systems using casework samples. Therefore, the STR analysis at Cellmark Diagnostics was initially implemented only on evidence samples where a sufficient amount of DNA was extracted to complete both DQα and PM analysis which was then followed by STR analysis.

To date we have analyzed twelve cases (approximately 44 samples) using this approach. Approximately one half of these samples were evidence samples and one half are known samples. One half of the evidence samples are sexual assault samples. DNA extractions were done using phenol/chloroform or Chelex following Cellmark's standard operating procedures. DNA quantification using the Quantiblot kit was performed on all samples.

Samples containing as little as approximately 0.1 ng of DNA, as measured by the slot blot, produced results in all three systems. All samples which produced results using DQα and PM also produced results using the CTT triplex. Additionally, as would be predicted from validation studies, equivalent sensitivity for detection was found in both the PM and STR systems. Use of the two detection methodologies together helped confirm the presence of minor mixture components.

Thus far, all inclusions using DQα and PM have also been inclusions using the CTT triplex. Bases on a database of 105 Caucasians and 98 African Americans, the most common frequencies expected for these combined systems is 1 in 46,000 for Caucasians and 1 in 156,000 for African Americans. The STR CTT triplex is as robust as the DQα/PM systems. The use of two types of detection systems provides additional information for the interpretation of mixed samples.

Analysis of Mexican, Guatemalan and El Salvadoran populations for non-random association of alleles

Donald T Jones, San Bernardino Sheriff's Scientific Investigation Division, 200 South Lena Road, San Bernardino, CA 92415-0056

A total of 260 Hispanic samples (100 Mexican, 72 Guatemalan and 88 El Salvadoran) were typed for seven genetic markers: esterase D, phosphoglucomutase I, erythrocyte acid phosphatase, adenosine deaminase, adenylate kinase, group specific component and DQα. The results were examined for non-random association of alleles using the observed variance in the number of heterozygous loci for each subgroup as well as the entire data set. No association was detected among the seven markers in any subgroup or in the total:

S^2 MEXICAN = 1.252, 95% limits 0.869–1.507
S^2 EL SALVADORAN = 0.955, 95% limits 0.834–1.432
S^2 GUATEMALAN = 1.296, 95% limits 0.881–1.507
S^2 HISPANICS = 1.165, 95% limits 0.988–1.377

Prevention of selective probe signal reduction on amplitype PM and HLA DQα1 typing strips

Rebecca Reynolds PhD, Michael Grow and Vince Phillips, Roche Molecular Systems, 1145 Atlantic Avenue, Alameda, CA 94501

Alleles of the HLA DQα1, LDLR, GYPA, HBGG, D7S8 and GC loci can be detected using sequence specific oligonucleotide probes immobilized on a nylon membrane strip. In many cases, interpretation of typing results is dependent not only on the presence of probe signals, but also on their relative intensities. Because probe signal intensity is an essential consideration for interpretation, factors that can influence it need to be identified. Clearly, the time and temperature of the assay steps and the salt concentration in the typing solution can affect probe signal intensity. Also, if heat-denatured PCR products are allowed to cool for several minutes, the strands will reanneal and become unavailable for binding to the probes immobilized on the strips. However, the selective loss of GC B and HLA DQα1 4.1 probe signals observed after shorter cooling times cannot be explained by these factors. We demonstrate that following heat denaturation of PM PCR products there is sufficient residual Taq DNA polymerase activity to extend primers as the solution cools and that this primer extension occurs at a more rapid rate than PCR product reannealing. Primer extension across probe binding sites will prevent hybridization of the PCR product to complementary probes on the strip. We recommend a simple modification to the Amplitype™ typing protocol to ensure all probe binding sites will be available for hybridization.

Short Tandem Repeat (STR) polymorphism analysis using energy transfer fluorescent primers

J Atherton Wallin and George F Sensabaugh, PhD, University of California, Berkeley, Forensic Science Group, School of Public Health

Y Wang and R Mathies, University of California, Berkeley, Chemistry Department

We have demonstrated previously the advantage of Energy Transfer (ET) primers in STR analysis using both slab gel

electrophoresis and capillary array electrophoresis. ET primers excite at a single laser wavelength yet have distinctly non-overlapping and intense emissions, making unambiguous multi-color detection possible. This feature can be exploited in detecting sample and standard and/or multiple samples in the same lane or capillary. Fluorescence detection also allows electrophoretic patterns to be represented as electropherograms; comparison of peak intensities aids in the interpretation of mixed samples.

In this presentation, we extended our studies using the THO1 marker as a model system and additionally demonstrate multicolor detection of multiple STR loci, including CSF, D18S51, FES, TPO, VWF and D20S85. We have used a CAE unit with a confocal fluorescence detection system; fluorescence detection on slab gels has been achieved using the FluorImager 575 (Molecular Dynamics). The dyes used in primer labeling cause electrophoretic mobility shifts in PCR products; mobility shifts are observed between ET labeled PCR products and non-ET fluorescein dye-labeled THO1 ladder (Promega). Mixtures of heterozygote samples (both with and without common alleles) were analyzed quantitatively to determine parameters for the characterization of mixed samples.

Directionality determination of a bullet strike on soft body armor and identifying a bullet from a barrel that was damaged after firing

Scot Armstrong, California Department of Justice, 440A Airport Boulevard, Watsonville, CA 95076

In the spring of 1994 an officer with a small central coast police department reported that he had been shot by an unknown assailant. The bullet that struck his vest was submitted to the laboratory so that a list of weapons that may have fired it could be compiled. The list included three weapons, one of which was his backup pistol. His pistol was submitted with a part missing, the barrel hammered out of round and deep scratches in the muzzle end of the barrel. The pistol was restored to operating condition. Test fired ammunition could not be identified to the pistol. However, by pushing a lead bullet into the undamaged breech end of the barrel and pushing it back out, an identification was made. The officer's soft body armor showed evidence of directionality and indicated that the bullet would have had to pass through the unbroken windshield of the patrol unit. The damage to the vest was consistent with it being shot when it was hanging up with no backing material. The hole in the shirt did not line up with the hole in the vest. Finally, the wound received by the officer was not consistent with a bullet striking a vest. The officer pleaded guilty and received a felony conviction.

Observation of .22 caliber gunshot residue on fabric targets by scanning electron microscopy

Bryan Burnett, Forensic Science Consulting Group, 2037 San Elijo Avenue, Cardiff, CA 92007

Preparation of fabric for viewing in the scanning electron microscope can now be accomplished in minutes without the need for carbon or metal coating. Fabric associated particles, smaller than 0.2 microns, can be observed and analyzed for elemental composition using this technique.

Burnett showed that for close-range pistol shots to fabric using jacketed 9 mm and .357 ammunition, much of the gunshot residue (GSR) impacting the target is molten. However, a study with .22 LR ammunition fired by pistol shows a different pattern. Almost all the .22 ammunitions observed produce an abundance of submicron GSR which is not molten when these particles impact the target. For some .22 ammunitions, the GSR particle size and form is remarkably uniform whereas with others there is a diversity of particle sizes and forms. Distance between the muzzle and target is also a factor on the nature of the particles deposited on the target. Most .22-generated GSR is likely to become airborne during normal shooting activities and, due to the submicron size, is respirable to the lower airways. Such ammunition can pose a serious health hazard if the shooter or bystanders are downwind or stay within a GSR cloud.

Jin Bu Huan poisoning: a toxicological analysis of tetrahydropalmatine

Daniel T Anderson, MS, Los Angeles County, Dept. of Coroner 1104 N. Mission Road, Los Angeles, CA 90033

Tetrahydropalmatine (THP) or Jin Bu Huan anodyne is an alkaloid found in plants of the *Carydalis* species. The drug is considered to be a natural herb; part of traditional Chinese medicine with sedative and tranquilizing effects. It has been around for more than 1000 years as an analgesic, but only available in the United States for the past ten years.

Documentation of the drug's involvement in patients is limited only to toxicity, not lethality. This paper will present a case study report involving a female who committed suicide with a lethal ingestion of tetrahydropalmatine.

The analysis of tetrahydropalmatine consisted of an n-butylchloride extraction procedure; screening and quantitation on a GC/NPD. Confirmation of the drug was performed on a GC/MS by comparison to a purified Jin Bu Huan anodyne tablet.

The tetrahydropalmatine concentrations detected in postmortem blood of a 50 year old female were 7.2 mg/L. The gastric levels suggested that there was 3.8 g or 141 pills ingested. To my knowledge, this is the first reported fatality involving the drug, tetrahydropalmatine.

Mitochondrial DNA typing using sequence–specific oligonucleotide probes

J Varlaro, BS, E Clark, Ph.D., T Marschak, BA, S Rubin, MA, & R Reynolds, Ph.D. Roche Molecular Systems, Human Identity Group, 1145 Atlantic Avenue, Alameda, CA 94501

Mitochondrial (mt) DNA analysis is extremely valuable to the characterization of biological samples containing a small amount of DNA or highly degraded nuclear DNA. Since mt DNA occurs in high copy number (1,000 –10,000 copies per cell), it is possible that a mt DNA type can be obtained when a nuclear DNA type cannot. Mitochondrial DNA is also useful in the identification of human remains (e.g., missing persons, mass disaster) not only because of its high copy number but also because it is inherited matrilineally; all maternal relatives will share the same mt DNA sequence as the individual in question and can serve as a reference. We are developing an immobilized sequence-specific oligonucleotide probe-based assay to detect sequence polymorphisms in the hypervariable region II (HVII) segment of the mt DNA control region. A panel of fourteen HVII probes, immobilized as lines rather than dots, yields a power of discrimination comparable to DQα (approximately 0.94). Consequently, this HVII test can serve not only as a screen prior to mt DNA sequencing but also as a valuable PCR-based DNA typing test for the routine analysis of bloodstains and hairs. An instrument recently developed for automated strip hybridization, wash and color development, has greatly reduced the amount of "hands on" time of assay. Future developments will provide for computer scanning and typing of the HVII strip to further simplify (automate) the assay and maximize throughput. Results will be presented that demonstrate the assay's high sensitivity and its ability to amplify samples containing Taq DNA polymerase inhibitors, as well as typing results generated from the automated typing procedure.

Yiwen Wang[1*]
Jeanette M. Wallin[2**]
Jingyue Ju[1***]
George F. Sensabaugh[2]
Richard A. Mathies[1]

[1]Department of Chemistry
[2]Forensic Science Group, School of Public Health, University of California, Berkeley, CA USA

High-resolution capillary array electrophoretic sizing of multiplexed short tandem repeat loci using energy-transfer fluorescent primers

Short tandem repeat regions (STRs) from the polymorphic loci VWFA, THO1, TPO and CSF were amplified by the multiplex polymerase chain reaction (PCR) and analyzed by capillary array electrophoresis with fluorescence detection of energy transfer (ET) labels. The fluorescent ET primers are labeled with one fluorescein at the 5′ end and a second fluorescein at the position of the 7th or 9th (modified) base to produce fragments that fluoresce in the green (λ_{max} = 525 nm). M13 A-track sequencing fragments, used as an internal sizing standard, were generated with a universal primer that has a donor fluorescein at the 5′ end and a rhodamine acceptor at the position of the 11th (modified) base to produce fragments fluorescing in the red (>590 nm). The labeled DNA fragments were excited at 488 nm, and the fluorescence was detected with a two-color confocal fluorescence scanner. Separations were performed on arrays of hollow fused silica capillaries filled with denaturing and replaceable hydroxyethyl cellulose sieving matrices. Separations were complete in less than 50 min, and single base resolution as well as reproducible STR sizing was achieved. The relative standard deviation in sizing was below 0.6%. This work establishes the feasibility of high-resolution, high-speed and high-throughput STR typing of single-stranded DNA fragments using capillary array electrophoresis.

1 Introduction

DNA sequences containing di-, tri-, tetra- and pentanucleotide repeats are often genetically polymorphic [1–3]. Over two thousand of these short tandem repeat (STR) polymorphisms have been mapped on the human genome and it is estimated that thousands more remain to be discovered [4, 5]. Because of the abundance of this type of polymorphism and the relative ease of STR detection following amplification by the polymerase chain reaction (PCR), STRs or microsatellites have found widespread use as markers in gene mapping studies [4] and are emerging as potential markers for use in testing for paternity, personal identity [6–9] and genetic diseases [10, 11]. The analysis of STR markers in gene mapping and in DNA profile population data banks is typically performed by slab gel electrophoresis. As the number of interrogated STR markers increases and as one desires to have a more extensive polymorphism data base, the speed, throughput and sample handling associated with slab gels becomes limiting. We first demonstrated that capillary array electrophoresis (CAE) can be used to increase the speed and throughput of CE-based DNA analysis [12–14]. Several alternative methods for detecting electrophoretic separations performed on capillary arrays have subsequently been presented [15, 16]. In this technique, the speed of electrophoretic separations is increased ~10-fold through the use of high electric fields, and the throughput is increased by using large numbers of capillaries in an array. High sensitivity detection of fluorescently labeled DNA fragments is achieved through the use of a laser-excited, confocal-fluorescence scanner [17, 18]. An additional advantage of CAE is that samples are easily loaded in parallel through electrokinetic injection techniques. Thus far, we have demonstrated the utility of CAE for DNA sequencing [12–14, 19], for DNA fragment sizing [20, 21] and for STR sizing of double-stranded DNA fragments [22].

In our previous work, we performed rapid size analysis of human tyrosine hydroxylase locus (THO1) fragments by performing separations of ds-DNA on capillaries filled with replaceable nondenaturing hydroxyethylcellulose (HEC) sieving solutions. We also demonstrated the application of energy-transfer (ET) dye-labeled primers, recently developed by Ju and co-workers [23–25], to amplify the THO1 allelic ladder and unknowns for two-color sizing experiments. With ET primers we can excite the standard and unknown samples with a common laser wavelength while detecting intense and nonoverlapping emissions. Thus, accurate allele assignments are facilitated. However, the use of nondenaturing separation matrices can have several disadvantages. First, single base-pair resolution cannot be routinely achieved, making it difficult to identify single base insertion/deletion variants. Second, extra peaks due to the formation of heteroduplex structures sometimes make interpretation and multiplexing difficult. In this paper, we demonstrate that single base resolution and accurate multi-

Correspondence: Prof. Richard A. Mathies, Department of Chemistry, University of California, Berkeley, CA 94720, USA (Tel: +510-642-4192; Fax: +510-642-3599; E-mail: rich@zinc.cchem.berkeley.edu)

Nonstandard abbreviations: **CAE,** capillary array electrophoresis; **ET,** energy transfer; **FAM** or **F,** 6-carboxyfluorescein; **HEC,** hydroxyethylcellulose; **nt,** nucleotide; **ROX** or **R,** 6-carboxy-X-rhodamine; **STR,** short tandem repeat; **TBE,** Tris-borate-EDTA buffer

Keywords: Capillary array electrophoresis / Short tandem repeat / Energy transfer primers / Microsatellites

* Performed as a part of the Ph. D. thesis research of YW in the Graduate Group in Biophysics

** Current address: Perkin-Elmer Corp., Applied Biosystems Division, 850 Lincoln Center Drive, Foster City, CA 94404, USA
*** Current address: Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304, USA

0173-0835/96/0909-1485 $10.00+.25/0

plexed STR sizing can be routinely achieved by using denaturing, replaceable sieving matrices and an M13 A-termination ladder as the standard [7, 26, 27]. Multiplexed STR samples are amplified with ET primers that are detected in the green channel, and the M13 A-extension is produced with an ET primer detected in the red channel. The denaturing sieving matrices exhibit separations as fast those performed under native conditions, single base resolution, and the absence of interference from heteroduplex structures.

2 Materials and methods

2.1 Instrumentation

CAE separations were detected using the laser-excited, confocal-fluorescence scanner previously described by Huang *et al.* [12–14]. Briefly, excitation light at 488 nm from an argon ion laser is reflected by a long-pass dichroic beam splitter, passed through a 20 ×, N.A. 0.5 microscope objective, and brought to a 10 μm diameter focus within the 75 μm ID capillaries in the capillary array. The fluorescence is collected by the objective, passed back through the first beam splitter to a second dichroic beam splitter that separates the red (λ>565 nm) and green (λ<565 nm) detection channels. The emission is then focused on 400 μm diameter confocal pinholes, spectrally filtered by a 590 nm long-pass filter (red channel) or a 20 nm band-pass filter centered at 520 nm (green channel), followed by photomultiplier detection. The output is amplified, filtered, digitized, and then stored in an IBM 386 computer. A computer-controlled stage is used to translate the capillary array past the optical system at 10 mm/s. The fluorescence is sampled bidirectionally at 1000 Hz/channel. The scanner construction and operation have recently been described [18]. Postacquisition image processing was performed with the programs IPLab, KaleidaGraph and Canvas.

2.2 Capillary electrophoresis

Polyacrylamide-coated, fused-silica capillaries were prepared using a modification of the procedure described by Hjertén [28]. A 2–3 mm wide detection window was produced by burning off the polyimide coating with a hot wire followed by cleaning the external surface with ethanol. The detection window was placed 25 cm from the injection ends of the 75 μm ID, 350 μm OD, 50 cm long fused silica capillaries (Polymicro Technologies, Phoenix, AZ). The inner walls of the capillaries were washed with 1 N NaOH for 45 min at room temperature, followed by rinsing with acetonitrile for 5 min. Then, 40 μL γ-methacryloxypropyltrimethoxysilane was added to 10 mL acetonitrile with 15 μL acetic acid. The solution was forced slowly through the capillaries under N_2 pressure for 2 h. The capillaries were then rinsed again with acetonitrile. Freshly made 3%T acrylamide solution in 1/2 × TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8.3) was filtered with a 0.2 μm syringe filter and degassed under vacuum for 30 min. Five μL tetraethylmethylenediamine (TEMED) and 10 μL of 10% ammonium persulfate (APS) solution were added to 1 mL of gel solution. The acrylamide was forced slowly through the capillaries with N_2 pressure for 45 min. Then the capillaries were flushed with deionized water and filled with buffer consisting of HEC dissolved in 1 × TBE. To prepare the sieving buffer, the HEC is initially purified by stirring over ion-exchange resin [29]. A 50 mL solution is prepared by dissolving 1.25 g HEC (M_n = 90 000–105 000, Polysciences, Warrington, PA) and 22.5 g urea in 32.5 mL water. This solution is stirred overnight with 0.5 g Amberlite TM MB-1 ion-exchange resin (Alfa, Ward Hill, MA). The solution is then transferred to 25 mL tubes, centrifuged in a table-top centrifuge for 30 min, and the matrix is decanted off of the pelleted resin. To 24 mL of the purified matrix, 3 mL of 10 × TBE and 3 mL formamide are added; the final solution is stirred and degassed under vacuum for 1 h. The matrix is then dispensed into 1.5 mL microcentrifuge tubes and centrifuged for 10 min at 14 000 rpm. The top 800 μL of the solution in each tube is then transferred to the buffer reservoir. Capillaries were prerun at 200 V/cm for 3 min before each experiment. PCR samples mixed with an M13 A-termination standard were injected by inserting the capillary in a 3 μL sample volume held in an Eppendorf tube followed by electrokinetic injection (200 V/cm for 20 s). The sample tubes were replaced with tubes containing 2% HEC, 1 × TBE, 6 M urea, 10% formamide buffer after injection. Electrophoresis was performed at 200 V/cm using 5-capillary arrays held at ambient temperature (22°C). Capillaries are rinsed with water and methanol between the runs. When the experiments were complete, capillaries were flushed with water, then with methanol, and then dried.

2.3 PCR amplification of multiplexed STR loci

DNA was isolated from blood by using standard methods [30]. PCR multiplex amplifications for VWFA, THO1, TPO, and CSF loci were performed in 25 μL volumes using 10 ng genomic DNA template, 0.5 μM each of forward fluorescent and reverse blank primers for VWFA, THO1, TPO and CSF loci, 2.5 units of *Taq* DNA polymerase, 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl, pH 8.3, and 200 μM dNTPs (final concentrations indicated). The PCR cycle protocol using a Perkin Elmer Cetus Model 480 was: (1) melting at 95°C for 3 min, (2) 28 cycles at 95°C for 1 min, then 59°C for 1 min, then 72°C for 1 min, and (3) 72°C for 2 min to complete extension. The samples used in this study were drawn from a reference collection that was independently typed in single locus amplifications at 10 or more loci; single locus typings were performed essentially as described [2, 6, 31] with the modification that slab gels were stained with SybrGreen and detected by fluorescent imaging on a Molecular Dynamics FluorImager 575. The three samples selected for study include most of the common alleles at the four loci in the multiplex; one sample also illustrates the THO1 9.3, 10 type where the allelic PCR products differ by one base [31].

2.4 M13 sequencing ladders

M13 sequence ladders were prepared using M13mp 18 template DNA and modified T7 DNA polymerase [24]. Sequenase buffer (2 μL, 200 mM Tris-HCl, 100 mM

VWFA-A (F8F) FAM-5'-GAAAGCCCT*AGTGGATGATAAGAATAAT-3'
|
$(CH)_2(CO)$-NH-$(CH_2)_6$-NH-FAM

VWFA-B 5'-GGACAGATGATAAATACATAGGATGGATGG-3'

THO1-A (F6F) FAM-5'-GTGGGCT*GAAAAGCTCCCGATTAT-3'
|
$(CH)_2(CO)$-NH-$(CH_2)_6$-NH-FAM

THO1-B 5'-ATTCAAAGGGTATCTGGGCTCTGG-3'

TPO-B (F8F) FAM-5'-GGAGGAACT*GGGAACCACACAGGT-3'
|
$(CH)_2(CO)$-NH-$(CH_2)_6$-NH-FAM

TPO-A 5'-ACTGGCACAGAACAGGCACTTAGG-3'

CSF-A (F8F) FAM-5'-AACCTGAGT*CTGCCAAGGACTAGC-3'
|
$(CH)_2(CO)$-NH-$(CH_2)_6$-NH-FAM

CSF-B 5'-TTCCACACACCACTGGCCATCTTC-3'

M13 (F10R) FAM-5'-GTTTTCCCAGT*CACGACG-3'
|
$(CH)_2(CO)$-NH-$(CH_2)_6$-NH-ROX

Scheme 1. Structures of the PCR primers used for the amplification of the VWFA, THO1, TPO and CSF loci. The fluorescent primers are labeled with a common fluorescein donor (F) at the 5′ end and either a second fluorescein or a rhodamine (R) acceptor at the indicated locations of a modified T in the sequence. The number of nucleotides between the two fluorophores is indicated in the primer designation.

$MgCl_2$, 250 mM NaCl, pH 7.5) was combined with 2 μL of M13 universal primer labeled with energy-transfer dyes (0.4 μM) and 4 μL of template DNA (0.2 μg/μL). The mixtures were annealed by heating at 65°C for 2 min and slowly (~35 min) cooling to <30°C. Three μL of ddATP/dNTP mix was then added and the reaction mixture was warmed to 37°C for 2 min. Then, 4.5 μL of a freshly diluted mixture of T7 DNA polymerase (2 units μL) containing yeast pyrophosphatase (0.23 units/mL) was added and incubation continued at 37°C for 30 min. The reaction mixture was then stopped by adding 20 μL of 50 mM EDTA, 40 μL of 4 M ammonium acetate and 300 μL of ethanol. After 2 h at −20°C, the precipitated DNA was collected by centrifugation (12000 *g*) for 15 min and washed twice with 150 μL 70% ethanol. The precipitated DNA was vacuum dried and resuspended in 3 μL 98% formamide containing 1 mM EDTA. The M13 DNA A-extension was mixed with the multiplexed PCR sample (final solution 70–80% formamide), heated at 90°C for 2 min, cooled on ice for 30 s and electrokinetically injected into the capillary.

2.5 Design and synthesis of PCR primers

Chemicals were purchased from Applied Biosystems (Foster City, CA). Oligodeoxynucleotides were synthesized by the phosphoramidite method on an Applied Biosystems 392 DNA synthesizer. The structures of the blank and energy-transfer dye-labeled PCR primers for VWFA, THO1, TPO, and CSF loci are presented in Scheme 1. Primer sequences followed published sequences (THO1 [2], TPO [32], CSF [6], VWFA-B [7]) except VWFA-A, which was redesigned to avoid hairpins and dimer formation; as a result, VWFA products are 5 bp longer than those reported by Kimpton *et al.* [7]. The nomenclature and procedures for preparing and purifying energy-transfer fluorescent primers have been described by Ju *et al.* [23–25]. The energy-transfer dye-labeled primers are advantageous for two-color fragment sizing because the 488 nm exciting light is optimally absorbed by the 6-carboxyfluorescein (FAM) chromophore in these primers followed by enhanced emission at the FAM wavelength for the THO1 locus (amplified with F6F) and for the VWFA, TPO and CSF loci (amplified with F8F primers), or very distinctively Stokes-shifted emission following energy transfer in the case of the M13 ladder (generated with the F10R primer) [22, 24]. Primers were dissolved in 10 mM Tris-HCl, 1 μM EDTA buffer at a final concentration of 10 μM for PCR reactions and 0.4 μM for the M13 sequencing reaction.

3 Results and discussion

3.1 Two-color sizing of multiplexed STRs

Figure 1 presents CAE separations of the VWFA – THO1 – TPO – CSF tetraplex from three different individuals. Denaturing and replaceable HEC separation sieving matrices were used to achieve rapid (<50 min) separations. The VWFA, TPO and CSF loci were amplified with an F8F primer and the THO1 locus was amplified with an F6F primer. The data in Fig. 1 demonstrate that the four loci are evenly amplified. Since they are amplified with the F8F or F6F primers, they are only detected in the green channel. The cross-talk in the red M13 channel is low and can be easily removed by subsequent computer analysis. Each locus is heterozygous except the CSF locus in trace C. In these runs, each allele is represented by doublet peaks; the first peak of each doublet is the nominal allelic PCR product and the second peak is a non-templated one-base addition product [33]. Nontemplated base addition is a well-recognized PCR phenomenon. Although not employed in this study, two approaches can be taken to by-pass these two peak patterns. Nontemplated base addition can be partially to completely reversed by treatment of PCR products with Klenow enzyme prior to electrophoresis [33, 34]. Alternatively, nontemplated base addition can be driven to completion by extending the final PCR extension reaction to 2 h or more (Walsh, Fourney, personal communication).

In this study, we have investigated the use of an M13 sequencing ladder as a universal fragment sizing calibration standard; this standard is easily prepared from commercially available M13 templates. M13 A-, T-, C- and G-track ladders were individually prepared and tested singly and in combinations. The single-track and pair-track ladders were evaluated for eveness of peak distribution over the sizing range of interest (100–350 nucleotides), position of landmark peaks to put the ladder into register, and linearity of the correlation of size

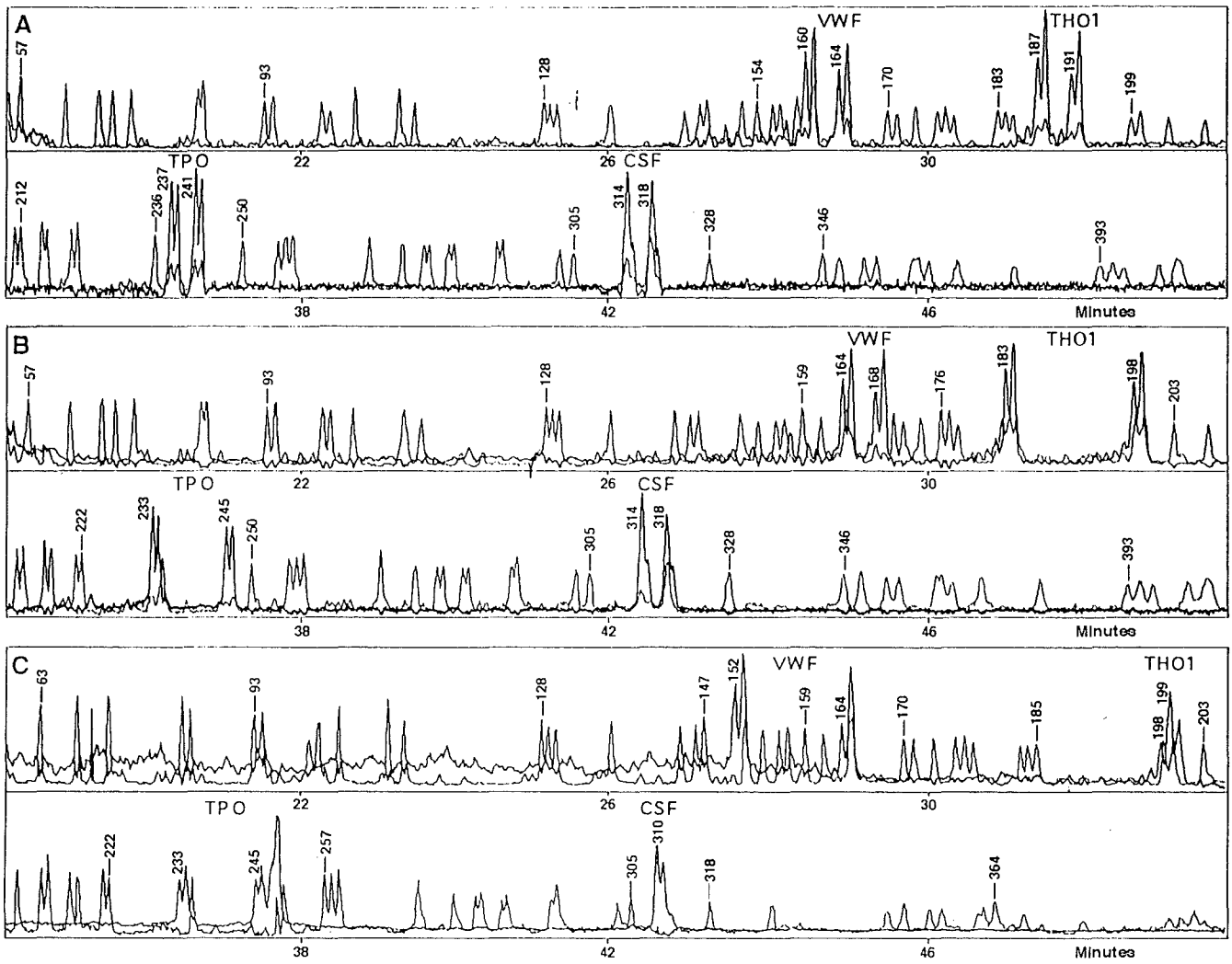


Figure 1. Representative electropherograms of three different multiplexed STR samples typed for VWFA, THO1, TPO and CSF loci (green). Each set of the fragments is sized against an M13 A-termination standard generated with an F10R primer. Electrophoresis was performed with replaceable 2% HEC, 1 × TBE, 6 M urea, 10% formamide sieving matrices at 200 V/cm. This figure has been processed with matrix transformation to correct for the spectral cross-talk between channels. The samples in traces (A), (B), and (C) had been previously typed at the four loci as follows: (A) VWF 17,18; THO1 7,8; TPO 9,10; CSF 11,12; (B) VWF 18,19; THO1 6,9.3; TPO 8,11; CSF 11,12; (C) VWF 15,18; THO1 9.3,10; TPO 8,11; CSF 10,10. Nominal sizes in bp (based on primary sequence data) for the allelic products at each locus are listed in Table 1.

to mobility. The M13 A-track was found to provide the best combination of peak spacing and linear correlation of size to mobility. Some compressions appeared in the C- and G-track ladders although the deviation from a linear correlation was small. The T-track had large gaps over the size range of interest. The M13 A-track was produced using an F10R-labeled sequencing primer and was detected in the red channel.

Single base resolution is routinely achieved for the separation of M13 A-fragments up to 400 bp. The importance of single base resolution is illustrated in the THO1 locus typing in Trace C. This individual has the THO1 alleles 9.3 and 10; the 9.3 and 10 allele amplicons differ by a single base at 198 and 199 bp, respectively [31]. Trace C shows three clearly resolved peaks for THO1 at ~198, 199, and 200 bp as determined against the M13 A-track ladder. The first peak is the 198 bp product of the 9.3 allele. The more intense middle peak is the overlap of the nontemplated one-base addition product of the 9.3 allele and the nominal 199 bp product of the 10 allele. The last peak is the one-base addition product of the 10 allele. This triplet peak pattern contrasts with the doublet peaks seen for allelic products separated by 4 bp or more. The individual profiled in Trace B, for example, has THO1 alleles 6 and 9.3; the 9.3 allele is represented by the doublet at *ca.* 198 and 199 bp, the nominal allelic amplicon and its one-base addition product. Thus the triplet structure clearly distinguishes the 9.3, 10 heterozygote type from the 9.3 and 10 homozygote types. Trace C also shows an extra peak ~2 bp bigger than the 245/246 bp doublet at the TPO locus; this is an artifact of unknown origin which appeared in this particular multiplex amplification but not in other single TPO amplifications from this individual.

The relationship between fragment size and migration time for the M13 A-track is illustrated in Fig. 2; this plot is derived from the data from trace A in Fig. 1 and is typical of the plots obtained in other runs. The plot

Table 1. Statistical analysis of multiplexed STR fragment sizing using CAE

Locus	Length bp/rept[a)]	Determinations	Mean size[b)]	SD[c)] %	Mean size[d)]	SD[c)] %
VWFA	147/15	3	152.3	0.83 (0.54)	151.0	0.38 (0.25)
(AGAT/C)	155/17	3	160.6	0.20 (0.12)	160.0	0.14 (0.09)
	159/18	9	164.3	0.80 (0.49)	163.9	0.60 (0.37)
	163/19	3	168.4	0.97 (0.58)	167.9	0.71 (0.42)
THO1	183/6	3	183.2	0.34 (0.18)	183.6	0.20 (0.11)
(AATG)	187/7	3	187.2	0.04 (0.02)	187.9	0.01 (0.01)
	191/8	3	191.1	0.03 (0.02)	191.9	0.07 (0.04)
	198/9.3	6	197.9	0.18 (0.09)	198.8	0.11 (0.06)
	199/10	3	198.8	0.04 (0.02)	199.6	0.07 (0.04)
TPO	232/8	6	232.6	0.39 (0.17)	233.9	0.48 (0.21)
(AATG)	236/9	3	236.8	0.01 (0.01)	238.2	0.05 (0.02)
	240/10	3	241.1	0.07 (0.03)	242.4	0.11 (0.05)
	244/11	6	245.0	0.38 (0.15)	246.6	0.78 (0.32)
CSF	307/10	3	310.0	0.50 (0.16)	309.6	0.52 (0.17)
(AGAT)	311/11	6	314.4	0.39 (0.12)	314.1	0.37 (0.12)
	315/12	6	318.6	0.39 (0.12)	318.2	0.34 (0.11)

a) Fragment lengths as determined by sequence analysis. The number of tandem repeats of the indicated sequence is also indicated.
b) Mean PCR product size as determined by global linear regression using all M13 A-termination fragments in the range 136-355 nt as the sizing standard.
c) Standard deviation in terms of base pairs for the indicated number of determinations. The percent relative SD is given in parentheses.
d) Mean PCR product size as determined by local linear regression using 3–7 M13 A-termination fragments immediately surrounding the unknown fragment.

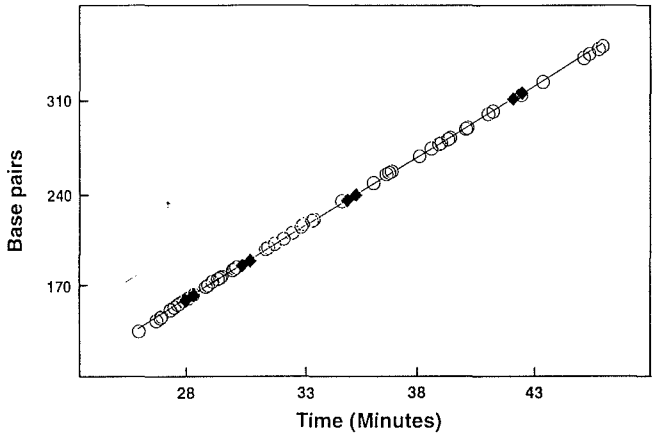


Figure 2. Demonstration of the linear relationship between fragment size and their migration time in the region of 136 bp to 355 bp for trace (A) in Fig. 1. The M 13 A-track ladder fragments (red channel) are represented by the open circles (O) and the 4 locus multiplex fragments (green channel) are represented by the filled diamonds. Allele fragment sizing can be done using a global regression line on all the M 13 ladder fragments as shown here or on a set of M 13 fragments surrounding the allele fragment in question.

shows near-perfect linearity over the fragment size range 136–355 nt with a linear correlation coefficient r = 0.99984; small deviations from linearity are present for small fragments (<160 nt) and for large fragments (>300 nt). Using a global regression over the 136–355 nt range, the determined sizes of the allelic fragments from trace A (in nt) are 160.6 and 164.4 for VWFA; 187.2 and 191.1 for THO1; 236.9 and 241.1 for TPO and 314.5 and 318.7 for CSF. Sizes were also determined using a local regression on the 3–7 M13-A fragments immediately surrounding the unknown. Using this approach, the determined fragment sizes were close to the global values: 160.0 and 164.2 for VWFA; 187.8 and 191.9 for THO1; 238.2 and 242.4 for TPO and 314.3 and 318.4 for CSF. Sizing determinations using both global and local regression analysis are summarized in Table 1. Both approaches yield comparable values with regard to sizing accuracy and reproducibility. The local approach is marginally better with regard to sizing precision but not so much so as to outweigh the inconvenience of computing multiple local regressions. Both approaches yield sizes for the allelic fragments at the VWFA and CSF loci that are *ca.* 5 and 3 nt larger, respectively, than the true fragment sizes; the THO1 and TPO fragment sizes are closer to the true values. This may be due to migration differences associated with the sequence of the repeat, AGAT/C in the case of VWFA and CSF and AATG for THO1 and TPO. For the practical objective of identifying genetic types at a locus, however, reproducibility in fragment sizing ± <1 nt coupled with single base resolution are more important than sizing accuracy; these two conditions allow definition of criteria for the reliable identification of alleles.

4 Concluding remarks

This work demonstrates the advantages of CAE with ET primers for high-resolution, high-speed and high-throughput typing of polymorphic short tandem repeat loci. We routinely achieve single base resolution using denaturing HEC; thus this technology and method should be equally useful for sizing di- and trinucleotide repeats. Second, the replaceable HEC solutions used as the sieving media provide high speed (<50 min) separations on columns that can be easily refilled and rerun. Third, we can amplify multiple PCR targets with ET primers. This provides an additional increase in the speed and throughput of STR typing. Fourth, ET primers provide higher sensitivity and better color discrimination than that possible with conventional single dye-labeled fluorescent PCR primers. Finally, since the CAE format

provides the ability to run up to 50 or more separations in parallel [35], this work establishes the feasibility of generating forensic data bases with high throughput CAE STR sizing.

We thank the members of the Berkeley High-Sensitivity DNA Analysis Group for their support and Alexander N. Glazer for valuable discussions. This research was supported by a grant from the Director, Office of Energy Research, Office of Health and Environmental Research of the US Department of Energy under contract DE-FG-91ER61125 to RAM, and by a grant from the National Institute of Justice (93-IJ-EX-0010) to GFS. J. Ju was supported by a Human Genome Distinguished Postdoctoral Fellowship sponsored by the US Department of Energy, Office of Health and Environment Research, and administered by the Oak Ridge Institute for Science and Education. Support from Amersham Life Sciences Inc. is also gratefully acknowledged.

Received March 20, 1996

5 References

[1] Tautz, D., *Nucleic Acids Res.* 1989, *17*, 6463–6471.
[2] Edwards, A., Civitello, A., Hammond, H. A., Caskey, C. T., *Am. J. Hum. Genet.* 1991, *49*, 746–756.
[3] Charlesworth, B., Sniegowski, P., Stephan, W., *Nature* 1994, *371*, 215–220.
[4] Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M., Weissenbach, J., *Nature Genetics* 1994, *7*, 246–339.
[5] Beckmann, J. S., Weber, J. L., *Genomics* 1992, *12*, 627–631.
[6] Hammond, H. A., Jin, L., Zhong, Y., Caskey, C. T., Chakraborty, R., *Am. J. Hum. Genet.* 1994, *55*, 175–189.
[7] Kimpton, C. P., Gill, P., Walton, A., Urquhart, A., Millican, E. S., Adams, M., *PCR Methods Applicat.* 1993, *3*, 13–22.
[8] Frégeau, C. J., Fourney, R. M., *BioTechniques* 1993, *15*, 100–119.
[9] Alford, R. L., Hammond, H. A., Coto, I., Caskey, C. T., *Am. J. Hum. Genet.* 1994, *55*, 190–195.
[10] Cawkwell, L., Bell, S. M., Lewis, F. A., Dixon, M. F., Taylor, G. R., Quirke, P., *Br. J. Cancer* 1993, *67*, 1262–1267.
[11] Mao, L., Schoenberg, M. P., Scicchitano, M., Erozan, Y. S., Merlo, A., Schwab, D., Sidransky, D., *Science* 1996, 659–662.
[12] Huang, X. C., Quesada, M. A., Mathies, R. A., *Anal. Chem.* 1992, *64*, 967–972.
[13] Huang, X. C., Quesada, M. A., Mathies, R. A., *Anal. Chem.* 1992, *64*, 2149–2154.
[14] Mathies, R. A., Huang, X. C., *Nature* 1992, *359*, 167–169.
[15] Ueno, K., Yeung, E. S., *Anal. Chem.* 1994, *66*, 1424–1431.
[16] Takahashi, S., Murakami, K., Anazawa, T., Kambara, H., *Anal. Chem.* 1994, *66*, 1021–1026.
[17] Quesada, M. A., Rye, H. S., Gingrich, J. C., Glazer, A. N., Mathies, R. A., *BioTechniques* 1991, *10*, 616–625.
[18] Mathies, R. A., Scherer, J. R., Quesada, M. A., Rye, H. S., Glazer, A. N., *Rev. Sci. Instrum.* 1994, *65*, 807–812.
[19] Kheterpal, I., Ju, J., Radhakrishnan, A., Brandt, G. S., Ginther, C. L., Clark, S. M., Scherer, J. R., Sensabaugh, G. F., Mathies, R. A., *The Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE)* 1996, Vol. 2680, pp. 204–213.
[20] Clark, S. M., Mathies, R. A., *Anal. Biochem.* 1993, *215*, 163–170.
[21] Zhu, H., Clark, S. M., Benson, S. C., Rye, H. S., Glazer, A. N., Mathies, R. A., *Anal. Chem.* 1994, *66*, 1941–1948.
[22] Wang, Y., Ju, J., Carpenter, B. A., Atherton, J. M., Sensabaugh, G. F., Mathies, R. A., *Anal. Chem.* 1995, *67*, 1197–1203.
[23] Ju, J., Ruan, C., Fuller, C. W., Glazer, A. N., Mathies, R. A., *Proc. Natl. Acad. Sci. USA* 1995, *92*, 4347–4351.
[24] Ju, J., Kheterpal, I., Scherer, J. R., Ruan, C., Fuller, C. W., Glazer, A. N., Mathies, R. A., *Anal. Biochem.* 1995, *231*, 131–140.
[25] Ju, J., Glazer, A. N., Mathies, R. A., *Nature Medicine* 1996, *2*, 246–249.
[26] Urquhart, A., Oldroyd, N. J., Kimpton, C. P., Gill, P., *BioTechniques* 1995, *18*, 116–121.
[27] Smith, R. N., *BioTechniques* 1995, *18*, 122–128.
[28] Hjertén, S., *J. Chromatogr.* 1985, *347*, 191–198.
[29] Bashkin, J., Marsh, M., Barker, D., Johnston, R., *Appl. Theor. Electrophor.* 1995, *4*, 141–146.
[30] Sambrook, J., Fritsch, E. F., Maniatis, T., *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 1989.
[31] Puers, C., Hammond, H. A., Jin, L., Caskey, C. T., Schumm, J. W., *Am. J. Hum. Genet.* 1993, *53*, 953–958.
[32] Huang, N. E., Schumm, J., Budowle, B., *Forens. Sci. Int.* 1995, *71*, 131–136.
[33] Clark, J. M., *Nucleic Acids Res.* 1988, *16*, 9677–9686.
[34] Savill, M. G., Sensabaugh, G. F., in: Bar, W., Fiori, A., Rossi, U. (Eds.), *Advances in Forensic Haemogenetics*, Springer-Verlag, Berlin 1994, Vol. 5, pp. 575–577.
[35] Bashkin, J. S., Roach, D., Rosengaus, E., Barker, D. L., *Human Genome Program, Contractor-Grantee Workshop III, U.S. Department of Energy* 1993, pp. 116.

Indu Kheterpal[1]
James R. Scherer[1]
Steven M. Clark[1]
Arun Radhakrishnan[1]
Jingyue Ju[1]*
Charles L. Ginther[2]
George F. Sensabaugh[2]
Richard A. Mathies[1]

[1]Department of Chemistry
[2]Forensic Science Group, School of Public Health, University of California, Berkeley, CA, USA

DNA sequencing using a four-color confocal fluorescence capillary array scanner

The design, construction and operation of a four-color capillary array electrophoresis scanner are presented. The use of sensitive energy transfer primers facilitates four-color detection of the DNA sequencing fragments following excitation at a single laser wavelength (488 nm). This scanner collects fluorescence data from up to 25 capillaries in parallel. The resulting four-color image files are automatically reduced to four-color line plots, and a base-calling program (Sax) is used to call the sequence. The performance of this system for DNA sequencing is demonstrated by examining twelve different motifs of the hypervariable region I of human mitochondrial (mt) DNA obtained from a Sierra Leone population.

1 Introduction

The goal of the Human Genome Project is to determine the complete nucleotide sequence of the 3 billion base-pair human genome [1]. High speed, high throughput and extremely sensitive methods are required to complete this challenging task in a reasonable amount of time. Capillary electrophoresis (CE) has emerged as an excellent analytical tool for fast and sensitive DNA analysis [2–4]. The high surface area-to-volume ratio due to the small size of capillaries facilitates efficient heat dissipation. This allows the use of high voltages to achieve high speed and high resolution separations [5]. A number of groups have now developed methods for DNA sequencing using CE [2–4, 6–13].

All of the capillary electrophoresis DNA sequencing methods cited above use only one capillary, resulting in low sample throughput. To address this limitation, Mathies and Huang [14] developed the first capillary array electrophoresis (CAE) system using a confocal fluorescence scanner. Its application to DNA sequencing was demonstrated by running 25 capillaries simultaneously, increasing the production rate of sequencing data by 25-fold [14–17]. Since then, other CAE instruments have been developed utilizing either the multiple sheath flow [18] or on-column detection [19] using charge-coupled devices.

The Sanger dideoxy chain termination method [20] is most widely used for large-scale DNA sequencing. This method produces four different chain termination ladders, one for each base. Due to the run-to-run variation in the electrophoretic mobility of DNA fragments in CE, it is necessary to mix sequencing fragments ending in all four dideoxynucleotides and separate them in the same capillary run [4]. Therefore, the application of CE methodology to DNA sequencing requires detection methods capable of identifying four distinct terminal nucleotides. Even though a large amount of progress has been made on the development of one- [6, 13] and two-color methods [6–8, 11, 16], four-color sequencing is most commonly used in large-scale DNA sequencing projects. This method uses four fluorophores, one for each base, and sequence is read based on the color of the peaks. It results in the longest read lengths and highest accuracy rates for sequencing. However, the four fluorophores commonly used for DNA sequencing do not have high molar absorbances at a single excitation wavelength. Hence, the four-color instruments developed thus far generally require the use of two excitation wavelengths [3, 6, 9, 18] resulting in complicated optical systems and generally suboptimal excitation.

We have addressed the problem of optimizing the dye labels for DNA sequencing by developing energy transfer (ET) primers containing a common donor at the 5'-end of the primer and an acceptor dye several nucleotides away [10, 21–23]. The presence of a common donor facilitates the excitation of all four fluorophore pairs with one laser at 488 nm. The use of these primers leads to simple instrumentation using only one excitation wavelength and dramatically enhances fluorescence signal strengths. The application of ET primers to sequencing [10, 21–25] and short tandem repeat sizing [26–28] has been described before and these primers are now used by many genome sequencing centers.

We describe here the design and operation of our current four-color confocal fluorescence CAE scanner. This scanner is capable of running up to 25 capillaries in parallel, and the sequence data is collected, processed and called automatically. We have tested this instrument by sequencing each of the 12 motifs of the hypervariable region I of human mitochondrial (mt) DNA from a West African (Sierra Leone) population (C. L. Ginther *et al.*, unpublished) using ET primers. The mtDNA is of great interest in evolutionary studies [29] and is a valuable tool for maternally linked individual identification [30]. This sequence region provides a real world test for the CAE scanner because different patterns of sequence variation are present in the 12 motifs. This work also provides a basis from which to predict the performance of 48 capillary [31] or 96 capillary array machines.

Correspondence: Prof. Richard A. Mathies, Department of Chemistry, University of California, Berkeley, CA 94720, USA (Tel: +510-642-4192; Fax: +510-642-3599; E-mail: rich@zinc.cchem.berkeley.edu)

Nonstandard abbreviations: CAE, capillary array electrophoresis; ET, energy transfer; HVR, hypervariable region; mt, mitochondrial; PMT, photomultiplier tube; TBE, Tris-boric acid-EDTA

Keywords: DNA sequencing / Capillary electrophoresis / Confocal fluorescence detection / Energy transfer primers / Mitochondrial DNA polymorphisms

* Current address: Incyte Pharmaceuticals Inc., 3174 Porter Drive, Palo Alto, CA, USA

0173-0835/96/1212-1852 $10.00+.25/0

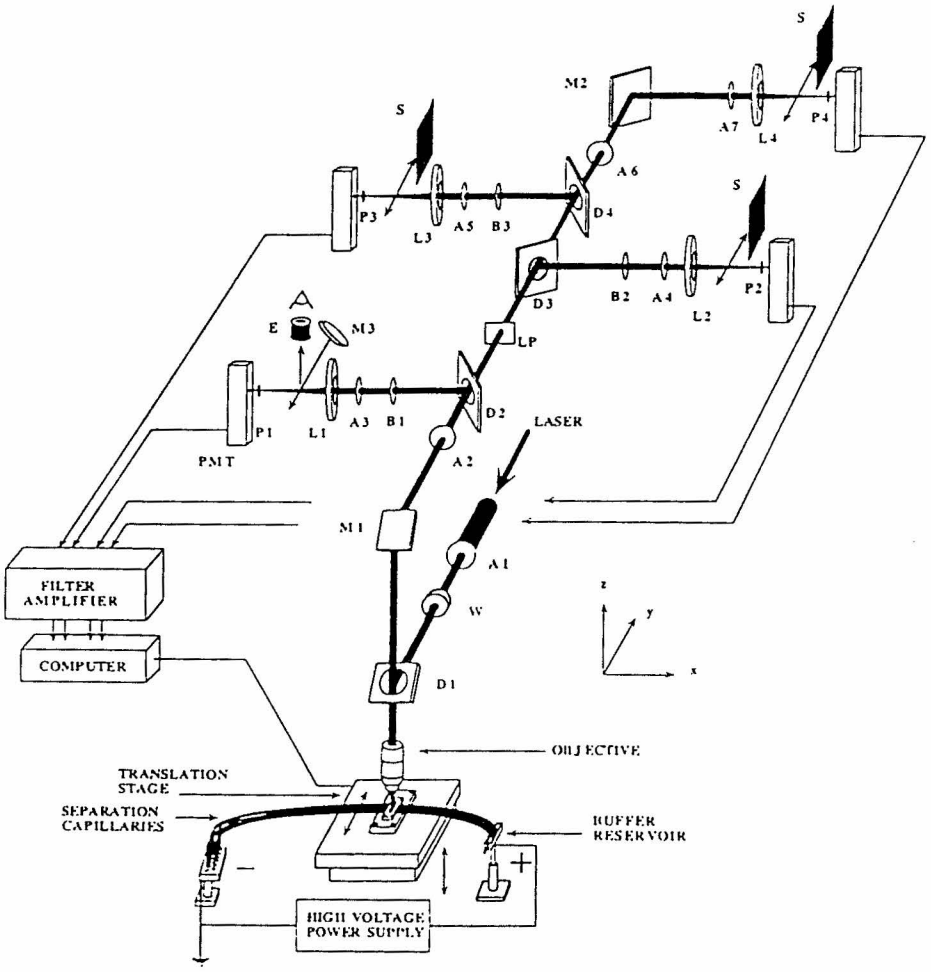


Figure 1. Schematic of the four-color capillary array electrophoresis scanner. A: iris diaphragm; B: band-pass filter; D: dichroic beamsplitter; E: eyepiece; L: lens; M: mirror; P: pinhole; S: shutter; W: window.

2 Instrumentation

2.1 Four-color capillary array scanner design

A schematic of the four-color capillary array electrophoresis scanner is shown in Fig. 1. The scanner design is based on our single capillary four-color electrophoresis system [10], with the addition of a translation stage (105021P-LH, Daedal Division, Parker, Harrison City, PA) and a microstep indexer (SX57-83, Compumotor Division, Parker). An argon ion laser (Spectra-Physics, Model 2020, Mountain View, CA) beam is used for excitation (488 nm, ~ 3 mW). The laser beam is expanded to 10 × and collimated before introduction into the optical system.

The four-color instrument consists of five connected compartments: one for sample excitation and fluorescence (called the objective compartment), and four sequential modular color-detection units. The whole instrument is supported by four posts on a movable base (guided by fixed rails) and can be moved and repositioned reproducibly for convenient capillary array loading and instrument alignment. The beam is directed under the color detection units of the instrument and enters the objective compartment through an iris diaphragm, A1. It passes through a $^1/_2$ inch thick glass window, W, and is deflected downward (-z) by a dichroic beam splitter, D1 (Omega Optical, Brattleboro, VT)*. All dichroic beam splitters are used at 45°. The beam is focused into the capillary array through an achromat microscope objective (20×, 0.5 NA, Rolyn Optics, Covina, CA) used with an infinite conjugate. The capillaries are placed in a grooved holder, insuring that they all lie in the same focal plane with the same periodicity. Grooves (90°, 110 μm deep) were made by a computer-controlled milling machine with lateral translation accuracy of 5 μm. The computer alignment program allows the user to visually define on screen the individual capillary positions. The translation stage which supports them rests on a second vertical (z) stage to allow the array to be brought into the focus of the objective.

The capillary surface has a high curvature at the glass/air interface. Consequently, rays entering the objective along the x direction are focused to a region near the lower portion of the inside volume of the capillary and rays entering along the y direction are focused near the upper portion of the inside volume. The difference between the foci can be as much as 50–75 μm

* Beamsplitter designation (D1, D2, D3, D4); % R @ 488 nm (98, 74, 20, ~26); Range nm > 90% R (425–490, 497–548, 520–567, 495–585); nm > 50% T (505, 560, 578, 596); Bandpass filters: (B1, 525DF30), (B2, 555DF30), (B3, 580DF24).

depending on the inside and outside diameter of the capillary and the numerical aperture (NA) of the objective. Contrary to the usual correlation between NA and diffraction-limited spot size when focusing through plane surfaces, smaller excitation volumes are achieved by decreasing the NA.

Fluorescence from the capillaries is collected by the objective (at full NA), passed through D1, and deflected by mirror M1 (through iris A2) into the first of the four-color detection compartments. Dichroic beamsplitter D2 reflects wavelengths 497–548 nm and the bandpass filter B1 limits the wavelength range to 510–540 nm. Iris A3 and the location of pinhole P1 (Melles-Griot, Boulder, CO) are used to define the unfocused beam path and appropriate dichroic position. An achromat lens L1 (105 mm focal length, Rolyn Optics), provided with xyz adjustment, focuses the color-filtered beam on a 200 μm confocal pinhole P1. A diagonal mirror, M3, which can be moved into the beam path between L1 and P1, creates an image plane at the top cover of the compartment, and a 20X Ramsden eyepiece E (Rolyn Optics) is used to view the image plane. Light passing through the pinhole, P1, is detected by a photomultiplier tube (PMT; Hammatsu HC120-05, Photonics Research Corp., San Jose, CA). The beam that passes through D2 enters the second compartment, which is a mirror image of the first detection unit. A removable longpass glass filter LP (OG530, Omega Optical) is used for additional blocking of the laser wavelength. Each successive compartment measures a longer wavelength range defined by the respective bandpass filters. Compartments 2, 3 and 4 have a sliding shutter, S, in place of a diagonal mirror to block light from entering the photomultipliers. Compartment 4 contains a mirror in place of a dichroic beamsplitter and no bandpass filter is used to limit the long wavelength detection range. Plasma lines in the primary laser beam are significant beyond 590 nm and are removed using a line filter (10LF10-488, Newport, Irvine, CA). The specific transmission and reflection characteristics of dichroic beamsplitters and bandpass filters are listed in the footnote. The outputs of the photomultiplier tubes (PMTs) are amplified and filtered using 500 Hz low-pass filters (SR 640, Stanford Research Systems Inc., Sunnyvale, CA). The signal is digitized using a 16-bit ADC operating at 1 KHz (DAS-HRES, Keithley Instruments, Inc. Taunton, MA). The intensity data are collected bidirectionally and stored in a PC.

2.2 Alignment

The primary laser beam is adjusted to parallel the y direction and intersect the center of the iris A1 at the entrance to the objective compartment. A far aperture is placed in the laser beam path about two feet before entering A1. The objective is removed and the instrument is moved away from the translation stage so that the laser beam strikes the optical table (our primary reference for the xy plane). A front surface mirror with parallel top and bottom surfaces is placed at the beam position. D1 is adjusted so that the laser beam returns on itself (all dichroics and mirrors are provided with orthogonal adjustments). The unaligned return beam can be seen on the instrument side of the far aperture. The front surface mirror is removed and the position of the beam striking the table is marked. The objective is replaced, iris A1 reduced to minimum, and the position of the beam is compared with the mark. The beam position is moved to match the mark by tilting the Gimbal mounted plane window, W. This adjustment ensures that the beam enters the center of the objective. The mirror is placed on top of the translation stage and the instrument is moved forward ($-y$) so that the focused laser beam strikes the mirror. The focus is adjusted until the return beam is slightly larger than the far aperture. The translation stage is now moved along y. If the return beam does not expand or contract on the far aperture, the stage is translating parallel to y. If this condition is not met, the stage must be shimmed. If all the adjustments are made correctly, the beam on the far aperture will expand and contract symmetrically when the mirror is driven in and out of focus (z).

The objective and the LP filter are removed. M1 is adjusted to make the return beam (transmitted by D1) from the front surface mirror center on A6. L1 and B1 are removed and D1 is positioned to make the beam pass through the middle of A3 and center on pinhole P1. The PMT, fastened on the outside of the compartments, is removed and L1 is replaced. A white card is placed a small distance beyond the pinhole and L1 is adjusted (yz) to pass the focused beam through the pinhole. The pinhole diameter (200 μm) is much larger than the spot size produced by L1 and can be used, as in the standard "knife edge test", to establish the focal distance for L1. The focus of L1 (position of minimum beam waist) is moved (x) until small translations of L1 in the y or z direction produce sudden complete cut-off or turn-on of the beam. Since the laser beam in front of L1 is parallel, the objective will be forced to function as if it had an infinite conjugate (even though it has no corrected infinite conjugate), thereby ensuring that the subsequent compartments will be sampling the same confocal region within the capillaries. The beam is left centered on the pinhole. The optics in the remaining compartments are adjusted in a similar manner and the PMTs are refastened on the exterior housings. Bandpass filters B1–3 and longpass filter LP are replaced.

Final adjustments on the optics are made using a 1 nM fluorescein solution, in 1 × TBE (90 mM Tris – 90 mM boric acid – 2 mM EDTA) buffer, pH 8.0, flowing through a stationary 100 μm ID (200 μm OD) capillary. The capillary is placed in an array holder and the laser beam brought to a focus within the flowing fluorescein solution. The diagonal mirror M3 is moved to intercept the beam between L1 and P1, and the eyepiece is positioned to bring the bright fluorescence cross, caused by the two capillary focii, into sharp focus. In a darkened room, with the covers of the compartments removed and voltage applied to the PMTs, the signal from compartment 1 is maximized by first adjusting the objective focus (z), and second, lateral adjustments (yz) on L1. When properly adjusted, signals from the four compartments will vary synchronously as the capillary is moved in and out of focus. Once the overall alignment has been established, reoptimization of the system is performed daily in < 20 min.

2.3 Data acquisition

A capillary array, with twelve separation capillaries and an alignment capillary on both sides, is placed on the computer-controlled translation stage (linearity: < 5 μm deviation over 1"-travel). Programs written in Quick C facilitate the array alignment and the movement of the stage during data collection. The alignment capillaries contain a fluorescein solution (1 nM, pH 8.0, 1 × TBE) drawn through the capillaries by house vacuum. The laser is focused on one of the alignment capillaries and the program takes 15 scans to create an image of the capillary array. The user marks the left and right inner boundaries of one of the fluorescein capillaries and the inner boundaries of the remaining capillaries are automatically located based on the 220 μm periodicity of the array holder.

During an electrophoresis run, the capillary array is scanned continuously at 1 cm/s in the $\pm y$ direction. An image of the migrating bands in the detection zone is created with bidirectional periodic 0.5 s sweeps. The transit time for the DNA fragments through the detection zone varies from 10–15 s and therefore at least 20 sweeps are made through a peak. A line image of the intensities collected during one sweep of the stage across the capillary array is displayed horizontally on the screen. Each subsequent scan is displayed below the previous scan, thereby producing an image of fluorescence across a line cross-section of the array with time. A 16-bit image file for each of the detection channels is recorded for each electrophoresis run. The fluorescence in each channel is sampled at 1000 Hz. The image resolution is 10 μm / pixel; therefore 10 pixels cover the entire 100 μm ID of the capillary in one sweep of the stage. The output is filtered using a 500 Hz electronic low-pass filter. Data processing, performed post run, automatically sums the 10 pixels across each capillary interior and the reduced four-color trace files are stored in a format readable by the base-calling program.

3 Materials and methods

3.1 Preparation of capillary arrays

The 50 cm long, 100 μm ID, 200 μm OD fused-silica capillaries (Polymicro Technologies, Phoenix, AZ) with a detection window 30 cm from the injection end were pretreated [17] and filled with linear polyacrylamide according to the procedure described by Zhang *et al.* [12]. Briefly, the capillaries were derivatized with a 2% solution of 3-(trimethoxysilyl)propylmethacrylate (Aldrich, Milwaukee, WI) in 95% ethanol at room temperature for 30 min. A linear polyacrylamide solution was prepared by adding 20 μL of 10% APS (ammonium persulfate, Bio-Rad, Hercules, CA) and 2 μL of TEMED (tetraethylmethylenediamine, Bio-Rad) to 5 mL of the freshly degassed 5% acrylamide in 7 M urea, 1 × TBE buffer. The capillaries were air dried and filled with the polyacrylamide solution and stored overnight at 4°C. Gel-filled capillaries were preelectrophoresed for 30 min while raising the voltage from 0–10 kV. The DNA sequencing samples were electrokinetically injected for 20 s at 150 V/cm and electrophoresed at 200 V/cm. The failure rates of injection and columns were 8% and 14%, respectively.

3.2 Energy transfer primers

The design and synthesis of the energy transfer primers for sequencing have been presented in detail by Ju *et al.* [10, 21, 22]. The primers used here were specifically synthesized for sequencing the light strand of the hypervariable region I (HVR I) of the D-loop of mtDNA and are denoted as L16040. The primer sequence is TCTGTTCTTT*CATGGGGAAG. These primers contain 6-carboxy-fluorescein (FAM) as the common donor at the 5'-end and an acceptor dye at the modified thymidine residue (T*) nine bases away from FAM. The acceptors used are 2', 7'-dimethoxy-4', 5'-dichloro-6-carboxy-fluorescein (JOE), *N,N,N',N'*-tetramethyl-6-carboxyrhodamine (TAMRA) and 6-carboxy-X-rhodamine (ROX).

3.3 DNA sequencing samples

The mitochondrial DNA samples had been previously sequenced as a part of a larger study on mtDNA sequence variation in Sierra Leone (C. L. Ginther *et al.*, unpublished). The samples were provided as single- and double-stranded hypervariable region I (15978–16517) templates captured on magnetic beads (Dynabeads, Dynal, Oslo, Norway) [32]. DNA sequencing fragments were produced using the Sanger dideoxy terminator method. The detailed protocols for the preparation of sequencing samples have been described before [25]. Briefly, 0.4 pmoles of the primer was mixed with amplified DNA (0.1–0.4 μg per reaction tube), ThermoSequenase reaction buffer and dNTP/ddNTP mix (Amersham Life Science Inc., Arlington Heights, IL). ThermoSequenase was added to the preheated mixture followed by cycle sequencing (15 ×). The samples were pooled, ethanol-precipitated and resuspended in 98% formamide and 1 mM EDTA. A 2 μL aliquot was used for injection.

3.4 Data processing

The four-color trace files obtained from the CAE sequencer were in a format readable by the base-calling program Sax written by Anthony J. Berno [33]. This program was modified by Alan B. Greenfield and Jeffrey R. Sachs (Daniel H. Wagner Associates, Sunnyvale, CA) to be operated on a Power Macintosh 7100/80. The Sax program uses a graph-theoretic approach to optimize the set of called bases. The process of automatic base calling is divided into several steps including low-pass Gaussian filtering to remove the noise, color separation to remove the cross-talk due to spectral overlap, mobility shift corrections from the use of different fluorophores, baseline subtraction, second and fourth derivative Fourier filtering to deconvolute overlapping peaks followed by Gaussian filtering of the output, and finally the base calling. An estimated base spacing parameter provided by the user is utilized in low-pass Gaussian filtering, baseline subtraction and deconvolution processes, and, most importantly, as a starting point for event editing to accurately determine the base calls [33]. The cross-talk matrix used for color separation can be calculated in advance or determined from the data itself. Mobility shift and mobility stretch factors affecting the peak spacing can also be modified by the user.

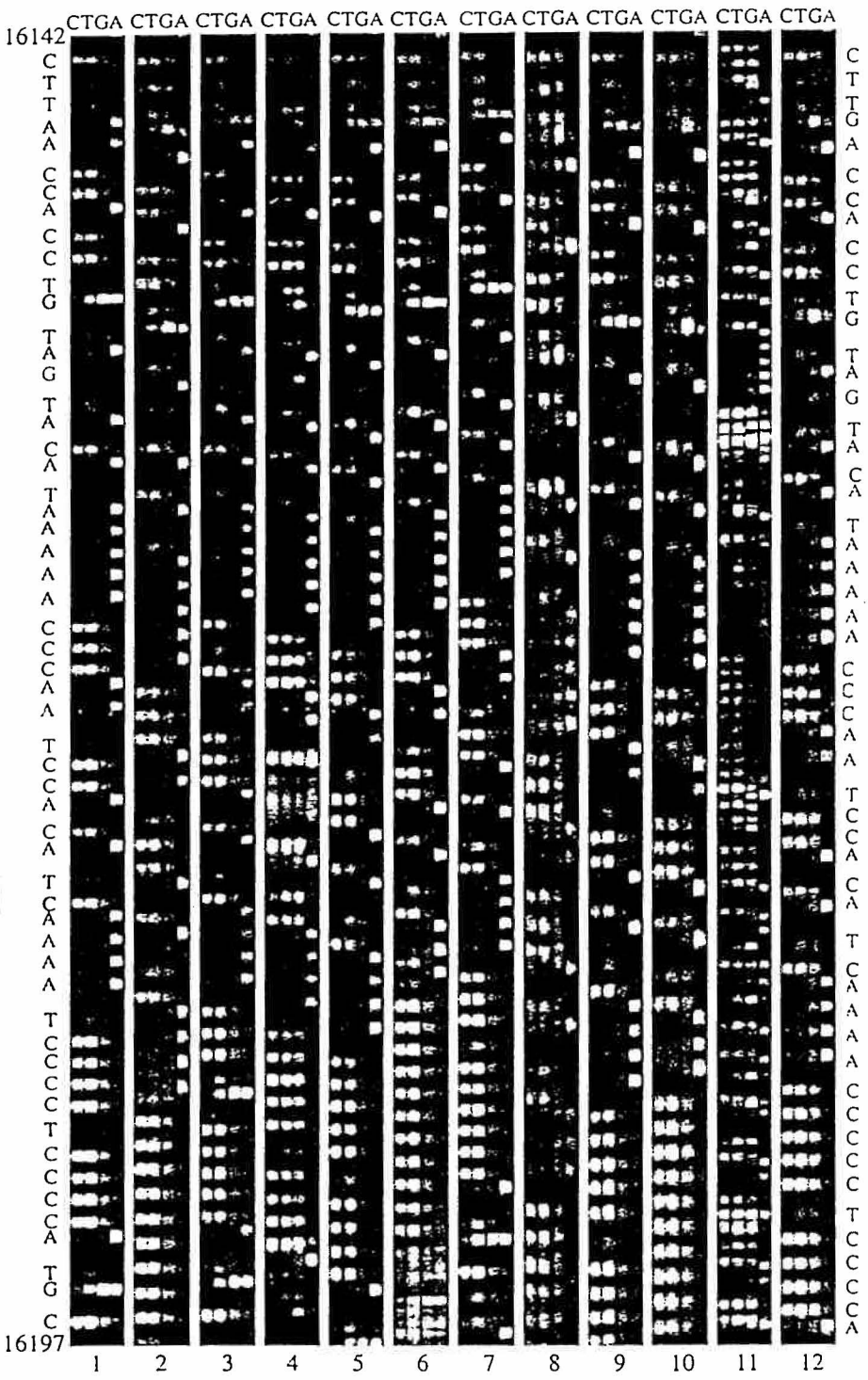


Figure 2. Image of the four-color DNA sequencing data from twelve different motifs of the hypervariable region I of mtDNA using a twelve-capillary array. All the sequences have been aligned at position 16142 (C). ssDNA templates and dITP were used to synthesize sequencing fragments except for samples in capillaries 4 and 11, which used dsDNA template and 7-deaza-dGTP.

4 Results and discussion

The image data from a twelve capillary array is presented in Fig. 2 for a segment of the light strand sequence from HVR I of human mtDNA. The image presented here shows a representative sequence from each of the twelve HVR I motifs found in Sierra Leone. Each capillary is represented with four different lanes; one for each detection color channel. The capillary images have been background-subtracted, filtered using a 5-point filter and color-contrasted for display. Thirteen minutes of data from each capillary, aligned at the top-most base (C, 16142) are shown in the image to display comparable sequences in each capillary. All the bases in each capillary are well resolved. The bands are observed in more than one lane due to spectral overlap in the fluorescence emission spectra of the primers. This overlap is constant and therefore is easily removed using the multicomponent analysis in Sax. The highest intensity band represents the terminating nucleotide and can be read manually in the images and electropherograms.

Table 1. Summary of mtDNA base calling parameters and results using Sax

Capillary number	Base spacing	A Stretch	G Stretch	C Mobility shift	T Mobility shift	Ambiguous calls (%)	Insertion errors (%)	Deletion errors (%)	Substitution errors (%)	Unedited calling rate (%)	Edited accuracy rate (%)
1	33	50	35	10	10	0.22	0	0.22	0	99.57	100
2	33	50	35	10	10	1.31	0	0	0	98.7	100
3	33	50	35	10	10	0.43	0.22	0	0	99.35	100
4	32	50	35	10	10	1.57	0	0	0	98.43	100
5	36	50	35	10	10	14.38	2.88	0.44	0	82.30	98.89
6	33	50	35	5	5	0.67	0	0	0	99.33	100
7	34	55	40	5	5	0.64	0.21	0.21	0	98.94	100
8	34	55	40	5	5	32.14	0	0	0	67.86	99.29
9	34	55	40	5	5	20.04	0.85	1.28	0.21	77.61	99.15
10	33	55	40	5	5	3.36	0	1.34	0	95.30	100
11	34	0	0	5	5	1.55	0.22	1.55	0	96.69	99.56
12	35	55	40	10	10	4.96	0.43	0.22	0.22	94.18	99.57

Capillaries 1–4: cross-talk matrix obtained from capillary # 1
Capillaries 7–11: cross-talk matrix obtained from capillary # 8

The sequence data obtained using the capillaries shown in Fig. 2 were called automatically using the Sax program and compared to the known sequence. The cross-talk matrix used to eliminate the spectral overlap was first calculated dynamically by the program. The program allows the user to save the last matrix used and therefore, depending on the performance of the program, matrices of other capillaries were tried. The source of the matrices used for final analysis are indicated in Table 1. Due to the run-to-run variation in the electrophoretic mobility of DNA fragments, the program allows the user to manually change default values of input parameters such as estimated base spacing, mobility shifts due to the use of different fluorophore pairs, and mobility stretch factors affecting spacing of the fragments close to the primer. These parameters influence the base-calling accuracy of the program. Various parameter settings were tried for each run and the settings resulting in the best calling rates were selected for the analysis. Table 1 lists the parameters used for each capillary run and the performance of the program in analyzing the twelve sequences presented here.

Ambiguous calls contribute most to the unedited calling rates obtained with Sax. One possible reason could be that the program as currently operated uses only one average base spacing for the entire run whereas the base spacing in a sequencing run changes according to the length of the fragments. The unedited base-calling rates vary from 68% to 99.6%. The performance of the program for capillaries 5, 8, and 9 is especially low. The data quality of these runs was comparable to the rest; the cause of this increase in ambiguous calls is therefore unclear. The algorithms used in this program made very few insertion, substitution and deletion errors. The edited accuracy rates given in Table 1 correspond to the sequencing accuracy rate obtained after all the ambiguous calls and the errors were visually inspected and, where possible, corrected. The time involved in editing the sequences presented here was ~5 min per run.

The color-separated and base-called files obtained from the Sax program can be plotted as electropherograms. Figure 3 presents an example of the electropherogram obtained for the sequencing run of capillary 7 in the array shown in Fig. 2. The HVR I template DNA is obtained by PCR amplification using 20 base long primers ending at positions 16498 and 15997. The sequencing primers extend from 16020–16040. Therefore, a total of 476 (16041–16517) bases can be sequenced. The sequence shown in Fig. 3 contains base information from positions 16043 to 16514. The entire PCR product with the exception of five bases is sequenced with single base resolution and 100% accuracy. The sequence run was complete in 130 min from the time of injection. An unedited base-calling rate of 99% was obtained for this sequence. The automatically called sequence contained three ambiguities, one deletion, and one insertion error, marked as N, D, and I, respectively, in Fig. 3. All five mistakes were easily identified in the electropherogram and corrected.

Several of the 12 motifs in the Sierra Leone population studied here contain mutations at bases 16183 and/or 16189, where A or T, respectively, is replaced by a C. These mutations result in a sequence containing 10 or more consecutive C's. The DNA polymerase can not extend through these C-rich regions and falls off. Therefore, it is necessary to sequence both strands of the mtDNA in order to obtain the complete sequence. The samples in capillaries 2, 6, 8 and 10 in the Fig. 2 contain a row of 10 or more C's. For this reason only ~ 150 base long sequences were obtained from these samples. The length of the sequences obtained in the rest of the capillaries varies from 460 to 473 bases from the primer.

We can envision several improvements that will make CAE sequencing even more practical. Preliminary work with the heavy strand sequence of the HVR I of mtDNA shows that the dITP analog of dGTP does not work well for G-rich sequences. It is necessary to use 7-deaza-dGTP and to separate sequencing fragments at elevated temperature to eliminate compressions and obtain high quality sequence [12]. We would like to sequence both strands of the HVR I of mtDNA with the same sequencing protocols and therefore plan to use 7-deaza-dGTP and high separation temperatures in the future. Also, we are currently using pre-filled capillaries which have to be removed after ~5 runs. Large-scale sequencing instruments using 25–100 capillaries would require the use of replaceable polymer matrix. The separation of sequencing fragments using replaceable

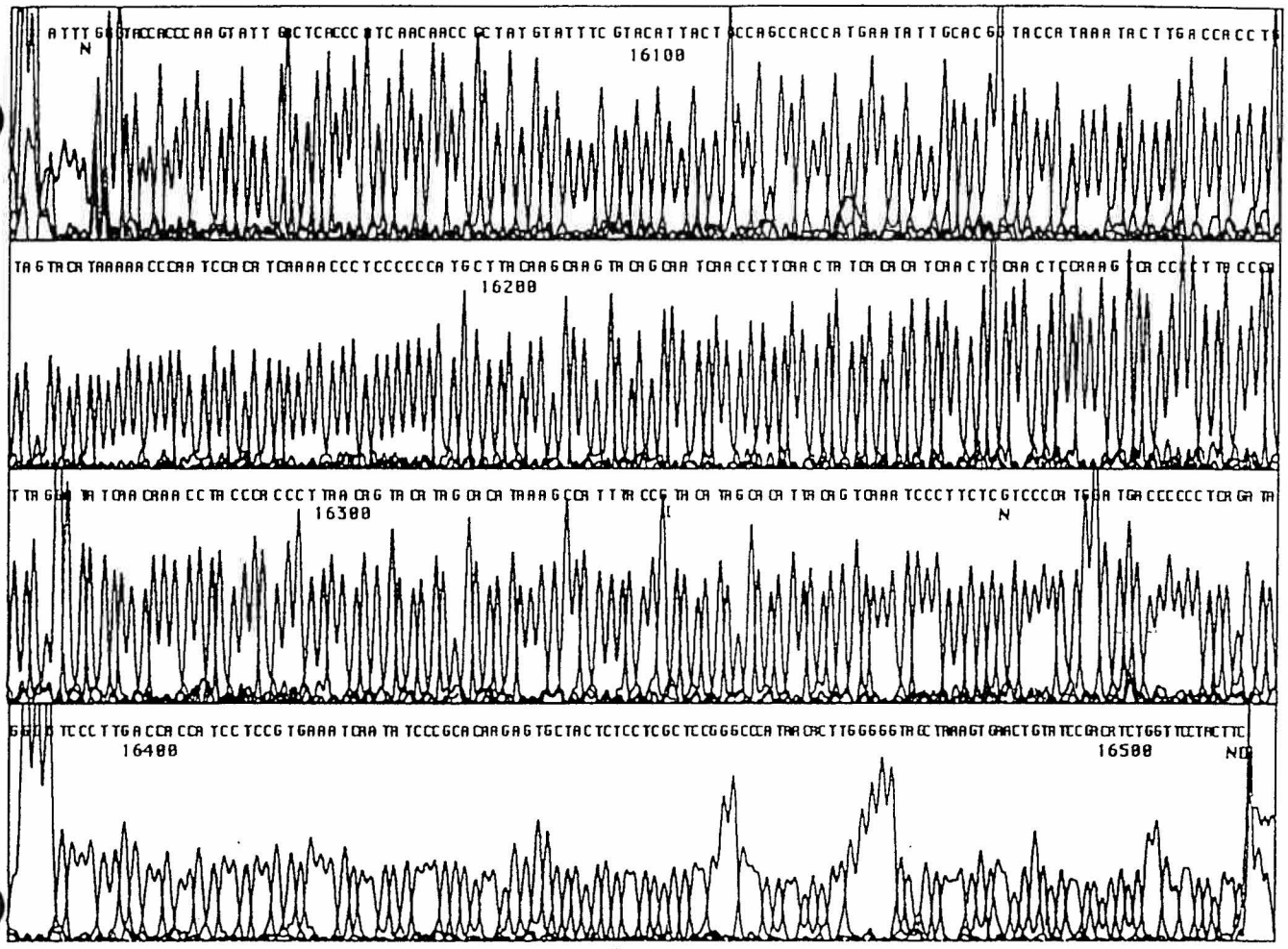


Figure 3. DNA sequence of the light strand of the hypervariable region I of the mitochondrial D-loop using CAE and ET primers. The primers used for each base were: C, Blue (F9F); T, Green (F9J); G, Black (F9T); A, Red (F9R). The electrophoresis time from the injection to the end of this run is 130 min. Ambiguous calls are marked by N, insertion errors by I and deletion errors by D. These data are from capillary 7 in Fig. 2.

matrices has been shown using linear polyacrylamide [8], mixed polyethylene oxide [11], and hydroxyethylcellulose [34] solutions. Finally, needed improvements in the base-calling program are currently being made (A. Greenfield and J. Sachs, personal communication). The new version of the program will contain improved default values and base-calling algorithms.

In summary, we have demonstrated that arrays of capillaries can be run in parallel and high quality sequence data obtained using our four-color confocal fluorescence capillary array scanner and ET primers. The light strand of the hypervariable region I of human mtDNA has been successfully sequenced with an average sequencing accuracy of 99.7% for the twelve samples examined here. Currently, we have two CAE scanners in operation that are being used for DNA sequencing. These scanners can detect up to 25 capillaries at a time and should be capable of a raw sequencing rate of ~ 6 kilobases/hour/scanner.

We thank Gabriel S. Brandt for technical assistance. We also thank Alan B. Greenfield and Jeffrey R. Sachs (Wagner Associates, Sunnyvale, CA) for the modification of base-calling program. This research was supported by the Director, Office of Energy Research, Office of Health and Environmental Research of the U.S. Department of Energy under Contract DE-FG-91ER61125, by the National Institutes of Health (HG01399-01), and by the National Institute of Justice (93-IJ-EX-0010). We also acknowledge generous additional support from Amersham Life Sciences Inc.

Received June 29, 1996

5 References

[1] Guyer, M. S., Collins, F. S., *Proc. Natl. Acad. Sci. USA* 1995, *92*, 10841–10848.
[2] Swerdlow, H., Gesteland, R., *Nucleic Acids Res.* 1990, *18*, 1415–1419.
[3] Luckey, J. A., Drossman, H., Kostichka, A. J., Mead, D. A., D'Cunha, J., Norris, T. B., Smith, L. M., *Nucleic Acids Res.* 1990, *18*, 4417–4421.
[4] Cohen, A. S., Najarian, D. R., Karger, B. L., *J. Chromatogr.* 1990, *516*, 49–60.
[5] Jorgenson, J. W., Lukacs, K. D., *Anal. Chem.* 1981, *53*, 1298–1302.
[6] Swerdlow, H., Zhang, J. Z., Chen, D. Y., Harke, H. R., Grey, R., Wu, S. Dovichi, N. J., Fuller, C., *Anal. Chem.* 1991, *63*, 2835–2841.
[7] Chen, D., Harke, H. R., Dovichi, N. J., *Nucleic Acids Res.* 1992, *20*, 4873–4880.

[8] Ruiz-Martinez, M. C., Berka, J., Belenkii, A., Foret, F., Miller, A. W., Karger, B. L., *Anal. Chem.* 1993, *65*, 2851–2858.
[9] Carson, S., Cohen, A. S., Belenkii, A., Ruiz-Martinez, M. C., Berka, J., Karger, B. L., *Anal. Chem.* 1993, *65*, 3219–3226.
[10] Ju, J., Kheterpal, I., Scherer, J. R., Ruan, C., Fuller, C. W., Glazer, A. N., Mathies, R. A., *Anal. Biochem.* 1995, *231*, 131–140.
[11] Li, Q., Yeung, E. S., *Appl. Spectrosc.* 1995, *49*, 1528–1533.
[12] Zhang, J., Fang, Y., Hou, J. Y., Ren, H. J., Jiang, R., Roos, P., Dovichi, N. J., *Anal. Chem.* 1995, *67*, 4589–4593.
[13] Williams, D. C., Soper, S. A., *Anal. Chem.* 1995, *67*, 3427–3432.
[14] Mathies, R. A., Huang, X. C., *Nature (London)* 1992, *359*, 167–169.
[15] Huang, X. C., Quesada, M. A., Mathies, R. A., *Anal. Chem.* 1992, *64*, 967–972.
[16] Huang, X. C., Quesada, M. A., Mathies, R. A., *Anal. Chem.* 1992, *64*, 2149–2154.
[17] Huang, X. C., Mathies, R. A., in: Adams, M. D., Fields, C., Craig Venter, J. (Eds.), *Automated DNA Sequencing and Analysis Techniques*. Academic Press, San Diego 1994, pp. 17–28.
[18] Takahashi, S., Murakami, K., Anazawa, T., Kambara, H., *Anal. Chem.* 1994, *66*, 1021–1026.
[19] Ueno, K., Yeung, E. S., *Anal. Chem.* 1994, *66*, 1424–1431.
[20] Sanger, F. Nicklen, S., Coulson, A. R., *Proc. Natl. Acad. Sci. USA* 1977, *74*, 5463–5467.
[21] Ju, J., Ruan, C., Fuller, C. W., Glazer, A. N., Mathies, R. A., *Proc. Natl. Acad. Sci. USA* 1995, *92*, 4347–4351.
[22] Ju, J., Glazer, A. N., Mathies, R. A., *Nature Medicine* 1996, *2*, 246–249.
[23] Ju, J., Glazer, A. N., Mathies, R. A., *Nucleic Acids Res.* 1996, *24*, 1144–1148.
[24] Woolley, A. T., Mathies, R. A., *Anal. Chem.* 1995, *67*, 3676–3680.
[25] Kheterpal, I., Ju, J., Radhakrishnan, A., Brandt, G. S., Ginther, C. L., Clark, S. M., Scherer, J. R., Sensabaugh, G. F., Mathies, R. A., *Proc. Soc. Photo-Opt. Instrum. Engineers* 1996, *2680*, 204–213.
[26] Wang, Y., Ju, J., Carpenter, B. A., Atherton, J. M., Sensabaugh, G. F., Mathies, R. A., *Anal. Chem.* 1995, *67*, 1197–1203.
[27] Atherton, J. M., Ju, J., Wang, Y., Carpenter, B., Mathies, R. A., Sensabaugh, G. F., *Proceedings from the 6th International Symposium on Human Identification*, Promega Corporation 1996, pp. 72–80.
[28] Wang, Y., Wallin, J. M., Ju, J., Sensabaugh, G. F., Mathies, R. A., *Electrophoresis* 1996, *16*, 1485–1490.
[29] Stoneking, M., *J. Bioenerg. Biomembr.* 1994, *26*, 251–259.
[30] Gill, P., Ivanov, P. L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., Sullivan, K., *Nature Genetics* 1994, *6*, 130–135.
[31] Bashkin, J., Bartosiewicz, M., Roach, D., Leong, J., Barker, D., Johnston, R., *J. Capil. Electrophor.* 1996, *3*, 61–68.
[32] Hopgood, R., Sullivan, K. M., Gill, P., *BioTechniques* 1992, *13*, 82–92.
[33] Berno, A. J., *Genome Res.* 1996, *6*, 80–91.
[34] Bashkin, J., Marsh, M., Barker, D., Johnston, R., *Appl. Theor. Electrophor.* 1995, *4*, 141–146.

Int J Legal Med (1997) 110:36–38

TECHNICAL NOTE

K. Yoshida · K. Sekiguchi · K. Kasai · H. Sato · S. Seta
G. F. Sensabaugh

Evaluation of new primers for CSF1PO

Received: 5 August 1996 / Received in revised form: 28 October 1996

Abstract We describe new primers for the detection of the STR polymorphism at the CSF1PO locus. These primers have been designed to produce shorter amplicons (150–182 bp) than the primers in standard use (295–327 bp). The reliability of the new primers for CSF1PO typing has been demonstrated by testing on known samples and by sequence analysis. These primers are superior to the original primers with regard to electrophoretic resolution and utility for typing of severely degraded DNA.

Key words CSF1PO · STR polymorphism · Forensic identification · Degraded DNA

Introduction

The tetranucleotide tandem repeat polymorphism at the CSF1PO locus (5q33.3-q34) is one of a number of short tandem repeat (STR) polymorphisms under consideration as a core marker for forensic identification purposes [1, 4–8]. The CSF1PO typing primers described by Hammond et al. [4] (and used in the GenePrint STR System typing kit marketed by Promega, Madison Wisc. [8]) yield amplicons in the size range 295–327 bp. Because fragments in this size range are less well resolved than fragments of smaller size and because smaller fragments are better amplified from samples containing severely degraded DNA [3, 5, 11, 15], we have investigated the use of redesigned CSF1PO primers giving smaller PCR products(150–182 bp). The use of these primers results in improved electrophoretic resolution of allelic fragments and enhances typing of degraded samples.

G. F. Sensabaugh
Forensic Science Group, School of Public Health,
University of California, Berkeley, CA 94720, USA

K. Yoshida (✉) · K. Sekiguchi · K. Kasai · H. Sato · S. Seta
National Research Institute of Police Science,
6 Sanbancho, Chiyoda-ku, Tokyo 102, Japan
FAX: +81 (3) 3221 1245

Materials and methods

DNA extraction

DNA extracted from blood samples was used for reference typing purposes. Aged bloodstains, 16 years old, were from a reference collection maintained at the National Research Institute of Police Science. DNA was extracted from whole blood and aged bloodstain samples by using a standard phenol-chloroform method [13]. The concentration and quality of extracted DNA was determined by UV absorbance using spectrophotometric readings at 260 and 280 nm and by yield gel electrophoresis, respectively [13].

Primer design

Prospective primer sequences were generated from the CSF1PO gene sequence (GenBank accession number X14720) using the program Designer PCR (Research Genetics, USA). The primer pair selected for this study was:

CSF-3F, 5′ GTTGCTAACCACCCTGTGTCTC 3′ (GenBank sequence 11874–11895)

CSF-3R, 5′ TTCCTGTGTCAGACCCTGTTC 3′ (GenBank sequence 12043–12023)

The Tm values calculated for these primers are 58.1°C and 58.6°C respectively. This primer set was expected to yield products in the size range 150–182 bp.

PCR conditions and detection of amplified products

PCR amplifications were performed on a Perkin Elmer 480 thermocycler. Reference samples were typed for CSF1PO using a commercial kit (GenePrint STR Systems, Promega, USA) according to the manufacturer's instructions.

Amplifications using the CSF-3 primers were done in 25 µL reaction mixes containing 5 ng template, 0.5 µM of each primer, 0.25 Unit Taq polymerase (AmpliTaq, Perkin Elmer, USA), 1 × Taq polymerase buffer (50 mM KCl, 10 mM Tris-HCl pH9.0, 1.5 mM $MgCl_2$, 0.1% Triton X-100), and 200 µM each of dATP, dCTP, dGTP and dTTP. The best combination of amplification efficiency and specificity was obtained using a basic PCR protocol similar to that described by Puers et al. [12] with initial denaturation at 96°C for 2 min, then 10 cycles of denaturation 94°C 1 min, annealing 58°C 1 min, extension 70°C 1.5 min, followed by 20 cycles of denaturation 90°C 1 min, annealing 58°C 1 min, extension 70°C 1.5 min, and ending with a terminal elongation at 70°C for 5 min. Raising the annealing temperature to 60°C and 64°C resulted in

the appearance of stronger stutter bands, i.e. N+4 and N-4 bands. Accordingly, 58° C was chosen as the standard annealing temperature.

Amplified DNA products were separated by electrophoresis through a 0.4 mm-thick, 30 cm long, 6% denaturing polyacrylamide gel (C = 5%) in 1 × TBE buffer containing 7 M urea. Samples of 2.5 µl amplified DNA were mixed with 2.5 µl of 2 × loading solution (10 mM NaOH, 95% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol FF; Promega, USA) and denatured at 95° C at 2 min. Electrophoresis was conducted at 35 W for approximate 1.5 h. The PCR products were stained by SYBR-GREEN (Molecular Probes, USA) and the bands were detected on a FluorImager (Molecular Dynamics, USA).

Sequence analysis

CSF1PO amplicons were sequenced using cycle sequencing (fmol DNA Cycle Sequencing System, Promega, USA). The purification of PCR products was performed by filtration though Ultrafree-MC 100,000 NMWL filter cups (Millipore, USA). Primers were end-labeled with ^{32}P using T4 polynucleotide kinase; procedures otherwise followed manufacturer's instructions. Sequence determination entailed electrophoretic separation on 5.5% acrylamide gels containing 7 M urea followed by autoradiography using standard protocols [13].

Results and discussion

Typing reliability

The accuracy of genetic typing using the CSF-3 primers was tested by typing 40 samples previously typed using the GenePrint kit. An allelic ladder based on the CSF-3 primers was constructed by amplifying a mixture of DNA samples of known type. K562 DNA (CSF1PO type 9,10) was used as a common reference for comparing the amplicons of the two primer sets. The typings using the CSF-3 and conventional primers were completely concordant. Figure 1 shows the relative positions and overall appearance of the allelic bands using the GenePrint kit and CSF-3 primers. Note that the samples amplified using GenePrint kit were loaded 1 h before the CSF-3 primer samples; the samples amplified with the GenePrint kit were not adequately resolved if they were loaded at the same time as the CSF-3 sample set. It is clear that even with an extra hour separation time, the larger amplicons are less well resolved than the CSF-3 primer samples.

The sizes of the amplicons obtained with the CSF-3 primers were determined by measuring end-labeled primer extension products against an M13 sequencing ladder, under denaturing gel conditions. Amplicons sized 1–2 bp longer than predicted based on primary sequence data and repeat number determined by conventional typing. For example, the allelic fragment with 12 repeats sized at 170 bp rather than the predicted 169 bp. Size shifts of one base can be attributed to non-templated base addition [2, 9]. This was demonstrated by treating amplicons representing different allelic repeats with Klenow enzyme, the 3′-5′ exonuclease activity of which removes the non-templated added base. Treated samples exhibited a 1 bp size reduction relative to the untreated controls. Any residual difference in the sizes of the CFS1PO amplicon relative to the M13 sequencing ladder may be due to sequence specific mobility shifts [14].

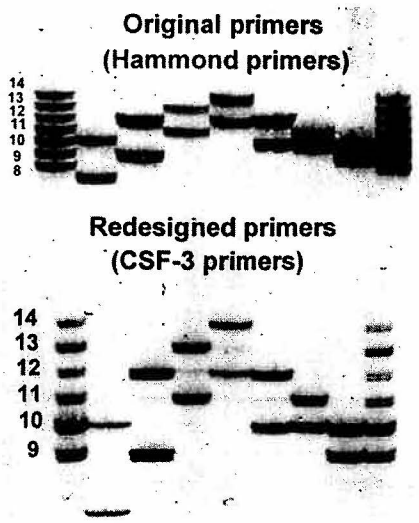


Fig. 1 Comparison of STR patterns obtained with the Promega primers (upper) and the CSF-3 primers (lower). Electrophoretic separation was performed as described in materials and methods. Samples amplified using Promega primers added to the gel 1 h before the CSF-3 primed samples; nevertheless, the CSF-3 primed samples migrate further and are better resolved than the Promega primers samples. The samples in each set are CSF1PO types (from left to right) ladder; 7,10; 9,12; 11,13; 12,14; 10,12; 10,11; K562(9,10); ladder. The ladder used with the Promega primer was supplied with GenePrint kit; the ladder used with the CSF-3 primers was constructed from a mixture of samples

The accuracy of CSF1PO typing using the CSF-3 primers was confirmed by sequence analysis. CSF-3 primed amplicons representing the 9, 10, 11, and 12 alleles (as determined by conventional typing) were sequenced. The sequences obtained were completely concordant with expectation, i.e. the sequence determined by sequence analysis matched the sequence expected based on conventional typing.

Degraded DNA

Because STR amplicons are relatively small in size, usually less than 350 bp, STR typing is particularly useful in the analysis of forensic samples containing severely degraded DNA [3, 5, 11, 15]. To determine whether the smaller CSF-3 amplicon might extend the value of this marker in typing degraded DNA, two sets of experiments were performed. First, we tested DNA artificially degraded by heating at 100° C for 30–150 min [10]. For degraded DNA samples containing very low molecular weight DNA (< 600 base pairs), amplification using the GenePrint kit gave very weak or no PCR products. CSF1PO types could be determined, however, when the CSF-3 primers were used. Second, we tested degraded DNA extracted from five 16-year-old bloodstain samples. Figure 2 shows the results obtained with the two primer sets; 40 ng degraded DNA(determined spectrophotometri-

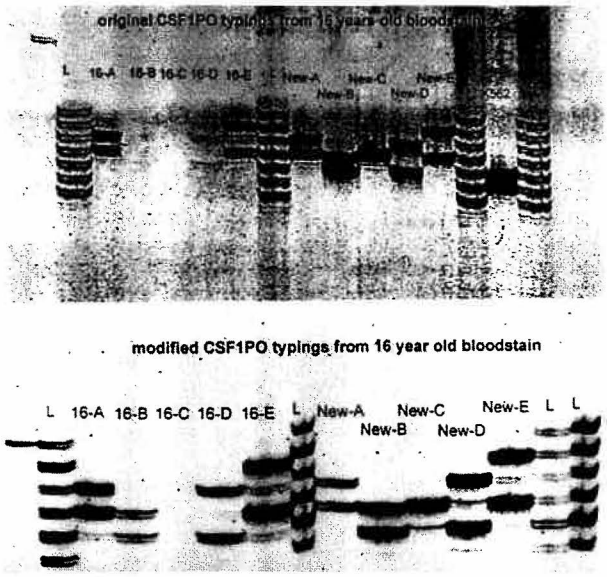


Fig. 2 CSF1PO typing on 16-year-old bloodstain samples using Promega (upper) and CSF-3 (lower) primers. The PCR products from the degraded samples are shown on the left and the PCR products from fresh reference samples from the same individuals are shown on the right

cally) were added to each reaction mix. Samples A and E contained DNA fragments larger than 1 Kb and could be typed with both primer sets. The DNA in samples B, C, and D was more severely degraded with most of the DNA fragments less than 500 bp. With the GenePrint kit, sample D amplified weakly and samples B and C yielded no detectable PCR products. With the CSF-3 primers, samples B and D gave strong PCR products; sample C again failed to amplify. These experiments indicate that the CSF-3 primers may be of value for CSF1PO typing of samples that cannot be typed using the standard primers.

In conclusion, we have demonstrated that the CSF1PO polymorphism can be typed reliably using a PCR primer set that yields shorter amplicons than the standard primers. The shorter PCR products are better resolved on slab gels and can be used to advantage in the typing of highly degraded DNA.

Acknowledgements We thank Dick Rubin and Molecular Dynamics for the loan of the FluorImager. Portions of this work were supported by National Institute of Justice grant 93-IJ-EX-0010 to GFS

References

1. Alford RL, Hammond HA, Coto I, Caskey CT (1994) Rapid and efficient resolution of parentage by amplification of short tandem repeats. Am J Hum Genet 55:190–195
2. Clark JM (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic Acid Res 16:9677–9686
3. Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K (1994) Identification of the remains of the Romanov family by DNA analysis. Nat Genet 6: 130–135
4. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 55:175–189
5. Hochmeister MN, Budowle B, Borer UV, Rudin O, Bohnert M, Dirnhofer R (1995) Confirmation of the identity of human skeletal remains using multiplex PCR amplification and typing kits. J Forensic Sci 40:701–705
6. Hochmeister MN, Budowle B, Schumm JW, Sprecher CJ, Borer UV, Dirnhofer R (1995) Swiss population data and forensic efficiency values on 3 tetrameric short tandem repeat loci -HUMTH01, TPOX and CSF1PO- derived using a STR multiplex system. Int J Legal Med 107:246–249
7. Hochmeister MN, Budowle B, Eisenberg A, Borer UV, Dirnhofer R (1996) Using multiplex PCR amplification and typing kits for the analysis of DNA evidence in a serial killer case. J Forensic Sci 41:155–162
8. Huang NE, Schumm J, Budowle B (1995) Chinese population data on three tetrameric short tandem repeat loci-HUMTH01, TPOX and CSF1PO- derived using multiplex PCR and manual typing. Forensic Sci Int 71:131–136
9. Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Methods Appl 3:13–22
10. Klintschar M (1995) Validation of the system FES/FPS for forensic purpose in an Austrian population sample. Int J Legal Med 108:162–164
11. Lygo JE, Johnson PE, Holdaway DJ, Woodroff S, Whitaker JP, Clayton TM, Kimpton CP, Gill P (1994) The validation of short tandem repeat (STR) loci for use in forensic casework. Int J Legal Med 107:77–89
12. Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW (1993) Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG]n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 53:953–958
13. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, a laboratory manual, 2nd edn Cold Spring Harbor Laboratory Press, Cold Spring Harbor
14. Wang Y, Wallin JM, Ju J, Sensabaugh GF, Mathies RA (1996) High-resolution capillary array electrophoretic sizing of multiplexed short tandem repeat loci using energy-transfer fluorescent primers. Electrophoresis 17: 1485–1490
15. Whitaker JP, Clayton TM, Urquhart AJ, Millican ES, Downes TJ, Kimpton CP, Gill P (1995) Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. Biotechniques 18:670–677

✓ **Amplified Restriction Fragment Polymorphism (AFLP) analysis of human genomic DNA**

Kevin Andera, Jenny Liu, and GF Sensabaugh, School of Public Health, University of California, Berkeley, CA 94720-7360

Amplified Restriction Fragment Polymorphism (AFLP) analysis is a new technique developed for the identification of strain variation in plant and animals (Vos et al., Nuc. Acids Res. 1995; 23: 4407–4414). AFLP employs PCR to amplify a selected set of restriction fragments prepared from total genomic DNA; the result is a complex 'fingerprint' type pattern which characterizes the restriction fragment set. By using different combinations of selective primers, distinct and non-overlapping restriction fragment sets can be visualized. Considerable variation has been observed between cultivars within plant species.

We have tested a commercial AFLP kit (Life Technologies) for potential usefulness in human differentiation. Genomic DNA samples from four individuals were restricted with EcoRI (E) and MseI (M). Restriction fragments were ligated to E and M linkers and E-M fragments were amplified using primers specific for the E and M linker sequences. Twelve sets of E-M fragments were selectively amplified using combinations of E and M primers with different 3 base additions at their 3' ends. Fragments were separated on non-denaturing 5% acrylamide gels, stained with Sybr Green, and visualized on a FluorImager (Molecular Dynamics). Very little variation was observed between individuals with any of the selective primer combinations tested, suggesting AFLP typing has limited usefulness for human differentiation.

✓ **CSF1PO Polymorphism: evaluation of new primers and allele frequencies in the Japanese population**

Kanako Yoshida, Kazumasa Sekiguchi, Kentaro Kasai, Hajime Sato, Sueshige Seta and George F Sensabaugh, Forensic Science Group, School of Public Health, University of California, Berkeley, CA 94720, National Research Institute of Police Science, Tokyo, Japan

STR typing is promising for forensic DNA analysis of aged evidential samples containing severely degraded DNA because of the short amplified fragment size. The CSF1PO locus shows an informative polymorphism and is considered a useful STR loci for forensic identification [Hammond et al, Amer. J. Hum. Genet. 55: 175–189. 1994]. Since degraded DNA is more likely amplified with the primers which give low molecular weight fragments, CSF1PO primers have been redesigned to give shorter PCR fragment. The redesigned primer sequences are:

CSF-3F 5'gtt gct aac cac cct gtg tct c 3'	22mer
CSF-3R 5'ttc ctg tgt cag acc ctg ttc3'	21 mer

PCR products obtained using the new primers are approximately half the size (149–181bp) of fragments produced using the conventional primer (295–327bp). The id the new PCR products was verified by sequence and by comparison to typing results obtained using ventional primers. CSF1PO typings from 16 year ol stains using the new and the original primers shov the new primers gave typing results on samples that amplify with the original primers. This demonstra the redesigned CSF1PO primers are more suitable t original primers for STR typings from stain samp contained degraded DNA.

Population studies on the CSF1PO polymorph Japanese (n=142) have also been conducted. All quencies are as follows: *7-0.014; *8-0; *9-0.03 0.201; *11-0.229; *12-0.387; *13-0.120; *14-0.00 0.004. These allelic frequencies show no deviatio Hardy-Weinberg equilibrium.The discrimination po CSF1PO in the Japanese population is 0.893.

Differential extractions using a Fluorescence-Acti Cell Sorter

Matt Piucci, Armand Tcheong, CA Dept. of Justic Lab, 626 Bancroft Way, Berkeley, CA

Don Hesler, Gill Reinin and Chris Spenner, B Dickinson Immunocytometry Systems, 2350 Qume San Jose, CA

The use of DNA databases to produce suspects in oth suspectless sexual assault cases is one of the main go forensic DNA analysis. The use of robots to perform mated DNA analysis has proven to be very help achieving the aim of developing these databases. Ho there is no associated protocol for performing high-th put separation of sperm and epithelial cells to produc files for searching against felon databases. One p solution to this problem may be the use of Fluores Activated Cell Sorters (FACS) to perform these separ Cell sorters can process up to 7000 events (cells) per m and in this application the processing of a typical swa about ten to fifteen minutes. This suggests that a throu of twenty five swabs per day is not unreasonable.

Simulated sexual assault kits were prepared using b swabs that were spiked with diluted semen. The swabs incubated in PBS at 4°C for one hour, minced and th uid was removed from the substrate. These were stained with Propidium Iodide (a general DNA stain 0.1% Triton 100 solution. The sperm cells were sep from the epithelial cells based on total DNA content (a sure of ploidy or n number) and particle size by a Be Dickinson FACS Vantage. The sorted cells were st with Nuclear Fast Red and Picroindigocarmine for n scopic examination. DNA was extracted from the cell was typed via PCR at the DQA1 locus. Complete sepa was confirmed by microscopic examination of the frac and by the typing results. Negative controls produce types.

in press

GENETIC VARIATION AT 6 STR LOCI IN THE JAPANESE POPULATION

K. Yoshida[1], K. Kasai[1], H. Sato[1], C.H. Brenner[2], and G.F. Sensabaugh[3]

[1] National Research Institute of Police Science, 6,Sanbancho, Chiyoda-ku, Tokyo 102, Japan
[2] 2486 Hilgard Ave., Berkeley, CA 94709 USA
[3] Forensic Science Group, School of Public Health, University of California, Berkeley, CA 94720 USA

Genetic typing has been done on 187 individuals from six geographic regions in Japan. The populations tested were:

Population **H** - Sapporo, Hokkaido district (n=16)
Population **T** - Sendai, Tohoku district (n=26)
Population **M** - metropolitan Tokyo (n=80)
Population **I** - Nara, Kinki district (n=20)
Population **Y** - Saga, Kyushu district (n=24)
Population **O** - Naha, Okinawa district (n=21)

Six loci were typed: CSF1PO, TPO, THO1, D3S1358, vWFA31, and FGA. CSF1PO, TPO, THO1 and vWFA31 typing used commercial kits (Gene Print STR System, Promega Corp., USA) following the kit protocol. Following electrophoretic separation on polyacrylamide slab gels under denaturing conditions, PCR products were stained with SYBRGreen (Molecular Probes, Inc, USA) and the bands were detected on a FluorImager 575 (Molecular Dynamics, USA). Alleles were typed by direct comparison to an allelic ladder. D3S1358, vWFA31, and FGA were typed using the AmpFlSTR Blue kit (Perkin Elmer, USA) according to the manufacturer's protocol on an Applied Biosystems 377 DNA sequencer. Allele fragment sizes were calculated automatically from the mobility of the GS350 internal standard marker and typed by comparison to an AmpLlSTR Blue allelic ladder loaded on the same gel. vWFA31 typings by the two methods were concordant.

Genetic types at three or more loci were obtained for each of the 187 individuals in the study; complete 6-locus profiles were determined for 150 individuals. No previously undescribed alleles were detected. Allele frequencies were estimated by gene counting. Statistical analysis for fit to Hardy-Weinberg (HW) equilibrium expectations and linkage equilibrium were performed using the computer program, DNA•VIEW, developed by one of us (C. Brenner); this program uses the exact test algorithms described by Guo and Thompson [1] and by Zayki et al. [2]. Allele frequency data and summary statistics for the total sample set are shown in **table 1**. The observed data were in good agreement with HW equilibrium expectations for all but CSF1PO; however, with a Bonferroni correction for multiple tests, the deviation of CSF1PO was not significant. Five of the six loci showed a slight heterozygote excess; vWFA31 was the exception.

Pairwise comparison of loci using Fisher's Exact Tests showed no evidence of linkage disequilibrium. A multi-way independence test on all six loci also showed no departure from independence expectations. The combined probability for a random match across all six loci is 1.2×10^{-6}.

The allele frequency values found in this Japanese population are similar to allele frequencies reported previously for the Japanese population but exhibit some differences from frequencies reported for non-Asians [3,4].

Statistical analysis of the allele frequencies in the regional populations showed only two deviations from HW equilibrium expectations at the $p<0.05$ level and none at the $p<0.01$ level. Some differences from the total population mean frequencies were noted; these could be due to statistical

TABLE 1: STR ALLELE FREQUENCIES IN THE JAPANESE POPULATION

	CSF1PO		TPO		TH01		D3S1358		VWF		FGA	
6					85	0.241						
7	5	0.014			102	0.290						
8			147	0.425	17	0.048						
9	13	0.037	48	0.139	131	0.372						
9.3					14	0.040						
10	74	0.209	12	0.035	3	0.009						
11	80	0.226	118	0.341								
12	145	0.410	21	0.061								
13	33	0.093										
14	3	0.008					12	0.035	66	0.176		14
15	1	0.003					136	0.400	9	0.024		15
16							101	0.297	71	0.190		16
17							68	0.200	107	0.286	2	0.006 17
18							20	0.059	101	0.270	6	0.018 18
19							3	0.009	16	0.043	18	0.055 19
									4	0.011	35	0.107 20
											42	0.129 21
											69	0.212 22
											2	0.006 22.2
											73	0.224 23
											40	0.123 24
											23	0.071 25
											15	0.046 26
											1	0.003 27

N=	354	346	352	340	374	326
obs het	0.76	0.70	0.74	0.72	0.76	0.8
exp het	0.73	0.68	0.72	0.71	0.78	0.85
PD	0.88	0.84	0.87	0.86	0.91	0.96
HWE p=	0.03	0.7	0.8	0.7	0.5	0.5

sampling artifact or to underlying population substructure. **Figure 1** compares the frequency values for each allele at each locus in each of the regional populations (circles) with the mean frequency value for the total population (cross bars). The "whiskers" above and below the cross bars represent a ± 1 standard deviation range around the mean given the size of the regional population sample. Most of the observed allele frequencies fall within the ± 1 S.D. window. However, 8 of the allele frequency values, including 3 from Okinawa, are outside a ± 2.33 S.D. range; these are noted on the figure by the arrows. Thus most but not all of the observed variation in frequencies can be accounted for simply by sampling variation.

Evidence of underlying population sub-structure is provided by F_{ST} analysis [5,6]. The jackknife calculation of F_{ST} over all loci and all regional populations is 0.009 ± 0.003; the possibility that $F_{ST} \leq 0$ is rejected ($p = 0.001$). Genetic distances between the six regional populations were assessed using pairwise F_{ST} values [7]. The distances were small but indicated the Okinawa population to be separated from the other regional populations. This picture is consistent with the deviations in allele frequencies noted in this study, with analyses based on classical blood group and protein marker typing, and with current theory on the peopling of Japan [8].

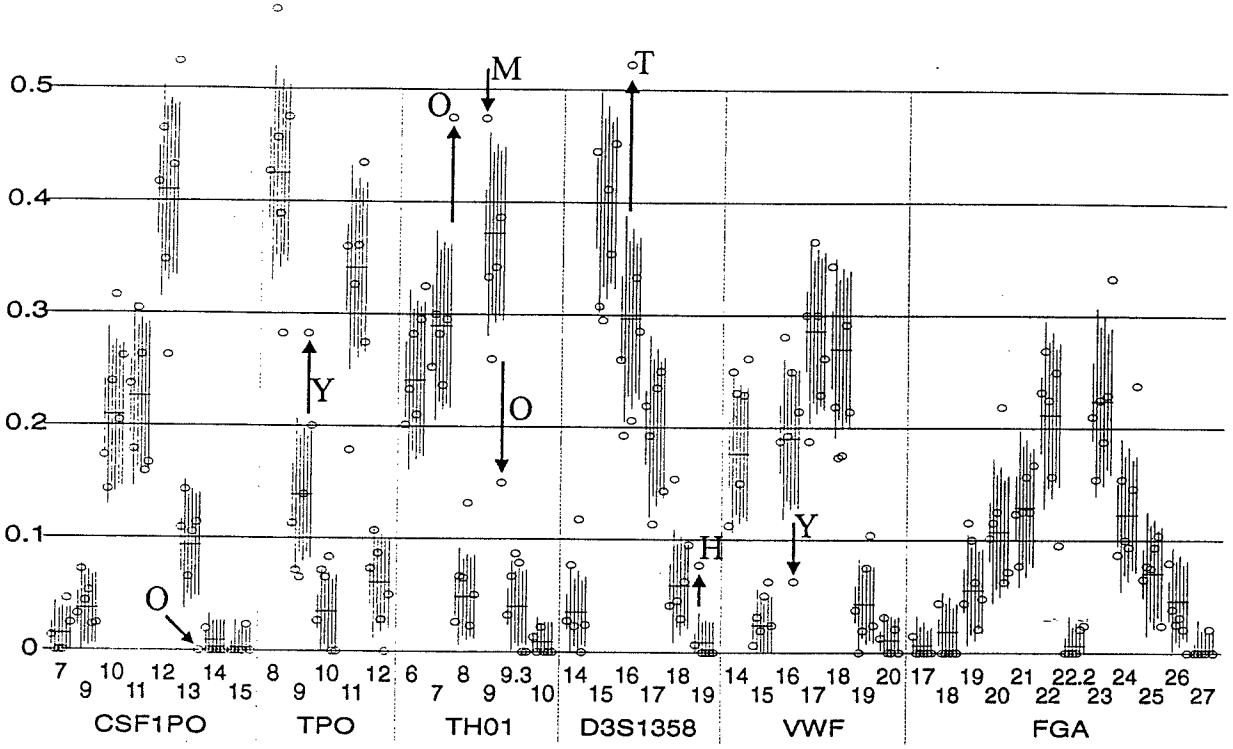


Figure 1: Allele frequency plot by locus, by allele, and by population in order MHTIYO

REFERENCES

1. Guo SW, Thompson EA (1992) Performing the exact test of Hardy Weinberg proportion for multiplex alleles. Biometrics 48: 361-372

2. Zaykin D, Shivotovsky L, Weir BS (1995) Exact tests for association between alleles at arbitrary numbers of loci. Genetica 96: 169-178

3. Nagai A, Yamada S, Watanabe Y, Bunai Y, Ohya I (1996) Analysis of STR loci HUMF13A01,HUMFXIIIB, HUMLIPOL, HUMTH01, HUMTPOX and HUMVWFA31 in a Japanese population. Int J Leg Med 109: 34-36

4. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 55: 175-189

5. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370

6. Goudet J (1995) FSTAT (version 1.2): A computer program to calculate F-statistics. J Hered 86: 485-486

7. Reynolds J, Weir BS, Cockerham CC (1983) Estimation of coancestry coefficient: Basis for a short term genetic distance. Genetics 105: 767-779

8. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The History and Geography of Human Genes. Princeton University Press, Princeton, NJ