ALLELIC FREQUENCIES OF TWO DNA RESTRICTION FRAGMENT

LENGTH POLYMORPHISM SYSTEMS COLLECTED FROM AN

UNDEFINED POPULATION : APPLICATION TO DNA PROFILING

by

LORI L. C. (NEE CHEUNG) BONNYCASTLE

B.Sc. (Microbiology), University of British Columbia, 1984
M.Sc. (Med. Microbiology), University of British Columbia, 1987

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF PATHOLOGY

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March 1992

© Lori L.C. Bonnycastle

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of _Pathology_

The University of British Columbia
Vancouver, Canada

Date _March 27/92_

TABLE OF CONTENTS

ABSTRACT

Restriction fragment length polymorphisms (RFLPs) within the deoxyribonucleic acid molecule were analysed to determine the extent of their ability to provide discriminating human DNA profiles for forensic identification. The two types of RFLPs studied were variable number of tandem repeats (VNTR), the technique commonly used by North American forensic laboratories, and dimorphic restriction endonuclease recognition site (RES) developed in this laboratory.

Data bases showing allele frequencies were collected for four VNTR hypervariable regions, D2S44, D16S85, INS, and D14S13 from an undefined population group obtained from the Greater Vancouver Area. Allele frequency distributions from this group were similar to the larger and more statistically defined allele distributions obtained by other laboratories, the RCMP Molecular Biology Unit in Ottawa and Promega associated laboratories in all of which analysed Caucasian populations.

An alternative technology to VNTR was based on the selection of five regions within the human genome that contain dimorphic restriction endonuclease recognition sites (RES). A nucleotide sequence of approximately one kilobase that contained the RES was selected from the genes of the following proteins : adenine phosphoribosyltransferase (APRT), prealbumin (PALB), adenine deaminase (ADA), carbonic anhydrase II (CAH) and lipoprotein lipase (LPL). After amplification of the nucleotide sequence by the polymerase chain reaction (PCR), the presence or absence of the internal RES was analysed by restriction endonuclease digestion followed by agarose gel electrophoresis. From the pattern (size and number) of the restriction fragments generated, it is possible to make genotype assignments with respect to the RES. Restriction endonuclease activity was monitored by the presence of a control DNA fragment which is also cleaved by the restriction endonuclease used for each specific analysis. Statistical calculations demonstrated that by using 16 unlinked dimorphic RES, rapid individual identification to a certainty of one in 6.5 million is possible.

Analysis of casework and/or laboratory simulated forensic specimens indicate that both the VNTR and RES system are discriminating and can be applied to non-ideal specimens. However, the RES system offers a much simpler and more reliable methodology.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Dr. J.A.J. Ferris, for providing me with the opportunity to carry out this project in his laboratory, and for his support and guidance throughout the course of this study. I would also like to thank the members of my supervisory committee, Drs. A.P. Autor, G.R. Gray, B. McGillivray, and G. Tener, for their guidance and helpful suggestions. My sincere thanks also go to Mr. K.McNeil for his valuable discussions and technical assistance during this project, to Dr. M. Altamirano-Dimas for the generous use of his laboratory equipment and materials, and to Drs. M. Schultzer and E. Wijsman both of whom have helped tremendously with the statistical aspects of this analysis.

My love and thanks goes to my parents who have always had great confidence in my abilities and have urged me forward to meet new challenges. I would also like to express my love and gratitiude to my husband Jeffrey who not only assisted with drawing of figures and tables in this thesis, but has shown great patience and has provided unwavering support during my trying but rewarding journey through this thesis project.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

BRL : Bethesda Research Laboratories

DTT : dithiothreitol

EDTA : ethylenediamine tetraacetic acid

EtBr : ethidium bromide

Het_{obs} : observed heterozygotes

Hom_{obs} : observed homozygotes

Het_{exp} : expected heterozygotes

Hom_{exp} : expected homozygotes

H.W.E. : Hardy-Weinberg Equilibrium

LB : Luria-Bertani

Mol. Wt. Std. : molecular length standard

PCR : polymerase chain reaction

RES : restriction endonuclease recognition site

RFLPs : restriction fragment length polymorphisms

SDS : sodium dodecyl sulfate

Tris : tris(hydoxymethyl) aminomethane

UV : ultraviolet light

VNTRs : variable number of tandem repeats

GENERAL INTRODUCTION

At the University of Lyon in France in 1910 Edmund Locard formulated his exchange principle in which he postulated that the criminal always leaves something behind at the scene of his crime which was not there before and similarly carries away with him something which was not on him when he arrived [1]. One of the most important applications of the Locard exchange principle for the forensic scientist is to either associate or disassociate biological specimens such as blood and semen stains with persons under suspicion of committing the crime under investigation. Forensic science may be defined as the application of scientific techniques to the investigation of crime. Forensic scientists are often required to examine materials associated with an incident, either suspected or known to contravene the law.

Until fairly recently, forensic serology routinely involved the application of immunological, serological and biochemical methods to blood and body fluids and stains of blood and body fluids [2-4]. These tests are possible due to genetic markers in human populations which are detected as inherited protein variants. The five major classes of genetic markers are blood group antigens, isoenzymes, serum group antigens, haemoglobin variants, and the human leucocyte antigens (HLA).

In most instances, the analysis is based upon the presence of structural differences of specific proteins in vivo. Hence, the success of such testing is critically dependent on the preservation of these proteins in samples. Factors such as age, size, and storage condition (environment) of the biological specimen influences the preservation of the original structure [5]. Tissue-specific gene expression will also determine the presence of certain proteins thereby limiting the types of specimen suitable for analysis. Furthermore, protein and other cellular components from bacterial contamination of samples [6], mixed body fluid samples [7], or samples originating from more than one individual may make interpretation difficult or even give inaccurate results.

Even in the absence of these problems, while exclusion is absolute, a conclusive match is almost impossible. The limited number of measurable variations in the majority of these protein systems means that a large percentage of the population will share any one specific protein variant.

In most instances, the "inclusion" of an individual (positive result) may only be an indication that the specimen is consistent with a possible origin from that individual [8] and not positive proof of association. In the last five to six years, extensive efforts have been made to remedy the problems of uncertain identification and to increase the discriminative potential of forensic specimen testing by utilizing the rapid advances in the analysis of the cellular genetic material otherwise identified structurally as deoxyribonucleic acid (DNA) [9-18].

Genetic Variation In The Human Genome:

The study of individual genetic variation at the genotypic level as opposed to the phenotypic level has been made possible by the introduction of recombinant DNA technology. The direct analysis of DNA not only allows for the determination of specific base sequences responsible for altered proteins, but also permits the detection of sequence variation in non-expressed DNA [19]. The human genome, which is 6×10^9 nucleotides per diploid genome, consists of both coding and non-coding regions. Coding regions function in determining the composition and sequence of amino acids making up proteins [20]. The non-coding regions are DNA sequences which either have regulatory roles such as controlling the degree of gene expression in specific tissues, or have no known function.

It is believed that only a small percentage of the genome serves a coding function so that the majority of the non-coding sequences may be in fact non-functional [21]. However, the significance of most of the sequences will probably remain unclear until the Human Genome Project has been completed. It is in these non-coding sequences that the potential for sequence variation is the greatest since such sequences would not be under the influence of evolutionary constraints placed on coding sequences [20].

Due to selection pressures and genetic drift, a 400 amino acid protein changes only approximately once every 200 thousand years [22]. In electrophoretic polymorphism studies of 71 different types of proteins in European populations, the frequency of polymorphic loci found was 0.28 [23]. Based on this value, Botstein et al. have estimated that the frequency of DNA sequence

polymorphism for protein-coding sequences to be about 0.001 per base pair [24]. However, analysis of polymorphism at the level of the protein does not provide an adequate measurement of variation of the total genome. Compared to coding regions, the extent of sequence polymorphism is far greater in the human genome as a whole (both coding and non-coding regions) [25]. Cooper et al. calculated that unique DNA sequence heterozygosity in human autosomes is 0.0039 which suggests that the majority of the sequence variation occurs in non-coding regions at a frequency of about one every two to three hundred base pairs [25].

These findings support the conclusions of Alec Jeffreys in his analysis of the ß-related globin gene complex in man [26]. Three DNA sequence variants were detected within this complex of genes and surprisingly, all three variants were found in the intervening (non-protein coding) sequences which made up only 3.4 kilobases out of the total of 41 kilobases. As it seemed unlikely that this observation occurred purely by chance, Jeffreys suggested that there was a preferential accumulation of DNA sequence variation within the intervening sequences. This would indicate that any possible role played by the intervening sequences would not require fixed sequences. Jeffreys further suggested that if the amount of polymorphism associated with the ß-related globin gene complex is indicative of the degree of variation found in the rest of the genome, then the frequency of variation for the human genome would be about one in a hundred base pairs.

However, the results of a more recent study suggests that sequence variation in coding sequences contribute significantly to sequence polymorphism of the entire genome [27]. Comparison of protein-coding sequences between human and rodent homologous genes indicate that the rate of nucleotide substitution is highly variable with regards to 1) type of substitution (synonymous and nonsynonymous), 2) different gene regions and 3) different genes [27]. In this study of 17 protein coding sequences the mean rate of synonymous substitution (no change in amino acid) was found to be five fold higher than that of nonsynonymous substitutions (change in amino acid). This is a consequence of selection against any nonsynonymous mutations which cause deleterious effects on the function of a protein whereas synonymous mutations are tolerated.

Different regions within a protein (coded by different gene region) have various functions such that some regions are more constraint than others leading to differing substitution rates in the coding sequence (i.e. substitution rate is seven times higher in the C-peptide of the proinsulin gene compared to the A- and B- chain which make up the active portion of the hormone) [27]. Synonymous and nonsynonymous substitution rate variation also occurs among genes. This difference may be due to differing functional constraints or differences in the rate of mutation at different loci and the differences in the chromosomal position of the gene [27]. In a comparison between cows and goats, fourfold degenerate sites within the ß- and α-globin gene had nucleotide substitution rates comparable to that observed with regions not subject to functional constraints such as introns, 3'flanking sequences and ß-globin pseudogenes [27]. In light of these results, the frequency of sequence heterozygosity in coding sequences may be much higher than previously assumed from proteins studies and limited coding sequence studies.

Restriction Fragment Length Polymorphisms (RFLPs):

The method used by some of the above mentioned researchers such as Cooper et al. [25] and Jeffreys [26], for their studies of DNA sequence polymorphism is based on the utilization of bacterial restriction endonucleases. There are different types of restriction endonucleases and the most commonly used type is type II restriction endonuclease. These enzymes recognize specific base sequences and will cut DNA only at a specific site either within or adjacent to that site [28]. Type I restriction endonucleases bind DNA at specific sequences but cleave DNA at random sites when the DNA loops back to the bound enzyme [29]. These type I restriction endonucleases are not applicable to VNTR analysis and will not be considered here.

Restriction endonuclease digestion of genomic DNA results in multiple DNA fragments of various sizes. The sizes generated are dependent upon the distances separating any two base sequences recognized by the specific endonuclease. The resulting fragments are then organized according to size by agarose gel electrophoresis and detection of specific sequences (fragments) is possible using Southern blotting [30] or variations of this blotting procedure. This involves the

transfer of DNA fragments from agarose to a solid support membrane such as nitrocellulose or nylon which is much more easily manipulated than agarose, followed by hybridization with specific nucleic acid probes. The discovery that a DNA probe for one region of the genome detected restriction fragments of different sizes in different individuals resulted in the exploitation of restriction fragment length polymorphisms or "RFLPs" in molecular biology.

Restriction fragment length polymorphisms are due to changes in DNA which alter the distance separating two restriction endonuclease recognition sites. This can be accomplished by point mutations leading to changes in base sequences which either create or eliminate a restriction endonuclease recognition site (RES) [19]. Point mutations in DNA are responsible for a number of inherited disorders. While these genetic changes within some RES are deleterious, changes in other RES are apparently harmless to the affected individual. These non-deleterious mutations are inherited in a Mendelian fashion and occur at high frequencies.

RFLPs have become valuable tools as markers for genetic mapping, and for genetic disease research [31]. They have been used to determine the approximate location of genes responsible for diseases such as Huntington disease [32], Alzheimer's disease [33], Duchenne muscular dystrophy [34] and cystic fibrosis [35]. By following the inheritance of linked markers, predictive tests are available to determine individuals at risk for certain genetic diseases. Analysis of this kind has made possible the prenatal diagnosis of several inherited disorders such as sickle cell anemia [36,37], phenylketonuria [38], Duchenne muscular dystrophy [39] and Hemophilia A [40].

Inherited alterations in restriction fragment size can also be a result of insertion of DNA into an area or deletion of DNA from an area bracketed by two sequences recognized by the same endonuclease. This type of polymorphism was first detected on chromosome 14 by Wyman and White who used a segment of single-copy human DNA as a probe [41]. This locus exhibited 8 alleles and family studies indicate that these alleles are inherited in a Mendelian fashion. Similar polymorphisms detectable with single copy (or single locus) probes have since been found near genes such as the insulin [42,43], and alpha globin [44,45] loci. These loci are hypervariable due to

the presence of core sequences tandemly repeated a variable number of times. Such polymorphisms have been termed "variable number of tandem repeats" or " VNTRs". VNTRs are very informative for genetic analysis because each locus has the potential for a large number of alleles resulting from differences in the number of tandem repeats [41]. These differences are presumably due to recombination by DNA slippage during replication or by unequal exchanges during mitosis or meiosis [9].

Before proceeding further, it is necessary to discuss the use of "allele" in the context of restriction fragment length polymorphisms (RFLPs). Historically, "allele" refers to an alternate form of a gene which serves some coding or regulatory function [46]. Others have argued that in modern times, the term "gene" is used to describe any heritable unit of the genome regardless of coding potential and since restriction DNA fragments are inherited as codominant traits according to Mendel's laws, the term "allele" should be defined as alternate forms of any identifiable DNA variation in the genome [47,48]. In light of the fact that published material from the forensic DNA typing community utilizes the term "allele" for RFLPs, this author will also adopt this term for the purposes of consistency.

In contrast to the VNTRs described above which exist as unique sequences, there are also VNTRs found in multiple copies throughout the human genome. These VNTRs or "minisatellites" were first described by Jeffreys et al. who utilized a 33 base pair sequence from an intron of the human myoglobin gene [9]. A tandemly repetitive probe (probe 33.15) made with this 33 base pair core sequence was able to simultaneously detect many variable regions within the human genome. The resultant autoradiograph exhibited numerous bands and the banding pattern was highly variable from individual to individual. Band comparisons between pairs of randomly selected individuals showed minimal band sharing [10].

These minisatellite banding patterns were highly specific to an individual such that the probability of finding the same pattern, using probe 33.15, in a second unrelated individual was 3 X 10^{-11} or 1 in 30 trillion [10]. Further analysis with other such minisatellite probes demonstrated

that a combination of two of these probes would be almost totally individual specific. The only exception to this specificity is between monozygotic twins in which case the DNA patterns are identical. The high level of discrimination of the minisatellite probes led to the term "DNA fingerprinting" which is a trade name registered by the British company Imperial Chemical Industries (ICI).

All fragments from an individual can be traced to one or the other parent with approximately half of the fragments being of paternal origin and the remaining fragments of maternal origin. This is due to the fact that these hypervariable fragments are stably inherited and segregate in a Mendelian fashion [9]. This observation led to the eventual application of such multi-locus (minisatellite) probes to create individual-specific "fingerprints" for human identification, including paternity determinations.

The first official use of DNA genetic typing for human identification was in a 1985 immigration case [49]. The objective was to determine the maternity of a boy who was attempting to re-enter the United Kingdom on the premise that his mother was already a resident of the United Kingdom. Immigration authorities suspected that a switch had occurred and that the boy was in fact the nephew of the woman involved. Conventional blood testing could not differentiate between the two possible relationships. DNA testing with two of the Jeffreys' minisatellite probes showed beyond a reasonable doubt that the relationship of the boy and woman was one of mother and son.

Human identification for immigration purposes is only one application for DNA genetic typing. Forensic identification of unidentified bodies or of crime-associated specimens can also be greatly facilitated and improved with this advanced technology. Initial studies to assess the feasibility of the application of DNA genetic typing to forensic specimens were performed at the Home Office Forensic Science Service Central Research Establishment in the U.K. [11]. The results of these studies confirmed the potential of the procedure. DNA prints were obtained from specimens such as semen stains, four year old blood stains, and fresh hair roots. This was further

substantiated by a subsequent evaluation of a blind trial involving 43 blood samples, 11 bloodstains and 11 semen stains [16]. These promising results eventually lead to the introduction of DNA profiling to its first criminal case in 1987 [50,51].

This historic case not only identified the individual responsible for two rape/murder cases but it also conclusively excluded another individual of the same crimes. A young man was charged with the 1986 sexual assault and murder of a school girl in Leicestershire, England. A similar unsolved murder had occurred in 1983 and conventional testing of semen stains from both victims did not rule out the possibility that this youth was also responsible for the earlier murder. However, the results of these tests showed that the semen could have been deposited by at least 10% of the male population. Hence, DNA genetic typing of the semen stains and blood from the accused male was performed, initially by Dr Alec Jeffreys and subsequently repeated by Home Office forensic scientists. These results excluded the accused but did confirm that both crimes had been committed by one individual with a chance association of the different seminal stain DNA patterns of only 5.8×10^{-8} or 1 in 580 billion. The accused male was released and a intensive blood testing program was undertaken by the Leicestershire Constabulary.

More than 5000 males were analyzed with an initial screen by conventional forensic serological tests which eliminated approximately 86% of the population. DNA typing was then performed on the remaining 14%. It was discovered that one individual, Colin Pitchfork, had paid a workmate to provide the police with a blood sample on his behalf. When confronted by the Police, Pitchfork confessed to both of the rape/murders in question. Subsequent DNA testing of Pitchfork's blood showed a pattern identical to that of semen stains. The successful conviction of Pitchfork was only possible with the help of DNA fingerprinting which excluded a wrongly accused individual and in turn identified the real culprit.

This case clearly illustrated the potential power of DNA typing as an investigative tool. The Home Office Forensic Service and the British criminal courts have since approved the use of DNA fingerprinting in forensic cases [52]. The British civil courts have also officially approved

DNA fingerprinting for paternity disputes and as of 1989, the British immigration authorities were considering the routine use of DNA typing for immigration applicants [53].

Single Locus VNTR Probes Verse Multi-locus Minisatellite Probes:

Although DNA genetic typing originally started with the Jeffreys minisatellite probes, the majority of the North American laboratories are using only the single locus probes. In fact, the RFLPs (minisatellite regions) detected by the minisatellites probes, have since been isolated individually to create probes which detect only one minisatellite region at a time and thus serve as single locus probes [54,55]. The reason for the switch in strategy was the fact that a number of complications involving minisatellite probes had been encountered.

Unfortunately, the multi-banding pattern which affords a minisatellite probe its tremendous discriminatory power is also one of its downfalls because of the complexity of the numerous DNA bands. These probes simultaneously detect many loci and the recombination rate at these regions is high [9,56] such that the number and/or sizes of loci detected are not constant from one individual to the next. Since there are no expected patterns, there may be some difficulties in determining which bands are allelic [14]. Extra bands may appear due to partial or incomplete digestion of DNA or bands may be missing due to inadequate amounts of DNA [16]. These errors could go undetected and lead to false positives or false negatives.

Difficulties with interpretation of DNA prints will also occur if the DNA source is a mixed sample i.e. contributed by more than one individual [57]. In fact, contamination of a sample with non human blood from sources such as birds, dogs, cats and mice, will also produce banding patterns specific to the contaminant [58-61].

Another drawback is the absence of chromosomal localization of the minisatellites [14]. The highly polymorphic nature and complexity of the minisatellites bands hinders attempts at localization of such bands [62] to specific areas within specific chromosomes. Difficulties arise when attempts are made to determine which band in each of two unrelated individuals actually represent the same locus. Finally, the size of the alleles detected by minisatellite probes range

from 4 kb to 20 kb [57]. Since the likelihood of obtaining intact full length DNA from old and degraded samples is low, the large fragments may be missing leading to incomplete DNA prints.

In contrast, single locus VNTR probes detect RFLPs in only one defined region and the number of alleles or bands is known before the analysis. As the VNTR regions chosen are ones which follow Mendelian segregation, each locus will have two alleles with one allele of paternal origin and the other of maternal origin. If both alleles are identical or have similar sizes, such that their difference is beyond the resolution of the electrophoretic system, then only one band will be visible on the autoradiograph. The presence of more than two alleles for any one of these single locus probes is indicative of some methodological error such as incomplete digestion of the DNA or contamination of the DNA sample. Recently however, there has been a report of individuals with three band patterns when probed with a single locus VNTR probe [63]. The three loci so far shown to demonstrate this phenomenon resulting from restriction site polymorphisms are D3S46, D4S139, and D7S22. The nomenclature for DNA segments are as follows :

- "D" refers to DNA

- number after "D" refers to the chromosomal assignment

- "S" refers to unique DNA segment

- numbers after "S" refers to different sites on the same chromosome   In the case of locus D4S139, the abnormal pattern is due to the presence of an internal Hae III site within the VNTR flanked by the two expected Hae III sites. This results in the production of two bands for one allele such that a normal (2 Hae III sites) allele and an abnormal (3 Hae III sites) allele produce a three band pattern. Nevertheless, the concept of having a predetermined number of bands serves as a primary screen for errors which could lead to inaccurate interpretations. Since one single locus probe only detects differences in one region of the genome it does not have the discriminating power of one minisatellite probe.

In order to approach the level of discrimination afforded by the multi-locus probes, a number of non-linked single locus probes must be used, essentially recreating the multiple banding pattern of the minisatellites. However in this case, every band has a known origin.

The frequencies of alleles from the different loci can be used together to calculate the chance occurrence of such a DNA banding pattern. The simplicity of the banding patterns obtained with the single locus probes also allows linkage analysis studies to ensure that allele frequencies of one locus are not influenced by the frequencies of other hypervariable loci used in the identification procedure. Another advantage of using single locus probes is that they can detect a smaller amount of DNA than that of the multi-locus minisatellite probes which require one microgram of genomic DNA in order to produce a DNA fingerprint [57]. This fact is especially important since the quantity of forensic samples is often very limited.

Application of DNA Genetic Typing To Forensic Samples:

Single locus probes have been used in forensic case work in North America by both law enforcement agencies and private commercial companies. Such services for both forensic and paternity testing have been made available since 1987 by the first two private DNA typing laboratories in the United States, Lifecodes Corporation (Valhalla, NY) and Cellmark Diagnostics (Germantown, MD). These two companies have reported tremendous success with this new technology. The first conviction using DNA evidence in the United States was prosecuted in Orlando, Florida in 1987 [64]. Since then, there have been numerous criminal trials in which DNA profiling has been admitted as evidence [65,66]. The United States Federal Bureau of Investigation has aggressively developed forensic DNA analysis techniques and have imparted their knowledge to state and local crime laboratories in the U.S. [67]. As well as on going research, the FBI are presently providing a forensic DNA typing service for criminal case materials.

Forensic laboratories in Canada including the Royal Canadian Mounted Police Central Forensic Laboratory in Ottawa, and the Centre of Forensic Sciences in Toronto have also put extensive efforts into this revolutionary technique [8,68]. Canada's first trial experience with DNA

evidence in December 1988 resulted in the exoneration of an accused, James Alexander Parent, who was charged with a number of related sexual offenses [69]. DNA evidence indicated that Parent was not the assailant in several of the offenses that he had been charged with. In April 1989, DNA evidence was used in court to positively identify an individual as the assailant in the sexual assault of a 68 year-old woman [8,69]. Comparison of DNA extracted from semen stains found on the victim's nightgown and bedspread with that from the accused's blood sample, showed identical DNA banding patterns. The judge presiding over the "voir dire" in the District Court of Ottawa, ruled that the DNA evidence was admissible. RCMP experts testified that the chance occurrence of the DNA print from these forensic samples matching those of the accused was found to be less than 1 in 70 billion people. Faced with such overwhelming odds, the accused Paul McNally, plead guilty to the charges.

Controversies Surrounding DNA Genetic Typing:

In many of the earlier cases, DNA evidence was not challenged in the courts because defense attorneys had difficulty finding scientists to question the accuracy of DNA typing [70]. However, in 1989 the scientific community started to question the accuracy and validity of DNA analyses performed by the various DNA typing laboratories in North America. In addition to several other cases involving questionable DNA evidence, the highly publicized New York murder case of People v. Castro in August 1989, proved that these concerns were justified [71,72]. The DNA analysis of the Castro case was performed by Lifecodes which is the largest private forensic DNA laboratory in the United States. Defence expert witness Eric Lander, from the Massachusetts Institute of Technology in Cambridge, along with the two experts for the prosecution, signed a statement declaring that the DNA evidence presented by Lifecodes was not "scientifically reliable enough" to support its conclusions which incriminated the accused [71,73]. Based on this declaration, the presiding Judge Gerald Sheindlin ruled that the DNA evidence was not admissible in this case. The issues raised in the Castro case by Eric Landers forced both the forensic and scientific community to examine more critically the entire concept and technology of DNA analysis.

Although DNA genetic typing with single locus VNTR probes has made a major impact in forensic science as well as in other disciplines concerned with individual identification, several issues still need to be dealt with by the "experts" in this field [72,74,75]. One of the major issues is the unpredictability of the sizes of the DNA fragments after restriction endonuclease digestion. A hypervariable (VNTR) locus may have 50 to 100 alleles such that it is impossible to predict the sizes of the DNA fragments [71]. DNA extracted from forensic samples are frequently of poor quality (i.e. degraded) thus eliminating the detection of the large alleles which require the presence of intact DNA molecules. Individuals heterozygous for a probed VNTR locus would be erroneously typed as being homozygous for a small allele. The poor quality of the forensic samples may also result in band shifting to such an extent that false negatives may occur. Furthermore, VNTR alleles may differ in size by as little as one base [76]. Assignment of a particular band to a discrete allele would be difficult if the sizes of bands differ by only a few base pairs (i.e. 1000 bp +/- 50 bp may have two possible allelic assignments). For DNA fragments >1 kb, slight changes in DNA length may not be resolved by agarose gel electrophoresis. Both of these problems interfere with the acquisition of reliable quality data and accurate interpretation of results.

Another important issue revolves around the need for implementation of standard quality control and quality assurance procedures by various DNA laboratories [75,77]. The problem in this area is that different laboratories, especially the commercial ones, tend to keep their work a secret and only reluctantly reveal their procedures when forced to by the courts [74]. In addition, there are as yet no government or scientific agency-enforced proficiency testing programs to accredit forensic laboratories for DNA analysis [70]. However, within the last two years, the FBI have published several articles concerned with issues of quality assurance, proficiency testing and statistical standards for DNA analysis [78-80]. Furthermore, members from several agencies across North America have recently formed a group called "TWGDAM" or Technical Working Group on DNA Analysis Methods. The purpose of TWGDAM is to set up guidelines for DNA typing laboratories so as to promote consistency wherever possible at the different facilities. The

Canadian Society for Forensic Science has set up a similar group (subcommittee) for the same purpose in Canada.

At the present time, the United States Congress is considering the DNA Proficiency Testing Act of 1991 [81]. This bill would promote a step towards the requirement of standard guidlines and regular proficiency testing in laboratories performing DNA forensic testing.

The last major issue is one concerning the calculation of the probabilities of chance "matches" between DNA restriction patterns from non-related individuals. The highly publicized claims of astronomical odds against "matches" arising at random such as 67,500,000,000:1 odds in Georgia v. Caldwell [71], and 5,000,000,000:1 odds in Florida v. Andrews [82], are undoubtedly impressive to trial juries. However, the statistical basis for such calculations may not be applicable for some VNTR loci if the reference population used to build the data base consists of heterogenous subpopulations or if the combination of loci used are in linkage disequilibrium [70-72]. The method currently used for probability calculations is based on The Hardy-Weinberg Equilibrium which is actually an extension of Mendel's laws of inheritance [83].

The Hardy-Weinberg Equilibrium (which gave rise to the Hardy-Weinberg Law) was proposed in 1908 independently by G.H Hardy in England and by Wilhelm Weinberg in Germany [84]. This equilibrium describes " the expected relationship between the frequencies of alleles in local populations and the frequencies of individuals of various genotypes in these same populations" [83]. In the context of RFLPs, the Hardy-Weinberg Equilibrium relates the frequency of the various bands (alleles) in a population to the frequency of individuals possessing pairs of bands (genotypes) in a population [73].

The Hardy-Weinberg Law states that "under certain conditions of stability both allelic frequencies and genotypic ratios remain constant from generation to generation in sexually reproducing populations" [85]. Hence, using the Hardy-Weinberg formula, the probability of a person having a specific banding pattern can be computed once the frequencies of the individual alleles have been established. In order to increase the discriminating power of single locus probes,

the genotype frequencies of each loci probed can be multiplied together to give the probability of concurrent incidence. This "product rule" [86] is only valid if the combination of loci analyzed are unlinked (at linkage equilibrium) i.e. the presence of certain alleles at one locus has no influence on the alleles at another locus [71].

The controversy surrounding the application of the Hardy-Weinberg formula stems from the fact that with respect to certain loci, mixed populations may not be in the Hardy-Weinberg Equilibrium. This means that allele frequencies may differ from one ethnic group to another such that population data bases developed with one population may be inaccurate for another population. To explain this discrepancy, one must realise that in a continuously evolving world, the four conditions necessary to achieve the Hardy-Weinberg Equilibrium ie. a large population size, no mutations or mutational equilibrium, no immigration or emigration and random reproduction, are never met [85]. Although appreciable changes in allele frequencies due to violations of these four conditions may occur quickly with respect to the evolutionary time scale, it is slow with respect to real time. Therefore, it is still possible to observe populations in Hardy-Weinberg Equilibrium despite breaches of the "specific conditions". However, due to the existence of small and especially isolated communities in which there is inbreeding and/or reduced gene flow (no addition of new alleles by inter-racial matings), there would be an increase in the number of individuals homozygous at various loci because the alleles they carry are identical by descent or parentage [83]. As a result, individual communities may have different gene frequencies for various alleles.

Evidence for allelic heterogeneity between American blacks, Caucasians, and Hispanics is available for at least four loci [87,88]. Even within these racial subpopulations there may be a mixture of ethnic subgroups leading to deviations from the Hardy-Weinberg Equilibrium. This deviation was observed by Eric Lander in his analysis of the Hispanic data base used as the reference population in the problematic Castro case [71,73]. "Subgroups" may by themselves be in Hardy-Weinberg Equilibrium but the pooling of subgroup data before the calculation of the Hardy-Weinberg expectations could result in a divergence from the equilibrium (Wahlund's Principle)

[83]. If the total population data base does not take in to account heterogenous subpopulations, there could be an under-estimation of the frequencies of some alleles. The error is further complicated by the possibility of disequilibrium between loci since loci in heterogenous populations may not be in equilibrium even if the loci are on separate chromosomes [71]. This could result in estimated probabilities of matched DNA profiles being lower than the true probabilities and hence unjustly more incriminating to the suspect [89]. It is clear therefore that allele frequency data bases for each subpopulation are required to avoid erroneous results.

When dealing with RFLPs (VNTRs), another avenue for error leading to non Hardy-Weinberg expectations is the limit of resolution of bands in agarose gel electrophoresis. As mentioned previously, VNTR regions have a continuum of allele sizes and slight differences may not be resolved so that two similar sized alleles may appear as one allele. This would lead to a similar over representation of individuals homozygous for certain loci and cause a population which is in Hardy-Weinberg Equilibrium for the loci studied, to appear to not be in equilibrium.

All the issues related to the use of RFLPs in forensic identification discussed here are the major problems associated with this technology and are not by any means trivial. Due to the tremendous potential of this technology, numerous laboratories both government and commercial, are actively in search of new strategies and methods to overcome some of the existing problems as evidenced by the vast amount of literature concerning DNA genetic typing. The recent discovery of the polymerase chain reaction [90-92] has provided more avenues to deal with the identification of forensic specimens.

THESIS HYPOTHESIS

Forensic identification is a critical component of any criminal investigation. Information that associates an object or a person to a specific crime may be used to corroborate other evidentiary material or may provide the law enforcement agencies with a "starting point" from which the investigation proceeds. Unfortunately, with the exception of fingerprints, a forensic "match", using conventional techniques, of an individual to a biological specimen(s) found at a crime scene, will in most cases only indicate that the individual is a possible contributor of the specimen. The biological components, proteins, tested are not polymorphic or variable enough within the population to provide sufficient discrimination to distinguish one individual from another at a statistically significant level. In order to observe greater variation, the analysis must proceed to a more fundamental level, i.e. the basic heritable material of the cells (deoxyribonucleic acid or DNA) some of which codes for the proteins and some of which is phenotypically silent.

In 1985, with the use of DNA probes, Dr. Alec Jeffreys of the University of Leicester (England) reported the detection of tandem-repetitive "minisatellite" regions dispersed throughout the human genome [9]. These minisatellites (multi-locus variable number of tandem repeats) were highly polymorphic and were capable of creating individual-specific DNA "fingerprints". The DNA banding pattern obtained by analysis with two such minisatellite probes was specific such that the probability of two individuals with the same pattern was $<< 5 \times 10^{-19}$ [10]. These impressive results caught the immediate attention of the forensic community and the media. It appeared that the opportunity had arrived for the positive identification of criminals by analysis of specimens such as hair bulbs, blood, semen and stains of blood or semen.

Although the DNA typing (fingerprinting) procedure described by Jeffreys utilizes a combination of techniques that have been already established in the scientific and medical community, it cannot be assumed that the reliability of these techniques is good enough to make life and death judgements on a suspect. There must be extensive research into many areas

such as DNA polymorphism, effects of environmental conditions on DNA integrity particularly during the subsequent typing procedure, and the reliability of the techniques (reproducibility and accuracy).

The aim of this thesis is to determine the extent to which hypervariable regions, specifically the areas containing the single locus variable number of tandem repeats (VNTRs), within the human genome can be used as a means of distinguishing between individuals in a forensic capacity. This investigation also focuses on the validity of data collected by non-conventional population sampling procedures (i.e. from undefined population). In addition, another DNA typing system involving dimorphic restriction endonuclease recognition sites is offered as an alternative for the existing hypervariable VNTR (single locus and multi-loci) systems that may solve some the obvious problems of the present VNTR technology such as questionable allelic assignments and inability to determine allele frequency equilibrium by the Hardy-Weinberg test.

Objectives :

1. To build a population data base for four hypervariable VNTR regions utilizing DNA samples from undefined individuals of the Greater Vancouver Area.

2. To establish technical conditions for VNTR analysis of genomic DNA.

3. To apply VNTR technology to forensic samples and to determine its applicability to non-laboratory samples.

4. To compare the frequency distribution of VNTR alleles collected from defined and undefined populations.

5. To set up another system of RFLP analysis utilizing dimorphic restriction endonuclease recognition sites (RES) and the polymerase chain reaction (PCR).

6. To build a population data base for several dimorphic RES utilizing DNA samples from undefined individuals and from Orientals in the population of the Greater Vancouver Area.

7. To test the discriminatory potential of this dimorphic RES system by the analysis of two families and a number of undefined individuals.

MATERIALS

Solutions :

1. Alkali transfer buffer - 1.5 M NaCl
                              0.25 M NaOH

2. Denaturation buffer - 1.5 M NaCl
                           0.5 M NaOH

3. Denhart's (100X) - 2% ficoll
                       2% polyvinylpyrrolidone
                       2% bovine serum albumin

4. HTE buffer - 0.100 M tris-base
                0.040 M EDTA
                (pH 8.0)

5. LTE buffer - 0.010 M tris-HCl
                0.001 M EDTA
                (pH 7.5)

6. Luria-Bertani medium (LB) - 0.5% Bacto-yeast extract
                                    1.0% Bacto-tryptone
                                    1.0% NaCl
                                    (pH 7.5)

7. Post hybridization wash
        Solution I   - 2X SSPE
                       0.1% SDS

        Solution II - 1X SSPE
                       0.1% SDS

        Solution III - 0.1X SSPE
                       0.1% SDS

8. SSC (1X) - 0.150 M NaCl
              0.015 M $Na_3C_6H_5O_7(2H_2O)$
              (pH 7.0)

2. Insulin 5'HVR (detects INS) - kindly provided by Dr. Graeme I. Bell, Department of Biochemistry and Biophysics, University of California

3. YNH24 (detects D2S44) - kindly provided by Dr. Y. Nakamura, Howard Hughes Medical Institute and Department of Human Genetics, University of Utah

4. CMM101 (detects D14S13) - kindly provided by Dr. Y. Nakamura, Howard Hughes Medical Institute and Department of Human Genetics, University of Utah

5. Haup15 - Adenine phosphoribosyltransferase (APRT) probe kindly provided by Dr. Peter J. Stambrook, Department of Anatomy and Cell Biology, University of Cincinnati College of Medicine

6. ADAN16 - Adenine deaminase (ADA) probe purchased from the American Type Culture Collection (ATCC), catalogue # 57227.

7. H25-3.8 - Carbonic anhydrase II (CAH) probe purchased from the American Type Culture Collection (ATCC), catalogue # 57263.

8. HLPLG18 - Lipoprotein lipase (LPL) probe kindly provided by Dr. S. S. Deeb, Department of Medicine, Division of Medical Genetics, University of Washington

CHAPTER 1

VNTR ANALYSIS

Section I - Data Base Generation

Section II - Forensic Case Samples

INTRODUCTION

Prior to 1985, DNA genetic typing for individual identification was virtually unknown to the forensic community. The first successful use of DNA fingerprinting in a criminal case, in the United Kingdom in 1987 [50], catapulted this revolutionary typing system into the public and forensic arena. The obvious potential of DNA genetic typing and the excitement fuelled by the media, resulted in a great demand by law enforcement agencies for such a forensic service. However, as the system was still in its infancy at that time, a great deal of research and development was required to establish the reliability and validity of the typing results.

At the onset of this thesis project in late 1987, most North American forensic laboratories were just beginning to focus more on single locus VNTR probes as opposed to minisatellites probes as developed by Jeffreys in the United Kingdom [9,10]. One of the major tasks was to build population data bases for the VNTR loci utilized in the typing procedure. Each laboratory began generating its own data base because of problems of consistency in allelic frequencies of data bases collected by sampling different geographic locations. It was not known whether allelic frequencies generated by sampling one population would be applicable to another population due to variations in subpopulation (i.e. ethnic subgroups) frequencies. Another reason for the separate data bases was that there was no unanimity in the choice of probes among all the forensic laboratories. Since the forensic application of DNA analysis was new, there was no consensus as to which particular probes would be most appropriate.

The complete absence of research into this novel field in British Columbia was the main impetus for this thesis project. With the dubious applicability of data bases generated elsewhere, it was important to build data bases for the local population (Greater Vancouver Area). Another important consideration was to determine and identify problems which may make the promising DNA genetic typing system (by VNTR) invalid and/or unreliable. Potential problems could be envisioned with respect to :

1) data bases

2) variability of the VNTR loci

3) technical procedures

4) applicability to forensic specimens which is influenced by the

collection and storage of such specimens

METHODS

1. PREPARATION OF DNA

Isolation of DNA From Liquid Blood (Protocol A) :

Deoxyribonucleic acid was extracted from approximately 6 ml of EDTA blood according to the preferential cell lysis method (David Hoar - Childrens' Hospital, Calgary Alberta : personal communication). Briefly, red blood cells were osmotically lysed with 5 volumes of a solution containing 0.017 M of Tris-HCl (pH 7.65) and 0.139 M of NH_4Cl. The remaining intact nucleated cells were washed twice with saline and resuspended in 2 ml of HTE buffer (100 mM Tris-HCl, 40 mM EDTA, pH 8.0). Two milliliters of WBC lysis buffer (100 mM Tris-HCl, 40 mM EDTA, pH 8.0, 0.2% SDS) were immediately injected into the nucleated cell suspension to lyse the cells instantaneously. The DNA released was purified by two phenol/chloroform/isoamyl alcohol (25:24:1) extractions and one chloroform/isoamyl alcohol (24:1) extraction. The aqueous layer containing the DNA was then precipitated with two volumes of 95% ethanol. The precipitated DNA was pelleted and finally resuspended in LTE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5).

Isolation of DNA From Liquid Blood (Protocol B) :

Deoxyribonucleic acid was extracted according to the protocol described by the FBI [93]. Briefly, liquid blood samples were frozen at $-70^{o}C$ in 700 ul aliquots. A frozen sample was thawed at room temperature, mixed with 800 ul of 1X SSC (150 mM NaCl, 15 mM sodium citrate, pH 7.0), and centrifuged for one minute in a Eppendorf microcentrifuge. One milliliter of the supernatant was discarded and replaced with one milliliter of 1X SSC. This mixture was centrifuged for one minute in the microcentrifuge after which all the supernatant was discarded. The pellet was resuspended with a 405 ul solution of 0.18 M sodium acetate, 0.62% SDS and 0.25 mg/ml Proteinase K. This suspension was then incubated at $56^{o}C$ for one hour, extracted once with 120 ul of phenol/chloroform/isoamyl alcohol (25:24:1), and once with 400 ul chloroform/isoamyl alcohol (24:1). The DNA in the aqueous layer was precipitated with one

milliliter of cold absolute ethanol, and centrifuged for 30 seconds before the supernatant was discarded. The DNA pellet was resuspended in 180 ul of TE^{-4}, incubated at 56°C for 10 minutes, and reprecipitated with 0.2 M NaAcetate and 500 ul of cold absolute ethanol. The precipitate was pelleted with a 10 second centrifugation in the microcentrifuge, washed with one milliliter of room temperature 70% ethanol, and vacuum or air dried before it was resuspended in 100-200 ul TE^{-4} and incubated overnight in a 56°C water bath.

Isolation of DNA From Forensic Samples (Protocol C) :

Deoxyribonucleic acid was extracted from bloodstains, hair shaft cells and tissue fragments according to the protocol described by Gill et al. [15]. Briefly, hair bulbs or bloodstains cut into small pieces, or chopped up tissue fragments were incubated on a shaker overnight at 37°C in extraction buffer (10 mM Tris-HCl, 10 mM EDTA, 100 mM NaCl (pH 8), 2% SDS, 100 ug/ml proteinase K, and 39 mM DTT). The sample was then extracted with one volume of liquified phenol and the aqueous layer was then extracted one volume chloroform/isoamyl alcohol (24:1). The aqueous layer was removed and treated with ethanol overnight at -20°C to precipitate DNA. After centrifugation, the DNA pellet was resuspended in LTE at room temperature.

Purification of DNA Extracted From Stains :

The resolubilized DNA was placed on a Millipore disc filter (0.05 uM) and dialysed for 2-4 hours in LTE.

Quantitation of DNA :

Various concentrations of calf thymus DNA (Hoefer Scientific Instruments) were used as standards for fluorometric quantification of linear DNA. A small volume (4 ul) of the DNA sample prepared by the methods described above, was diluted with (1000 ul) Hoechst 33258 (bis-benzimidazole) stain (0.1 ug/ml in TNE). The fluorometric density of this diluted sample was obtained with a Hoeffer TKO 100 mini fluorometer.

2. AMPLIFICATION AND PREPARATION OF DNA PROBES

Transformation of Recombinant Plasmids into Competent E.Coli

(DH5α) Cells :

Recombinant plasmids were amplified by transformation into DH5α bacterial cells and subsequent culturing of the transformed cells. The transformation procedure was performed as described in the protocol provided with the E.coli cells purchased from BRL. Briefly, a 20 ul aliquot of competent cells was removed from the -70°C freezer and thawed on ice. Five nanograms of recombinant plasmid was added to the tube of thawed cells and mixed gently with the pipette tip. The tube mixture was left on ice for 30 minutes and then heat shocked at 37°C for 20 seconds. The tube was placed on ice and 0.95 ml of Luria-Bertani (LB) medium was added before it was incubated for one hour at 37°C with aeration (on top of a horizontal platform shaker moving at 200 rpm). The cells from this culture were then spread on to LB/X-gal agar plates containing a specific antibiotic. The choice of the antibiotic used was dependent upon which antibiotic resistant gene was present within the recombinant plasmid used in the transformation. X-gal (5-bromo-4-chloro-3-indolyl-ß-D-galactoside) is a chromogenic substrate of ß-galactosidase and cleavage of this substrate results in the production of a blue color. The plates containing the transformed bacterial cells were incubated at 37°C until colonies appeared. Cells containing either the recombinant or non-recombinant plasmids formed colonies on the antibiotic plates due to the presence of the antibiotic resistant gene contained in the plasmid. Recombinant plasmids which contained a defective ß-galactosidase gene due to the insertion of the foreign DNA of interest, produced white colonies when transformed into these E.coli (DH5α) cells because X-gal is not cleaved to produce the blue color. These white colonies were selected for subsequent manipulations (i.e. plasmid isolation.

Small Scale Amplification and Isolation of Plasmid DNA :

The small scale amplification of plasmid was performed as described in the Cold Spring

Harbor manual [94] except for minor alterations. Bacterial cells containing recombinant plasmids

were inoculated into 5 ml of LB medium containing antibiotic corresponding to the antibiotic

resistance gene within the recombinant plasmid. The inoculated medium was incubated overnight

at $37^{o}C$ with aeration. A 1.5 ml volume of this culture was centrifuged at 13,000 xg for one minute

and the resulting pellet was resuspended in 100 ul of an ice-cold solution of 50 mM glucose, 10

mM EDTA, 25 mM tris-HCl (pH 8.0). This suspension was left at room temperature for 5

minutes before adding 200 ul of an ice-cold solution of 0.2 N NaoH and 1.0% SDS. After mixing,

the suspension was stored on ice for 5 minutes and then 150 ul of an ice-cold solution of 5 M

potassium acetate (pH 4.8 was added. The resulting suspension was mixed by vortexing the tube in

an inverted position for 10 seconds. After storing on ice for 5 minutes, the suspension was

centrifuged for 5 minutes at 13,000 xg to pellet out the precipitated cellular DNA and debris. The

supernatant was then extracted twice with one volume of phenol/chloroform/isoamyl alcohol

(25:24:1) and then ethanol precipitated at room temperature for 5 minutes. The pellet from this

precipitation was washed once with 70% ethanol, dried in a vacuum desiccator, dissolved in LTE

and treated with RNase to remove contaminating RNA.

Excision of Foreign DNA Insert From Recombinant Plasmids :

Recombinant plasmids dissolved in LTE were digested with the appropriate restriction

endonuclease to cleave at the sites immediately flanking the insert of interest. The digest

conditions were those recommended by the suppliers of respective restriction endonucleases. DNA

fragments in the digest were separated by electrophoresis through 0.8% low melting point agarose

gel (containing 1 ug/ml ethidium bromide) in 1X TBE or 1X TAE buffer. The insert band as

determined by size, was excised from the gel and placed in a 1.5 ml tube containing LTE. The tube

was placed in a $55^{o}C$ heating block until the agarose melted (approximately 5 minutes). The

resulting suspension was extracted once with one volume of liquified phenol, once with one volume

of phenol/chloroform/isoamyl alcohol (25:24:1), once with chloroform/isoamyl alcohol (24:1) and then ethanol precipitated to purify and collect the insert DNA. The purified DNA was finally resuspended in LTE.

Labelling and Purification of DNA Probes :

The DNA probes were labelled using the random primers DNA labelling system kit purchased from BRL. The kit protocol was followed except that the labelling reaction was allowed to proceed for 4-6 hours. Purification of the labelled probe was achieved with a Sephadex G-50 column. The column was prepared by packing Sephadex G25-80 into a commercially available Quik-Sep disposable chromatography column (Isolab Inc.). The packed column was washed twice with LTE. The probe was then passed through the Sephadex column by centrifugation for three minutes at 5000 xg to remove the unincorporated dNTPs. A portion of the eluted probe (2 ul/300 ul of eluent) was mixed with 10 ml of Aquasol (Dupont's universal LSC cocktail) and analyzed using a liquid scintillation counter (LKB 1217 RackBeta) to determine the specific activity of the probes.

3. VNTR ANALYSIS

Restriction Endonuclease Digestion of DNA Samples and Agarose Gel Electrophoresis :

Five micrograms of DNA were digested with 12.5 - 30 units of either Pvu II or Hae III. Digestion conditions for these restriction endonucleases were those recommended by the supplier, BRL, using a digest incubation period of 4 hours. DNA digests were separated by electrophoresis through 1.2% agarose gel in 1X TBE buffer (0.089 M of Tris-borate, 0.089 M boric acid, 0.025 M EDTA, pH 8.3). The progression of the DNA through the gel was detected with 0.04% bromophenol blue and 0.4% xylene cyanol. Electrophoresis was stopped when the xylene cyanol dye had migrated to 1 1/2 - 2 centimeters from the end of the gel (approximately 4 hours at 80

volts and 35 mAmps). The gel was then stained for 10 minutes in EtBr (1 ug/ml), soaked in distilled water for 10 minutes and viewed over an ultraviolet light box.

Molecular length Standards and Controls :

Molecular length standards or markers (BRL 1 kb ladder / Pharmacia ϕX174 RF-Hae III) and a control DNA sample were included in every gel. (Initial experiments utilized lambda-Hind III / ϕX174 RF-Hae III fragments as molecular length standards.) The molecular length standard lanes bracketed all the digested genomic DNA lanes eg. in a 14 lane gel, standards in lanes 1, 7 and 14. The standard in lane 7 was included to correct for variations in the middle of the gel. One or two DNA sample with known VNTR patterns i.e. sample analyzed several times previously, was also included to serve as a control sample for the quality of the typing procedure. A minimum of two control DNA samples, one male and one female, were included in each gel during analysis of case material.

Transfer of DNA From Gel to a Solid Support :

A modified Amersham alkali transfer protocol was used to transfer the electrophoretically separated DNA fragments from the gel to a Hybond nylon membrane. Briefly, the agarose gel was soaked in denaturation buffer (1.5 M NaCl and 0.5 M NaOH) for 30 minutes at room temperature. The gel was then placed in a capillary blotting apparatus with an alkali transfer buffer (1.5 M NaCl and 0.25 M NaOH) for 4-6 hours. The hybond membrane containing the transfer DNA was washed in 5X SSC for 2-5 minutes and baked at 80°C for 2 hours to fix the DNA.

Hybridization of ^{32}P-DNA probes to membrane-fixed DNA :

The baked Hybond membrane was incubated with prehybridization solution (5X SSPE, 5X Denhardt's solution, 1% SDS and 1% powdered non-fat milk) in a 65°C shaking water bath for 30-60 minutes. The DNA probe (prepared as previously described) was first denatured by heating in a 100°C water bath for 10 minutes, cooled on ice for 5 minutes and then added to the prehybridization solution containing the Hybond membrane. The concentration of probe in the

hybridization solution was 1.5 - 2.5 ng/ml. Hybridization was allowed to proceed in a 65^oC water bath (Blue M) for at least 12 hours.

Post hybridization washes :

Following hybridization, the membrane was washed twice with solution I (2X SSPE and 0.1% SDS) at room temperature for 10 minutes (each time), once with solution II (1X SSPE and 0.1% SDS) at 65^oC for 15 minutes, and a final wash with solution III (0.1X SSPE and 0.1% SDS) at 65^oC for 5 - 15 minutes. The incubation period of the final wash varied depending upon the probe used in the hybridization mix. The decreasing salt concentration of each successive solution and the increasing temperature of incubation, increased the stringency of hybridization. The washed membrane was wrapped in Saran-Wrap before it was exposed to Kodak X-OMAT AP or AR X-ray film at -70^oC for various time periods.

Probe removal from membrane for reprobing :

In order to increase the level of discrimination, more than one VNTR region was analyzed. Hence a membrane was sequentially probed up to four times with different probes. Between each reprobing, the filter was striped of the previous probe. Removal of the bound probe was achieved by an incubation at 45^oC for 30 minutes in 0.4 M NaOH followed by another incubation at 45^oC for 30 minutes in a solution of 0.1 X SSC, 0.1% (w/v) SDS, and 0.2 M Tris-HCl pH 7.5.

Determination of the Molecular length (Sizes) of DNA Fragments :

The sizes of the genomic DNA restriction fragments were determined by comparing the distance migrated by these DNA fragments from a fixed point (arbitrary origin) at the top of the gel, with the distances migrated from the same fixed point by various fragments of the molecular length standards. Distance measurements were made to the middle of the band. The molecular length standards have been sequenced by the commercial suppliers and hence are of known sizes. A line was generated by plotting the distance migrated by the molecular length standards against their respective sizes. This line was then used to extrapolate the sizes of the genomic DNA

fragments according to their migration distances from the arbitrary origin. The measurements were manually performed by a single person. The extrapolation was performed by a computer using a program model [95] based on Southern's proposal that the reciprocal of DNA mobility plotted against fragment length is linear [96]. However, a correction factor was added to each measured distance value for the genomic DNA fragments before they were entered into the computer. This was to correct for the skewness of the molecular length standard lanes within the same gel (i.e. identical sized fragments at progressively higher or lower positions in the gel going from one lane to the next). This skewness is due to internal problems in the electrophoretic system such as, unequal temperatures in centre of the gel as opposed to the outer edges because of heating effects of the electric current, and minute aberrations in the gel due to impurities in the agarose. The correction factor was calculated by dividing the distance differences between two known identical fragments in two flanking molecular length standards lanes by the number of lanes between the standards lane plus one of the standard lanes.

Example: Molecular length standard lanes - 1, 7 and 14

Number of lanes between standard lanes 1 and 7 = 5

Measured distance difference between 6.04 kb marker in
lanes 1 and 7 = 1.0 mm

Therefore, correction factor = 1.0 mm / (5+1) = 0.17 mm

Hence, lane 4 distance for allele approximately

6 kb in size = measured distance + (number of lanes
from first marker lane x 0.17 mm)

= measured distance + (3 x 0.17 mm)

Assessment of the Precision of the Typing System :

A mixture of DNA samples from three different individuals was digested and separated by electrophoresis as previously described for VNTR analysis. The DNA samples were chosen so that when probed sequentially with two probes (3'HVR/YNH24), a number of bands (alleles) were

detected at regularly spaced intervals throughout the length of the gel. By having the DNA bands at regular intervals spread from the top to the bottom of the gel, it was possible to get a range of DNA fragment sizes (alleles) representative of the alleles of the four VNTR regions probed. This digested DNA sample mixture was run in six lanes on the same gel and repeated in 15 gels. Molecular length standards (1 kb ladder / ϕX174 RF-Hae III) lanes were also included in each of the 15 gels and were situated such that the molecular length standards and the genomic DNA mixtures were in alternating lanes (i.e. lanes 1,3,5,7,9,11,13 - standards, lanes 2,4,6,8,10,12 - genomic DNA) (Figure 1). Lanes 1, 7 and 13 were used as standard lanes for the genomic DNA mixtures to simulate the conditions used in the regular typing procedure. Lanes 1, 7 and 13 were used as standard lanes for the determination of the fragment sizes of the other standard lanes (lanes 3, 5, 9, and 11) which were in this situation treated as fragments of unknown sizes. Determination of DNA fragment sizes were performed as described in the previous section (page 32-33).

Within each group, i.e. the lanes containing genomic DNA (Figure 1 - even number lanes) or the molecular length standard lanes (Figure 1 - odd number lanes), the experimental size variation for identical fragments in the respective lanes within a gel and between gels were used to calculate the confidence intervals for the specific size range. The confidence intervals for each size range were in turn used to create bins for allelic assignments.

The rationale for analyzing the degree of size variation of identical fragments not only in genomic DNA lanes (fragments not sequenced such that their true DNA size was unknown) but also in the lanes containing the molecular length standards (commercially sequenced DNA of known size), was to establish whether there was a difference in variability between commercially obtained DNA and the more complex genomic DNA extracted in this laboratory.

Assessment of the Accuracy of Experimental Size Determinations :

The molecular length standard lanes (Figure 1 - odd numbered lanes) in the gels described in the preceding section, were also used to assess the accuracy of the DNA typing procedure. As

Figure 1 : Assessment of Accuracy and Precision of
Electrophoretic Separation

Probe 3'HVR α-globin was used to detect DNA fragments digested with Hae III. Odd

number lanes contain molecular length standard DNA (BRL 1 kb/ϕX174 RF-Hae III).

Even number lanes contain Hae III digested genomic DNA from three different individuals

in order to produce simultaneously a number of fragment of various sizes. Lanes 1, 7

and 13 were used as molecular standard lanes

i.e. lanes 1 and 7 were used as standards for lanes 2-6

lane 7 and 13 were used as standards for lanes 8-12

The molecular length standard DNA in lanes 3, 5, 9, and 11 were treated as unknowns so

that accuracy and precision measurement could be determined with both commercially bought

non-genomic DNA and genomic DNA extracted from this laboratory.

Figure 1 : Assessment of Accuracy and Precision of
Electrophoretic Separation
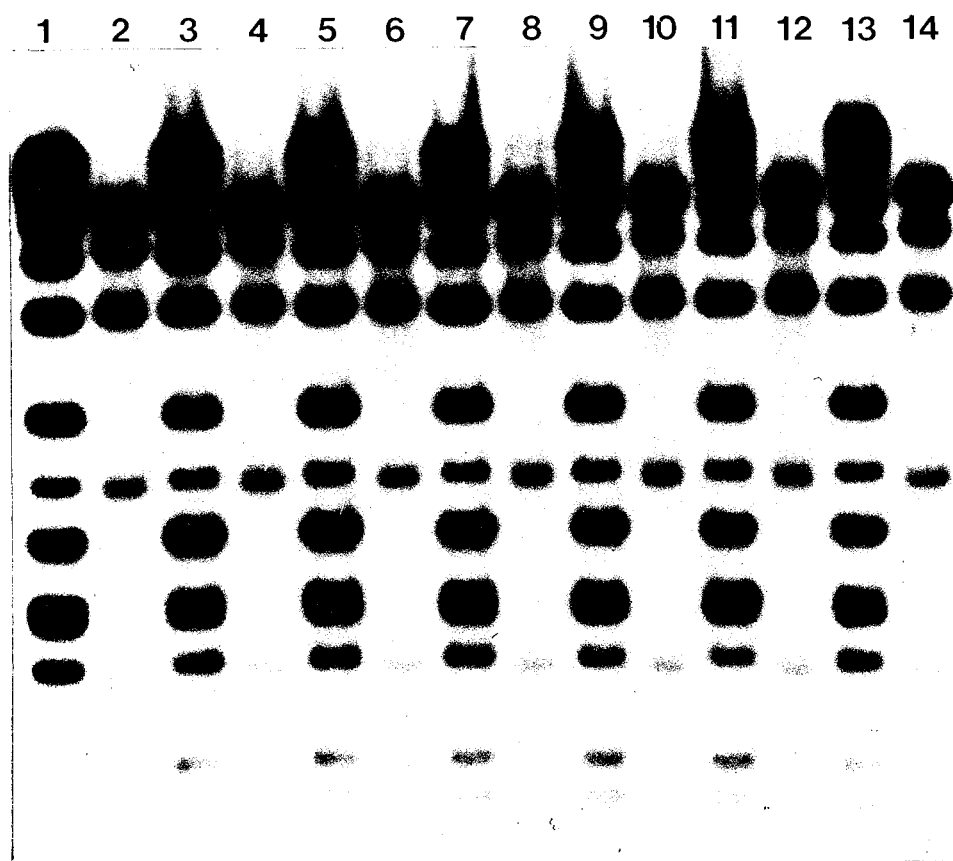
stated previously lanes 1, 7 and 13 were used as standard lanes and the other standard lanes (3, 5, 9 and 11) were treated as fragments of unknown sizes. Comparison of the experimentally calculated size with the true sizes provided an indication of the accuracy of the typing system in this laboratory.

<u>SECTION I</u>

RESULTS AND DISCUSSION

Population Samples :

The VNTR data in this thesis was collected from an undefined population. This population sample consist of individuals admitted to The Vancouver General Hospital (live patients or morgue autopsy patients). Patients admitted to the hospital had blood drawn for routine tests and the leftover blood was used for DNA typing analysis. Blood drawn from morgue patients at autopsy were also used in this analysis. No other selection criteria was placed on the samples. Details are not known about these samples as hospital policy does not allow questions such as ethnic or racial origins to be included in the information sheets unless the blood samples were initially designated for research purposes.

In 1987 when this project was being planned, the issue of collection and identification of population control samples was a matter of great concern to my supervisory committee. At that time in Vancouver, there was a great deal of concern within the medical and biotechnology fields as to the issue of confidentiality of samples and the public press concerns had been raised about the accessibility of results of any form of DNA genetic typing to police, governments and other agencies. The supervisory committee decided that I would be allowed to use blood samples from the Haematology department which had been collected for other purposes, tested and were to be discarded. I would be allowed access only to the names of the individuals for the determination of racial origin provided that all subsequent information regarding the patients was destroyed and that I would not be allowed to persue the issue of racial identification beyond this process. Access to autopsy material was limited by the fact that the B.C. Coroners Act specifically forbids the use of Coroners' autopsy material for research or teaching purposes. This has resulted in limited and conservative statistical interpretation of these population studies. However, within the context of a

forensic investigation, this type of limitation is more often the rule and as a result, these population studies are very relevant to practical forensic case investigations.

Duplication of samples is not a significant problem in this sampling because samples from live patients who may provide blood more than once, were checked by patient names before the records of the names were destroyed for patient confidentiality. Errors in checking could have occurred but should not be large enough to significantly alter the frequency values. Morgue patients have blood drawn only once and hence no sample could have been repeated.

Technical Conditions For VNTR Analysis of Blood DNA :

Two different methods for DNA isolation from liquid blood were utilized in this study. Protocol A was used for the majority of the DNA isolations for VNTR analysis because it :

1) provided large quantities of high molecular length DNA

   suitable for DNA typing

2) was technically undemanding

3) it had been verified by long use in this laboratory

However, protocol B was used to extract DNA from blood samples processed during the latter part of this study. The first method begins with a primary osmotic lysis of the red blood cells (anucleated), followed by lysis of the white blood cells (nucleated) with a detergent (SDS). The second method is similar except that there is an initial freezing/thawing step before the osmotic lysis of the red blood cells. Other than the differences in methods of cell lysis between the two protocols, there is also a difference in the quantity of blood processed. Protocol B is designed to process a much smaller volume of whole blood: ie. 700 ul compared with the usual volume of 5-7 ml by protocol A. The amount of DNA extracted from 700 ul of whole blood is usually sufficient for 5 VNTR analyses. Excess DNA from the large scale extraction by protocol A is usually not required for the average VNTR analysis. Protocol B is a more efficient method to use since less materials are required for this small scale extraction. In the event that more DNA is needed, additional blood stored frozen in 700 ul aliquots can be extracted.

Furthermore, since protocol B is used by the FBI laboratory and it seemed appropriate to adopt procedures developed by the FBI laboratory as they are the predominant group in the United States responsible for the development of forensic DNA analysis techniques. Standardization of forensic DNA analysis techniques would greatly facilitate efforts towards quality control and quality assurance. Data gathered using standardized techniques where possible will be much more valuable for comparative purposes.

Technical conditions for preparation of DNA from forensic specimens :

The methodology described for protocol C was the basic procedure followed. Since forensic specimens can come in a variety of forms, additional procedures were added or slight changes were made to existing procedures where appropriate. These alterations will be discussed individually for each case in section II of this chapter.

Restriction Endonuclease Digestion of Genomic DNA :

Originally restriction endonuclease Pvu II was used to cleave genomic DNA for VNTR analysis. Subsequently, Hae III was utilized in place of Pvu II because it was decided that Hae III would be more appropriate for VNTR analysis of forensic specimens [97]. This decision was based on that fact that :

1) Hae III recognizes a four base sequence G-G-C-C and will cut DNA

more frequently than five or six base cutters such as Hinf I and

Pvu II. This leads to smaller fragments (alleles) which enables

greater resolution of small size differences between fragments.

2) Hae III is not sensitive to methylation which could interfere with

cleavage of DNA.

3) Hae III maintains its activity under adverse conditions such as those

found in contaminated forensic samples.

4) Hae III is inexpensive compared to other endonucleases.

The success of this entire DNA typing procedure is dependent on the degree of DNA digestion by the restriction endonuclease Hae III. Incomplete digestion of DNA will result in more than the expected two DNA bands detected by each VNTR probe. This complicates the interpretation of the banding pattern seen on autoradiogram and may lead to inaccurate conclusions. A number of experiments were performed to establish the quantity of restriction endonuclease and the length of the incubation time required for complete digestion of DNA obtained from blood. Restriction endonuclease concentration was varied from 2.5 - 6.0 units/ug DNA and the incubation time was varied from 2 hours to 20 hours (overnight). Complete digestion of DNA was obtained in every case even with the lowest concentration of restriction endonuclease (2.5 units/ug DNA) incubated for the shortest time interval (2 hours) (Figure 2). Incomplete digestion would have been detected by the presence of multiple bands (more than 2) within each lane. Hence for subsequent digests, the incubation time was 4 hours with a restriction endonuclease concentration of 2.5 units/ug DNA.

Detection of Hypervariable Regions of DNA Using Multiple VNTR DNA ^{32}P-Probes :

Hybond membranes bound to electrophoretically separated digested genomic DNA were sequentially probed with four single locus hypervariable VNTR DNA ^{32}P-probes. Membranes were stripped between each successive probing as described in the methods (p44). The four hypervariable VNTR probes were:

1) YNH24 [98]

2) 3'HVR/α-globin [44,45]

3) phins 310 [42,43]

4) CMM101 [98]

These probes were selected because they were freely available during the on-set of this study. General properties of the four probes are listed in Table 1. Each of these single locus probes

Figure 2 : Conditions For DNA Digestion

Restriction endonuclease concentrations of 2.5 - 6.0 units/ug DNA were incubated with

5 ug genomic DNA for periods up to 16 hours (O/N). The electrophoretically separated

digested samples were hybridized with VNTR probes YNH24 (panel A) and phins 310 (panel B)

to determine the degree of digestion of DNA samples.

Lane   1   =   2.5 units/ug DNA

        2   =   4.5 units/ug DNA

        3   =   6.0 units/ug DNA

Figure 2 : Conditions for DNA Digestion
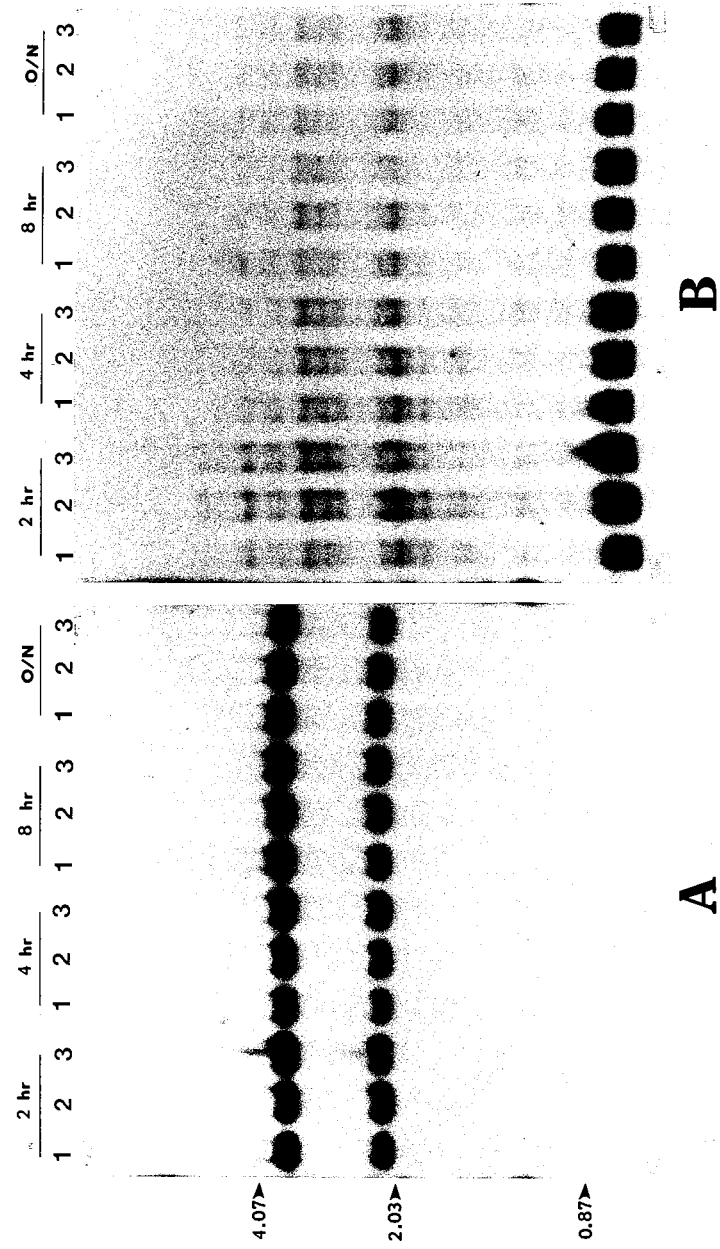
Table 1 : General Properties of VNTR Probes

Probe	Locus	Size (bp)	Core Repeat (bp)	Identified by
YNH24	D2S44	2000	31	HBV-2 oligonucleotide
3'HVR (α-globin)	D16S85	4000	17	Human a-globin pseudogene (ψç1)
phins 310 (5'HVR - insulin)	INS (chromo 11)	879	14 - 15	Human insulin cDNA
CMM101	D14S13	2200	15	Human Myoglobin-2 oligonucleotide

detected a maximum of two bands per lane (individual) representing the alleles for the particular locus (Figures 3-6).

Accuracy of the DNA Typing Procedure :

The validity of the DNA typing procedure is critically dependent upon the precision and accuracy of experimentally calculated allelic sizes. Precision (deviation from the mean size determination) plays a major role in the derivation of confidence intervals and will be discussed in the following section. The accuracy of this system was determined by comparison of the experimentally obtained value with the true fragment size.

Fragments of known sizes (molecular length standards) as determined by sequencing by the supplier, were run in four lanes and repeated in 15 gels resulting in a total of 60 intra- and inter-gel comparisons for each standard size. The mean size determination for each standard was compared to the true fragment size (Table 2). The greatest deviation was the 4.072 kb fragment whose mean size determination of 4.278 kb was 5.0% away from its actual size. However, a high degree of accuracy was obtained with some of the other fragment sizes i.e. 1.018 kb and 0.872 kb fragments. There does not appear to be any correlation between fragment sizes and degree of experimental accuracy. Furthermore, the direction of errors is not consistent in that some estimates are above and some are below the true value. Other than measurement errors , inaccuracy of this typing system may be a possible result of the use of an inappropriate method for the extrapolation of DNA size to migration distance from the origin of the gel [99].

The method used in this analysis is one proposed by Schaffer and Sederoff [95] which incorporates a full set of standards for the extrapolation of migration distance to DNA size. Elder and Southern proposed that only the four neighbouring standards rather than a whole range of standards, be used for extrapolation [99]. Factors such as heterogeneity in the gel and temperature and voltage gradients across the gel will affect the mobility of DNA fragments leading to deviations from the simple reciprocal relationship of migration distance and molecular size. Therefore,

Figure 3 : VNTR Analysis With Probe YNH24

Genomic DNA from 11 individuals was probed with YNH24 which detects VNTRs at the

D2S44 locus (lanes 1-11). Three molecular standard DNA lanes are present (lane M),

bracketing the genomic DNA lanes.

[6 days exposure to Kodak X-OMAT AP film at -70°C with screens]
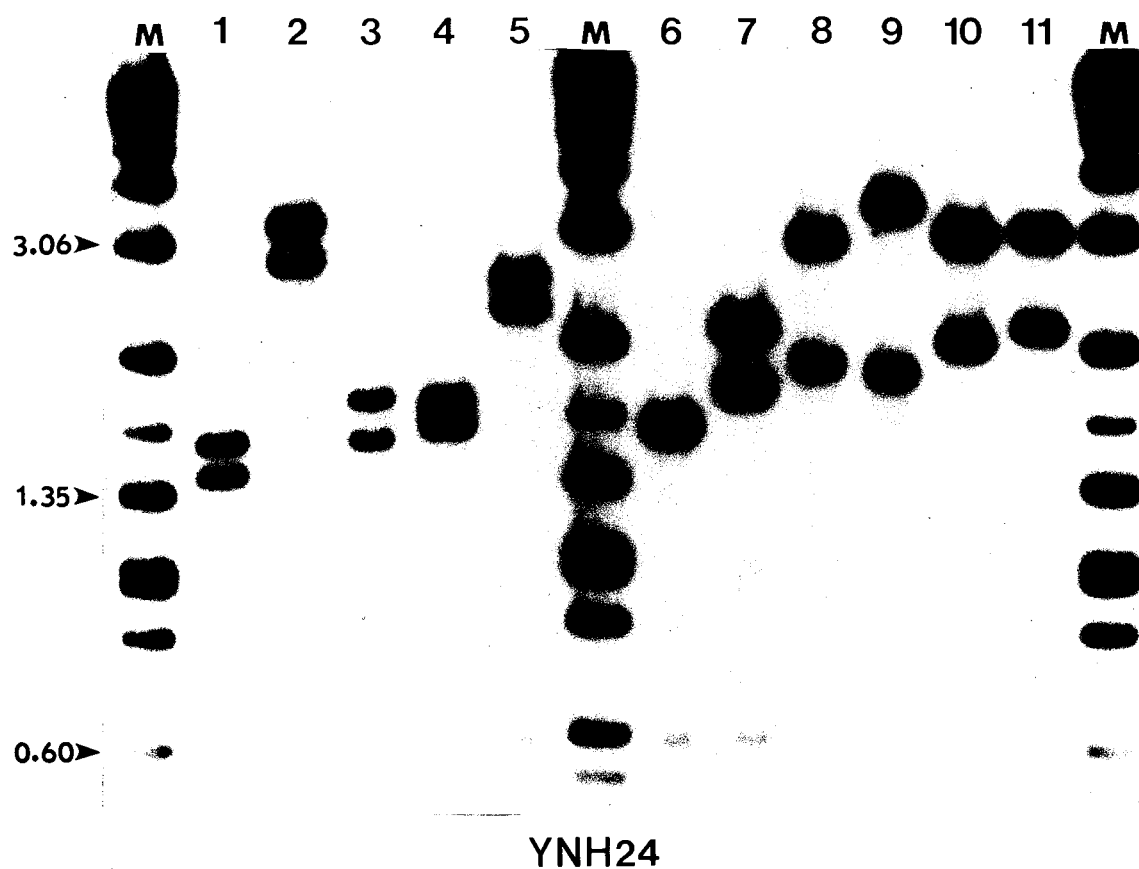
Figure 4 : VNTR Analysis With Probe 3'HVR α-Globin

Genomic DNA from 11 individuals was probed with 3'HVR α-globin which detects VNTRs at

the D16S85 locus (lanes 1-11). Three molecular standard DNA lanes are present (lane M),

bracketing the genomic DNA lanes.

[4 days exposure to Kodak X-MAT AR film at -70°C with screens]

Alpha-Globin
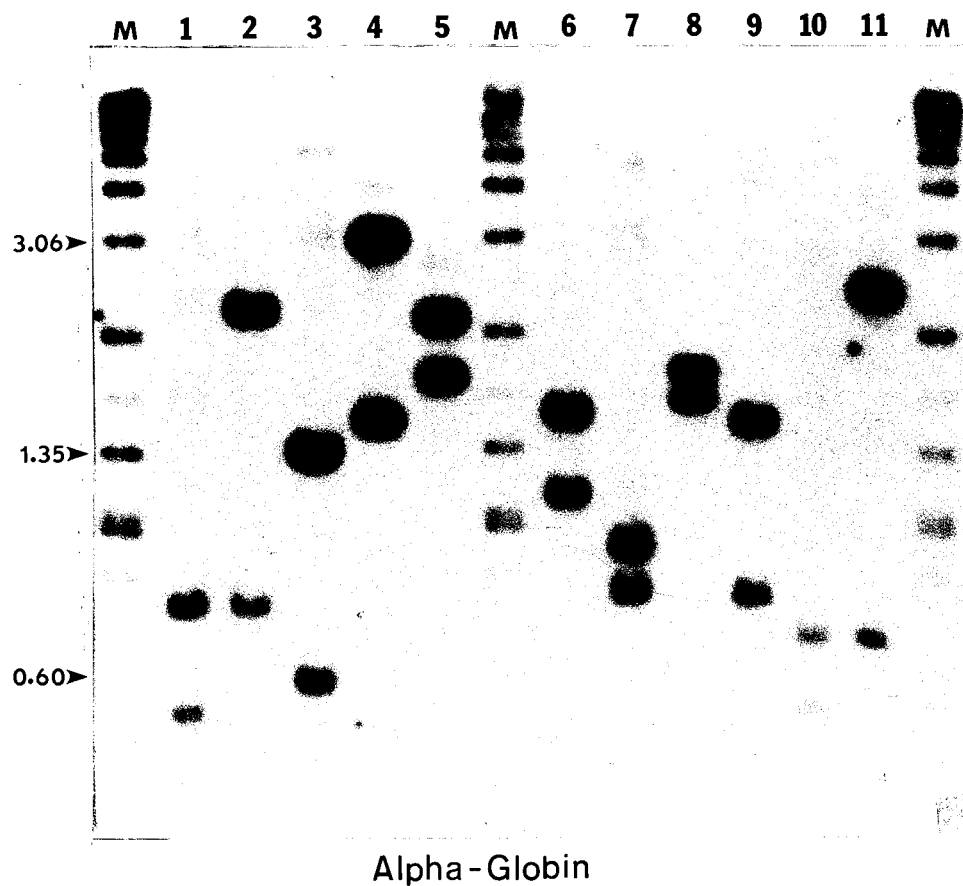
Figure 5 : VNTR Analysis With Probe phins 310

Genomic DNA from 11 individuals was probed with phins 310 which detects VNTRs at the

INS (insulin) locus (lanes 1-11). Three molecular standard DNA lanes are present

(lane M), bracketing the genomic DNA lanes.

[5 days exposure to Kodak X-MAT AP film at $-70^{o}C$ with screens]

p310
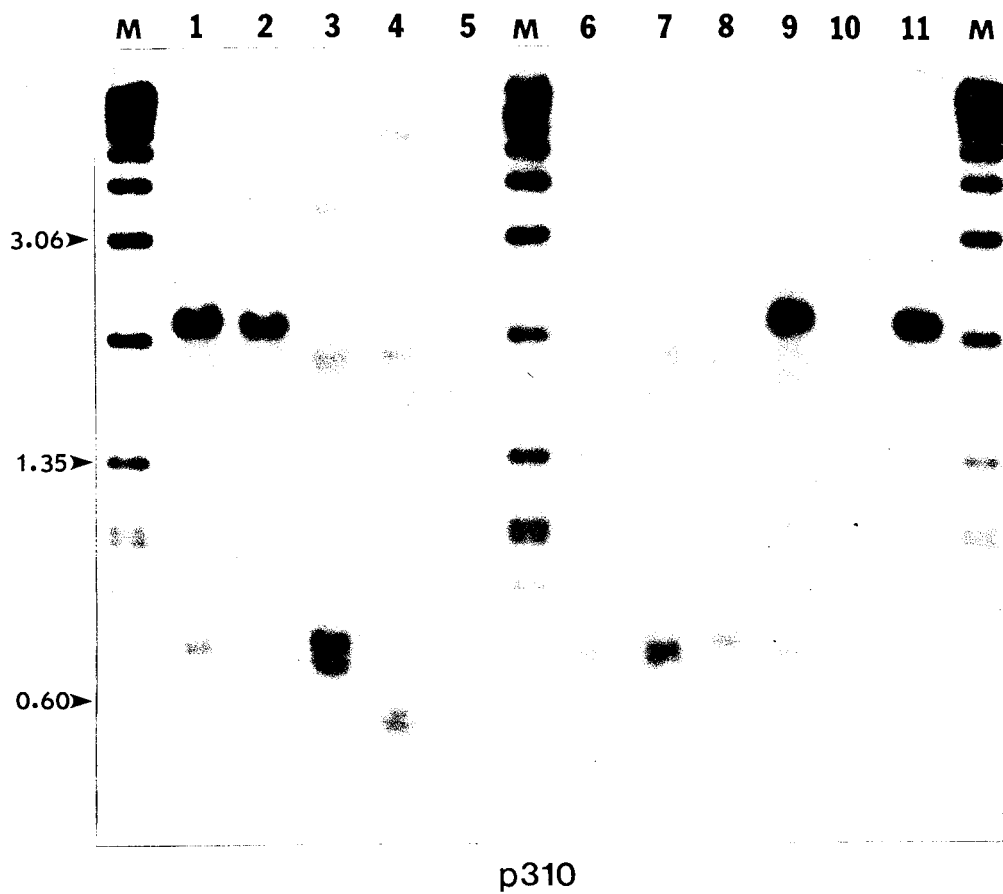
Figure 6 : VNTR Analysis With Probe CMM101

Genomic DNA from 11 individuals was probed with CMM101 which detects VNTRs at the

D14S13 locus (lanes 1-11). Three molecular standard DNA lanes are present (lane M),

bracketing the genomic DNA lanes.

[6 days exposure to Kodak X-MAT AP film at -70^{0}C with screens]

CMM101
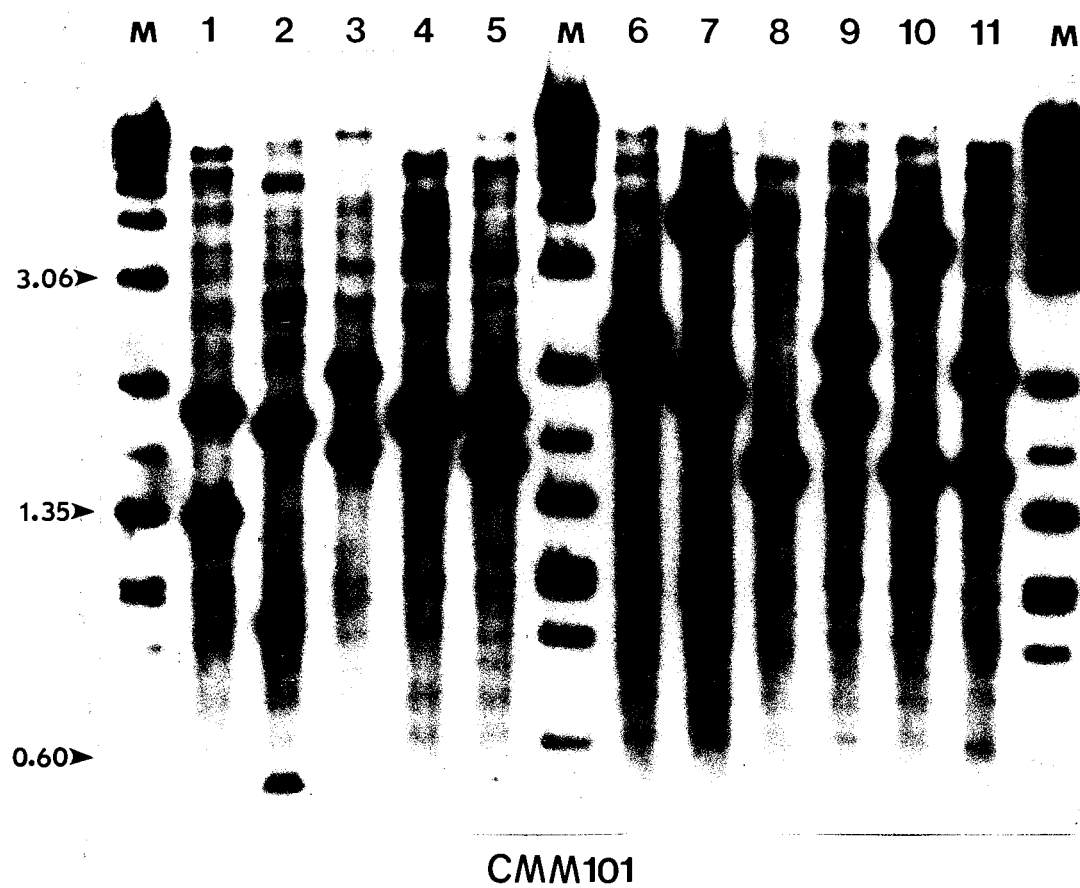
Table 2 : Accuracy of DNA Typing System

*Actual Size (kb)	Mean Determined Size (kb)	Variation from actual size (%)
6.106	5.992	1.9
5.096	5.231	2.6
4.072	4.278	5.0
3.064	3.187	4.0
2.030	2.052	1.1
1.636	1.606	1.8
1.353	1.318	2.6
1.078	1.052	2.4
1.018	1.011	0.7
0.872	0.864	0.9
0.603	0.621	3.0

n=60

* - Molecular weight standard DNA

according to Elder and Southern, incorporation of standards distant from the unknown to be measured (method of Schaffer and Sederoff) results in less accurate size determinations.

The published computer program for the method of Schaffer and Sederoff was the only program readily available to this author during the period of data analysis. Hence the method of Schaffer and Sederoff was utilized in the absence of a more suitable and accessible alternative.

Precision of DNA Typing System :

Assessment of the precision of the DNA typing system was accomplished by repetitive experiments with genomic DNA and molecular length standards. The experimental size variation of identical DNA fragments in different lanes within and between gels was determined by single factor analysis of variance, often abbreviated "ANOVA" [100]. The precision (deviation from the mean size determination of identical samples) of a system which determines bin sizes, is distinct from measurement imprecision (maximum size difference between identical samples) which defines the "match" criterion. Measurement imprecision will be discussed on pages 92-94.

For comparable sized fragments, the variance and thus the corresponding standard deviation was greater in the genomic DNA lanes than in molecular standards lanes (Table 3). This may suggest that the genomic DNA extracted in this laboratory is not as pure as commercially obtained molecular length standards. Impurities could affect the migration of the DNA fragments through the gel resulting in greater size variations. An alternate explanation may be the difference in amount of DNA electrophoresed through the gel. The molecular length standards lanes contain nanogram amounts of DNA whereas the other lanes contain 2.5 micrograms of digested genomic DNA. DNA mobility is affected by the total mass in the sample loaded in the well [101]. Increased mobility, band width and band distortion occurs with increased quantities (>0.2 μg per 0.15 cm^2 of surface area in the loading well) of DNA.

The magnitudes of variance and standard deviation are dependent upon the magnitude of the DNA fragment size [102]. Therefore, to compare the degree of precision for different sized DNA fragments, the "coefficient of variation" or "relative standard deviation" was calculated

Table 3 : Standard Deviation Comparisons of Molecular Weight Standards and Genomic DNA Fragments (Inter/Intra Assay Variation)

A

*MWS (kb)	Variance σ^2	Standard Deviation σ (kb)
5.992	0.00668	± 0.081
5.231	0.00400	± 0.064
4.278	0.00266	± 0.052
3.187	0.00121	± 0.035
2.052	0.00058	± 0.024
1.606	0.00030	± 0.017
1.318	0.00024	± 0.016
1.052	0.00010	± 0.010
1.011	0.00010	± 0.010
0.864	0.00006	± 0.007
0.621	0.00005	± 0.007

n = 4 lanes x 15 gels = 60

B

*Genomic DNA (kb)	Variance σ^2	Standard Deviation σ (kb)
5.376	0.01519	± 0.123
4.333	0.00610	± 0.078
3.961	0.00449	± 0.067
3.159	0.00212	± 0.046
2.610	0.00120	± 0.034
1.560	0.00036	± 0.019
1.490	0.00041	± 0.020
1.270	0.00031	± 0.017
0.830	0.00007	± 0.008
0.660	0.00006	± 0.007

n = 6 lanes x 15 gels = 90

MWS = molecular weight standard DNA

* = mean determinations

σ^2 = sample variance

σ = sample standard deviation

(Table 4). Relative standard deviations were $\leq$ 1.4% in the molecular length standards group and $\leq$ 2.3% for the genomic DNA fragment group. The values obtained for the latter group is of greater importance since genomic DNA samples are the unknowns in regular (actual) DNA typing analysis. There appears to be a trend, in the genomic DNA group, towards a greater deviation with increased DNA fragment size (Table 4 - panel B). This observation is not surprising since measurement error is usually a function of restriction fragment length [103]. A measurement error of a given distance results in a greater deviation in larger fragments than smaller fragments.

Precision may be influenced by a single factor or a combination of factors such as gel impurities, gel imperfections, and band measurement errors. Gel impurities should be minimal as BRL's Ultra Pure agarose and distilled filtered water was utilized. However, slight variations in the migration of DNA in outer and middle lanes did occur in some gels as a possible result of inconsistent heat dissipation. This curvature in the gel might have been eliminated with the use of a continuously flowing gel buffer system.

Distances of each band are determined from a fixed point to the centre of each band. Errors in band measurement may have affected the degree of precision since these measurements were made manually by one individual. Measurement errors may occur particularly with thick bands which make the determination of the centre of the band difficult especially without the aid of a densitometer, that is, by relying on visual inspection during band measurements. Despite the possibility of error introduced by manual assessments, the frequency distribution of alleles analyzed with the FBI's image analysis system was not significantly different (95% level) from the distribution of the same alleles measured manually. The greatest discrepancy between the two assessments was in bin 15 (alleles ranging in the size range of 1666-1795 bp) which had a frequency value difference of 3.1%. The validity of the typing results will be further discussed in one of the following sections (p94-97). This author would like to thank the FBI forensic laboratory for the use of their image analysis system which has made this comparison possible.

Table 4 : Relative Standard Deviation (Inter/Intra Assay Variation)

A

*MWS (kb)	V (%)
5.992	1.4
5.231	1.2
4.278	1.2
3.187	1.1
2.052	1.2
1.606	1.1
1.318	1.2
1.052	1.0
1.011	1.0
0.864	0.9
0.621	1.1

n=60

B

*Genomic DNA (kb)	V (%)
5.376	2.3
4.333	1.8
3.961	1.7
3.159	1.4
2.610	1.3
1.560	1.2
1.490	1.3
1.270	1.3
0.830	1.0
0.660	1.1

n=90

MWS = molecular weight standard DNA

* = mean determinants

V = relative standard deviation = coefficient of variation

= (σ / mean) x 100

The RCMP laboratory has assessed the precision of their typing results by repeated analysis of a 2.731 kb fragment [47]. The relative standard deviation of the mean size determination (2.724 kb) is 0.7%. Although typing results of the 2.610 kb genomic fragment from this study (Table 4 - panel B) exhibited a high degree of precision (1.3%), the RCMP laboratory's results with the 2.724 kb fragment is almost twice as precise. This discrepancy does not suggest that the procedures used in this laboratory are inadequate but merely that larger confidence intervals are required for the allelic bin system. Further comparisons with other fragment sizes were not possible due to the absence of published data.

Determination of Bin Sizes Using Confidence Intervals :

To account for intra-gel and inter-gel variabilities and measurement imprecision, alleles of the HVR for each locus were assigned to a bin corresponding to the appropriate size range (Table 5). Hence, several resolvable DNA fragments may contribute to the numbers generated for each bin. Since each bin was then considered as one allele, its frequency would always be an over-estimation of the true frequency for any one specific DNA fragment length.

In order to set the arbitrary boundaries for each bin, it was necessary to determine the confidence intervals for a range of DNA fragment sizes representative of the alleles of the four VNTR regions analyzed. The standard deviations obtained from the repetitive experiments (precision studies) were used to establish these intervals.

The mean determined sizes of the molecular length standards spanned a greater range of sizes than the genomic DNA fragments i.e. 5.992-0.621 kb versus 5.376-0.660 kb. Since the aim of these experiments was to create a binning system to encompass a wide range of allelic sizes, the mean molecular length standards sizes were used for confidence interval calculations. However, the variance of most similar sized genomic DNA fragment was used to derive standard deviation values of respective mean molecular length standards. It was necessary to use larger variances of the genomic DNA fragments because confidence intervals derived from these values were used to set up bins for genomic DNA alleles.

Table 5 : Allele Frequencies of Four VNTR Loci

BIN	Size Range (bp)	Frequencies (%)			
		D2S44	D16S85	INS	D14S13
1	0-629	0.1	8.9	17.2	0.2
2	630-677	0.0	1.4	15.2	0.0
3	678-726	0.0	2.2	29.4	0.0
4	727-774	0.0	7.4	12.1	0.0
5	775-822	0.2	13.9	1.5	0.0
6	823-904	0.5	8.7	0.2	0.1
7	905-986	0.4	4.2	0.0	0.2
8	987-1069	1.3	3.8	0.1	0.1
9	1070-1151	1.2	3.1	0.0	0.9
10	1152-1234	1.8	1.8	0.0	0.8
11	1235-1316	2.3	1.6	0.1	4.0
12	1317-1426	4.8	1.9	0.0	4.9
13	1427-1535	6.0	2.8	0.0	15.0
14	1536-1665	11.5	3.7	0.1	10.5
15	1666-1795	9.8	4.1	0.0	8.8
16	1796-1925	8.0	3.2	0.1	7.3
17	1926-2145	8.8	4.9	8.9	8.4
18	2146-2365	6.7	3.4	13.7	4.6
19	2366-2586	5.0	3.0	1.0	4.4
20	2587-2806	4.4	2.1	0.3	4.3
21	2807-3027	6.7	2.6	0.1	2.0
22	3028-3247	8.6	2.5	0.1	2.6
23	3248-3631	6.5	2.1	0.1	3.5
24	3632-4016	2.9	1.2	0.0	3.2
25	4017-4635	1.2	1.4	0.0	4.0
26	4636-5254	0.6	2.7	0.0	4.6
27	5255-5992	0.4	0.7	0.0	1.8
28	5993-7300	0.4	0.2	0.0	1.5
29	7301-	0.0	0.2	0.0	2.7

Example: (Values are listed in Table 3)

Mean mol. wt. std. = 5.992 kb   Mean genomic DNA = 5.376 kb

variance = σ^2 = 0.0066   variance = $\sigma 2$ = 0.0152

std. dev. = σ = ± 0.081   std. dev. = σ = ± 0.123

But use :   Mean mol. wt. std. = 5.992 kb

variance = σ^2 = 0.0152

As it was not practical to obtain the variances for an unlimited number of DNA sizes, an average variance was used for each region bracketed by the mean molecular length standard values. Confidence intervals (mean ± 3σ) for bins were established such that a genomic DNA fragment typed repeatedly in different lanes in the same gel or in different gels will be assigned to the same bin in 99.7% of the cases.

Example : (Values listed in Table 3)

1) Mean mol. wt. std. = 5.992 kb → → → Variance = σ^2 = 0.0152

2) Mean mol. wt. std. = 5.231 kb → → → Variance = σ^2 = 0.0152

3) Mean mol. wt. std. = 4.278 kb → → → Variance = σ^2 = 0.0061

Variance of region between 5.231 and 4.278 kb

= average of σ^2 of 5.231 and 4.278 kb

= (0.0152 + 0.0061) ÷ 2 = 0.0106

Standard deviation = σ = ± 0.103 kb

Confidence interval = δ = * 2 x 3σ

= 2 x 3(0.103)

= 0.6189 kb

* The 3σ value is multiplied by 2 because delta has to account for ± 3σ.

Hence, bin size for the region bound by the 5.2314 and 4.2778 kb standard fragments, should be 619 bp i.e. bins spaced 619 bp apart. The endpoints of each bin do not overlap to overcome the problem of bin assignment for alleles that are "on the line" of two bins.

Population Data Base For the Allelic Frequencies of the 3"HVR of the D2S44
Locus :

Genomic DNA from an undefined group of 578 individuals from the Greater Vancouver Area was analyzed for allelic types (sizes) at the hypervariable region of the D2S44 locus. This corresponded to the analysis of the alleles of 1156 chromosomes. Comparison of bin (allelic) frequencies of the first 578 chromosomes with the total frequency values obtained with 1156 chromosomes revealed no significant difference (at 5% significance level) between the two frequency distributions (Figure 7). This suggests that a sample size of only 289 individuals (578 chromosomes) is large enough to provide an adequate estimate of the frequencies of the DNA fragment sizes (alleles). However, because methodological errors are replicated in a single study, this is not a reflection on the accuracy of the DNA fragment sizes in relation to their true sizes which is unknown without DNA sequencing.

The frequency distribution for the alleles of D2S44 is shown in Table 5 and also graphically in Figure 8. For statistical analyses, the frequencies for bins 1-7 (DNA fragments ≤ 986 bp) will be combined as with bins 26-29 (fragments ≥ 4501 bp) in order to reduce the significance placed on these rare alleles. This is in accordance with the procedures followed by the FBI's DNA specialists who suggested that bins which have less than 5 observed events are not statistically significant [104]. In order to avoid situations where a DNA fragment, lying on or near the boundary of two bins, is assigned a frequency value lower than the true frequency of the sample population, the highest value of either the identified bin or its two flanking bins will be used in the calculations of the individuality of the DNA profile.

The frequencies of these D2S44 alleles show a slightly overlapping bimodal distribution with the first mode at bin 14 (1536-1665) and the second (smaller) mode at bin 22 (3028-3247 bp)

Figure 7 : Two Allele Frequency Distributions For the D2S44 Locus

Genomic DNA from an undefined group of 578 individuals from the Greater Vancouver Area

was analysed for alleleic types (sizes) at the hypervariable region of the D2S44 locus.

Allele frequency distributions of sample sizes (578 chromosomes and 1156 chromosomes)

are compared.

Figure 7 D2S44 (1156/578 Chromosomes)
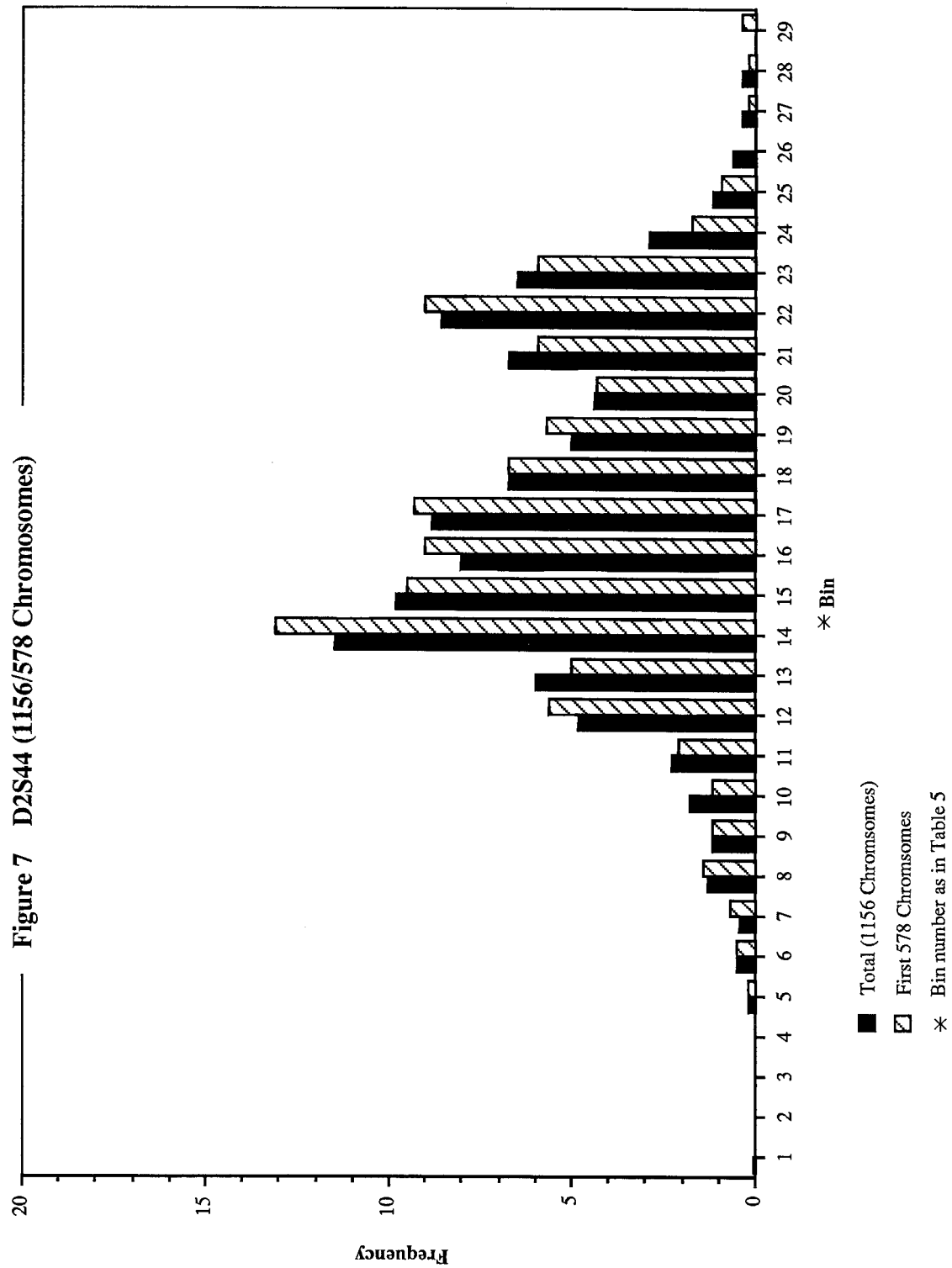
Figure 8 : Allele Frequency Distribution For the D2S44 Locus

Genomic DNA from an undefined group of 578 individuals from the Greater Vancouver Area

was analysed for alleleic types (sizes) at the hypervariable region of the D2S44 locus.

Allele frequency distribution of this locus is shown for 1156 chromosomes.

Figure 8   D2S44 (1156 Chromosomes)

Frequency (%)

* Bin	Frequency (%)
1	0.08
2	0
3	0
4	0
5	0.17
6	0.52
7	0.43
8	1.3
9	1.2
10	1.8
11	2.3
12	4.8
13	6
14	11.5
15	9.8
16	8
17	8.8
18	6.7
19	5
20	4.4
21	6.7
22	8.6
23	6.5
24	2.9
25	1.2
26	0.6
27	0.4
28	0.4
29	0
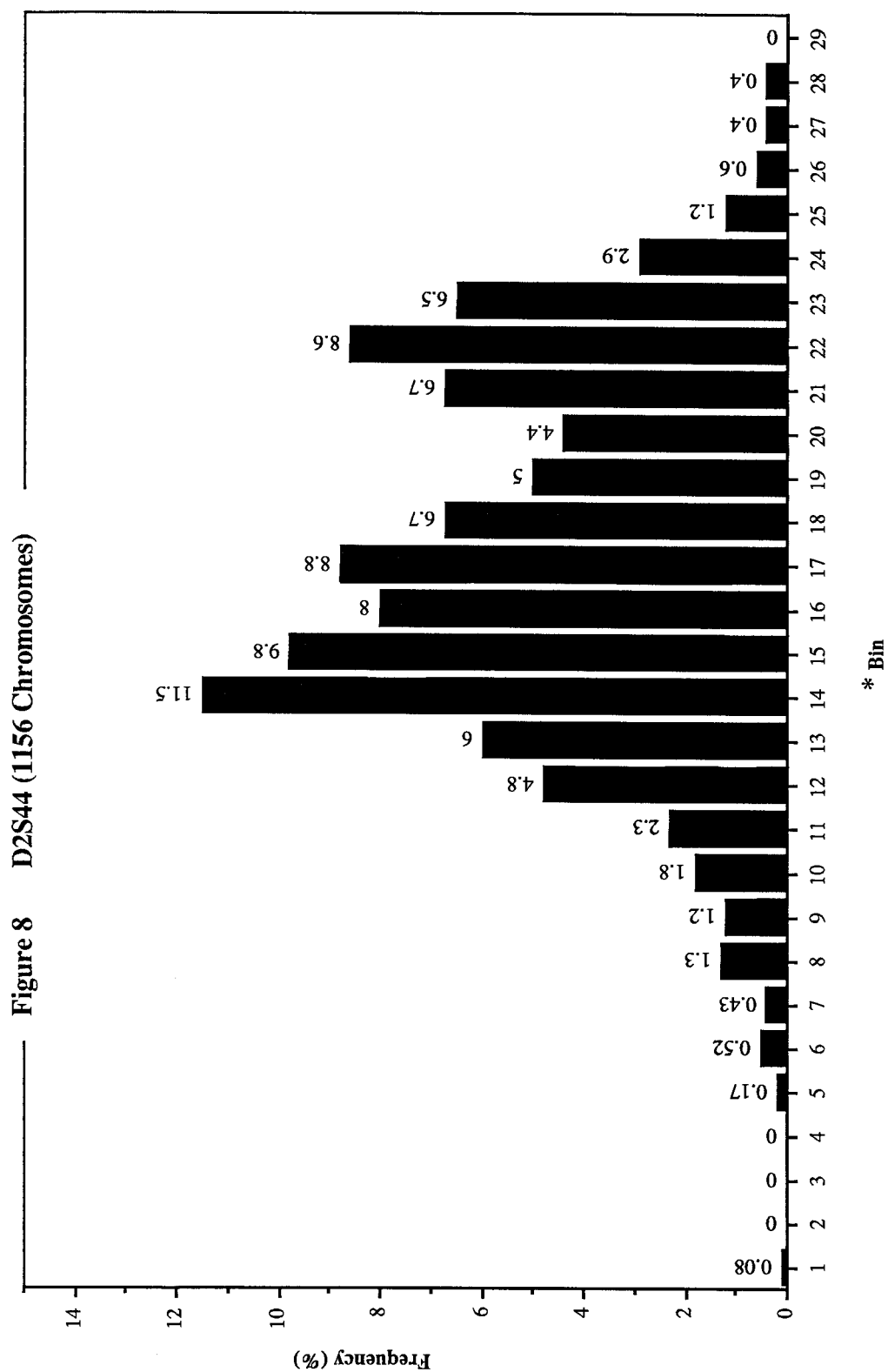
*   Bin number as in Table 5

(Figure 8). The distribution obtain in this study is very similar to those obtained by both the Promega Corporation (collection of data from other laboratories) [105] and the RCMP laboratory [47]. Promega is a Wisconsin based company committed heavily to the promotion and advancement of DNA typing. Its frequency distribution for D2S44 has one mode at approximately 1660 bp (Table 6) and another smaller mode at approximately 3030 bp. Similarly, the frequency distribution obtained by the RCMP showed the first mode at the size range of 1508-1637 bp (Table 6) and the second smaller mode at the size range of 2863-3033 bp.

The slight differences in the mode (most common allelic sizes) of the three distributions are due to the fact that Promega data collection does not use a binning system and the RCMP have different bins from the ones used in this study. Depending on the position of the arbitrary boundaries of the bins, the position of the mode for one distribution may be plus or minus one bin for another distribution. To solve this discrepancy, it may be more appropriate to compare the cumulative frequencies for a specific range of DNA fragment sizes for which the boundaries bisect at the low points of the distributions (Table 7). The frequency values obtained in this study indicate that approximately 50.8% of the D2S44 alleles are within the size range of 1427-2365 bp (bins 13-18) and 21.1% are within the range of 2807-3621 bp (bins 21-23). The Promega frequency values for two similar sizes ranges are approximately 45% and 21.5% and the RCMP values are approximately 48% and 25.2%. In each of the three analyses, these two size ranges account for about 2/3 to 3/4 of all D2S44 alleles. The comparable values of each of the two cumulative frequencies in all three distributions reaffirms the reliability of the sample data to provide an accurate estimate of true (population) allelic frequencies and true allelic sizes.

Population Data Base For the Allelic Frequencies of the 3'HVR of the D16S85 Locus :

Genomic DNA from an undefined group of 608 individuals from the Greater Vancouver Area were analyzed for their allelic types (sizes) at the 3'hypervariable region of the D16S85 gene. This corresponded to the analysis of the alleles of 1216 chromosomes. The various D16S85 alleles

Table 6 : Comparison of the Most Common Alleles

Locus	Lab	Sample Size [†]	Most Common	
			Size Range (bin)	Frequency
D2S44	UBC	578	1536 - 1665 bp (14)	11.5%
	Promega	487	~1660 bp [*]	9% [*]
	RCMP	565	1508 - 1637 bp (9)	13.8%
D16S85	UBC	608	775 - 822 bp (5)	13.9%
	Promega	--	--	--
	RCMP	596	0000 - 1077 bp (1 ➝ 5)	49.1%
5'HVR Insulin	UBC	586	678 - 726 bp (3)	29.4%
	Promega	--	--	--
	RCMP	--	--	--
D14S13	UBC	530	1427 - 1535 bp (13)	15.0%
	Promega	891	~1510 bp [*]	16% [*]
	RCMP	--	--	--

Bracketed numbers are bin numbers

† - UBC Sample Population : undefined
Promega Collection Sample Population : Whites (unknown origin)
RCMP Sample Population : Whites (from Ottawa, Canada)

* - Promega values are approximations only.

Table 7 : D2S44 Cumulative Frequency for a Specific Size Range

	Lab	Size Range (bp)	Cumulative Frequency
1st Mode	UBC	1427 - 2365 (13) (18)	50.0%
	Promega	~1501 - 2138 *	45 % *
	RCMP	1508 - 2351 (9) (13)	48.0%
2 nd Mode	UBC	2807 - 3631 (21) (23)	21.8%
	Promega	~2815 - 3637 *	22 % *
	RCMP	2863 - 3674 (17) (19)	25.2%

Bracketed numbers are bin numbers

* - Promega values are approximations only.

were assigned to bins as described for the D2S44 alleles (Table 5). Comparison of the bin (allelic) frequencies of the first 608 chromosomes with the total frequency values obtained with 1216 chromosomes revealed no significant difference (at 5% significance level) between the two frequency distributions (Figure 9). This suggests that a sample size of even only 304 individuals (608 chromosomes) is large enough to provide an adequate estimate of allelic frequencies, thus confirming the previous results with D2S44.

The frequency distribution of alleles (bins) is skewed to the right with the mode or the most common fragment in the size range of bin 5 (775-822 bp) (Figure 10). Bin 5 in combination with its two surrounding bins, 4 and 6, account for approximately one third of all the alleles of the 3'HVR of the D16S85. The remainder of the alleles are spread between the other bins and except for bin 1, the frequencies are all less than 5.0% for each bin.

Frequency values for bins with rare observations will be combined so as to reduce their significance during statistical analyses. The distribution obtained in this study for alleles of the hypervariable region of D16S85 is similar to those obtained by the RCMP [47]. (Promega does not have published data for this locus). The frequency distribution published by the RCMP laboratory group is also skewed to the right with the most common DNA fragment sizes in the range of 0000-1077 bp (Table 6) and the remainder of the alleles with sizes of up to approximately 7000 bp and occurring at frequencies of < 6.0%. Since the binning systems are different for the two distributions, the cumulative frequencies for a specific size range were compared (Table 8). The RCMP have combined all the alleles ≤ 1077 bp into one bin with a frequency of 49.1%. Similarly the cumulative frequency for the all the bins in this study which account for the DNA fragments sizes of ≤ 1069 bp (bins 1-8) is 50.5%. Bins 1-8 have not been combined in this study because:

1) the number of observed events in each bin is > 16 and is thus

   significant

2) the size of each bin was determined according to a 99% confidence

   interval ($\pm$ 3σ)

Figure 9 : Two Allele Frequency Distributions For the D16S85 Locus

Genomic DNA from an undefined group of 608 individuals from the Greater Vancouver Area

was analysed for alleleic types (sizes) at the hypervariable region of the D2S44 locus.

Allele frequency distributions of sample sizes (608 chromosomes and 1216 chromosomes)

are compared.

Figure 9   D16S85 (1216/608 Chromosomes)
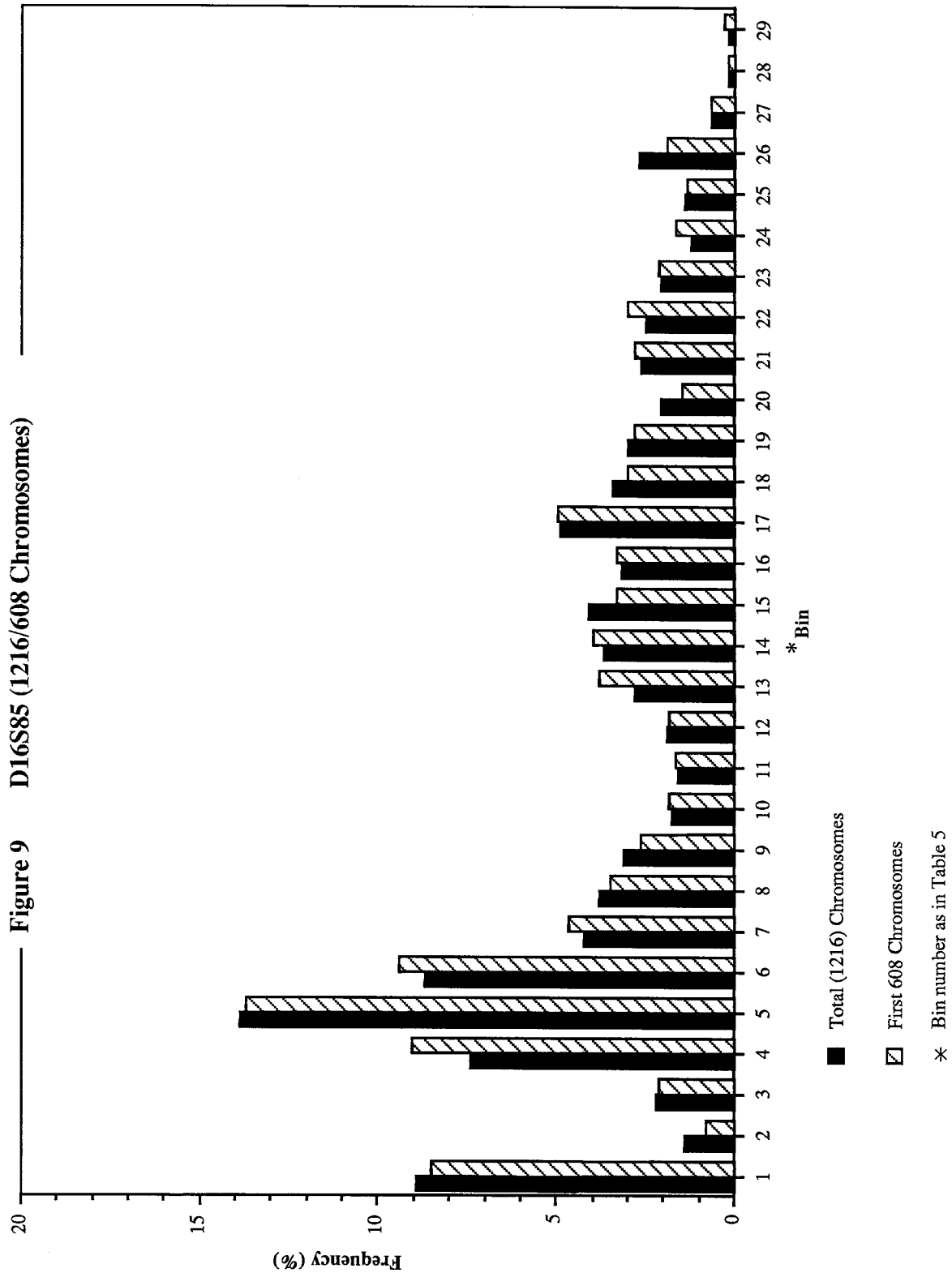
Figure 10 : Allele Frequency Distribution For the D16S85 Locus

Genomic DNA from an undefined group of 608 individuals from the Greater Vancouver Area

was analysed for alleleic types (sizes) at the hypervariable region of the D16S85 locus.

Allele frequency distribution of this locus is shown for 1216 chromosomes.

Figure 10   D16S85 (1216 Chromosomes)

* Bin
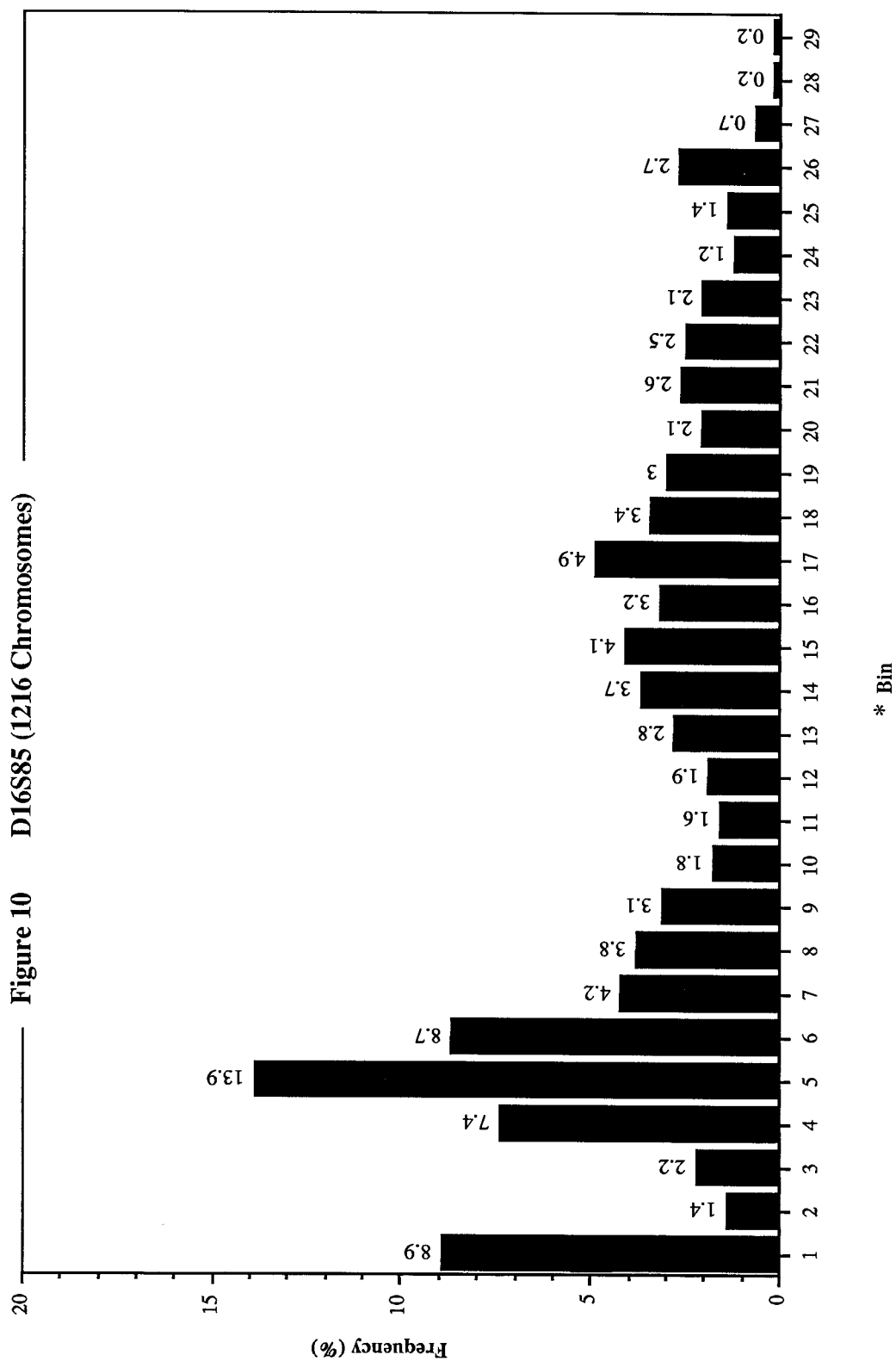
*   Bin number as in Table 5

Table 8 : D16S85 Cumulative Frequency for a Specific Size Range

Lab	Size Range (bp)	Cumulative Frequency
UBC	0000 - 1069 † (1) † (8)	50.5%
RCMP	0000 - 1077 * (1 ⟶ 5)	49.1%

Bracketed numbers are bin numbers

† - bins not combined
* - bins combined

3) the allele frequency for any one bin should be already an over

estimation of any one fragment size (allele).

Therefore, the grouping of all the alleles < 1069 bp would be an overly conservative estimate.

However, it would not be unreasonable to reduce the number of bins in this size region (as seen

with the FBI binning system) by merging two or more bins. Alternatively, for statistical calculations

of the individuality of a DNA profile, the bin frequencies of alleles three standard deviations (for a

99% confidence interval) from the measured allele size of interest may be combined to ensure an

over estimation of the true frequency.

For example: (refer to Table 5 for bin frequencies)

Frequency of D16S85 allele with size 2120 bp = X

standard deviation for this size range = σ = 36.7 bp

X + 3σ = 2120 + 3(36.7) = 2230 bp

X - 3σ = 2120 - 3(36.7) = 2010 bp

Frequency of 2120 bp = Frequency of alleles 2010-2230 bp

= Frequency of bin 17 + bin 18 = 8.3%

This approach combines the "fixed-bin" system used for the data base generation of this thesis and

the "floating-bin" system used by Lifecodes Incorporated. The "fixed-bin" system provides

conservative allele (bin) frequencies by assigning alleles of a specific size range to a particular bin

whose width is determined by the standard deviation of the alleles of that size category [104,106].

Lifecodes' "floating-bin" system involves measurement of alleles at 100 bp increments and the allele

frequency used in calculations were obtained by adding the values contained within two standard

deviations of the mean size determination [87]. (A disavantage of the "floating-bin" system is that

an assigned allele size of an evidence sample is assumed to be the mean value and such an

assumption may lead to a lower frequency estimate for the allele [104]. No further comment can

be made regarding this system due to the absence of published details.) The combination of these

two system makes the frequency estimates much more conservative than that obtained by each

method alone. This conservative method of calculation would also safeguard against concerns [107] that an allele will be assessed a lower frequency than its true value if the arbitrarily assigned boundaries of a bin bisects the peak of a distribution.

Population Data Base For the Allelic Frequencies of the 5'HVR of the Insulin

Gene :

Genomic DNA from 586 undefined individuals from the Greater Vancouver Area were analyzed for their allelic types (sizes) at the 5'hypervariable region of the insulin gene. This corresponded to the analysis of the alleles of 1172 chromosomes. Various insulin alleles were assigned to bins as described for the D2S44 alleles (Table 5). Comparison of the bin (allelic) frequencies of the first 586 chromosomes with the total frequency values obtained with 1172 chromosomes revealed no significant difference (at 5% significance level) between the two frequency distributions (Figure 11). This suggests that a sample size of only 293 individuals (586 chromosomes) is large enough to provide an adequate estimate of the allelic frequencies, again confirming previous results.

The frequencies of the alleles show a bimodal distribution with the first mode at bin 35 (678-726 bp) and the second (smaller) mode at bin 18 (2146-2365 bp) (Figure 12). As before, frequency values for bins with rare observations will be combined so as to reduce their significance during statistical analysis. Numerous bins with extremely low frequencies separating the two modes is clearly indicative of two very distinct size ranges of DNA. The small alleles (bin 1-5) account for approximately 75% of the total alleles and the larger alleles (bins 17-19) account for nearly 24%. Alleles with intermediate sizes (bins 6-16) are rarely observed. There is an absence of published data for the allelic distribution of the 5' hypervariable region of the insulin gene cleaved with restriction endonuclease Hae III. However, Bell et al. have collected data at this locus by cleaving genomic DNA with restriction endonuclease Bgl I [42]. Their allelic frequencies exhibited a comparable bimodal distribution with the smaller DNA fragments (approximately 2.7 - 3.3 kb) accounting for 75% of alleles and the remaining 25% made up of the larger alleles (approximately

Figure 11 : Two Allele Frequency Distributions For the INS Locus

---

Genomic DNA from an undefined group of 586 individuals from the Greater Vancouver Area

was analysed for alleleic types (sizes) at the hypervariable region of the INS locus.

Allele frequency distributions of sample sizes (586 chromosomes and 1172 chromosomes)

are compared.

Figure 11   5' HVR Insulin (1172/586 Chromosomes)

■ Total (1172 Chromosomes)

◪ First 586 Chromosomes

✳ Bin number as in Table 5
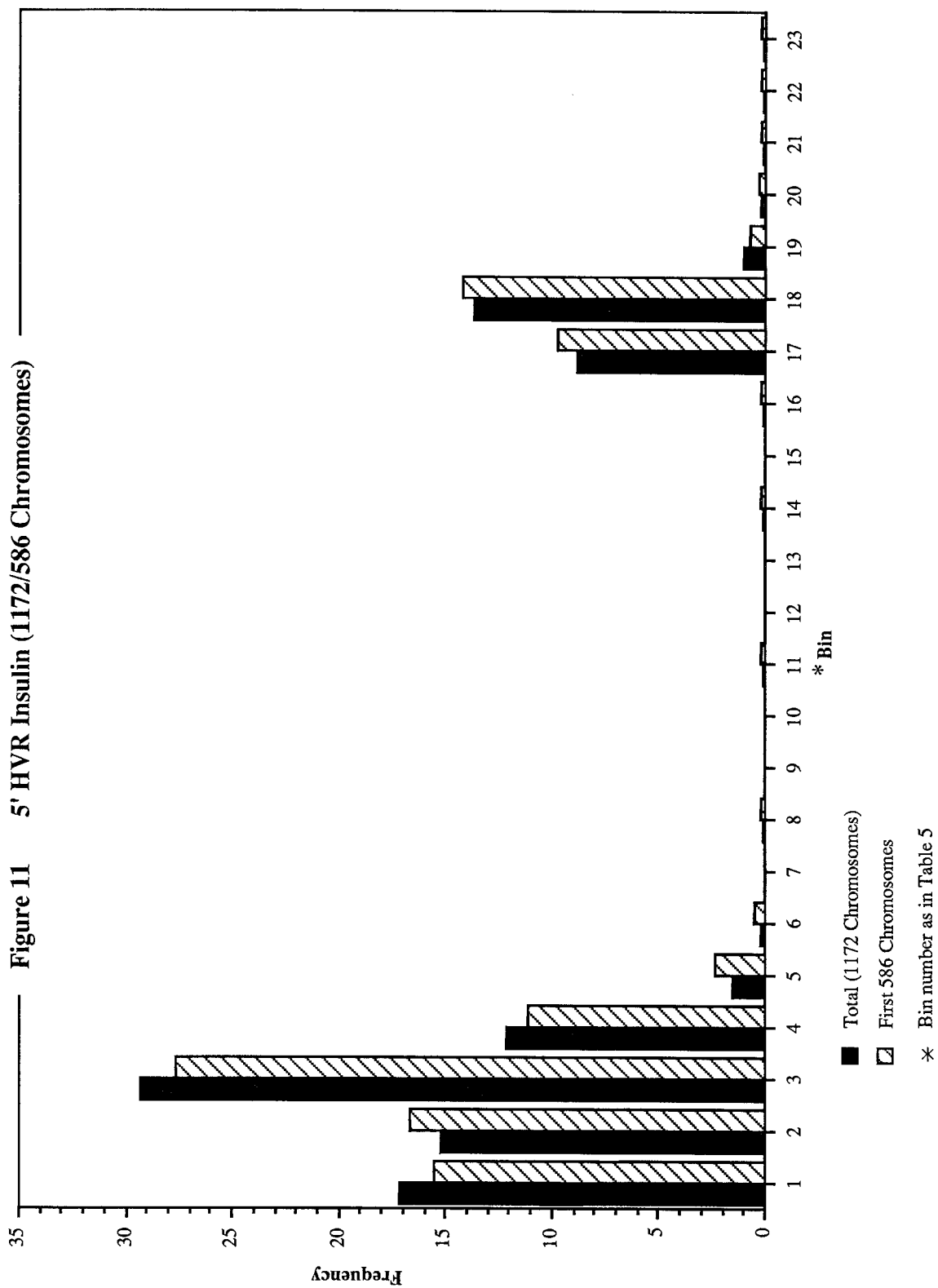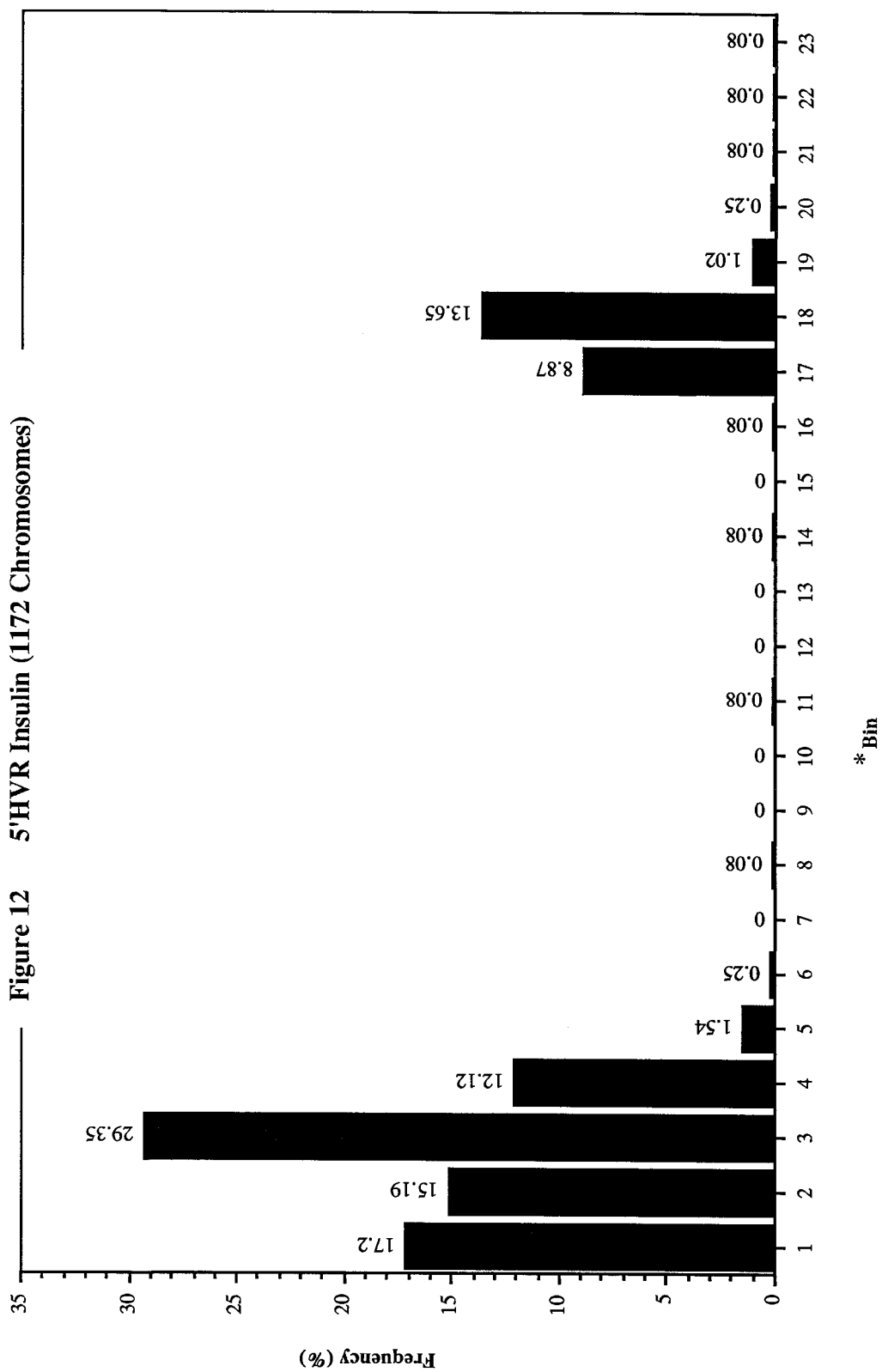
Figure 12 : Allele Frequency Distribution For the INS Locus

Genomic DNA from an undefined group of 586 individuals from the Greater Vancouver Area

was analysed for alleleic types (sizes) at the hypervariable region of the INS locus.

Allele frequency distribution of this locus is shown for 1172 chromosomes.

Figure 12   5'HVR Insulin (1172 Chromosomes)

* Bin number as in Table 5

4.3 - 4.9 kb) [43]. The DNA fragment sizes (alleles) from this laboratory and those of Bell et al., cannot be directly compared since different restriction endonucleases were used. Restriction fragment sizes are not only determined by the VNTR size but also the DNA length contributed by flanking regions which are dependent upon the particular restriction endonuclease used for the genomic DNA digest.

Contrasting results have been shown for a black American population sample using restriction endonuclease Sac I [108]. The allelic frequency distribution obtained with the black American data is unimodal compared with the bimodal distributions demonstrated for the two studies discussed above. This suggests that there are subpopulation differences at the 5'HVR of the insulin gene and that the undefined populations sampled by this author and that of Bell et al. are either similar or consist of a mixture of subpopulations that do not have significantly different allele frequencies.

Population Data Base For the Allelic Frequencies of the HVR of the D14S13

Locus :

Genomic DNA from 530 undefined individuals from the Greater Vancouver Area were analyzed for their allelic types (sizes) at the hypervariable region of the D14S13 locus. This corresponded to the analysis of the alleles of 1060 chromosomes. The various D14S13 alleles were assigned to bins as described for the D2S44 alleles (Table 5). Comparison of bin (allelic) frequencies of the first 530 chromosomes with the total frequency values obtained with 1060 chromosomes revealed no significant difference (at 5% significance level) between the two distributions (Figure 13).

The frequency distribution of the alleles (bins) is skewed to the right with the mode or most common DNA fragments in the size range of bin 13 (1427-1535 bp) (Figure 14). Frequency values for bins with rare observations will be combined so as to reduce their significance during statistical analysis. The distribution obtained in this study for the alleles of D14S13 is similar to that obtained by the Promega Corporation collection [109]. The RCMP laboratory group has not

Figure 13 : Two Allele Frequency Distributions For the D14S13 Locus

---

Genomic DNA from an undefined group of 530 individuals from the Greater Vancouver Area

was analysed for alleleic types (sizes) at the hypervariable region of the D14S13 locus.

Allele frequency distributions of sample sizes (530 chromosomes and 1060 chromosomes)

are compared.

Figure 13   D14S13 (1060/530 Chromosomes)

Frequency

* Bin

Total (1060 Chromosomes)

First 530 Chromosomes

* Bin number as in Table 5

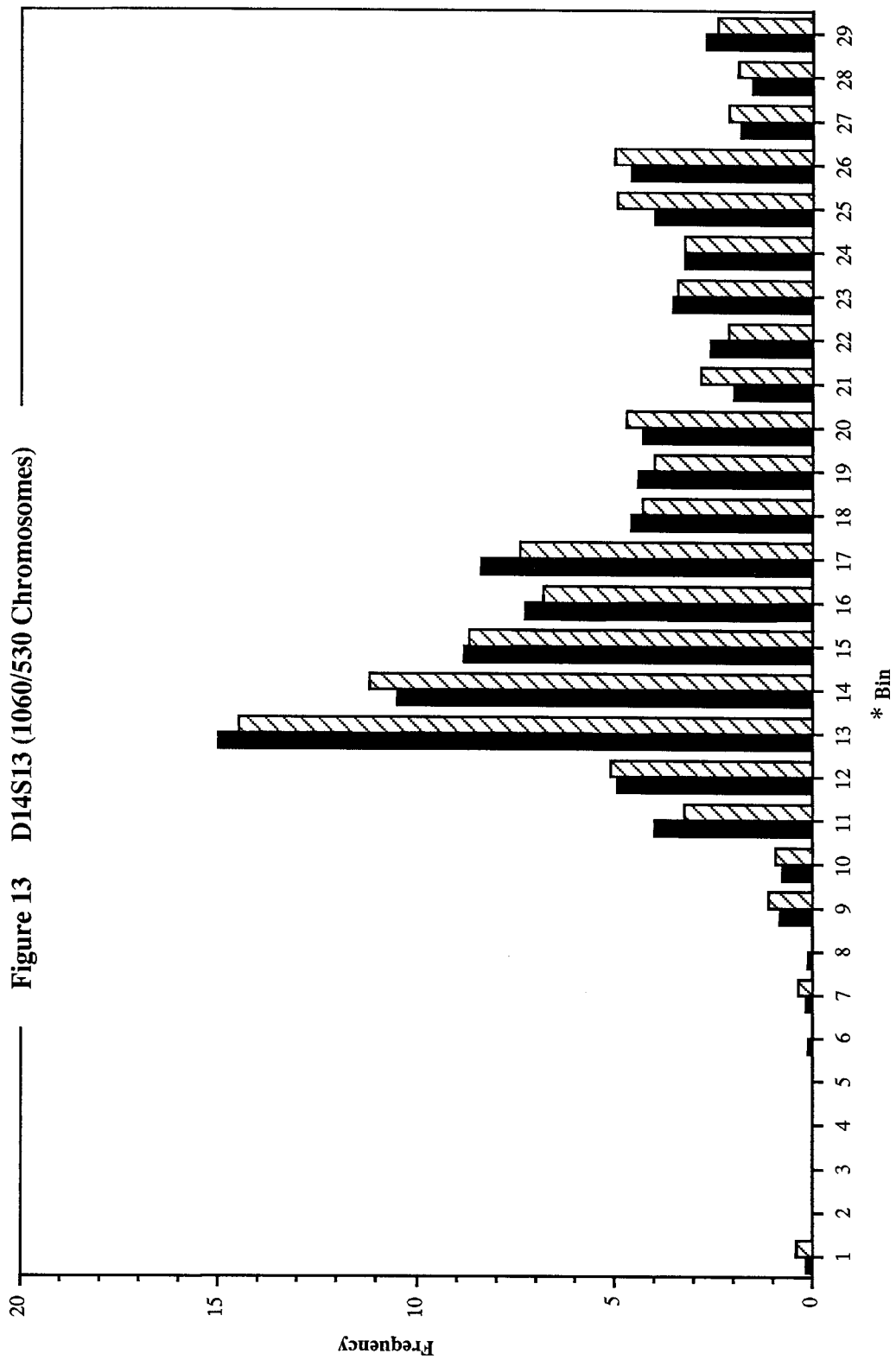Figure 14 : Alelle Frequency Distribution For the D14S13 Locus

Genomic DNA from an undefined group of 530 individuals from the Greater Vancouver Area

was analysed for alleleic types (sizes) at the hypervariable region of the D14S13 locus.

Allele frequency distribution of this locus is shown for 1060 chromosomes.
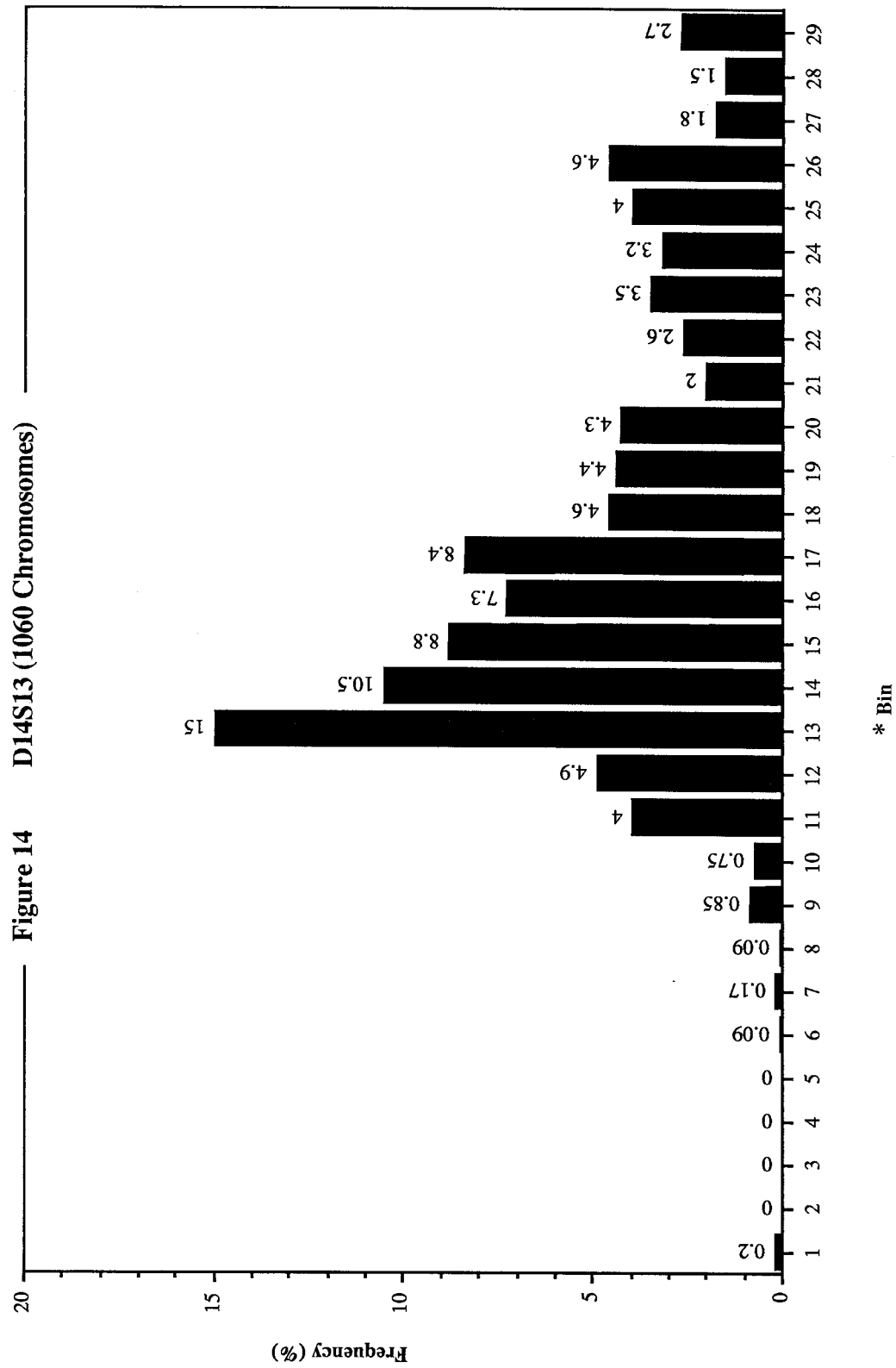
Figure 14   D14S13 (1060 Chromosomes)

* Bin

* Bin number as in Table 5

82

published data for this locus. The frequency distribution published by the Promega Corporation is also skewed to the right with the mode at approximately 1510 bp (Table 6). For comparison purposes, cumulative frequencies for a specific size range were calculated for two distributions (Table 9). The frequency values obtained in this study indicate that approximately 50% of the D14S13 alleles are within the size range of 1427-2145 bp (bin 13-17) and the remainder of the alleles occurring at frequencies of < 5%. For a size range of approximately 1374-2055 bp, a similar cumulative frequency value of 55% is seen in the Promega distribution with the remainder of the alleles occurring at frequency values of < 5%. The agreement between these two sets of data reinforces the reliability of the sample data to provide an accurate estimate of the true frequencies and true allelic sizes.

Comparison of the D2S44 and D14S13 distributions indicate that the first mode of D2S44 almost coincides with the mode of D14S13 (Figure 15). This would present an analytical problem if the two probes for these loci were used to detect simultaneously the alleles of the two loci. There would be a high probability (since the common alleles overlap) that the alleles of one locus would be of a very similar or identical size to the alleles of the other locus. Since overlapping bands cannot be detected, an individual heterozygous at both loci may be falsely categorized as being homozygous for one or both loci. Hence simultaneous probings of different loci should not involve loci with closely overlapping common alleles.

Simultaneous Detection of Two Single Locus VNTR Regions :

A combination of several single locus VNTR probes (to detect multiple regions) are usually required to create a discriminating DNA profile. Sequential probing of membrane-fixed DNA may be time consuming depending on the amount of fixed DNA and the efficiency of the labelled probe. It may also result in lower quality autoradiographs in the final probings. An alternative is to probe simultaneously for more than one locus at a time as seen in Figure 16. Two alleles each were detected for five individuals tested when one probe was used per hybridization reaction (Figure 16 - Panels A and B). When two probes were used per hybridization reaction, four alleles

Table 9 : D14S13 Cumulative Frequency for a Specific Size Range

Lab	Size Range (bp)	Cumulative Frequency
UBC	1427 - 2145 (13) (17)	50.0%
Promega	~1374 - 2055 *	55 % *

Bracketed numbers are bin numbers

* - Promega values are approximations only.

Figure 15 : Comparison of Allele Frequency Distributions of D2S44 and D14S13

Allele frequency distribution of D2S44 (1156 chromosomes) and D14S13 (1060 chromosomes)

are compared to determined the areas of overlapping alleles.

Figure 15   Distribution Comparison of D2S44 and D14S13
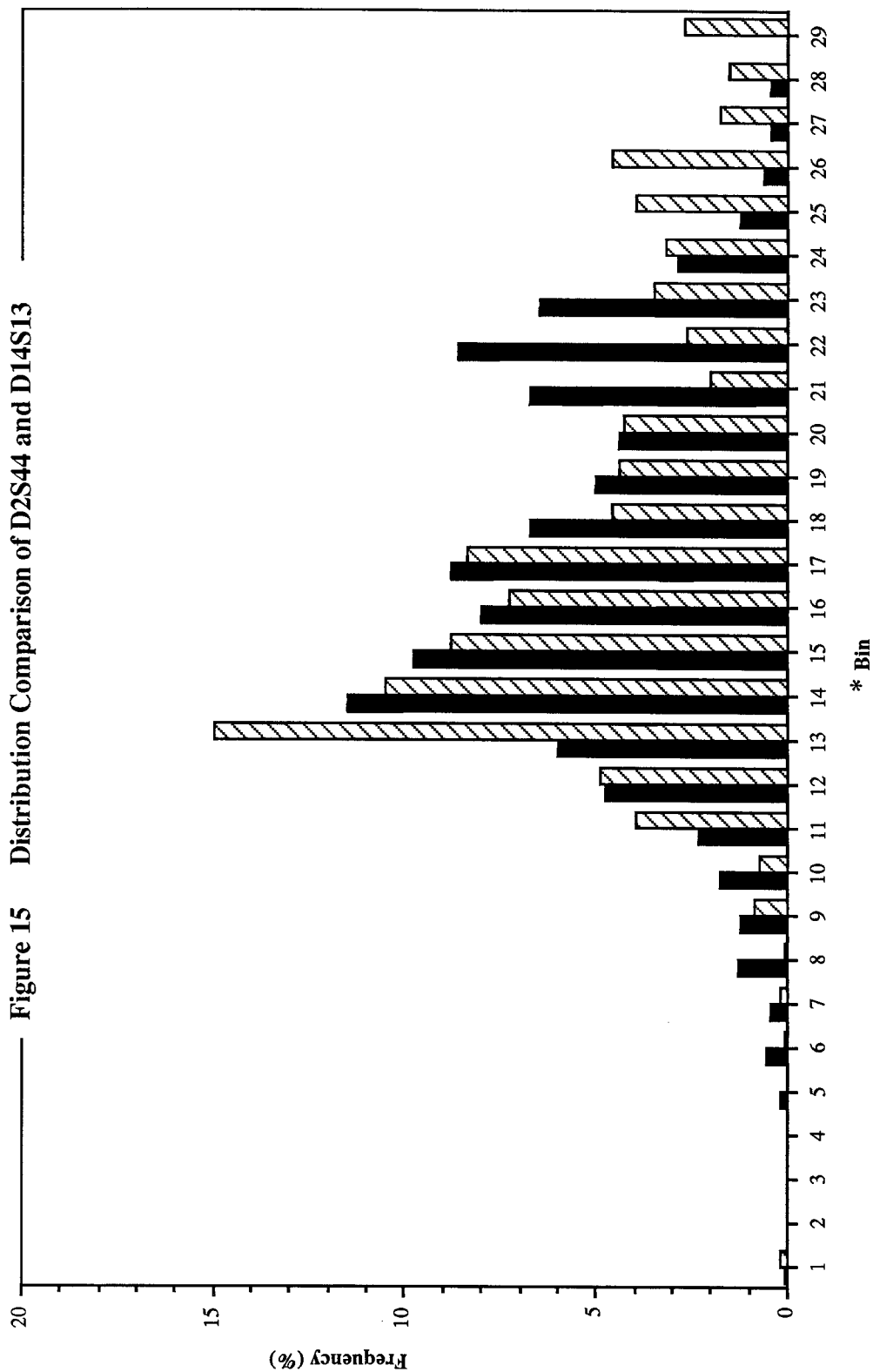
Figure 16 : Simultaneous Probing of Two Loci

Genomic DNA from 5 individuals (lanes 1-5) were probed with VNTR probes 3'HVR

α-globin(Panel A), YNH24 (Panel B) and combination of 3'HVR α-globin/YNH24 (Panel C).

[3 days exposure to Kodak X-MAT AR film at -70°C with screens]

Figure 16 : Simultaneous Probing of Two Loci
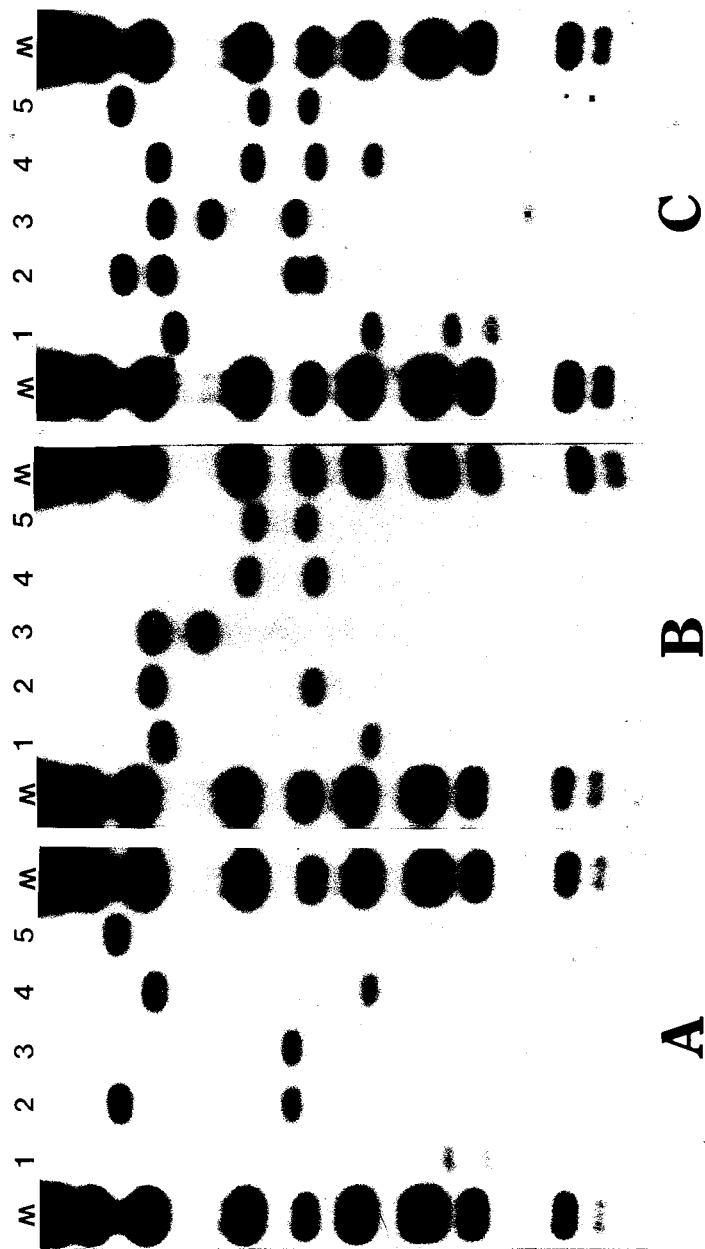
were detected for each individual, corresponding to two alleles per locus (Figure 16 - Panel C). The presence of four alleles for each individual (Panel C: lanes 1-5) eliminates the possibility of an allele(s) of one locus covering the allele(s) of the other locus; a result that would lead to an incorrect conclusions.

The combination of probes used were the 3'HVR/α-globin and YNH24 which detected the D16S85 and D2S44 hypervariable regions respectively. This combination was chosen because the size range of the most common alleles of D16S85 was distinct from that of D2S44 (Figure 17). Although these two loci do have some overlapping alleles, these alleles are less common. Thus there is a lower probability of an individual having rare alleles of the same size for both loci. In situations where < 4 alleles are detected for simultaneous hybridizations with two single-locus probes, the analysis would have to be repeated with sequential probings with one probe per hybridization reaction. Thus, more time and labour would be required than if the analysis were performed initially with sequential probings. The possibility of overlapping alleles increases in proportion to the number of regions detected simultaneously. Hence, the possible benefit in time saving of simultaneous probings (two or more) may not be great enough for multiple probings to serve as a viable alternative for sequential probings.

DNA Profiles :

Genetic typing of 470 individuals did not show any identical bin profiles at all four of the VNTR loci, D2S44, D16S85, INS and D14S13. Two individuals did exhibit the same bin pattern at 3/4 of the loci, but the bin numbers at the fourth locus were far apart. Identical bin profiles refers to alleles that fall into the same bin which in turn represents a range of allele sizes. Hence, even if bin numbers (alleles) at all four loci were identical, the two profiles may not constitute a real "match" of alleles. In order for a match to be declared, the two identical bin samples must be reanalysed on the same gel to provided a more accurate measurement. This concept will be discussed further in the following paragraphs regarding measurement imprecision.

Figure 17 : Comparison of Allele Frequency Distributions of D2S44 and D16S85

Allele frequency distribution of D2S44 (1156 chromosomes) and D16S85 (1216 chromosomes) are compared to determined the areas of overlapping alleles.

Figure 17   Distribution Comparison of D2S44 and D16S85
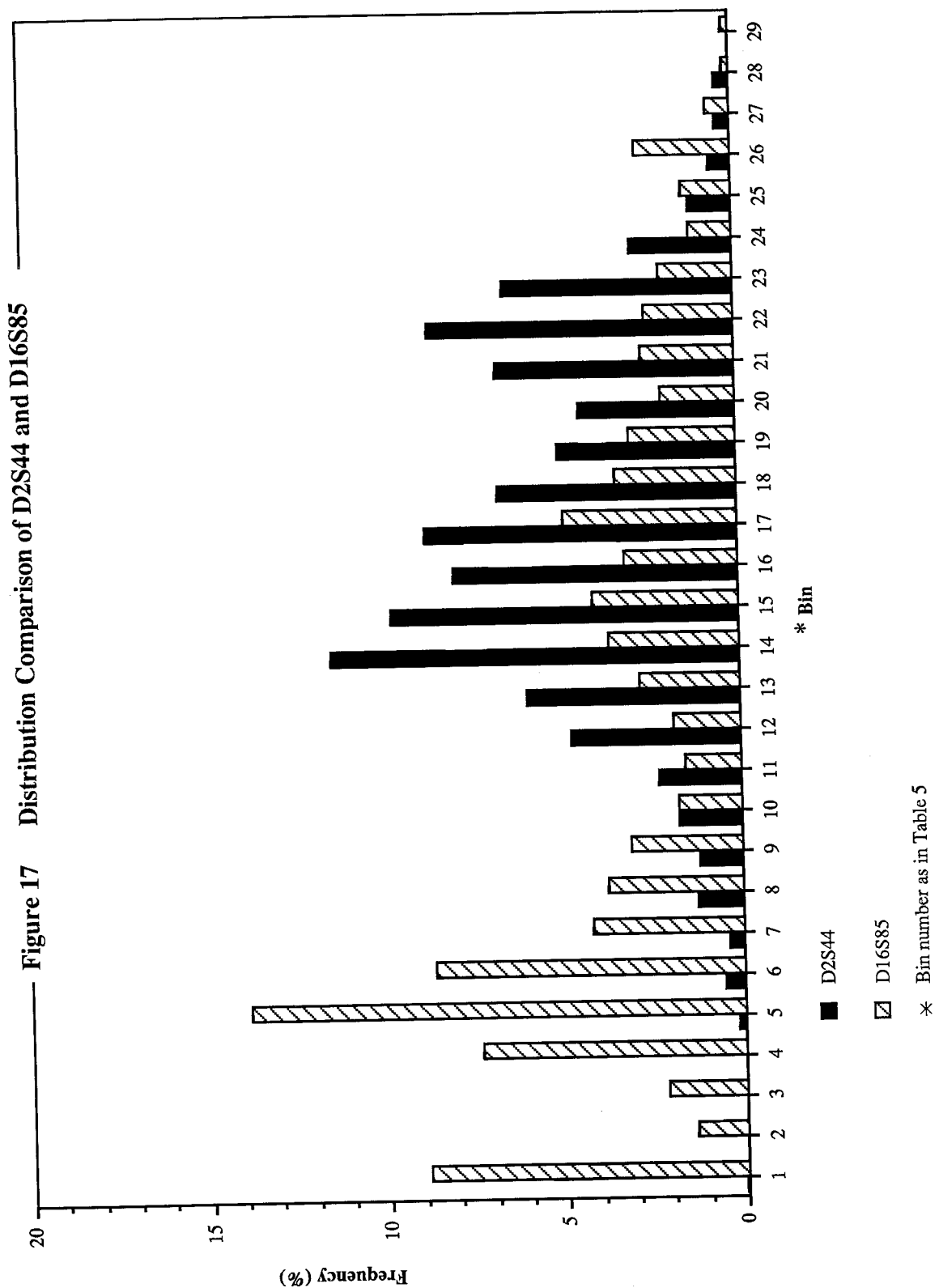
Measurement Imprecision :

The ultimate goal of the RFLP/VNTR DNA typing procedure is to determine whether two DNA banding profiles are identical and hence may be attributed to one individual. Two alleles, from two different lanes in the same gel, which fall into the same bin do not necessarily indicate a positive "match" for those DNA fragments. Conversely, two alleles with visually matching mobilities may fall into different bins depending on the arbitrary boundaries of the bins. The bin system only serves to provide conservative frequency estimates for alleles of the VNTR loci. These frequency values are used to calculate the uniqueness of a DNA profile once a "match" has been declared.

An objective basis for declaring a "match" between two samples is possible by incorporating an experimentally determined measurement imprecision value [47]. The degree of measurement imprecision is defined as the maximum size difference between identical DNA samples. The sizes of two DNA bands, from two different samples, must be within the window of measurement imprecision of each other before they are assessed as a "match". The degree of measurement imprecision for various sized genomic DNA fragments within or between gels was determined (Table 10). The degree of measurement imprecision within a gel appears to be relatively constant, ranging between 3.7 - 5.5%. However, there does not appear to be any correlation between measurement imprecision and fragment length especially in the measurements within a gel. Since actual casework usually involves comparison of samples within the same gel, it would be appropriate to use a value of 5.5% as the "window" for measurement imprecision for "match" determinations.

One can argue that if the "window" for measurement imprecision was too large, DNA fragments which are of different sizes will be falsely declared as "matched". As a consequence of the resolution of agarose gel electrophoresis and minute differences possible between VNTR alleles, false "matches" can occur even in the absence of measurement errors. The presence of "windows" is to protect against false negatives (non matches) resulting from measurement errors.

Table 10 : Measurement Imprecision

*Genomic DNA (kb)	Measurement Imprecision (within gel)	Measurement Imprecision (between gels)
5.376	4.9 %	9.0 %
4.333	4.1 %	9.8 %
3.961	5.5 %	7.1 %
3.159	4.8 %	6.5 %
2.610	4.4 %	6.0 %
1.560	4.2 %	6.6 %
1.490	5.5%	7.3 %
1.270	4.9 %	6.3 %
0.830	3.7 %	4.8 %
0.660	4.2 %	6.0 %

n=90

* = Mean determinations

$$\text{Measurement imprecision} = \frac{\text{Max - Min}}{\text{Mean}} \times 100$$

The occurrence of these false positive "matches" is equalized by the binning system of VNTR alleles. Bin sizes are all greater than the maximum measurement imprecision within gels (5.5%). Hence, as previously discussed, during data base generation there is a tremendous over estimation of the frequency of any one allele. Therefore, in the event of a false "match" with one VNTR, the frequency of the allele (bin) will be higher than its actual frequency which reduces the uniqueness of that allele and the corresponding value of the "match". Since a "matched" DNA profile is generated by analyzing several VNTR loci, the unlikely occurrence of multiple false "matches" for any two samples will reduce the likelihood of an inaccurate interpretation of the typing results [110]. Furthermore, a quantative match criterion is used only after an initial visual match has been made.

Imprecision within a gel for the 2.610 kb fragment (4.4%) in this study is higher than the value from the RCMP laboratory for the 2.724 kb fragment (approximately 2.9%). This difference may be because the RCMP laboratory value is based on the "maximum size difference encountered when two identical samples are compared relative to the same set of flanking size markers" [47] and more than one set of markers are present in one gel. In contrast, the values calculated from the data of this study included all measurements across the entire gel thus, accounting for more errors. The comparability of the imprecision values calculated from data from both laboratories is confirmed by the values obtained from measurements between different gels which are 6.0% (2.610 kb) for this laboratory and 5.5% (2.724 kb) for the RCMP laboratory [47].

Validity of Subjective (Manual) DNA Band Assessments :

One of the major concerns of any laboratory is the validity of generated data and the reliability of analyses. The allelic frequencies for the four VNTR loci analyzed in this laboratory were vulnerable to errors due to subjective DNA band measurements by a single individual. However, analyses and comparison of data with other laboratories which utilize automated (objective) assessments, confirm the validity and reliability of this system. This confirmation is based on the fact that :

1) -the frequency distribution (Figure 18) of 364 alleles

   objectively analyzed with the FBI's image analysing system was not

   significantly different (95% level) from the distribution of the same

   alleles measured manually (Figure 18)

   -the discrepancies ranged from 0.0% in bin 5 to 3.1% in bin 15

2) -frequency distributions of the alleles of the four VNTR loci are

   comparable to those of the Bell et al., RCMP laboratory and/or

   Promega Corporation

3) -a high degree of precision (or small relative standard deviation)

   was obtained

4) -the level of measurement imprecision for the 2.6 kb size range was

   comparable to that obtained by the RCMP laboratory

The validity of subjective DNA band measurements is corroborated by a European study in

which a comparison was made between fragment sizes calculated manually with a plotted curve or

calculated using a computer program and digitizing tablet [111]. Their analysis of 72 DNA

(fragments) bands indicated that there was no statistically significant differences between the two

alternative methods of DNA sizing.

Figure 18 : Comparison of Data From Computerized Versus
Non-computerized Measurements

Data for the D16S85 locus (364 chromosomes) was collected by using manual band

assements or the FBI laboratory's image analysis system in Quantico, Virginia. Allele

frequency distributions of the two sets of data are compared to determined whether

there is any significant difference between data sets.

Figure 18   Data From Non-Computerized vs Computerized Measurements
(364 Chromosomes)

* Bin

■ Non-Computerized Measurements

▨ Computerized Measurements

* Bin number as in Table 5

97
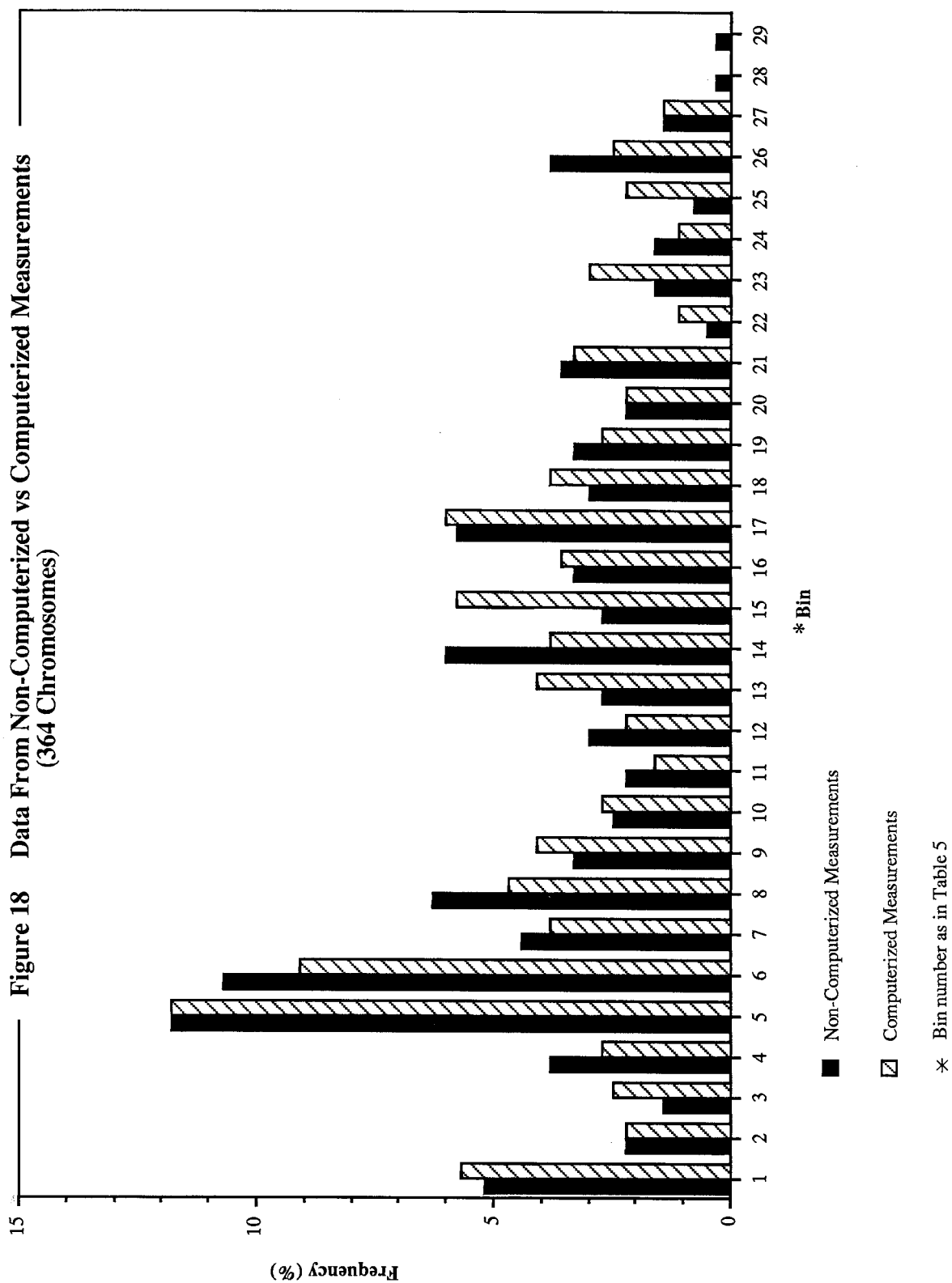
CONCLUSION

During the last four years, there has been a tremendous amount of work devoted to the research and development of DNA genetic typing for human identification. New technical refinements have appeared at a very rapid rate from government, corporate and university laboratories. Accordingly, various protocols in this laboratory have been refined or replaced with procedures developed either here or in other facilities.

The accuracy (amount of deviation from the actual size) of the typing results obtained could benefit from some improvement. However, the maximum level of inaccuracy (approximately 5.0 % of molecular size) was less than the window (greater than 13 % of molecular size) allotted for the bin of that size range. This would allow for a large enough "wobble" such that the assignment of DNA fragments to bins would not be drastically altered to cause appreciable changes in allelic frequencies.

The level of precision of this DNA typing system as measured by the relative standard deviation, is much higher than the level of accuracy. Although both aspects of the system are important, precision is more critical since the reliability of the results is dependent upon obtaining the same results in repetitive trials. The degree of accuracy is relative from one sample to the next and thus with a binning system, the allelic frequency distribution will not be altered. Bin sizes for allele assignments were determined according to the degree of precision of the typing system so that DNA fragments of the same size can be assigned to the same bin in 99.7% of the cases. The level of precision of this laboratory (maximum relative deviation of ± 2.3% - Table 4) is comparable with those of other forensic laboratories who quoted a value of ± 2.5% (or 5%) at the 1991 Promega International Symposium on Human Identification [A. Autor - personnel communication].

Measurement imprecision, defined by the maximum size difference between identical samples, serves as the basis for declaring a match of alleles from different samples. The maximum degree of measurement imprecision (within a gel) of this typing system for DNA fragments

between 5.3 - 0.6 kb is approximately 5.5% of the fragment size. Thus, fragment comparisons varying up to 5.5% can still be considered as "matched" (inclusion). In most instances, the values for measurement imprecision within the same gel will be used because casework material will more often than not be on one gel.

Allelic frequency distributions generated for four VNTR loci are comparable to published data from major laboratories such as that of the RCMP, and Promega Corporation associated laboratories. The population sampled by the RCMP and Promega associated laboratories are Caucasians, whereas this study analyzed an undefined population. Detailed comparisons of cumulative frequencies of specific size ranges of alleles at various loci obtained by the three groups have shown striking similarities with a estimated maximum difference of 5% at the D14S13 locus (Table 10). A more general representation of the similarities can be seen in Figures 19-21. The comparability of these different data bases may suggest that the :

1) undefined population samples consists of mainly Caucasians, and/or

2) undefined population samples consists of certain racial or ethnic

   subpopulations which have the same or very similar frequency values

   as those of Caucasians

3) frequency values are homogenous for all subpopulations

The first two alternatives are indistinguishable here because the ethnic or racial origin of individuals in the undefined populations unknown. However, the first suggestion is a reasonable possibility since the Vancouver population is composed of 75.6 % whites (non-Hispanic), 22.1% Asian, 0.3% Black, 0.5% Hispanic, and 1.5% Native North American [112]. The latter suggestion is probably invalid since analysis of three racial subpopulations (American Blacks, Caucasians, and Hispanics) have shown subpopulation heterogeneity for at least five loci (HRAS-1, D14S1, D2S44, D14S13 and 5'HVR insulin) [87,88,108]. Analysis of the D2S44 locus by the RCMP has revealed heterogeneity in the Caucasian and Native Indian Population [113]. Furthermore, there has been a report of slight differences in allelic frequencies of the same racial group (Blacks) in different

Fig. 19: D2S44 Allelic Distribution Comparisons

Size values represent midpoints of consecutive size ranges
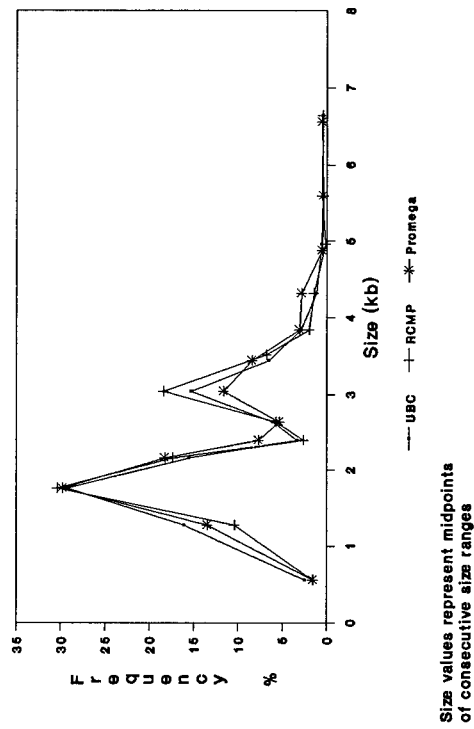
Fig. 19a: UBC D2S44 Allelic Distribution

Size values represent midpoints of consecutive size ranges

Fig. 19b: RCMP D2S44 Allelic Distribution

Size values represent midpoints of consecutive size ranges
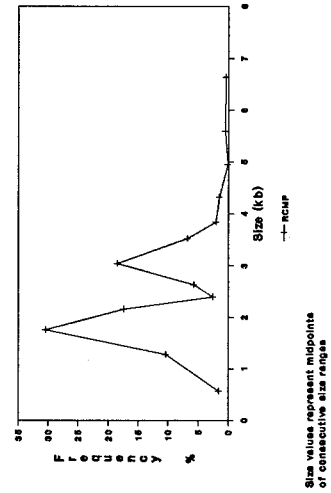
Fig. 19c: Promega D2S44 Allelic Distribution

Size values represent midpoints of consecutive size ranges
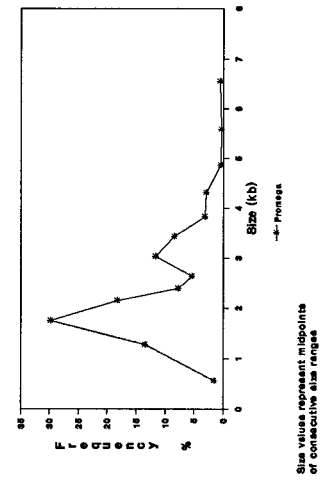
Fig. 20a: UBC D16S85 Allelic
Distribution

Size value represent midpoints
of consecutive size ranges
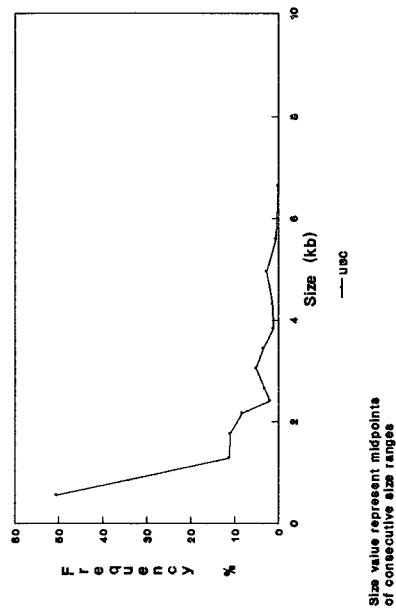
Fig. 20b: RCMP D16S85 Allelic
Distribution

Size value represent midpoints
of consecutive size ranges

Fig. 20: D16S85 Allelic Distribution
Comparison
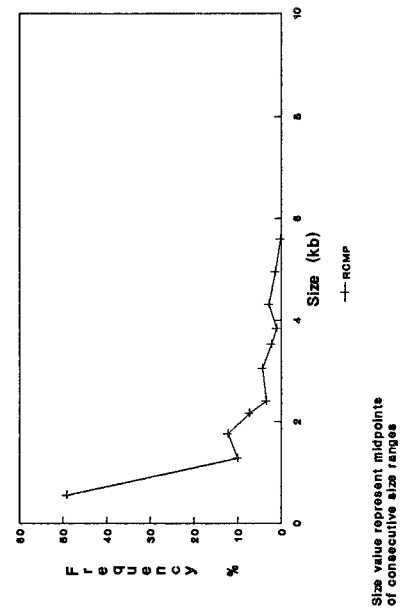
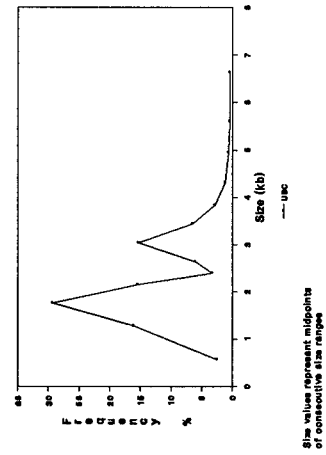Size value represent midpoints
of consecutive size ranges
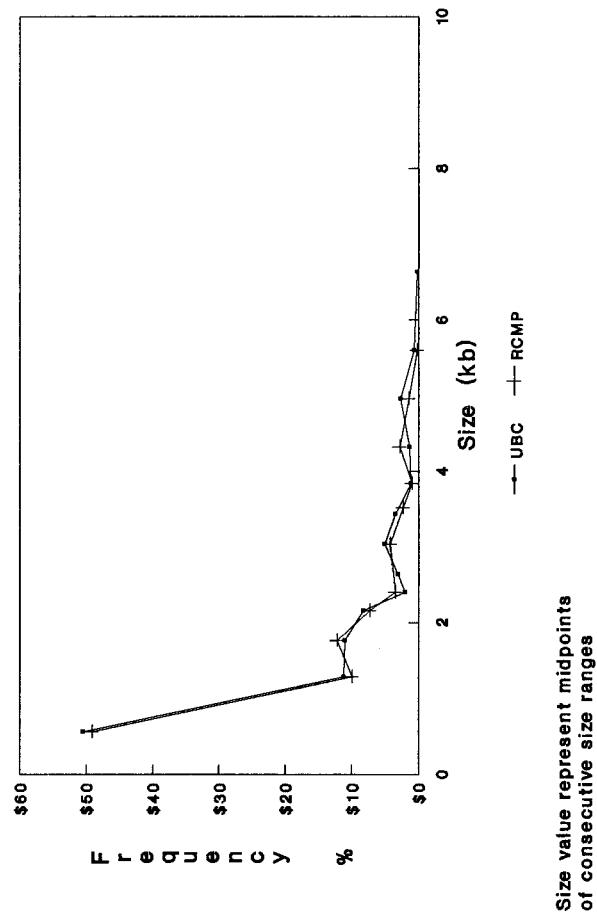
Fig. 21a: UBC D14S13 Allelic Distribution
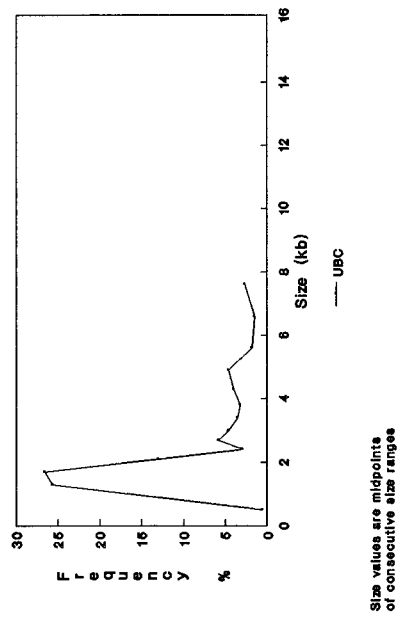
Fig. 21b: Promega D14S13 Allelic Distribution

Fig 21: D14S13 Allelic Distribution Comparisons

Size values are midpoints of consecutive size ranges
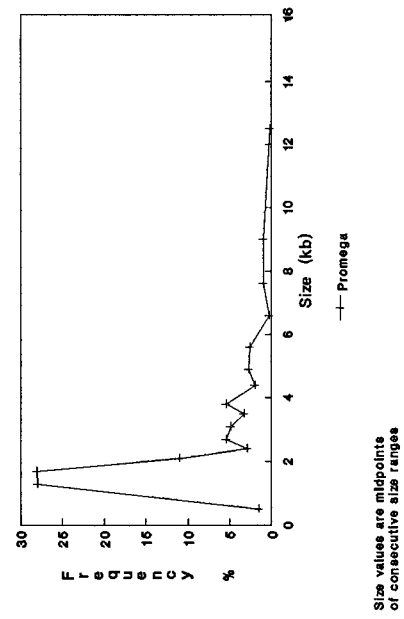
geographical locations across the U.S.A. [114]. This phenomenon of frequency differences in the same racial group was not observed in a comparison of Caucasian populations from Dusseldorf and Amsterdam [111].

The results obtained in this study suggest that this data collected by non-conventional sampling procedures such as sampling from an undefined population from the Greater Vancouver Area, would be applicable to other subpopulations where there is a majority of Caucasians and no significant restriction on gene flow due to barriers such as geographical isolation or religious beliefs. This would be true for loci which do not have large variations in allele frequency between racial subpopulations. Slight variations in allelic frequencies of the other racial subgroups present, would not significantly alter the overall frequencies because the contribution made by any one subgroup would be small. Furthermore, allelic frequencies generated in this thesis for the undefined population is based on a fixed bin system which minimizes potential subpopulation (racial or geographical) differences [110]. The application of this system to data from the Technical Working Group on DNA Analysis Methods participants, suggests that there are no dramatic differences in allelic (bin) frequencies for Caucasian (or black) geographical subpopulations across the U.S.A. and Canada [110].

Allelic frequency data bases generated for one geographical population would not always be applicable to another population due to differences in types and proportions of racial subpopulations. However, each geographically distinguishable population could use its own data base, or data collected from the racial subpopulation which account for a large majority of the specific test population in question, or data collected from a geographical population to which its racial subgroup makeup bears close resemblance. The possibility of utilizing databases collected from geographically distinct areas would eliminate the neccessity of building separate databases for each and every region. Another alternative would be to incorporate the highest value obtained for the specific allele from any one data base (geographical or racial subpopulation) in the calculations

for uniqueness of DNA profile. This conservative approach (used by the RCMP) will safeguard against biasing against individuals of any one subpopulation leading to false interpretations.

There has been considerable debate as to the validity of the data bases generated for the VNTR loci. Much of this controversy relates to the existence or non existence of Hardy-Weinberg equilibrium for the loci used in DNA genetic typing. It is highly probable that VNTR loci are in equilibrium but are perceived to deviate due to the lack of an adequate test for equilibrium at loci with non-discrete alleles [E. Wijsman - personal communication]. This opinion is in accordance with others who believe the Hardy-Weinberg test does not apply to VNTRs because of the continuum of allele sizes [B. Budowle - personal communication] and that the test is full of artifactual pitfalls [115]. With present analytical capabilities, such systems are predisposed to excess homozygosity [E. Wijsman-personal communication]. Excess of homozygotes have been shown for VNTR systems regardless of whether they are or are not in equilibrium [104]. The frequency of observed heterozygotes (Table 11) in this study is comparable to that obtained by other laboratories (as reported by the Budowle et.al. [106]) and by the RCMP laboratory [47]. Further analysis with respect to excess of homozygosity / heterozygosity, was not performed. Since present methods used to determine excess are inappropriate, the results from these methods would not be meaningful. This author does not feel that one should use a method that has been shown to be inappropriate (E. Wijsmen- unpublished results), to confirm the inconclusive results of other researchers.

More recently Devlin et.al. [103,116] have purposed another method for determining equilibrium by using only data consisting of heterozygotes and omitting the homozygotes ("contaminated data"). There have been some criticisms of this method [117]. However, this method has not yet been applied to a large amount of VNTR data or analysed by other researchers. Although data in this thesis may benefit from some analysis with this new method, this analysis was not performed since this computer program is not publically avaliable.

This thesis does not attempt to deal with the Hardy-Weinberg equilibrium because of the complexity of this problem. There is no universal agreement as to what valid statistical method

Table 11 : Comparison of Heterozygote Frequencies

Locus Laboratory	Frequency of Heterozygotes			
	D16S85	D2S44	INS	D14S13
UBC	87%	92%	75%	90%
RCMP	69%	91%	N/A	N/A
FBI	90%	91%	N/A	90%

UBC = data from this study

FBI = data from various laboratories
as reported by Budowle et. al. [107]

N/A = no available data

RCMP = data published in reference 47

could be utilized to test for equilibrium. Forensic DNA experts at the Promega DNA Symposium in April of 1991, have reported that judges have recently stopped requesting information for the Hardy-Weinberg test for VNTR loci in cases where DNA identification is presented in evidence [A. Autor - personal communication].

SECTION II

FORENSIC CASES

CASE I - Body Identification (1988)

Background Information :

On April 5 1988, two CF18 aircrafts scrambled from CFB Comox in the early hours to search for a fishing vessel in distress off the west coast of Vancouver Island. The master of the vessel had radioed a "May-Day" distress call giving his position as south of Brooks Peninsula. While one aircraft maintained an altitude of 23,000 feet, the other aircraft descended to approximately 2,000 feet to try to locate the fishing vessel visually. After repeated failed passes, the pilot of the low attitude aircraft told the master of the fishing vessel that he was going to engage his "after-burners" such that the vessel might see and/or hear the aircraft. No further contact was made with the pilot of that particular aircraft. The fishing vessel was rescued hours later by the Canadian Coast Guard.

Another search was initiated for the missing CF18 aircraft. Five days later, the crash site of the aircraft was found. The aircraft had hit the side of a mountain with its "after-burners" full on, impacting at a speed approaching mach 1 (600-700 mph). Very little remained of the aircraft and only approximately 2.5 kg of tissue from the pilot was recovered from the crash site.

Since the number of individuals capable of flying a CF18 aircraft is quite limited, the identity of the pilot was not particularly in doubt. However, to reduce chances of mistaken identity and hence the level of anxiety of the pilot's family, an attempt was made towards some means of identification. DNA genetic typing was utilized as no other means of identification was possible (i.e. no dental remains, fingerprints, or material suitable for radiological or anthropological techniques).

DNA Typing Results :

The parents of the missing pilot provided blood samples for comparison with the tissues recovered from the crash site. This analysis was performed to determine whether the profile of the tissue DNA is consistent with that of a offspring of the two individuals who provided the blood samples (parents of missing pilot). An exclusion (non-consistent DNA profile) is inferred if any particular sized DNA fragments found from the tissue is not equalled by a fragment from at least one tested parent.

The DNA samples were probed with three VNTR probes : YNH24, 3'HVR α-globin, and phins 310. The autoradiographs from two of the probings can be seen in Figures 22 and 23. In each case, DNA bands (fragments) found in the tissue DNA were also present in the profile of DNA from the parents of the missing pilot. These results do not exclude the tested parents as possible parents of the donor of the tissue. To evaluate such an inclusion, two hypotheses must be considered as discussed by D.A. Stoney [118]. These hypotheses are : a) that these tested parents passed the DNA alleles found in the tissues, and b) that a random couple have passed these alleles. In order to determine which hypothesis to accept, the probability of the observed alleles occuring under each of these hypotheses is calculated. The ratio of these two probabilities are then used to produce the relative likelihood of hypotheses "a" and "b".

Calculations : (according to formulations by D.A. Stoney [112])

$$R = \frac{\text{Probability parents would pass tissue DNA types (genotypes)}}{\text{Probability a random couple would pass tissue DNA types}}$$

a) The numerator is dependent upon the zygosity of the tested parents and can be a factor of 1.0, 0.5 or 0.25.

eg.	Couple	Possible tissue types and frequencies
1.	XX, XX	XX = 1.00
2.	XY, XY	XX = 0.25 ; XY = 0.50 ; YY = 0.25
3.	XX, XY	XX = 0.50 ; XY = 0.50
4.	XZ, YW	XY = 0.25 ; YZ = 0.25 ; ZW = 0.25 ; XW = 0.25

Figure 22 : Case I - Locus D16S85

Genomic DNA samples were probed with to detect 3'HVR α-globin alleles at the D16S85

locus.

Lane m = molecular length standards

1 = control genomic DNA sample with known alleles (female)

2 =   "   "   "   "   "   "   "   (male)

3 =   "   "   "   "   "   "   "   (female)

54 = genomic DNA sample (blood) from missing pilot's mother

53 = genomic DNA sample (blood) from missing pilot's father

52 = genomic DNA sample (tissue) found at crash site

Large arrow : allele possibly contributed by tested mother

Large arrow and two small arrows : allele possibly contributed by tested father

[1 day exposure to Kodak X-MAT AR film at -70°C with screens]

Figure 22 : Case I - Locus D16S85

028   **ALPHA – GLOBIN** (3′)

1 day
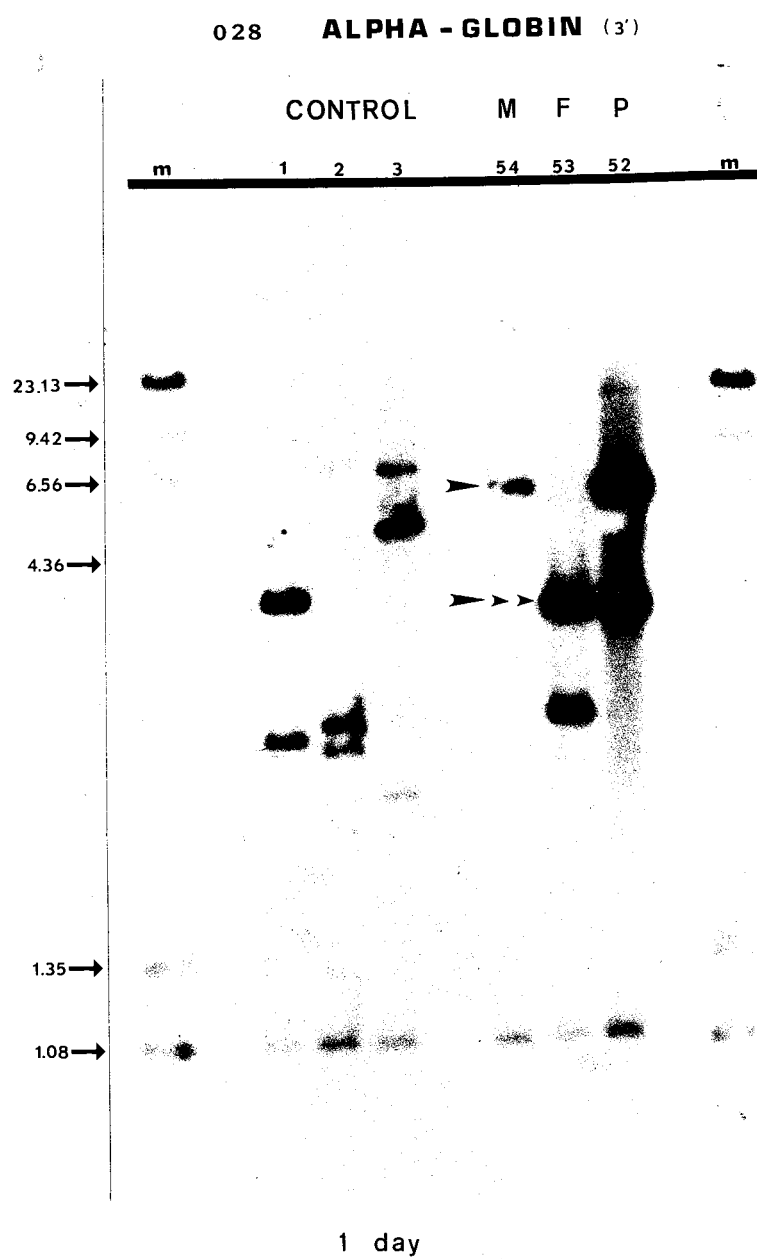
Figure 23 : Case I - Locus D2S44

Genomic DNA samples were probed with YNH24 to detect alleles at the D2S44 locus.

Lane m = molecular length standards

1 = control genomic DNA sample with known alleles (female)

2 = "   "   "   "   "   "   "   (male)

3 = "   "   "   "   "   "   "   (female)

54 = genomic DNA sample (blood) from missing pilot's mother

53 = genomic DNA sample (blood) from missing pilot's father

52 = genomic DNA sample (tissue) found at crash site

Large arrow : allele possibly contributed by tested mother

Large arrow and two small arrows : allele possibly contributed by tested father

[3 days exposure to Kodak X-MAT AP film at -70°C with screens]

Note : Molecular length markers are not visible in this autoradiograph.

Since the same membrane was probed each time, markers visible on the

D16S85 autoradiograph (Figure 19) was used for extrapolation of DNA sizes.

Figure 23 : Case I - Locus D2S44

028   YNH24

3 days
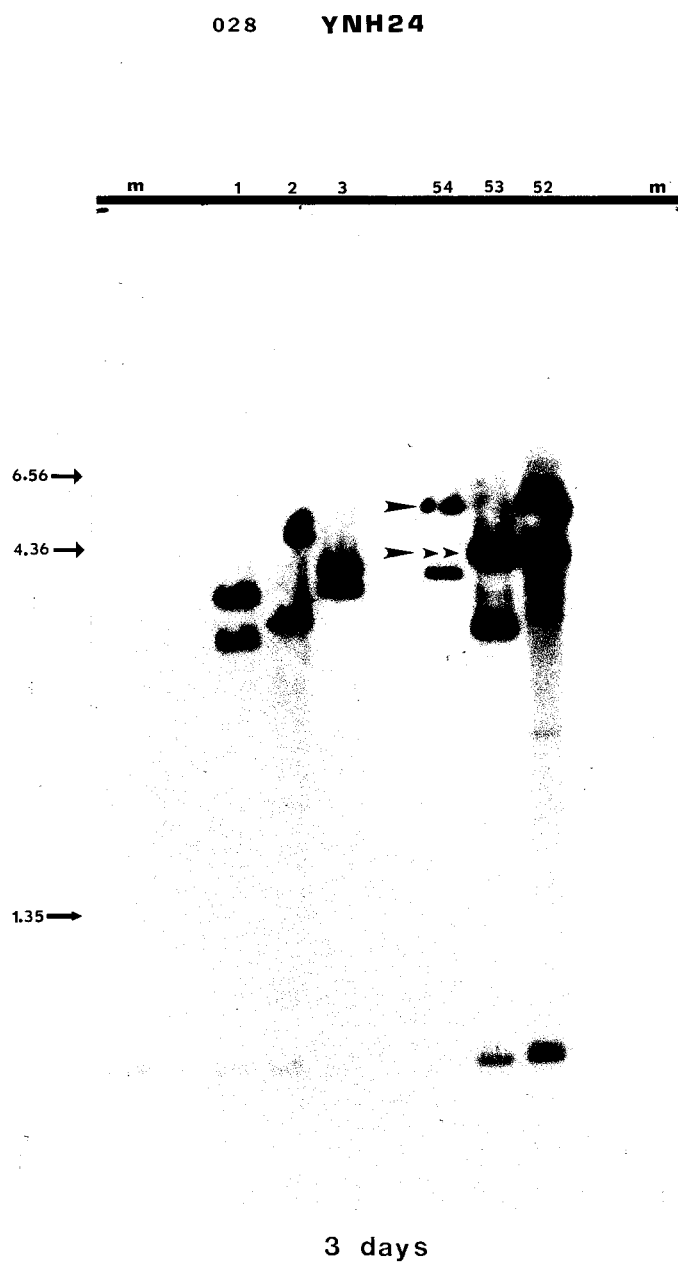
Since the tested parents were heterozygous for different alleles (analogous to couple 3 in the example above) at each of the three loci, the numerator in this case would be (0.25)(0.25)(0.25) or $(0.25)^3$.

b) The denominator is the product of the tissue DNA allele frequency in the population.

Locus	Allele Size	Frequency
D2S44	5.868 kb	* 0.154
	4.658 kb	* 0.154
D16S85	7.086 kb	* 0.012
	4.053 kb	* 0.050
INS	0.864 kb	* 0.373
	0.783 kb	0.373

*   - not the frequency value for that bin (allele)
    - the highest frequency value of either the identified bin or its two
      flanking bins is used to err on the conservative side

Freq. of occurrence = 2pq x 2pq x 2pq

= 2(0.154 x 0.154) x 2(0.012 x 0.050) x
2(0.373 x 0.373)

= 0.0000158

c) Likelihood ratio "R" = $(0.25)^3$ / 0.0000158 = 989

The tested parents are 989 times more likely as a randomly selected couple to pass this set of DNA types found in the tissue. Considering this statistic and the extremely low portion of couples with offsprings who are capable of operating a CF18 aircraft, the identity of the pilot was confirmed. This analysis was performed in 1988 during the infancy of DNA genetic typing and of this laboratory itself. Standard procedures were not outlined and this laboratory started out using

the restriction endonuclease Pvu II for the digestion of genomic DNA. Restriction endonuclease Hae III have since become the standard for the North American laboratories and was adopted by this laboratory in 1989. The genomic DNAs in this analysis were digested with Pvu II and the allele frequency values listed in the above calculations are not from Table 5 (Chapter 1) which represent values obtained with Hae III. The frequency values obtained with Pvu II are not listed in this thesis because analysis of the Pvu II data was minimal and these results cannot be corroborated with data from other laboratories. To establish some form of consistency between forensic laboratories, the Pvu II values have not been used since 1989. Although the results of this case analysis may not be acceptable by today's standards, it was adequate in 1988 in terms of the "state-of-the-art" for that time. This case serves to illustrate the applicability of DNA genetic typing to body identification.

CASE II - Associating Suspect With Victim (1988)

Background Information :

An assault victim was unwilling to identify his assailant, who was a gang member, because of fear of retaliation by the gang. The police had suspected a certain individual as the assailant and confiscated a pair of sweat pants from his room in his parents' house. The sweat pants contained 19 little dots (largest spot was < 5 mm in diameter) of what appeared to be bloodstains. DNA extracted from these 19 stains were pooled and compared DNA extracted from blood DNA of the victim. With conventional serological tests, one would not pool samples because of the possibility of more than one contributor. However, DNA typing with single locus VNTRs will only have a maximum of two bands on the autoradiograph for each probe (except in rare occasions as discussed in the introduction) so that multiple bands would be indicative of mixed samples. Since the quantity of recovered DNA was expected to be small, all the samples were used for one analysis.

DNA Typing Results :

Successful DNA banding patterns were obtained with three VNTR probes (YNH24, 3'HVR α-globin, and phins 310) and two of the probings are shown in Figures 24 and 25. In each instance, the bands of the DNA extracted from the sweat pants (Figures 24, 25 - lane SP 115) matched those of the blood DNA of the victim (Figures 24, 25 - lanes VB 116). Calculations incorporating the frequency of alleles from the three VNTR loci probed, were performed to determine the uniqueness of the DNA profile :

Probe	Allele Size	Frequency
YNH24	3.986 kb	0.195
	3.819 kb	0.195
3'HVR α-globin	6.305 kb	0.024
	5.512 kb	* 0.045
phins 310	0.617 kb	0.042
	0.602 kb	0.042

* - not the frequency value for that bin (allele)
- the highest frequency value of either the identified bin or its two
  flanking bins is used to err on the conservative side

Freq. of occurrence = 2pq x 2pq x 2pq

$$= 2(0.195 \times 0.195) \times 2(0.024 \times 0.045) \times 2(0.042 \times 0.042)$$

$$= 0.00000058 = 1/1,725,513$$

The probability of an individual having these six alleles (using these three probes) is approximately 1 in 1.7 million. Since the population of British Columbia is approximately 3.5 million, one can quite reasonably conclude that the suspect was present during the assault in order to be splattered with blood from the victim. This evidence was never brought to trial because the ownership of the sweat pants could not be proven. Initially, the brother of the suspect had claimed that the sweat pants belonged to the suspect but he changed his mind at a later date.

Figure 24 : Case II - Locus D16S85

Genomic DNA samples were probed with 3'HVR α-globin to detect alleles at the D16S85

locus.

Lane m   = molecular length standards

   1   = control DNA sample with known alleles - 5.0 ug (female)

   2   =   "   "   " "   " "   "   (male)

   115 = genomic DNA sample from suspect's pants - all

1st 116 = genomic DNA from victim's blood sample - 5.0 ug

2nd 116 = genomic DNA from victim's blood sample - 2.5 ug

[3 days exposure to Kodak X-MAT AR film at -70°C with

screens]

Figure 24 : Case II - D16S85

88-50 **ALPHA - GLOBIN**

3 days
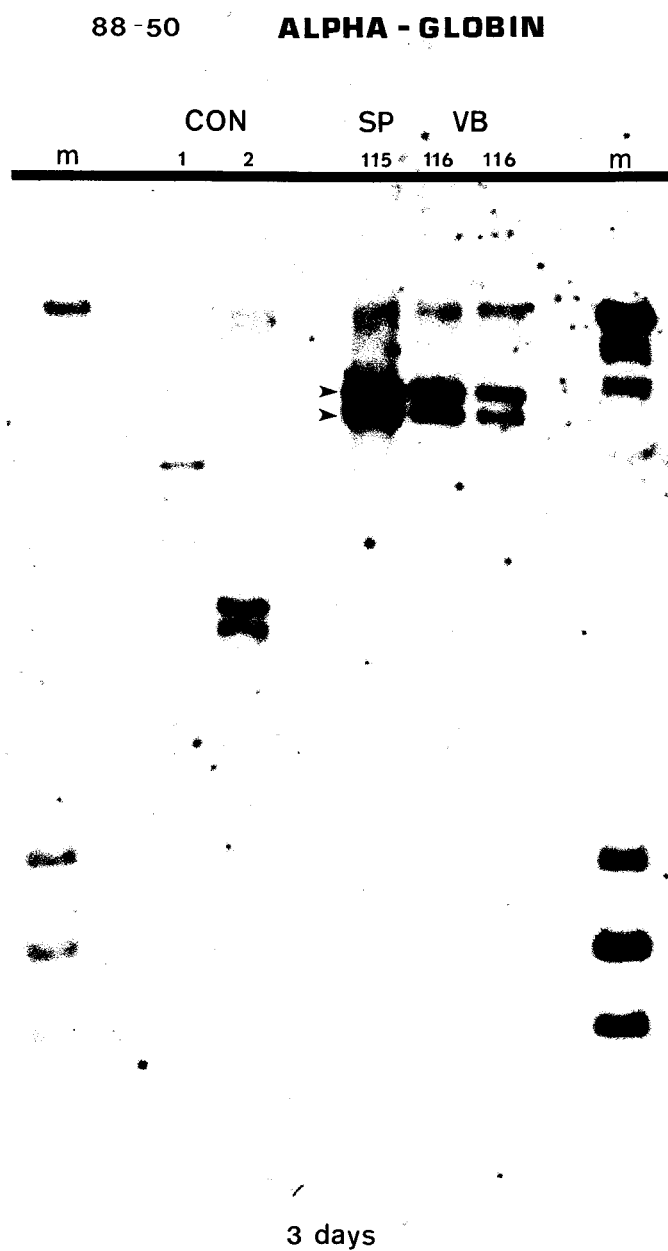
Figure 25 : Case II - Locus D2S44

Genomic DNA samples were probed with YNH24 to detect alleles at the D2S44 locus.

Lane m   = molecular length standards

      1   = control DNA sample with known alleles - 5.0 ug (female)

      2   =   "   "   "   "   "   "   "   (male)

     115 = genomic DNA sample from suspect's pants - all

1st   116 = genomic DNA from victim's blood sample - 5.0 ug

2nd   116 = genomic DNA from victim's blood sample - 2.5 ug

[3 days exposure to Kodak X-MAT AP film at $-70^{o}C$ with

screens]

Note : Molecular length markers are not visible in this autoradiograph.

    Since the same membrane was probed each time, markers visible on the

    D16S85 autoradiograph (Figure 21) was used for extrapolation of DNA

    sizes.

Figure 25 : Case II - D2S44
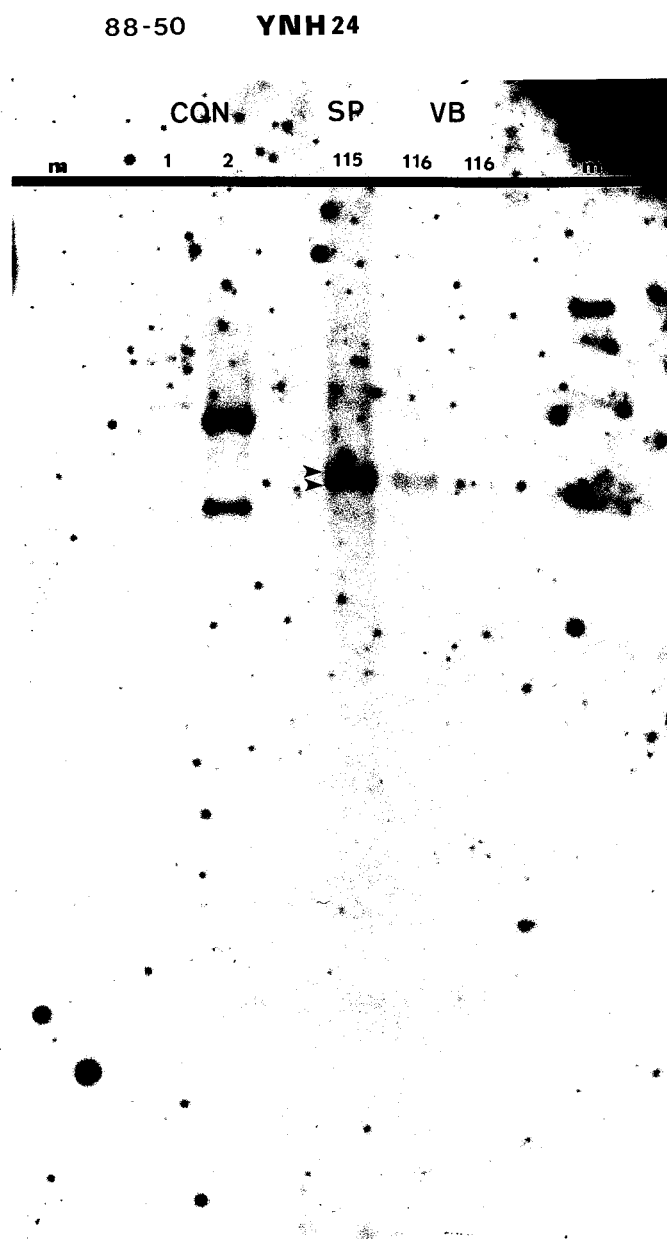
This analysis was also performed in 1988 with the restriction endonuclease Pvu II. Hence, comments from case I also apply in this case.

CASE III - Hair Analysis (1989)

Background Information :

A doctor was accused of having sexual relations with a female patient. The female patient had charged the doctor with this impropriety at the conclusion of the affair. She had apparently collected a number of pubic hairs which were claimed to be from the doctor and proof of the assault. Conventional hair fibre analysis failed to conclusively implicate the accused doctor. DNA typing was sought as a last resort at identification of the 16 pubic hairs.

Method (additions to that already mentioned in methods section) :

The pubic hairs had been mounted onto microscope slides for conventional testing and had to be place in an $80^{o}C$ xylene bath for 10 minutes to remove the cover slip and mounting medium. The nine extracted hairs were rinsed with 95% ethanol (to remove traces of xylene which may interfere with the typing procedure) and the two ends of each hair were cut-off and incubated in extraction buffer. Both ends of each hair were used to ensure that every hair shaft (which may have attached cells) was included. Microscopic examination showed that the unknown hair shafts (lower end) and follicles had degenerated, possibly due to the age of the hair and/or to the effects of the mounting on slides. Because of the dehydrated state of the hair bulbs, they were incubated in extraction buffer for 7 days instead of the usual overnight incubation. The remaining procedures for DNA extraction are as described previously (p27: protocol C).

Five day old plucked hair shaft ends from laboratory personnel were included in the analysis as controls for the extraction procedure of hair shafts. Half of these hairs (16 hairs) were treated with a $80^{o}C$ xylene bath to serve as a control for the effects of xylene on the typng procedure.

DNA Typing Results :

The most sensitive probe, 3'HVR α-globin, was used to determine whether a DNA pattern was obtainable for the hair DNA. Unfortunately, no DNA bands were observed for the unknown hair sample (Figure 26 - lane 2) after a four day autoradiographic exposure and did not appear even after a 14 day exposure (autoradiograph not shown). The control blood and hair DNA samples each produced two DNA bands with the lower band appearing quite faint in the print of the autoradiograph. The absence of band(s) in the unknown sample lane may be due to the quantity and/or quality of DNA. The entire sample (DNA extracted from cells attached to 16 hair shafts) is present in lane 2 (Figure 26) and although fluorometric readings indicate that the equivalent of approximately 0.16 ug of DNA is present, this may actually be an artifact due to fluctuations of the fluorometer when dealing with such a dilute concentration of DNA (0.009 ug/ul). Xylene treatment of DNA did not affect the DNA typing procedure as indicated by the equivalent amount of band intensities of lanes 3 and 4 (Figure 26). However, the amount of DNA recovered per hair bulb was noticeably less for the xylene treated control hair samples (Table 12). The recovery of DNA from the unknown hair samples was extremely low and could be due to a number of things such as :

1) degradation of DNA as a consequence of age

2) treatment with mounting media

3) treatment with xylene

4) small quantities of cells attached to lower end of hair shaft a result of shedding as opposed to plucking.

Although this analysis was not successful, it does demonstrate the applicability of DNA typing to hair fibre analysis as a valuable means of identification. Furthermore, the complexity of the affects of various components and substances on the entire typing procedure including DNA extraction, suggests that this system is not immune to problems and requires much more research and development.

Figure 26 : Case III - Locus D16S85

Genomic DNA samples were probed with 3'HVR α-globin to detect alleles at the D16S85

locus.

Lane 1 = control A (female) DNA sample from blood - 0.5 ug

2 = DNA from unknown hair sample - all

3 = control B (female) DNA sample from hair treated with xylene
0.5 ug (1/6 of extract from 16 hairs)

4 = control B (female) DNA sample from hair with no xylene treatment -
0.5 ug (1/9 of extract from 16 hairs)

5 = DNA from doctor's blood sample (male) - 0.5 ug

6 = control B (female) DNA sample from blood - 0.5 ug

7 = control C (male) DNA sample from blood - 0.5 ug

[4 days exposure to Kodak X-MAT AR film at -70^{0}C with screens]

Figure 26 : Case III - Locus D16S85
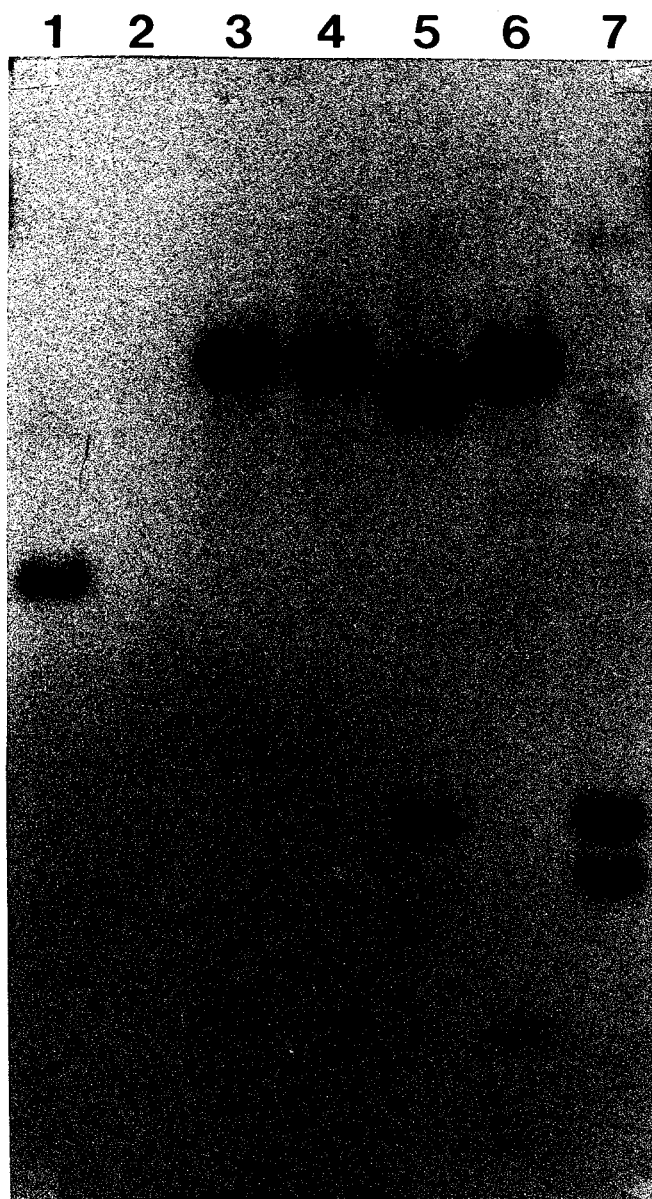
Table 12 : **Recovery of DNA From Cells Attached to Hair Shafts**

Sample	Age	Method of Removal	Treatment	Extracted DNA (ng/hair)
Control A	5 days	plucked	no treatment	275
			xylene	180
Unknown	Unknown (>> 5 days)	shed	-mounting media -xylene	11.2

DISCUSSION

The three forensic cases presented in this section serve to illustrate a few of the possible applications of DNA genetic typing. In cases 1 and 2, no mention was made of "matching" criteria (5.5% of fragment size) as determined by measurement imprecision within gels. The declared "matches" were only made visually (subjectively), without the objective boundaries described previously (section I - measurement imprecision). This was because these cases were analyzed with the restriction endonuclease Pvu II system (which also uses different size gels and molecular length standards) and the measurement imprecision experiments have not been and will not be performed by this laboratory. Once the switch was made to the Hae III system, it was more reasonable to fully concentrate on that system as the Pvu II system was not compatible with other North American Forensic laboratories. However, the calculated values of identity should not be significantly changed since the sizes of the bins for Pvu II system were greater than 5% of the fragment size.

The number of details to consider in DNA analysis of forensic specimens is enormous and the unknowns encountered in case III give an idea of the types of questions that have to be answered. For example, the effects of aging on DNA stability of biological samples is of such significance that, many laboratories have devoted much work to analysis of aged specimens with respect to DNA analysis for both sex typing and individual profiling.

Human sex determination of one week old bloodstains (equivalent to 10 ul of whole blood) by dot hybridization with a radioactive labelled Y-specific probe (pY3.4) was reported by Tyler et al. [119]. Furthermore, sex determination with a 4 year old blood stain (equivalent to 100 ul whole blood), a single hair, 5 ul whole blood and 2 ul semen was demonstrated with radioactive labelled probe pHY2.1, using either dot or blot hybridization [15]. More recently, Yokoi and Sagisaka have shown that another Y-specific radioactive labelled probe (pHY10) provides a more rapid (two day) dot hybridization sex determination test for blood and bloodstains [120]. The sensitivity of this

radioactive labelled pHY10 probe is at a range of 1 ul of blood and is 30-50 times greater than that of its photobiotin labelled counterpart.

With Jeffreys' minisatellite probe 33.15, Gill et al. have obtained DNA fingerprints from bloodstains up to 4 years old, 11 week old semen stain and hair roots [11]. However, blood and semen stains stored at room temperature for more than four years were found to be not typeable by this same laboratory [16]. Sperm DNA from a single vaginal swab taken at up to 36 hours post intercourse, have been typed by probe 33.15 [16]. Contamination of sperm DNA with vaginal cell DNA was eliminated by using a differential cell lysis method which preferentially lyses female cells [11]. Successful DNA typing with hypervariable single locus probes have been accomplished for 28 day bloodstains stored at room temperature and 2 to 3 year old stains stored in a cold room [12]. Studies have indicated that hypervariable single locus probes can also be used to successfully type semen stain DNA and sperm DNA isolated from vaginal swabs [13]. Additionally, DNA extracted from human bone has found to be suitable for DNA typing and DNA yields are approximately 10 to 20 fold higher in spongy tissue than in compact tissue (based on weight of tissue) [121].

Other than the quantity of DNA, the major determinant of the success of a DNA typing analysis of a forensic specimen is the quality of DNA. The presence of extractable DNA from aged specimens was demonstrated by the cloning of a 3.4 kb DNA fragment from a 2,400 year old Egyptian mummy [122]. However, agarose gel electrophoresis indicated that the most of the DNA was < 0.5 kb with only a minor portion of DNA > 5.0 kb. The rate of postmortem autodegradation of DNA in human rib bone was concluded to increase with increasing humidity and temperature of storage environment [123]. An extensive study of postmortem stability of DNA in human organs by Bar et al. demonstrated that the amount of degraded DNA correlated with the length of the postmortem period [124]. However, case histories indicated that this DNA degradation was also influenced by the environmental temperature at the site of death and/or the presence of infectious disease prior to death. DNA from different organs were stable for various lengths of time. Liver DNA was completely digested 24-36 hours postmortem, and DNA was stable

for up to five days in the kidney, spleen, and thyroid glands. Greatest stability was found in organs such as the brain, psoas muscle and lymph nodes. Blood DNA gave conflicting results with good stability in some cases and poor in others. This variability is thought to be a result of non-homogenous sampling of blood clots. Blood clots, even from decayed bodies, produced good yields of high molecular length DNA. Although environmental conditions such as humidity and temperature play a critical role with regards to the quantity and quality of DNA recovered from forensic specimens, there are undoubtedly other contributing environmental factors. The effects of ultra-violet radiation, rain, air pollution, soil, synthetic dyes (in clothes), oil, and gasoline, must all be taken into consideration and some of these issues will be discussed further in the Addendum.

CONCLUSION

The application of DNA genetic typing in forensic science has enormous potential. However, the nature of forensic specimens makes this system of analysis complex and it is important to determine the possible affects of a wide range of variables on the success, validity and reliability of the analysis. Results of such testing have tremendous implications on the lives of the victim and the accused in cases of assault. Although research in the last few years has been very informative, there is still an obvious need for continuing research and development of this system.

CHAPTER 2

DIMORPHIC RES ANALYSIS

Section I - Data Base Generation

- Reliability Studies

Section II - Discriminating Power Studies

INTRODUCTION

One of the greatest limitations of DNA technology as it applies to forensic work is the quantity and quality of DNA available for analysis. With the application of thermostable Taq DNA polymerase for use in DNA synthesis in the polymerase chain reaction (PCR) [125], the means became available to eliminate or substantially reduce these problems. The polymerase chain reaction allows enzymatic amplification of minute amounts of specific nucleotide sequences by more than a million fold in copy number (Figure 27). This powerful technique has simplified many working protocols in molecular biology and enabled genetic manipulations to be carried out with trace amounts of starting material. Amplification of nucleic acids has been applied in molecular medicine and has resulted in more efficient diagnostic tests for various genetic diseases [92,126-128] and more sensitive tests for the detection of microbial pathogens [129,130].

Currently, the polymerase chain reaction is being used to amplify polymorphic VNTR regions [131-133] which may be used for individual identification purposes. With this type of analysis, it is possible to complete the identification procedure within 24 hours [131]. In contrast, the current procedure used for non-PCR identification may require 6 to 8 weeks. Although PCR identification can be performed more quickly using minute amounts of DNA and is less labour intensive, several other problems previously discussed involving VNTRs still exist. Furthermore, the VNTR-PCR procedure has an obstacle of its own that may limit its application in forensic science. Since a hypervariable locus may have 50 to 100 alleles, it will be impossible to predict the allelic sizes of the amplified DNA fragments. Hence, nonspecific and/or unfaithful amplification products may be erroneously identified as true alleles leading to invalid interpretations.

In light of problems with methodology, interpretation, and statistics of RFLP analysis using the VNTR (or VNTR-PCR) system, it seemed appropriate to search for alternate strategies for RFLP analysis. This eventually lead to the idea of DNA profiling with PCR amplification of RFLPs created via a simpler mechanism, by utilizing point mutations (Figure 28). Such point mutations which could result in either the elimination or the creation of a restriction endonuclease

Figure 27 : Polymerase Chain Reaction
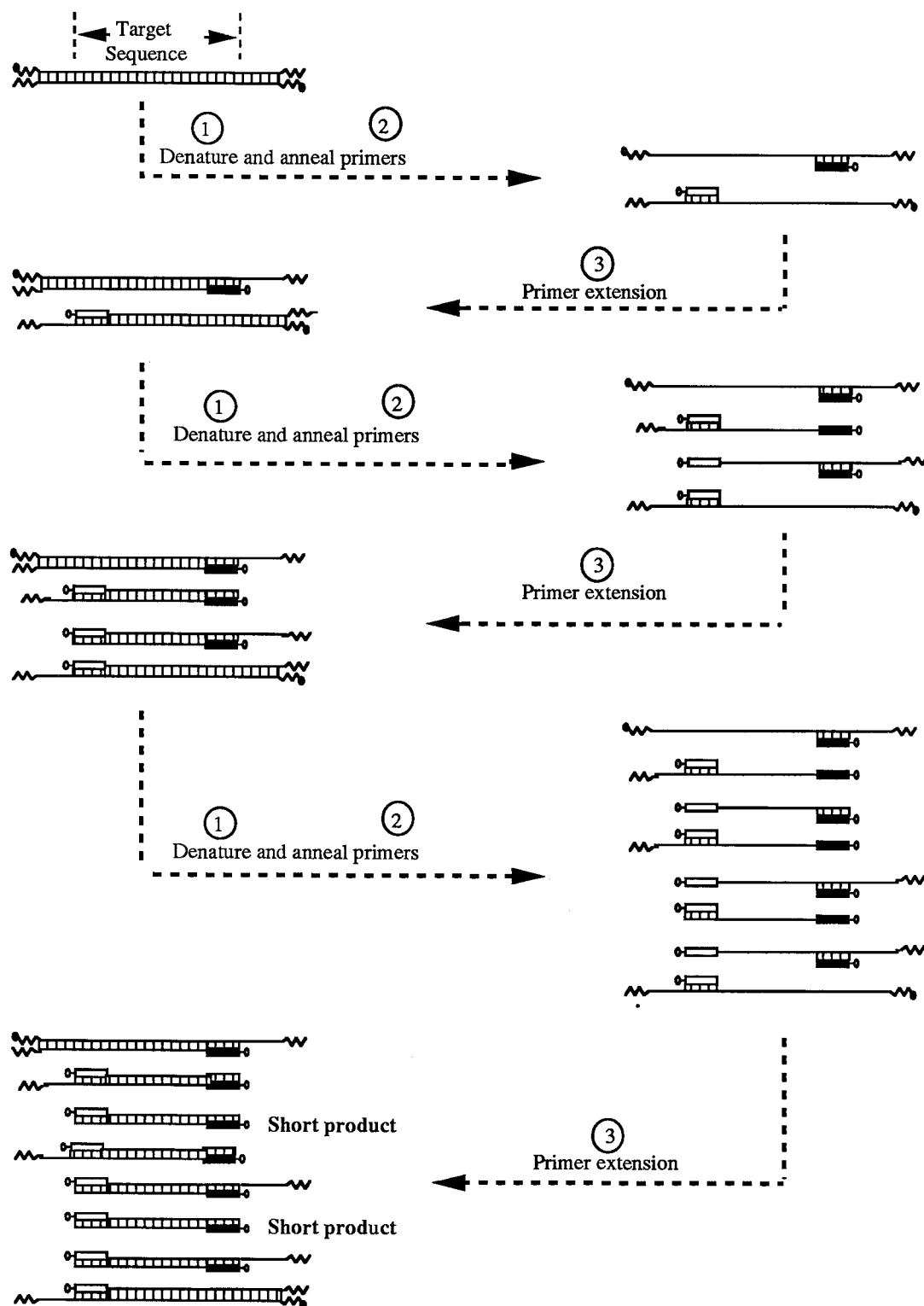
Figure 28 : Amplification of a Dimorphic Locus
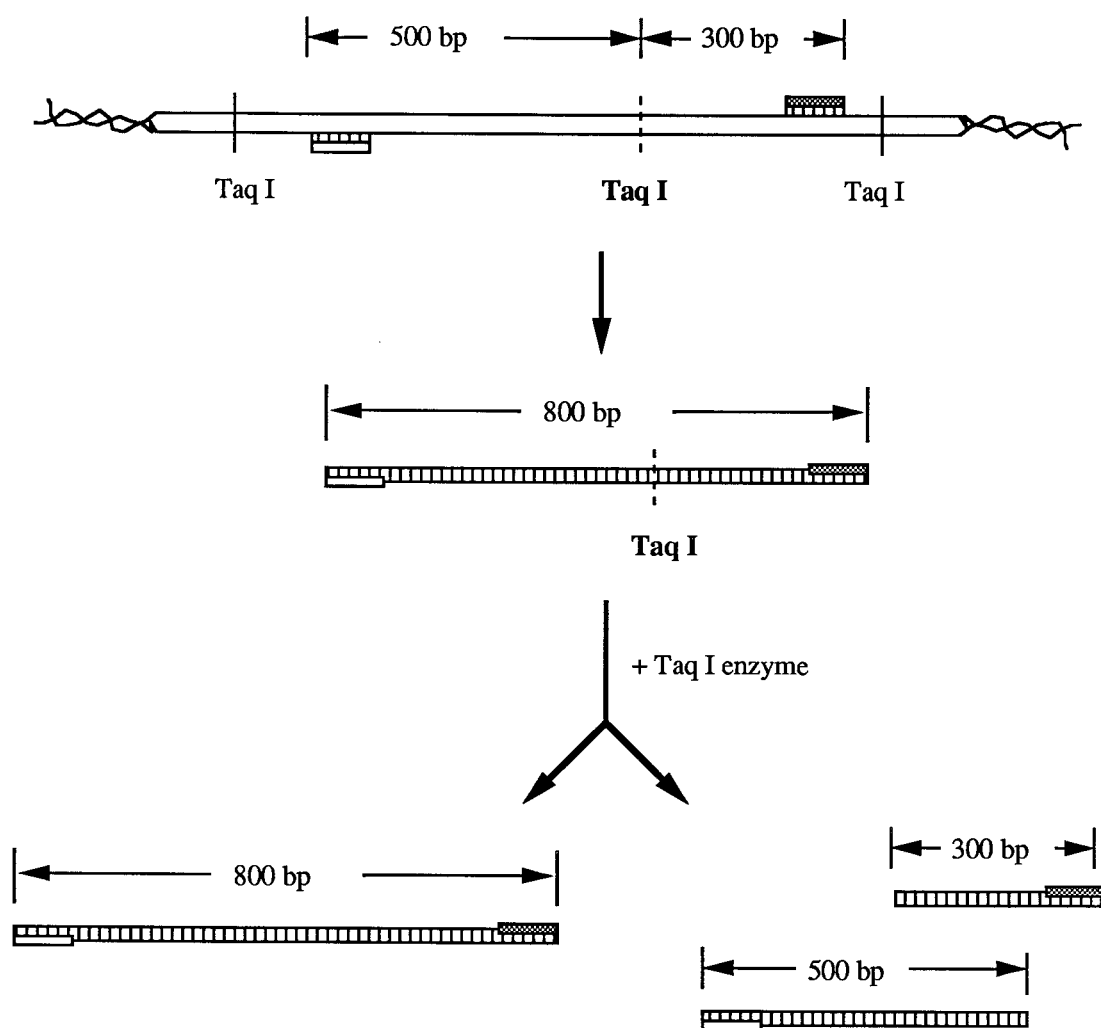
recognition site (RES), will have only two possible alleles (dimorphic). Thus, for a particular RES

that is located in a specific chromosomal pair, an individual may possess three genotypes:

homozygous negative, homozygous positive, and heterozygous.

The presence or absence of a specific RES may be considered as two alleles of one specific

locus. If the alleles of two particular loci ("X" and "Y") were unlinked, different combinations of

the alleles may occur leading to nine possible genotypes (eg. XXYY, XxYY, xxYY, XxYy, Xxyy

etc.) in the population. If a number of such unlinked genomic loci are available, analyses of their

respective allelic frequencies may provide a potential system for DNA profiling.

Amplification of such loci with PCR solves DNA quantity problems and simplifies the entire typing

procedure. In contrast to the VNTR system which identifies individuals according to the sizes of

the DNA fragments (alleles), this dimorphic RES system identifies by comparing specific DNA

sequences.

Polymerase chain reaction amplification of five dimorphic RES loci is analyzed to collect

allelic frequency data bases for these loci and to also determine their potential for generation of

discriminating DNA profiles. The five loci are : adenine phosphoribosyltransferase gene (APRT),

lipoprotein lipase gene (LPL), prealbumin gene (PALB), adenine deaminase gene (ADA), and

carbonic anhydrase II gene (CAH).

METHODS

Population Samples :

Undefined population sample are as described in chapter 1 for VNTR analysis (p38-39).

The "Oriental" population was also taken from the same group as that for the undefined population

but with an additional stipulation of an "Oriental-like" surname (i.e. Wong, Lee, Nyen). To reduce

the likelihood of "Non-Orientals" with "Oriental-like" surnames as a result of inter-racial marriages,

the females of this group also possessed an "Oriental-like" first name (i.e. Mae-Ling, Su Lan).

Extraction of DNA from liquid blood for PCR analysis:

DNA was extracted from liquid blood samples as described in protocol A and B for VNTR

analysis (chapter 1 - methods).

Quantitation of DNA :

Refer to chapter 1 - methods section

Loci With Dimorphic RES :

The selection criteria for loci were that each should contain an unique dimorphic RES :

1) with allelic frequencies of as close to 50-50 as possible and not more

divergent than 70-30

2) with available sequence information to reduce work load

3) which is known to be not associated with a disease.

Oligonucleotide Primers :

The specific primers (two for each locus) for the APRT, PALB, ADA, CAH and LPL loci

were designed by selecting oligomeric sequences (20 bases) from published sequence data [134-

138]. The selection criteria for the sequences were that each should:

1) flank a polymorphic RES and the distance separating each primer from the RES be <

0.70 kb

2) contain only a single cleavage site for the specific restriction endonuclease in the selected region

3) exhibit high specificity in terms of sequence homology with the locus of interest (low affinity with other loci)

4) not contain long stretches of polypurines or polypyrimidines.

The respective sequences selected for the construction of the primers were:

a) APRT 1 : 5'-TTGTGAGATTGAGCCCCCGA- 3'

APRT 2 : 5'-AGCTGAAAGGCCAGTGACAT- 3'

b) PALB 1 : 5'-TCTCAGCCTGATGGGAACCA- 3'

PALB 2 : 5'-GATCTGAGTGAGCCATACCA- 3'

c) ADA 1 : 5'-AGCAGCCAGCCAGTAAAATG- 3'

ADA 2 : 5'-AGCGTATCCTCCTCTTCCAA- 3'

d) CAH 1 : 5'-CACAGTGTCTCAGGTCCACA- 3'

CAH 2 : 5'-AAGGGCTCAGTCTCTGGTGT- 3'

e) LPL 1 : 5'-AAAGGAATGGTCGGAAAATG- 3'

LPL 2 : 5'-GTTTGAGACTAGCTTGGCCA- 3'

The APRT, primers were synthesized by the Biochemistry Department at the University of British Columbia, the PALB primers were synthesized by Pharmacia and the ADA, CAH and LPL primers were synthesized by this author using the PCRMate in the Pathology Department (Medical Microbiology) at the University of British Columbia.

When these primer sets were used individually in the PCR to amplify the APRT, PALB, ADA, CAH, and LPL loci, a 0.73 kb, 1.01 kb, 0.96 kb, 0.72 kb, and 1.03 kb product fragment respectively would be expected. The polymorphic restriction endonuclease recognition sites in the five loci were analyzed (Figure 29).

Figure 29 : Dimorphic RES Loci

APRT Locus

PALB Locus

ADA Locus

CAH Locus

LPL Locus
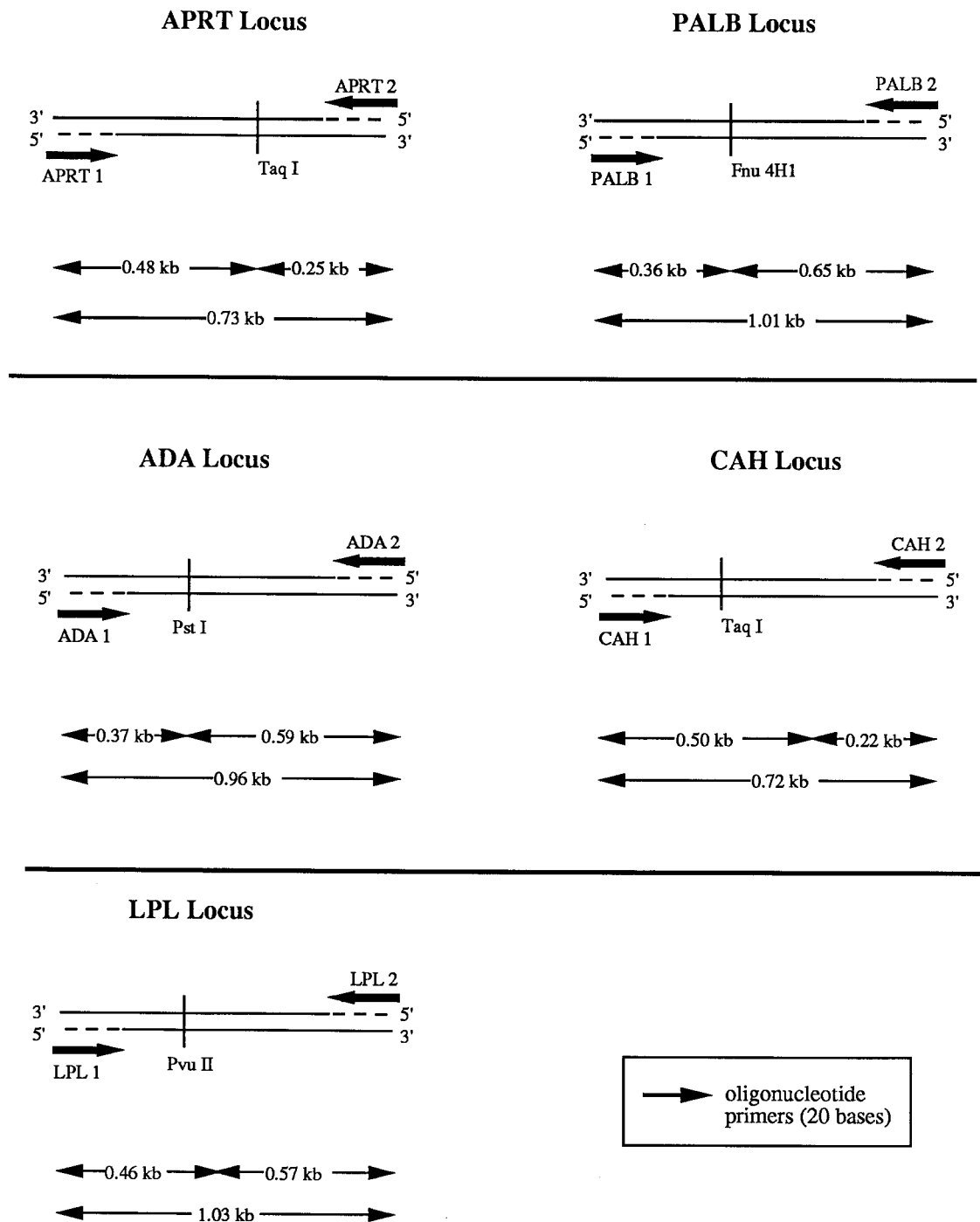
PCR for the APRT Locus From Genomic DNA Extracted From Blood :

DNA samples from undefined individuals from the population of the Greater Vancouver Area, were selected. Two hundred nanograms of each genomic DNA sample were added to 0.3 uM of each of the two APRT primers, 1.7 mM of $MgCl_2$, 200 uM of each of the four deoxynucleotide triphosphates (dNTPs) (Perkin Elmer Cetus/Pharmacia), and 1.25 units of Taq DNA polymerase (Perkin Elmer Cetus/BRL), in a total volume of 50 ul. An equal volume of mineral oil was layered onto the aqueous solution. The PCR was carried out in a thermal cycler (Perkin Elmer Cetus) for 31 cycles with an initial (before the first cycle) denaturation of 3 minutes at 94^OC. The reaction cycle conditions were: 94^OC (30 seconds) for denaturation, 61^OC (30 seconds) for primer annealing, 72^OC (60 seconds) for extension, and with an auto-extension period of 3 seconds (at 72^OC) at the end of each cycle. In the later experiments, the denaturation temperature was increased to 95^OC and the cyclic denaturation time was increased to either 40 or 60 seconds.

PCR For the PALB Locus From Genomic DNA Extracted From Blood :

Two hundred nanograms of genomic DNA from the same samples described above, or genomic DNA from Orientals, were added to 0.2 uM of each of the two PALB primers, 2.0 mM of $MgCl_2$, 200 uM of each of the four dNTPs, and 1.25 units of Taq DNA polymerase in a total volume of 50 ul. The PCR cycle conditions were carried out as described for the APRT locus.

PCR For the ADA Locus From Genomic DNA Extracted From Blood :

Two hundred nanograms of genomic DNA from the same samples described for the APRT locus, or genomic DNA from Orientals, were added to 0.3 uM of each of the two ADA primers, 3.0 mM $MgCl_2$, 200 uM of each of the four dNTPs, and 1.25 units of Taq DNA polymerase in a total volume of 50 ul. The PCR cycle conditions were carried out as described for the APRT locus.

PCR For the CAH Locus From Genomic DNA Extracted From Blood :

Two hundred nanograms of genomic DNA from the same samples described for the APRT locus, or genomic DNA from Orientals, were added to 0.3 uM of each of the two CAH primers, 1.5 mM $MgCl_2$, 200 uM of each of the four dNTPs, and 1.23 units of Taq DNA polymerase in a total volume of 50 ul. Equal volume of mineral oil was layered onto the aqueous solution. The PCR was carried out in a thermal cycler (Perkin Elmer Cetus) for 31 cycles with an initial (before the first cycle) denaturation of 3 minutes at 95^oC. The reaction cycle conditions were: 95^oC (40 seconds) for denaturation, 61^oC (30 seconds) for primer annealing, 72^oC (60 seconds) for extension, and with an auto-extension period of one second (at 72^oC) at the end of each cycle. In later experiments, the cyclic denaturation time (at 95^oC) was increased to one minute.

PCR For the LPL Locus From Genomic DNA Extracted From Blood :

Two hundred nanograms of genomic DNA aliquoted from the same samples described for the APRT locus, were added to 0.17 uM of each of the two LPL primers, 4.0 mM $MgCl_2$, 200 uM of each of the four dNTPs, and 1.23 units of Taq DNA polymerase in a total volume of 50 ul. An equal volume of mineral oil was layered onto the aqueous solution. The PCR was carried out in a thermal cycler (Perkin Elmer Cetus) for 31 cycles with an initial (before the first cycle) denaturation of 3 minutes at 95^oC. The reaction cycle conditions were: 95^oC (40 seconds) for denaturation, 59^oC (30 seconds) for primer annealing, 72^oC (60 seconds) for extension, and with an auto-extension period of one second (at 72^oC) at the end of each cycle. In the later experiments, the cyclic denaturation time (at 95^oC) was increased to one minute the primer annealing temperature was increased to 61^oC and the $MgCl_2$ concentration was decreased to 3.0 mM.

Restriction Endonuclease Digestion of the PCR Products and Gel Electrophoresis :

The PCR products were precipitated with 2 volumes of 95% ethanol and 1/10 volume of 0.3 M sodium acetate (pH 5.2). The resulting pellet was resuspended in 12 ul of sterile water and

separately digested with the appropriate restriction endonuclease. PCR products obtained from the APRT and CAH loci were digested with Taq I, PCR products obtained from the PALB locus were digested with Fnu 4H1, PCR products obtained from ADA locus were digested with Pst I, and PCR products obtained from the LPL locus were digested with Pvu II. The digest conditions were those recommended by the suppliers of the respective restriction endonucleases. DNA fragments in the digests were separated by electrophoresis through 1% agarose gel (containing 1 ug/ml ethidium bromide) in 1X TBE buffer (0.089 M Tris-borate, 0.089 M boric acid, 0.025 M EDTA, pH 8.3) or in 1X TAE buffer (0.04 M Tris-acetate, 0.002 EDTA). Molecular length standards included in PCR electrophoresis gels were a combination of lambda-Hind III/ϕX174 RF-Hae III (BRL).

Control For PCR Amplification :

The conditions for PCR amplification of the five loci were established using one stock DNA sample. This stock sample was used in each subsequent PCR run as a control to confirm PCR conditions. Since contaminated DNA would be present in extremely small quantities relative to the large amount (200 ng) of test sample DNA, negative controls were not included in all of the tests.

Control For the Activity of the Restriction Endonuclease :

Plasmid DNA fragments containing an unique restriction endonuclease recognition site (RES) were included as controls to confirm the activity of the restriction endonuclease used in each digest reaction of the PCR products. To avoid complicating the zygosity patterns, these plasmid fragments were chosen so that the digested plasmid fragments were either larger or smaller than the undigested or digested PCR products. Two different Taq I control fragments were used for the two loci (APRT/CAH) with the polymorphic Taq I RES. The control fragment (0.42 kb) for the APRT locus was prepared from pBR322 and digestion of this fragment with Taq I results in two fragments (differing in length by 10 bp) that appear as one band at the 0.21 kb range. The control fragment (2.48 kb) for the CAH locus was also prepared from pBR322 and digestion of this

fragment with Taq I results in two fragments (differing in length by 60 bp) that appear as bands in the 1.27 and 1.21 kb range. The only reason for having two Taq I control fragments was that the smaller control fragment was prepared initially and since it appeared to be easier to visualize larger fragments, another control fragment was made for the CAH locus. The control fragment (0.52 kb) for the PALB locus was prepared from pUC 19 and digestion of this fragment with Fnu 4H1 results in two fragments (differing in length by 60 bp) that is observed as bands at the 0.27 and 0.21 kb range. The control fragment (2.77 kb) for the ADA locus was prepared from pBR322 and digestion of this fragment with Pst I results in two fragments (differing in length by 30 bp) that appears as one band at the 1.40/1.37 kb range. The control fragment (3.05 kb) for the LPL locus was also prepared from pBR322 and digestion of this fragment with Pvu II results in two fragments (differing in length by 30 bp) that appear as one band at the 1.54/1.51 kb range.

Reliability of PCR Amplification :

The reliability of the amplification by PCR was monitored by one or more of the following methods:

1) DNA sequencing of the oligo primers

2) probing of the amplified products with a DNA probe that is homologous to the target locus

3) correlating the number and sizes of the restriction fragment(s) that were obtained experimentally with that of the calculated size.

Transformation of Recombinant Plasmids into Competent E.Coli (DH5α) Cells :

Recombinant plasmids containing inserts which probe for the amplification products at the APRT, ADA, CAH and LPL loci were transformed into E.coli (DH5α) cells for amplification.

Refer to chapter 1 - methods section

Small Scale Amplification and Isolation of Plasmid DNA :

Refer to chapter 1 - methods section

Excision of Foreign DNA Insert From Recombinant Plasmids :

Refer to chapter 1 - methods section

Labelling and Purification of DNA Probes :

Refer to chapter 1 - methods section

Hybridization of ^{32}P-DNA probes to membrane-fixed DNA :

Refer to chapter 1 - methods section

Post hybridization washes :

Refer to chapter 1 - methods section

Probe removal from membrane :

Refer to chapter 1 - methods section

Specificity of the Oligonucleotide Primers for Human DNA :

Attempts were made to amplify non-human DNA in order to establish the specificity of the

oligonucleotide primers. The different types of DNA used were: 1) bacterial DNA (E.coli.,

C.perfringens, P.syringe, and B.subtilis), 2) yeast DNA (C.albicans), 3) plasmid DNA (pACYC

184), 4) mouse DNA (cell line 3T3) and 5) DNA extracted from blood of a dog, cat, and chicken.

Amplification of these DNA was allowed to proceed for at least 35 cycles (cf. 31 cycles for DNA

extracted from human blood).

SECTION I

RESULTS AND DISCUSSION

Reliability of the PCR (one primer set / reaction):

Analysis of the electrophoretogram showed that only one band could be detected in each of the amplified DNA products and their respective sizes were as expected, i.e. 0.73 kb for the APRT locus, 1.01 kb for the PALB locus, 0.96 kb for the ADA locus, 0.72 kb for the CAH locus and 1.03 kb for the LPL locus. The results of the restriction endonuclease digestion of the respective PCR products agreed with that of theoretical calculations (Figures 30-34). The zygosity of each locus for an individual was determined by referring to the number and sizes of the restriction fragments generated in each case (Table 13). Individuals that are homozygous for the absence of the RES at a specific locus will show only one DNA band (locus DNA fragments intact) while those who are homozygous for the presence of the RES, will show two bands (result of complete cleavage of all locus DNA fragments at the respective RES).

Individuals that are heterozygous for the restriction endonuclease site will show three bands. One of these bands represents intact locus DNA fragments indicating that the enzyme site is absent in one chromosome. The other two bands represent cleaved DNA fragments indicating that the particular endonuclease enzyme site is present in the second chromosome.

In addition to correlating the number and sizes of the restriction fragment(s) that were obtained experimentally with that of the calculated size for each of the five loci analyzed, the specificity of the PALB amplification was ensured by DNA sequencing of the two oligo primers by their manufacturers (Pharmacia). Alternately the specificity of the APRT, ADA, CAH, and LPL amplifications were confirmed by hybridization of the amplification products of each loci with their respective genomic DNA probes. The Southern blots used in each hybridization contained the amplification products of all five dimorphic loci (Figure 35). Specific hybridizations (a single band corresponding to the calculated size) were obtained with the APRT (Haup15), CAH (H25-3.8), and

Figure 30 : APRT Locus

Human genomic DNA was amplified at the APRT locus and the amplified material (0.73 kb) was digested with restriction endonuclease Taq I. A DNA fragment (0.42 kb) with a recognition site for Taq I was included in the digest reaction to serve as an internal control for the activity of the restriction endonuclease.

Lane 1 = Homozygous positive (0.48 and 0.25 kb fragments)

        2 = Heterozygous (0.73, 0.48, and 0.25 kb fragments)

        3 = Homozygous positive

        4 = Homozygous positive

        5 = Undigested control fragment (0.42 kb)

        6 = Taq I digested control fragment (0.21 kb)
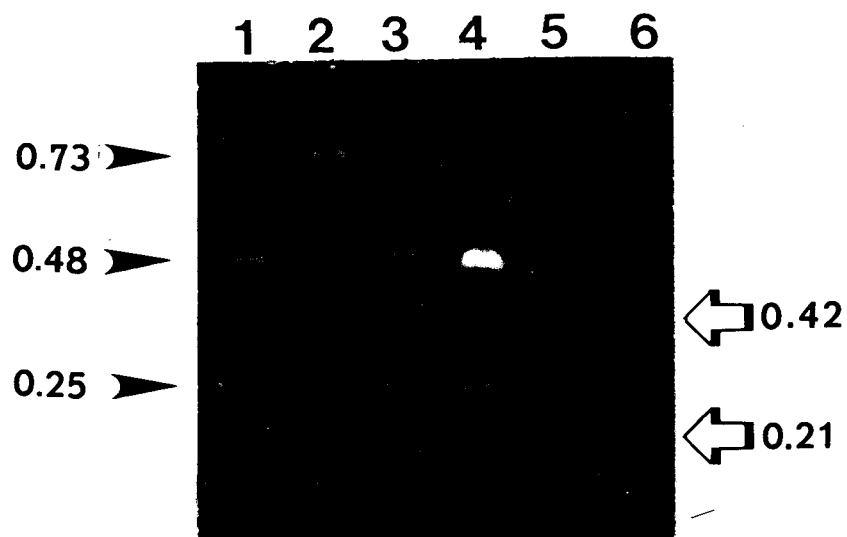
Figure 31 : LPL Locus

Human genomic DNA was amplified at the LPL locus and the amplified material (1.03 kb) was digested with restriction endonuclease Pvu II. A DNA fragment (3.05 kb) with a recognition site for Pvu II was included in the digest reaction to serve as an internal control for the activity of the restriction endonuclease.

Lane m = molecular length standards

      1 = Undigested control fragment (3.05 kb)

      2 = Pvu II digested control fragment (1.54/1.51 kb)

      3 = Homozygous positive (0.57 and 0.46 kb)

      4 = Heterozygous (1.03, 0.57, and 0.46 kb)

      5 = Homozygous negative (1.03 kb)

Note: The 0.57 and 0.46 kb fragments are fainter than that seen in the original gel because the print is a picture of a picture.
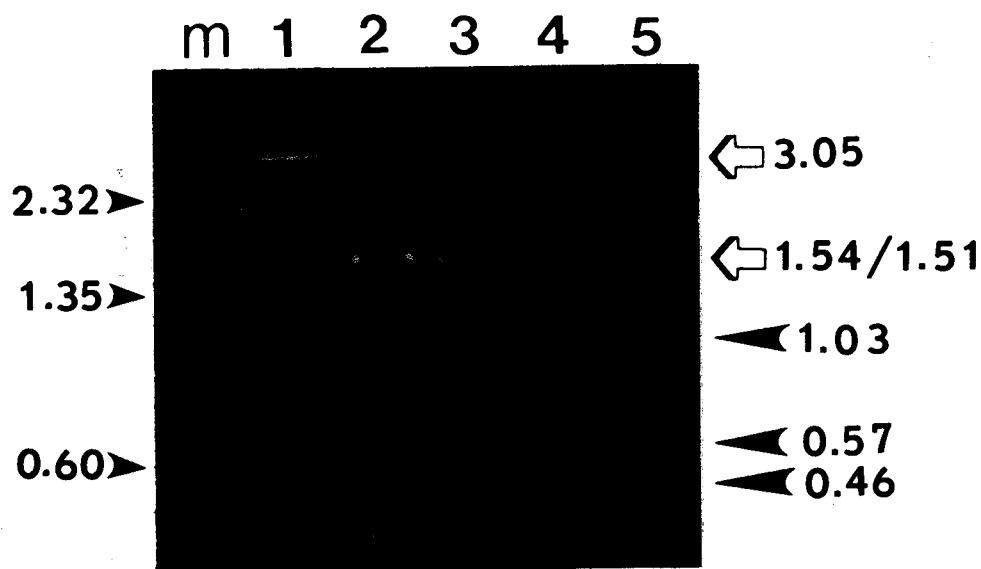
Figure 32 : PALB Locus

Human genomic DNA was amplified at the PALB locus and the amplified material (1.01 kb)

was digested with restriction endonuclease Fnu 4H1. A DNA fragment (0.52 kb) with a

recognition site for Fnu 4H1 was included in the digest reaction to serve as an internal

control for the activity of the restriction endonuclease.

Lane   M = molecular length standards

1 = Undigested control fragment (0.52 kb

2 = Fnu 4H1 digested control fragment (0.27 and 0.21 kb)

3 = Homozygous positive (0.65 and 0.36 kb)

4 = Heterozygous (1.01, 0.65, and 0.36 kb)

5 = Homozygous positive
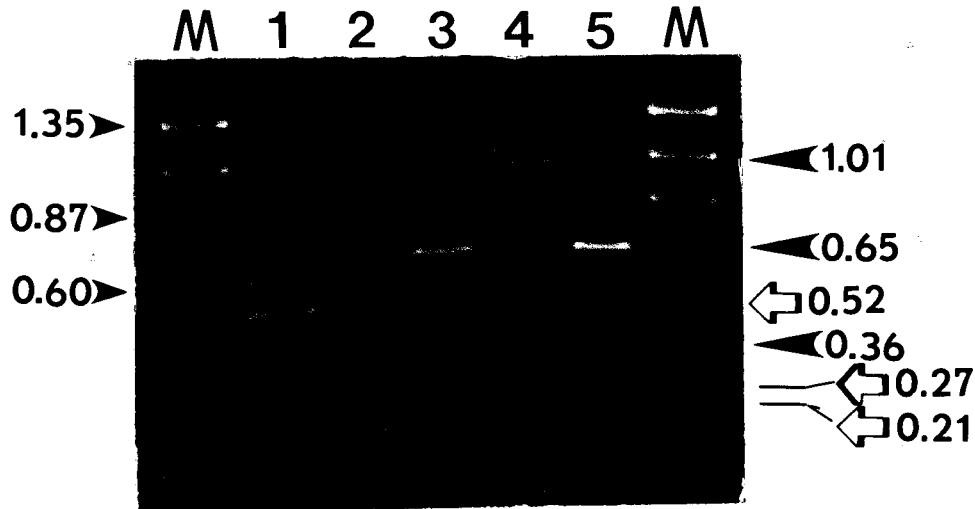
Figure 33 : ADA Locus

Human genomic DNA was amplified at the ADA locus and the amplified material (0.96 kb)

was digested with restriction endonuclease Pst I. A DNA fragment (2.77 kb) with a

recognition site for Pst I was included in the digest reaction to serve as an internal

control for the activity of the restriction endonuclease.

Lane m = molecular length standards

      1 = Undigested control fragment (2.77 kb)

      2 = Pst I digested control fragment (1.40/1.37 kb)

      3 = Homozygous positive (0.59 and 0.37 kb)

      4 = Homozygous negative (0.96 kb)

      5 = Heterozygous (0.96, 0.59, and 0.37 kb)

      Note: The 0.37 kb fragment is fainter than that seen in the original gel

         because the print is a picture of a picture.

m 1 2 3 4 5

2.32►   2.77

1.35►   1.40/1.37

        0.96

0.60►   0.59

        0.37
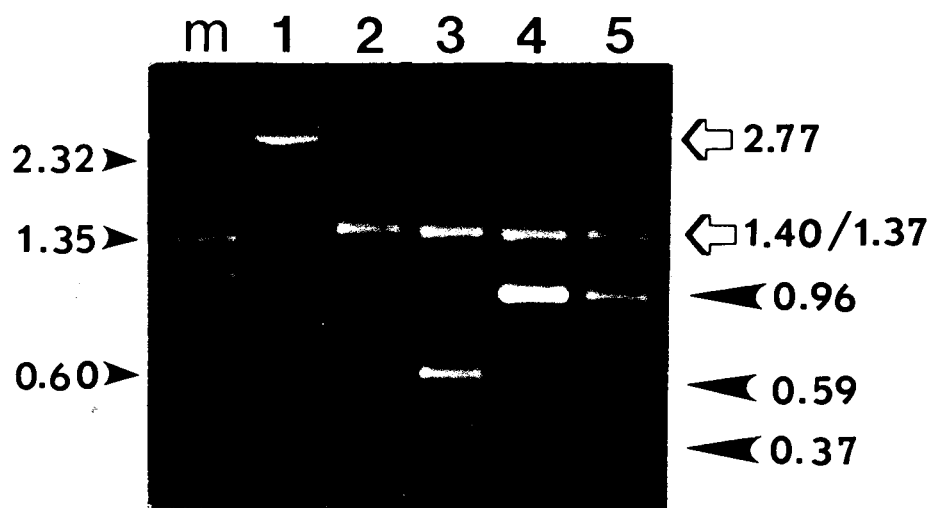
Figure 34 : CAH Locus

Human genomic DNA was amplified at the CAH locus and the amplified material (0.72 kb)

was digested with restriction endonuclease Taq I. A DNA fragment (2.48 kb) with a

recognition site for Taq I was included in the digest reaction to serve as an internal

control for the activity of the restriction endonuclease.

Lane   m = molecular length standards

     1 = Undigested control fragment (2.48 kb)

     2 = Pst I digested control fragment (1.27/1.21 kb)

     3 = Homozygous positive (0.50 and 0.22 kb)

     4 = Heterozygous (0.72, 0.50, and 0.22 kb)

     5 = Homozygous negative (0.72 kb)

m   1   2   3   4   5

2.32►

1.35►

0.60►

◁⊐2.48

◁⊐1.27
◁⊐1.21

◀0.72

◀0.50

◀0.22
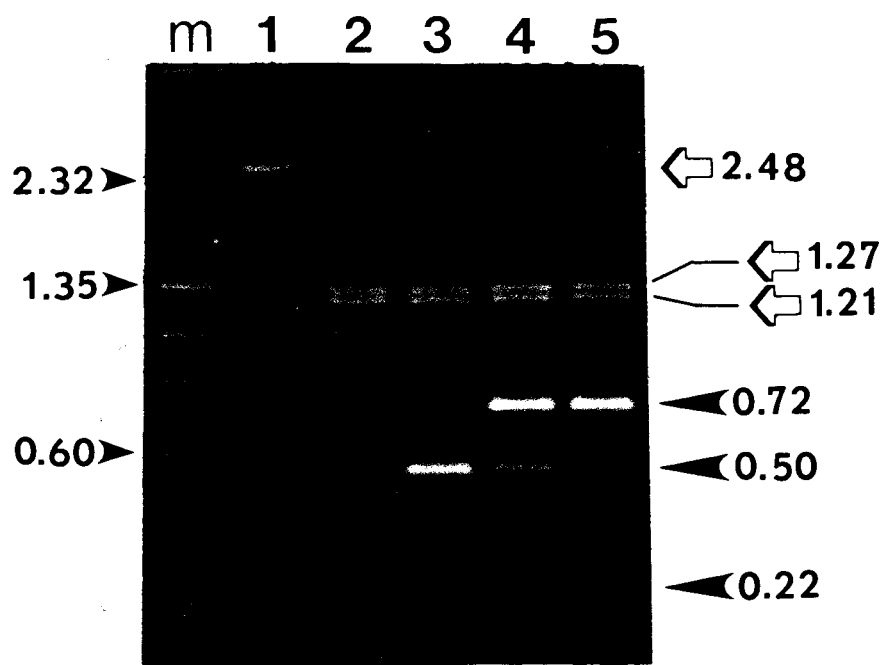
Table 13 : Calculated Sizes

LOCUS	ENZYME	SIZE OF AMPLIFIED REGION	SENSITIVITY TO RESTRICTION ENZYME		
			HOMOZYGOUS (1 band) -/-	HETEROZYGOUS (3 bands) -/+	HOMOZYGOUS (2 bands) +/+
APRT	Taq I	0.73 kb	0.73 kb	0.73 kb 0.48 kb 0.25 kb	0.48 kb 0.25 kb
PALB	Fnu 4HI	1.01 kb	1.01 kb	1.01 kb 0.65 kb 0.35 kb	0.65 kb 0.35 kb
ADA	Pst I	0.96 kb	0.96 kb	0.96 kb 0.59 kb 0.37 kb	0.59 kb 0.37 kb
CAH	Taq I	0.72 kb	0.72 kb	0.72 kb 0.50 kb 0.22 kb	0.50 kb 0.22 kb
LPL	Pvu II	1.03 kb	1.03 kb	1.03 kb 0.57 kb 0.46 kb	0.57 kb 0.46 kb

Figure 35 : Amplification Products of Five Loci

Human genomic DNA was amplified at the APRT, PALB, ADA, CAH and LPL loci.

Lanes M and m = different quantities of molecular length standards

1 = amplification products of APRT locus

2 =   "   "   " PALB "

3 =   "   "   " ADA "

4 =   "   "   " CAH "

5 =   "   "   " LPL "
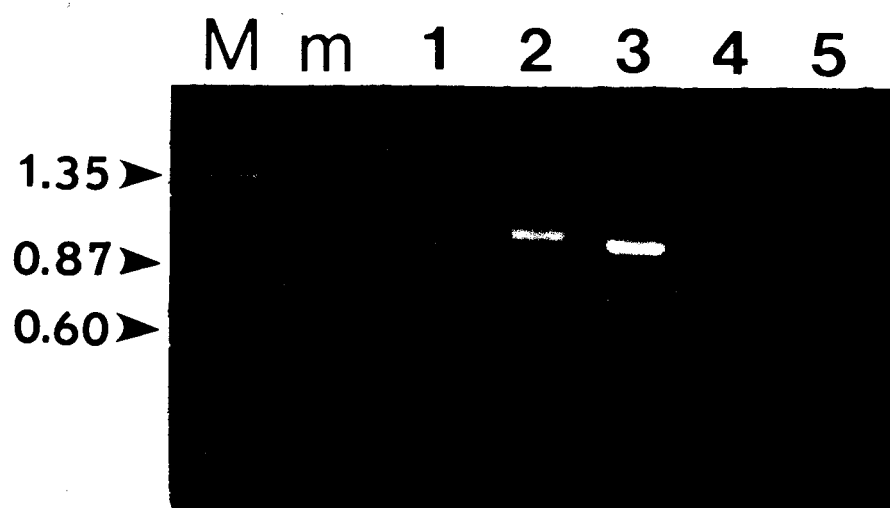
LPL (HLPLG18) DNA probes (Figure 36) which indicate that the appropriate regions of the genome have been amplified. The ADA (ADAN16) hybridized extremely well with the ADA amplification products (Figure 36 - lane 3) but the probe also hybridized weakly to the APRT amplification products (Figure 36 - lane 1). This shows that although the correct locus (ADA) has been amplified, the ADA genomic probe has a region that is partially homologous to the amplified APRT region.

It is of prime importance to include a DNA fragment of known size containing one specific restriction endonuclease site as an internal control to assess the activity of the restriction endonucleases (Figure 30-34). Addition of this control is necessary because the results obtained from an inactive or partially active restriction endonuclease might affect the subsequent zygosity assignment. Thus, by referring to the restriction fragments generated from the internal control DNA, anomalous restriction endonuclease activity can be detected. Accuracy of the results and appropriate interpretation of the data are thereby ensured.

Another important consideration is the fidelity of the Taq I DNA polymerase. It is estimated that the frequency of base misincorporation by the polymerase is approximately 1/400 [125]. A base substitution resulting from any misincorportion event will only be detectable if this event occurs during the first few cycles and if the initial copy number is low (less than 10) [139]. Furthermore, since these are random misincorporations, the amount of PCR products that contain a misincorporated base within dimorphic RES will be negligible (i.e. either the creation or elimination of the RES).

A final point to consider is the possibility that heterozygous individuals could be incorrectly categorized as homozygous negative individuals if the smaller bands were not visible as a result of low quanitities of PCR products. This would only occur if both of the two smaller bands (representing the digested fragment) were not visible. The presence of even one of the two fragments would infer the presence of the other small fragment. The fluorescence of the larger of the two small bands would be always at least 0.5 times as intense as the largest undigested

Figure 36 : Probing of Amplification Products

Lane 1 = amplification products using APRT primers

2 = "   "   " PALB   "

3 = "   "   "   ADA   "

4 = "   "   "   CAH   "

5 = "   "   "   LPL   "

Top left panel   : probed for the APRT locus

Top right panel   : probed for the LPL locus

Bottom left panel : probed for the ADA locus

Bottom right panel : probed for the CAH locus
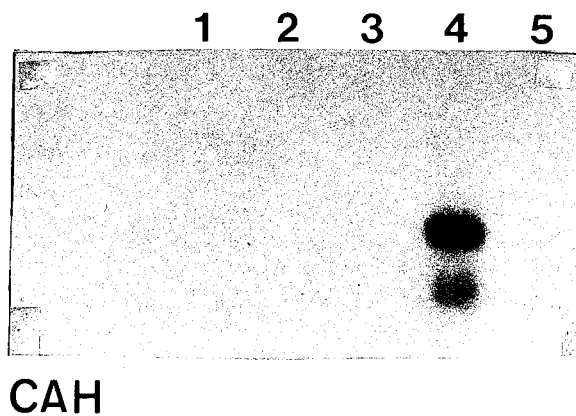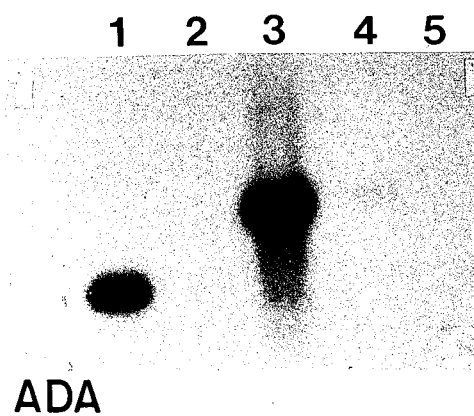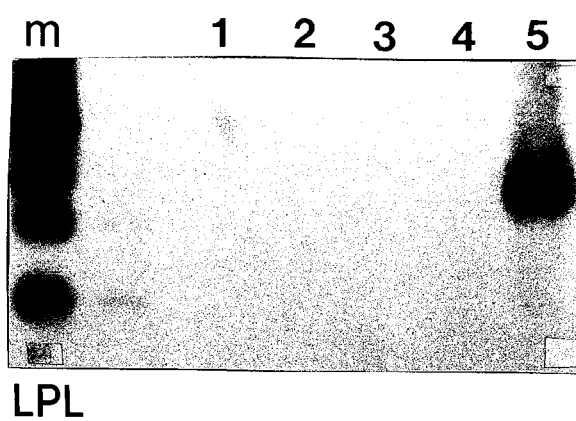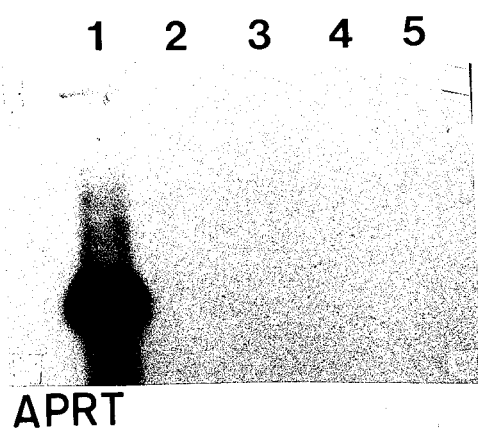
fragment. To ensure against such classification mistakes one may wish to consider southern blotting and hybridization techniques for checking low PCR product samples.

Allelic Frequencies of the Dimorphic Restriction Endonuclease Recognition Site at the APRT Loci:

The zygosity frequency distribution of a specific dimorphic restriction endonuclease recognition site (RES) in the second intron of the adenine phosphoribosyltransferase (APRT) locus was studied with genomic DNA samples that were obtained from an undefined population from the Greater Vancouver Area. The frequency values for the APRT locus were 75% Taq I site positive and 25% Taq I site negative (Table 14), and agreed with the previous report [140]. Chi square analysis [141] of the observed genotypes indicate that the dimorphic variation at the Taq I site of the APRT locus is in Hardy-Weinberg Equilibrium and that such frequency data should remain relatively stable in the absence of disturbing forces.

Allelic Frequencies of the Dimorphic Restriction Endonuclease Recognition Site at the LPL Loci:

The zygosity frequency distribution of a specific dimorphic restriction endonuclease recognition site (RES) in the sixth intron of the lipoprotein lipase (LPL) locus was studied with genomic DNA samples that were obtained from an undefined population from the Greater Vancouver Area. The frequency values for the LPL locus is 51% Pvu II site positive and 49% Pvu II site negative (Table 15). These values are not significantly different from those of Li et al. [142] who obtained frequency figures of 62% site positive and 38% site negative. The observed genotypes from this study indicate that the LPL locus for this population sample is in Hardy-Weinberg Equilibrium.

Allelic Frequencies of the Dimorphic Restriction Endonuclease Recognition Site at the PALB Locus:

The zygosity frequency distribution of a specific dimorphic RES in the third intron of the prealbumin (PALB) locus was studied with genomic DNA samples that were obtained from an

Table 14 : APRT Allelic Frequencies

LOCUS	ZYGOSITY				ALLELIC FREQUENCIES		Fitness to HWE	
	Number of people tested	Homozygous negative (-/-)	Heterozygous (-/+)	Homozygous positive (+/+)	RES negative	RES positive	χ^2	P (1 df)
APRT	173	12	63	98	0.25	0.75	0.22	>50%

$$\chi^2 = \text{"Goodness of fit" of data to null hypothesis} = \Sigma \frac{(obs_i - exp_i)^2}{exp_i}$$

where obs_i is the observed number of individuals with the "i" genotype
and exp_i is the expected number of individuals with the "i" genotype

P = the probability of obtaining the present results if the null hypothesis (H.W.E.) is correct

* $P < 5\%$ ⟶ reject hypothesis (locus not in equilibrium)

H.W.E. = Hardy-Weinberg Equilibrium   df = degrees of freedom

Calculations : (n = 173)

Phenotypes	+/+	+/-	-/-
Expected genotype frequency	0.562	0.375	0.062
Expected # of individuals	97.2	64.9	10.7
Observed # of individuals	98	63	12

$$\chi^2 = \left[\frac{(98 - 97.2)^2}{97.2}\right] + \left[\frac{(63 - 64.9)^2}{64.9}\right] + \left[\frac{(12 - 10.7)^2}{10.7}\right] = 0.2201$$

Table 15 : LPL Allelic Frequencies

LOCUS	ZYGOSITY				ALLELIC FREQUENCIES		Fitness to HWE	
	Number of people tested	Homozygous negative (-/-)	Heterozygous (-/+)	Homozygous positive (+/+)	RES negative	RES positive	χ^2	P (1 df)
LPL	152	40	69	43	0.49	0.51	1.29	>10%

$$\chi^2 = \text{"Goodness of fit" of data to null hypothesis} = \Sigma \; \frac{(obs_i - exp_i)^2}{exp_i}$$

where obs_i is the observed number of individuals with the "i" genotype
and exp_i is the expected number of individuals with the "i" genotype

P = the probability of obtaining the present results if the null hypothesis (H.W.E.) is correct

* P < 5% $\longrightarrow$ reject hypothesis (locus not in equilibrium)

H.W.E. = Hardy-Weinberg Equilibrium   df = degrees of freedom

Calculations : (n = 152)

Phenotypes	+/+	+/-	-/-
Expected genotype frequency	0.260	0.500	0.240
Expected # of individuals	39.5	76.0	36.5
Observed # of individuals	43	69	40

$$\chi^2 = \left[\frac{(43 - 39.5)^2}{39.5} \right] + \left[\frac{(69 - 76)^2}{76} \right] + \left[\frac{(40 - 36.5)^2}{36.5} \right] = 1.291$$

undefined population and a Oriental population of the Greater Vancouver Area. Individual

selected for the Oriental population were chosen on the basis of their surnames. Although it is

possible that some selected females were not Orientals, the proportion of such individuals should

be low so that the allele frequencies should not be significantly altered from their true frequencies.

Allelic frequencies obtained for a undefined population at the PALB locus were 66% Fnu 4H1 site

positive and 34% Fnu 4H1 site negative (Table 16), and chi square analysis of observed genotypes

demonstrated that the PALB locus is in Hardy-Weinberg Equilibrium. However, the allelic

frequencies obtained in this study were different from that reported by Yoshioka et al. [143]. The

samples analyzed by Yoshioka's report were from a Japanese population consisting of several

patients with familial amyloidotic polyneuropathy (FAP) and 10 "normal" people. The allele

frequencies of both Japanese subgroups were 50% for the presence or absence of the RES which

suggested that this variation was not responsible for FAP. The discrepancies between allelic

frequencies of this report and that of Yoshioka's, could be due to the differences in ethnic origins

of the selected sample population. As previously discussed in Chapter 1, it has been reported that

significant variations in the allelic frequency distribution of highly polymorphic loci occurred in

various ethnic groups, eg. American Blacks, Caucasians and Hispanics [87,88]. To test this

assumption, the sample population was altered such that Orientals were excluded from the

undefined sample and this group was re-categorized as the Non-Oriental group. The excluded

Orientals were then combined with another Oriental test sample and this new combined group was

designated the "Oriental group". The differences between the allelic frequencies of the two groups

are not significant at the 5% level. Chi square analysis of the genotypes also show that the PALB

locus for both groups are in Hardy-Weinberg Equilibrium (Table 17). Unless there are unique

differences within the Oriental subpopulation (i.e. Chinese, Japanese, Vietnamese) which can not

be determined here, it is unlikely that the observed discrepancies of the allelic frequencies of this

study and that of Yoshioka et al. are due to ethnic differences. A more likely explanation is the

Table 16 : PALB Allelic Frequencies

LOCUS	ZYGOSITY				ALLELIC FREQUENCIES		Fitness to HWE	
	Number of people tested	Homozygous negative (-/-)	Heterozygous (-/+)	Homozygous positive (+/+)	RES negative	RES positive	χ^2	P (1 df)
PALB	193	16	88	89	0.34	0.66	2.09	>10%

$$\chi^2 = \text{"Goodness of fit" of data to null hypothesis} = \Sigma \frac{(obs_i - exp_i)^2}{exp_i}$$

where obs_i is the observed number of individuals with the "i" genotype
and exp_i is the expected number of individuals with the "i" genotype

P = the probability of obtaining the present results if the null hypothesis (H.W.E.) is correct

* P < 5% ⟶ reject hypothesis (locus not in equilibrium)

H.W.E. = Hardy-Weinberg Equilibrium   df = degrees of freedom

Calculations : (n = 193)

Phenotypes	+/+	+/-	-/-
Expected genotype frequency	0.436	0.449	0.116
Expected # of individuals	84.1	86.6	22.3
Observed # of individuals	89	88	16

$$\chi^2 = \left[\frac{(89 - 84.1)^2}{84.1} \right] + \left[\frac{(88 - 86.6)^2}{86.6} \right] + \left[\frac{(16 - 22.3)^2}{22.3} \right] = 2.088$$

Table 17 : Subgroup Comparisons at the PALB Locus

PALB	ZYGOSITY				ALLELIC FREQUENCIES		FITNESS TO H.W.B.	
	Number of people tested	Homozygous negative (-/-)	Heterozygous (-/+)	Homozygous positive (+/+)	RES negative	RES positive	χ^2	P (1 df)
Undefined	193	16	88	89	0.34	0.66	2.09	>10%
Non-Orientals	180	16	83	81	0.35	0.65	1.98	>10%
Orientals	134	9	59	66	0.29	0.71	0.76	>10%

difference in the sample size of the two reports i.e. 193 individuals in this analysis versus < 20

individuals in the Japanese analysis.

Allelic Frequencies of the Dimorphic Restriction Endonuclease Recognition Site at the ADA

Locus:

The zygosity frequency distribution of a specific dimorphic RES in the second intron of the

adenosine deaminase (ADA) locus was studied with genomic DNA samples that were obtained

from an undefined population and a Oriental population from the Greater Vancouver Area. For

an undefined population, the allelic frequency values were found to be 73% Pst I site positive and

27% Pst I site negative (Table 18). These results are comparable to those obtained by Tzall et al.

who obtained frequency values of 67% and 33% for Pst I site positive and negative, respectively

[144]. Chi square analysis indicate that the allelic frequencies of the two studies are not significant

at the 5% level. However, in contrast with Tzall's study, data obtained for an undefined population

in this study demonstrated that the ADA locus is not in Hardy-Weinberg Equilibrium. Further

analysis was performed for comparison with an Oriental subgroup. The Non-Oriental and Oriental

groupings were organized as described for the PALB locus. The allelic frequencies of the Non-

Orientals and the Orientals are not significantly different, but the chi square values are noticeably

different (Table 19). The smaller the chi square value, the smaller the difference between the

observed and expected values. Although the data from the undefined and Non-Oriental groups

demonstrate non-equilibrium (P< 5%), the chi square value of the Oriental group (with P > 90%),

indicate that the ADA locus in Orientals is definitely in equilibrium as there is no difference

between the observed and expected values. The possibility exists that there are other subgroups

within the Non-Oriental group which have different allelic frequencies and by themselves, may

demonstrate equilibrium at the ADA locus but deviate from equlibrium when their frequencies are

pooled. The observed HWE in Tzall's study may be attributed to their population sample which

consisted of a more homogenous group i.e. 29 third generation Utah individual's and 16 individuals

of European extraction.

Table 18 : ADA Allelic Frequencies

LOCUS	ZYGOSITY				ALLELIC FREQUENCIES		Fitness to HWE	
	Number of people tested	Homozygous negative (-/-)	Heterozygous (-/+)	Homozygous positive (+/+)	RES negative	RES positive	χ^2	P (1 df)
ADA	175	7	80	88	0.27	0.73	4.70	* <5%

$$\chi^2 = \text{"Goodness of fit" of data to null hypothesis} = \Sigma \frac{(obs_i - exp_i)^2}{exp_i}$$

where obs_i is the observed number of individuals with the "i" genotype
and exp_i is the expected number of individuals with the "i" genotype

P = the probability of obtaining the present results if the null hypothesis (H.W.E.) is correct

* P < 5% ⟶ reject hypothesis (locus not in equilibrium)

H.W.E. = Hardy-Weinberg Equilibrium   df = degrees of freedom

Calculations : (n = 175)

Phenotypes	+/+	+/-	-/-
Expected genotype frequency	0.533	0.394	0.073
Expected # of individuals	92.8	68.8	12.8
Observed # of individuals	88	80	7

$$\chi^2 = \left[\frac{(88 - 92.8)^2}{92.8} \right] + \left[\frac{(80 - 68.8)^2}{68.8} \right] + \left[\frac{(7 - 12.8)^2}{12.8} \right] = 4.699$$

Table 19 : Subgroup Comparisons at the ADA Locus

ADA	ZYGOSITY				ALLELIC FREQUENCIES		FITNESS TO H.W.E.	
	Number of people tested	Homozygous negative (-/-)	Heterozygous (-/+)	Homozygous positive (+/+)	RES negative	RES positive	χ^2	P (1 df)
Undefined	175	7	80	88	0.27	0.73	4.65	<5%
Non-Orientals	161	6	75	80	0.27	0.73	5.30	<2.5%
Orientals	110	6	39	65	0.23	0.77	0.01	>90%

Allelic Frequencies of the Dimorphic Restriction Endonuclease Recognition Site at the CAH

Locus:

The zygosity frequency distribution of a specific dimorphic RES located one kilobase from

the 5' end of the carbonic anhydrase II (CAH) locus was studied with genomic DNA samples that

were obtained from an undefined population and a Oriental population of the Greater Vancouver

Area. For an undefined population, the CAH locus was found to be in Hardy-Weinberg

Equilibrium with allelic frequencies of 58% Taq I site positive and 42% Taq I site negative (Table

20). These frequency results are comparable to those obtained by Lee et al. who obtained values

of 50% for the presence or absence of the Taq I site [145]. Chi square analysis indicate that the

two data sets are not significant at the 5% level.

As with the PALB and ADA loci, comparison of allelic frequencies were made with an

Oriental sample group. The Non-Oriental and Oriental groupings were organized as described for

the PALB locus. Both groups demonstrate Hardy-Weinberg Equilibrium but their frequency values

are reversed (Table 21). These frequency differences of the two groups are highly significant with

a Chi square value corresponding to a $P < 0.005$. This suggests that there are subgroup differences

(at least between Orientals and Non-Orientals) at the CAH locus but these differences do not

disturb equilibrium. The number of Orientals in the undefined group is small (11 out of 93), hence

allele frequencies of the undefined group are identical to those of the Non-Oriental group.

Specificity of Oligonucleotide Primers for Human DNA :

Forensic specimens may be susceptible to contamination by a variety of agents, for

example : 1) yeast 2) bacteria and 3) body fluids of animals. Since such impure specimens may

produce inaccurate results, the specificity of the oligonucleotide primers for human DNA is of

utmost importance. Each set of primers were used in attempts to amplify DNA from 10 different

sources of possible contaminants. The PCR was allowed to proceed for 35 cycles, which exceeds

that for PCR with human DNA (31 cycles), to ensure that amplified products would not be missed

due to low levels of amplification. Amplification products of a size appropriate for each

Table 20 : CAH Allelic Frequencies

LOCUS	ZYGOSITY				ALLELIC FREQUENCIES		Fitness to HWE	
	Number of people tested	Homozygous negative (-/-)	Heterozygous (-/+)	Homozygous positive (+/+)	RES negative	RES positive	χ^2	P (1 df)
CAH	93	18	43	32	0.42	0.58	0.29	> 50%

$$\chi^2 = \text{"Goodness of fit" of data to null hypothesis} = \Sigma \frac{(obs_i - exp_i)^2}{exp_i}$$

where obs_i is the observed number of individuals with the "i" genotype
and exp_i is the expected number of individuals with the "i" genotype

P = the probability of obtaining the present results if the null hypothesis (H.W.E.) is correct

* P < 5% ⟶ reject hypothesis (locus not in equilibrium)

H.W.E. = Hardy-Weinberg Equilibrium   df = degrees of freedom

Calculations : (n = 93)

Phenotypes	+/+	+/-	-/-
Expected genotype frequency	0.336	0.487	0.176
Expected # of individuals	31.3	45.3	16.4
Observed # of individuals	32	43	18

$$\chi^2 = \left[\frac{(32 - 31.3)^2}{31.3} \right] + \left[\frac{(43 - 45.3)^2}{45.3} \right] + \left[\frac{(18 - 16.4)^2}{16.4} \right] = 0.288$$

Table 21 : Subgroup Comparisons at the CAH Locus

CAH	ZYGOSITY				ALLELIC FREQUENCIES		FITNESS TO H.W.E.	
	Number of people tested	Homozygous negative (-/-)	Heterozygous (-/+)	Homozygous positive (+/+)	RES negative	RES positive	χ^2	P (1 df)
Undefined	93	18	43	32	0.42	0.58	0.29	>50%
Non-Orientals	82	16	37	29	0.42	0.58	0.45	>50%
Orientals	99	33	50	16	* 0.59	* 0.41	0.18	>50%

Figure 37 : Amplification of Non-Human DNA

Amplification at the ADA(Panel A) and APRT (Panel B) loci was attempted with 200 ng

DNA from a variety of non-human sources.

Lane m = molecular length standards

   1 = human DNA

   2 = dog DNA

   3 = cat DNA

   4 = chicken DNA

   5 = mouse DNA (cell line 3T3)

   6 = C.albicans DNA (yeast)

   7 = pACYC (plasmid)

   8 = E.coli DNA (bacteria)

   9 = C.perfringens DNA (bacteria)

   10 = P.syringe DNA (bacteria)

   11 = B.subtilis (bacteria)

- amplification proceeded for 35 cycles

Figure 37 : Amplification of Non-Human DNA
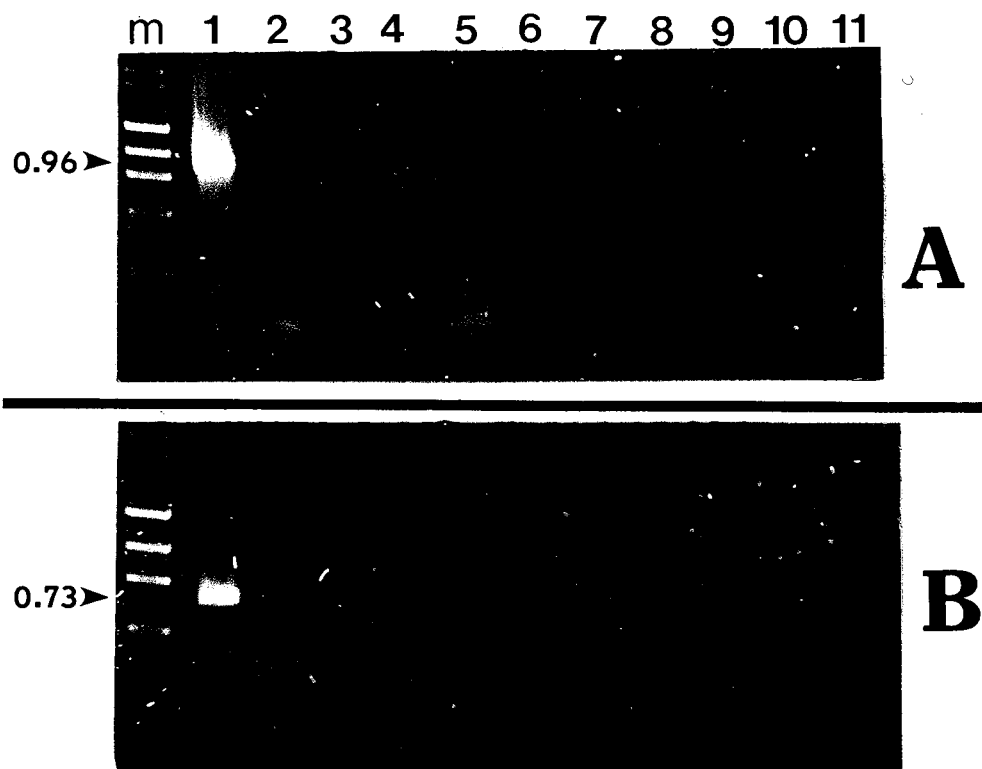
corresponding primer set (i.e. 0.96 kb for the ADA primers and 0.73 kb for the APRT primers)

were not observed for any of the reactions utilizing non-human target DNAs (Figure 37) (Data for

PALB, CAH, and LPL not shown). DNA which showed low levels of nonspecific amplification (i.e.

Figure 37 : Panel A-lanes 9 and 10, Panel B-lane 5) were reamplified with a 37 cycle reaction using

PALB primers. The absence of any observable DNA bands of the appropriate sizes even after 37

cycles confirmed the previous results with 35 cycles (data not shown).

To eliminate the possibility of inhibitors in the PCR which may have affected the activity of

the Taq DNA polymerase, reactions containing non-human DNA were "spiked" with human DNA.

Human DNA was successfully amplified in every case and reactions without spiked DNA were

devoid of amplified material (Figure 38). These results confirmed the specificity of the

oligonucleotide primers for human DNA.

Figure 38 : Amplification of Human DNA in Non-Human Samples

Human DNA was spiked into non-human DNA samples and were amplified at the PALB locus.

Lane   m = molecular length standards

       1 = human DNA

       2 = dog DNA

       3 = cat DNA

       4 = chicken DNA

       5 = mouse DNA (cell line 3T3)

       6 = dog DNA and human DNA

       7 = cat DNA and human DNA

       8 = chicken DNA and human DNA

       9 = mouse DNA and human DNA
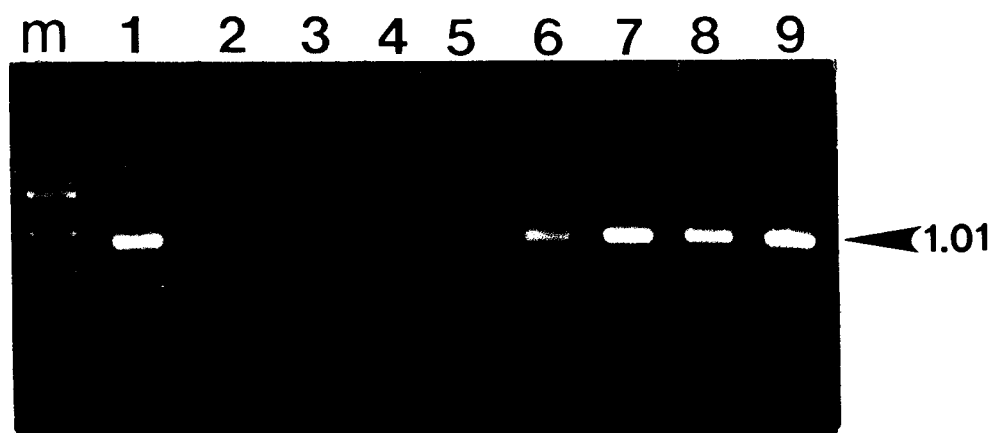
CONCLUSIONS

A DNA profiling system based on amplification of dimorphic restriction endonuclease recognition sites offers many advantages. Minimal population data is required for each loci since there are only two allelic types (the restriction endonuclease recognition site is either present or absent). Statistical problems with the Hardy-Weinberg test for equilibrium are absent as distinct from the complex statistical issues encountered when dealing with VNTRs. Interpretation of typing results is simple as the size of the DNA fragment amplified is predetermined by the flanking primers and the sizes of the digested fragments are also known. There are no requirements for time consuming manual or computerized assessments of DNA band sizes or need for binning systems.

Incorporation of an internal control DNA fragment in each digest reaction enables the analyst to monitor the activity of the restriction endonuclease which is critical to the whole dimorphic RES system. The validity and reliability of the system is also confirmed by the specificity of the oligonucleotide primers for human DNA at the designated (or appropriate) locus.

Data compiled from an undefined population from the Greater Vancouver Area demonstrated that four (APRT, LPL, PALB and CAH) of the five loci, with dimorphic RES, analyzed in this study are in Hardy-Weinberg Equilibrium. Since the ADA locus is found to be not in Hardy-Weinberg Equilibrium, this locus cannot be used in the context of forensic discrimination. Deviation from equilibrium suggests that the Hardy-Weinberg formula is not applicable for calculations of the probabilities of the occurence of specific genotypes in the population. Analysis of allelic frequencies at three loci (PALB, ADA, and CAH) of broadly categorized subpopulations (Orientals versus Non-Orientals), suggest that both subpopulations are in equilibrium at the PALB and CAH loci. Allelic frequency differences between subpopulations (even in the case of the ADA locus), are not statistically significant except for the CAH locus. Hence subpopulation data bases should be generated for each locus despite the presence or absence of Hardy-Weinberg Equilibrium.

Due to the possibility of name changes as a result of inter-racial marriages, categorization of racial origin by surnames may not always reflect the true racial orgins of an individual. However, allelic frequency values should not be significantly altered because the number of such name changes in the population study area is apparently very low and any frequency changes would be small compared to the high allelic frequency values of these dimorphic RES loci.

<u>SECTION II</u>

DISCRIMINATION POWER

The value of this dimorphic RES system may be determined by calculating the probability

of the system to discriminate between two samples. This ability or "discriminating power (DP) " is

defined as one minus the sum of the square of the phenotypic frequencies with the assumption that

the RES loci are independent [146]. The sum of the square of the phenotypic frequencies can be

defined as the probability of identity (PI) or the probability of randomly selecting two individuals

with the same phenotype (or genotype in this case) [141].

1) Hypothetical Examples : (Analyses of 16 unlinked loci)

I. Assuming that the frequencies of occurrence of the alleles at each of the 16 loci are 0.5

(enzyme site present) and 0.5 (enzyme site absent).

$$\text{For 1 locus : } DP = 1 - \Sigma(\text{frequency of phenotypes})^2$$

$$= 1 - (0.25)^2 + (0.50)^2 + (0.25)^2$$

$$= 1 - 0.375 = 0.625$$

$$\text{For 16 loci : } DP = 1 - (0.375)^{16} = 0.99999985$$

$$PI = \Sigma(\text{frequency of phenotypes})^2 = (0.375)^{16}$$

$$= 1.53 \times 10^{-7}$$

Hence there would be an one in 6.5 million $(1 / 1.53 \times 10^{-7})$ chance of randomly selecting

two individuals with the same phenotype (genotype) at these 16 loci.

II. Assuming that the frequencies of occurrence of the alleles at each of the 16 loci are 0.7

(enzyme site present) and 0.3 (enzyme site absent).

$$\text{For 1 locus : } DP = 1 - (\text{frequency of phenotypes})^2$$

$$= 1 - (0.49)^2 + (0.42)^2 + (0.09)^2$$

$$= 1 - 0.4246 = 0.575$$

For 16 loci : $DP = 1 - (0.4246)^{16} = 0.99999989$

$$PI = (0.4246)^{16} = 1.116 \times 10^{-6}$$

Hence there would be an one in a nine hundred thousand chance $(1 / 1.116 \times 10^{-6})$ of randomly selecting two individuals with the same phenotype (genotype) at these 16 loci.

2) Random DNA Profiling :

In order to test the discriminatory potential of the dimorphic RES system, two groups of 10 people were chosen at random (from the samples used to generate the allelic frequencies) and their DNA profiles were analyzed. In one group, 7/10 individuals had unique profiles (Table 22) and 8/10 individuals in the second group were distinguishable from one another (Table 23). This high level of discrimination was achieved using only 4 loci containing dimorphic RES. Two allele loci which have allelic frequencies of 50/50 will provide the highest probability of different alleles at the same locus of two different individuals. As only 1/4 loci used here (LPL) have allelic frequencies of 50-50, the feasibility of using loci with dimorphic RES is quite evident.

3) Family DNA Profiling :

Another important consideration was the ability of the proposed system to distinguish between close relatives. DNA profiles of individuals from two families were analyzed. Unique DNA profiles were obtained for all (6/6) individuals from one family (Table 24) but only 4/6 individuals were distinguishable in the second family (Table 25). For these loci, there are only two alleles occurring at high frequencies so that the probability of two people possessing the same allele is not too different regardless of whether they are related or not. However, once greater numbers of loci are used, discrimination between relatives would be reduced since there would be a greater chance of more alleles being identical by descent.

Table 22 : Allelic Profiles R1

Sample \ Locus	1	2	*3	4	5	*6	7	8	*9	10
APRT	+/+	+/+	+/+	+/+	+/+	+/+	+/-	+/-	+/+	-/-
PALB	+/-	+/+	+/-	-/-	+/-	+/-	+/-	+/-	+/-	+/+
CAH	+/+	-/-	+/-	+/+	-/-	+/-	+/-	+/-	+/-	+/-
LPL	+/+	+/-	-/-	-/-	-/-	-/-	-/-	+/+	-/-	-/-

* = non-unique profile for this group

Discrimination : 70 %

Table 23 : Allelic Profiles R2

Sample / Locus	1	2	*3	4	5	6	*7	8	9	10
APRT	+ / -	+ / +	+ / +	+ / -	+ / -	+ / +	+ / +	+ / +	+ / +	+ / +
PALB	+ / +	+ / -	+ / -	+ / -	+ / -	+ / +	+ / -	+ / +	+ / +	- / -
CAH	+ / +	+ / -	+ / +	+ / +	+ / -	+ / +	+ / +	+ / -	+ / -	+ / -
LPL	+ / +	+ / +	- / -	- / -	+ / +	+ / -	- / -	- / -	+ / +	+ / -

Discrimination: 80 %

Table 24 : Allelic Profiles F1

Sample / Locus	1	2	3	4	5	6
APRT	+/-	+/+	+/+	+/+	+/+	+/+
PALB	+/-	+/+	-/-	-/-	-/-	+/+
CAH	+/+	+/+	+/-	+/-	+/+	+/-
LPL	-/-	+/-	+/-	-/-	+/+	-/-

Discrimination: 100 %
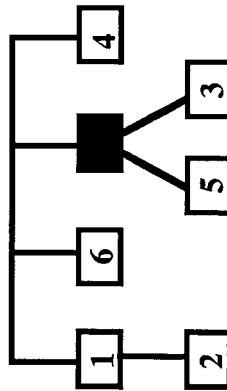
Table 25 : Allelic Profiles F2

Sample / Locus	*1	2	3	*4	5	6
APRT	+/-	+/-	+/-	+/-	+/+	-/-
PALB	+/-	+/-	+/-	+/-	-/-	+/-
CAH	+/-	+/+	+/+	+/-	+/-	+/-
LPL	-/-	+/-	-/-	-/-	+/-	-/-

Discrimination: 67 %

* = non-unique profile for this group
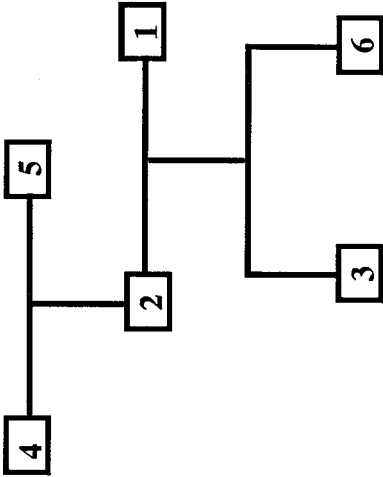
4) Discrimination Power With the Use of the Four RES Loci Studied Here :

$$\Sigma \text{ (phenotype frequencies)}^2 \text{ of APRT} = 0.4609$$

$$\text{LPL} = 0.3699$$

$$\text{PALB} = 0.4045$$

$$\text{CAH} = 0.3817$$

$$\text{DP} = 1 - \Sigma(\text{frequency of phenotypes})^2$$

$$= 1 - (0.4609) \times (0.3699) \times (0.4045) \times (0.3817)$$

$$= 1 - 0.0263 = 0.9737$$

$$\text{PI} = \Sigma \text{ (phenotype frequencies)}^2 = 0.0263$$

Hence there would be an one chance in thirty-eight (1 / 0.0263) of randomly selecting two individuals with the same phenotype (genotype) at these four loci.

Although discriminating power with only 4 loci is not very high compared with VNTR analysis, this can be increased with additional loci. Furthermore, the statistics for this system is much more reliable than for the VNTRs as there is no problem with differentiation of various alleles and their corresponding frequencies. The ADA locus was not used in the above examples because this locus was found to not be in Hardy-Weinberg Equilibrium.

CONCLUSIONS

The proposed dimorphic RES system is an extremely viable and attractive alternative to the VNTR system as a means for DNA profiling (genetic typing). Hypothetical calculations have shown that if 16 unlinked loci (each with allelic frequencies of 50% - 50%), were analyzed, only one person out of approximately 6.5 million would have a particular allelic pattern at the 16 loci tested (p170). Studies of DNA profiles from related and unrelated individuals demonstrated that a high level of discrimination is achievable with analysis of dimorphic RES at only 4 loci. The high numbers of loci required (10-20) should not be a deterrent to using this system with respect to either the availability of suitable loci or the amount of technical labour required.

Although extensive preliminary work is necessary to locate 10-20 loci, obtain sequence information, optimize PCR (for DNA from both blood and forensic specimens) and generate allelic frequency data bases, this obstacle can be removed by work sharing between several laboratories. Each group would be responsible for a less daunting number of loci. Once the system is complete, standard procedures could be easily established. Primers with compatible reaction temperatures could be put through the thermal cycler together (in different tubes) to improve time efficiency. By limiting the maximum size of the PCR products (prior to restriction endonuclease digestion) to no more than 2 kb, the separation of the DNA restriction fragments could be performed on polyacrylamide gels which allows for the possibility of automation.

In addition to the amount of labour involved, one must consider the feasibility of locating sequence information around 10-20 dimorphic RES. Since the human genome has been estimated to contain 1.4×10^7 heterozygous nucleotide sites due to neutral mutations [147], it should be possible to locate at least 20 unlinked loci with dimorphic RES. DNA sequence information for many loci should be available in the near future as a result of the Human Genome Project. The key factor is the determination of linkage equilibrium (random combinations of alleles at various loci) between 20 loci to ensure that the various loci are unlinked.

The five loci analyzed in this study are located on four different chromosomes to maximize the chance of linkage equilibrium between the five loci. This is based on the assumption that genes on different chromosomes assort independently during meiosis and are thus in linkage equilibrium [148]. However, disequilibrium between different loci may still occur for loci on different chromosomes [71]. Although simple methods are available to test for disequilibrium between as many as three loci [149-151], there is disagreement as to the best method for measurement of multi-loci associations [150].

Methods for analyses of multi-loci associations are complex and require in depth knowledge of statistical principles and extensive computer time [E. Wijsman - personal communication]. Since these requirements are beyond the capability and accessibility of this author, the five loci studied here were not tested for disequilibrium. This testing must be performed prior to application of this system to forensic casework in a service-orientated role.

GENERAL CONCLUSIONS

The application of DNA genetic typing to human identification is undoubtedly the highlight of the past decade in the field of forensic science. Studies of apparently random variations at the molecular level in DNA certainly amplifies the amount of information retrieved thus, facilitating greater discrimination among individuals. The distinctive power of hypervariable regions consisting of variable number of tandem repeats (VNTR), is evident by the large number of observed alleles or bins (29) despite the conservative organization of the binning system used in this study. However, this tremendous diversity of alleles which affords VNTR loci their high discriminating ability, also renders them susceptible to complex technical and statistical problems. Statistical problems in association with possible racial and geographical differences with respect to allele frequencies, demand that very conservative allele frequencies with high values to reduce chances of false conclusions of guilt, be employed in the calculation of uniqueness. The fixed bin approached (used in this thesis) was designed to minimize differences between subpopulatons (both ancestral and geographical) [106]. Hence with the use of conservative allele frequencies and the highly comparable results of data collected from defined and undefined populations suggest that population databases generated in one geographic location may be applicable to other geographically distinct populations.

The justified tendency towards conservatism leads to a reduction in the discriminative power gap between VNTRs and dimorphic restriction endonuclease recognition sites (RES). Although this gap is still large, the decrease in discrimination when moving from VNTRs to dimorphic RES, is not as great as it first appears. The disadvantage of the much higher frequency values of the dimorphic RES system is offset by its reliability and simplicity. Reliability of results is ensured as the amplified fragments and their digestion products are pre-determined so that typing errors are detected. The smaller number of possible alleles (three) makes interpretation and statistical analysis of results simple.

Allele frequency errors at one dimorphic RES locus would not dramatically alter the resultant profile uniqueness value (cf. errors at one VNTR locus). Since there are only two possible alleles, each occurring at high frequencies, the extent of the error relative to the large frequency value is small so that the level of inaccuracy is correspondingly small.

In contrast to VNTRs, Hardy-Weinberg Equilibrium can be easily tested for the dimorphic RES system. Four of the five dimorphic RES loci presented here have demonstrated Hardy-Weinberg Equilibrium with both the undefined group and the loosely defined racial subpopulations (Orientals and non-Orientals). However, there is evidence of small allelic frequency differences in racial subpopulations at the carbonic anhydrase II (CAH) locus.

The potential for generating characteristic DNA profiles by the analysis of a combination of 10 to 20 dimorphic RES loci, is solidly confirmed by the demonstrated high level discrimination of groups of both related and unrelated individuals by testing with only 4 dimorphic RES loci. The volume of work required for the analysis of 10-20 loci is greatly reduced with the aid of the polymerase chain reaction. Protocols for PCR can be developed and organised to handle this large number of analyses. In comparison with conventional VNTR (blot-hybridization) analysis, the DNA of amplified digestion products of dimorphic loci can be directly visualized on the gel without the use of probes. Time spent on the analysis of banding patterns is minimal in contrast to that of the VNTR system.

Use of the polymerase chain reaction technique in forensic science solves the important problem of sample size by increasing the quantity of DNA available for analysis . With this proposed system, reliable results could be obtained from degraded DNA since only the RES and a small portion of the flanking regions need to be intact. In situations where there is not enough DNA for separate amplification of several loci, two or more different primer sets could be used in the same reaction vial to amplify various regions simultaneously ("multiplex" amplification). Aliquots of the amplified products could serve as templates for secondary amplifications using only one specific primer set selected from the primary amplification. "Multiplex" amplification with 6 to

9 different primer sets have already been reported for the detection of Duchenne muscular dystrophy [152,153]. Jeffreys et al. have also simultaneously amplified 6 single locus minisatellite regions [154].

Preliminary environmental studies which are discussed in the addendum, indicate that forensic specimens can be successfully analyzed with both the VNTR (blot-hybridization) and the dimorphic RES (PCR) systems. Factors which affected the polymerase chain reaction amplification, with the exception of UV radiation, also affect the VNTR system. Samples exposed to the external environment are more likely to be successfully typed by the dimorphic RES system. Degraded DNA samples may be typed by this RES system because large sizes of intact DNA fragments are not required since the target fragments to be amplified are designed to be less than 1.5 kb. The results obtained from this study do not provide any definitive conclusion as to the best system for application to forensic specimens. The results of the dimorphic RES study are encouraging, however extensive research and development is still required for both systems.

The protocol described here for the proposed dimorphic RES system, provides a simple practical solution for individual identification in forensic science. Initial calculations showed that if 16 such unlinked loci (with allele frequencies of 70% -30%) were tested, only one individual out of nine hundred thousand individuals would have a particular allelic pattern at the 16 loci tested. This high discriminating value can be further enhanced by utilizing loci with allele frequencies closer to 50% - 50% and/or analyzing additional loci. Although, recent advancements in the amplification of VNTR loci (AMP-FLP) may solve many of the problems with VNTR analysis such as higher resolution leading to discrete alleles and less complicated statistics [155], the dimorphic RES system still appears to be a promising alternative to VNTR analysis. Since data to establish identity are frequently presented in court to non-scientists by non-scientist eg. lawyers, judges, and juries, the technique herein described for dimorphic RES allows the data to be presented in a simple and direct format that can be easily understood.

Recent criminal cases have highlighted the difficulties in the presentation of DNA evidence in court. Attempts have been made to discredit DNA evidence by making the technology appear complex, confusing, and unreliable. This dimorphic RES system makes possible the presentation of DNA evidence in court in a simple, clear and precise way and should help remove these doubts and difficulties that could inhibit valuable information necessary to convict or exonerate a defendant.

Although this thesis has focussed exclusively on the use of DNA profiling in a forensic context, restriction fragment length polymorphisms (RFLPs) has other practical applications. For example, DNA analysis can be applied to : 1) determination of twin zygosity [156], 2) human paternity disputes (in cases of both palimony and immigration issues) [157,158,53], 3) identification of domestic animals such as cattle and horses for breeding purposes and theft prevention, 4) prevention of theft and poaching of wild animals, 5) enforcement of fishing regulations with respect to specific species of fish [159], 6) selective fish breeding [160], 7) evolutionary studies of humans and animals, 8) screening of tissue compatibility for organ transplantations 9) assessement of clonality of tumors [161], and 10) epidemiology studies of microbial outbreaks. These numerous applications of DNA profiling through RFLP analysis necessitates and ensures that continuous efforts will be made towards the development and refinement of this discriminating procedure well into the 21st century.

SUMMARY CONCLUSIONS

1. VNTR allele frequency data bases collected by sampling an undefined population of the Greater Vancouver Area are comparable to those collected by other laboratories sampling Caucasian populations.

2. Hypervariable regions in DNA consisting of variable number of tandem repeats exhibit tremendous diversity and hence has the potential to provide high levels of discrimination suitable for forensic profiling.

3. A binning system is required to deal with the problems of a continuous distribution of VNTR alleles, inadequate gel resolution of closely spaced alleles and inter/intra gel variabilities.

4. PCR amplification of DNA sequences containing dimorphic restriction endonuclease recognition sites (RES) can be used to eliminate the use of DNA probes for detection of restriction fragment length polymorphisms (RFLPs).

5. Dimorphic RES allele frequency data bases collected by sampling an undefined population of the Greater Vancouver Area are comparable to those collected by other laboratories.

6. Hardy-Weinberg Equilibrium was demonstrated for the 4/5 dimorphic RES loci analyzed in this study, sampling either an undefined population (total) or two racial subpopulations (Orientals and non-Orientals).

7. Dimorphic RES allele frequency comparison of two racial subpopulations (Orientals versus non-Orientals) indicated that there are no significant differences at the PALB and ADA loci.

8. DNA profiling with only 4 dimorphic RES offers significant discrimination.

9. Both the VNTR and dimorphic RES systems can be used with varying degrees of success to obtain DNA profiles from forensic casework specimens and/or laboratory simulated forensic samples.

10. Although both DNA profiling systems described in this study still require extensive research and development, the dimorphic RES system offers a much simpler and reliable approach to DNA profiling.

ADDENDUM

Influences of Environmental Conditions

On DNA Profiling of Forensic Specimens

INTRODUCTION

The ultimate criteria of any system developed for forensic purposes, is that it be functional, reliable, and reproducible when applied to forensic specimens. As discussed in chapter 1 (section II), there has been extensive research into the types of specimens amenable to RFLP analysis. Correspondingly, studies have also been made to determine the types of forensic specimens suitable for amplification by the polymerase chain reaction.

Due to degradation of ancient DNA, amplification efficiency of aged samples was found to be inversely related to the size of the fragment to be amplified [162]. Fragments with maximum lengths of 100-200 bp have been successfully amplified from mitochondrial DNA extracted from a 7000 year old brain [163,164]. Amplification was found to be more efficient in partially degraded samples than in undegraded samples [165]. This effect was attributed to more efficient strand separation in samples that are partially degraded. Amplification of degraded DNA samples is possible as long as there are DNA fragments containing the entire target sequence [165].

The FBI has utilized the polymerase chain reaction to amplify a 242 bp region of the major histocompatibility complex (DQ-α), from DNA isolated from samples such as aged blood stains (1-9 months old), 1 ul of semen stains, single hair roots and vaginal swabs containing semen [166]. PCR amplification of the DQα locus from DNA isolated from either shed or plucked hairs has been reported, with more reliable results from plucked hairs [167]. DNA from two forensic cases involving aged specimens (a > 10 year old semen stain and skull cap reported to be 18 months old) have also been shown to be amenable to PCR analysis [168]. Recently, Walsh et. al. has published protocols for the extraction of DNA, suitable for PCR analysis, from a variety of forensic materials [169].

In addition to types of specimens, one needs to determine environmental factors that influence the success and validity of either the VNTR or the polymerase chain reaction systems. Several studies on the effects of ultraviolet light, humidity, soil, on bloodstains have already been

done [170-172]. The follow section serves as preliminary studies of environmental influences on

PCR dimorphic RES analysis and makes comparisons with that of the VNTR (blot-hybridization)

system.

METHODS

1. PREPARATION OF DNA

Preparation of Bloodstains :

One hundred microliter aliquots of whole blood were pipetted onto a piece of 100% cotton

cloth (washed and dried three times) and dried at room temperature. The dried bloodstains were

then treated as follows :

a) kept at room temperature without light (in a lab drawer) for various

time periods up to 11 months

b) hung on the external side of an east facing window (at 10th and

Heather Street) in Vancouver, for various recorded time intervals

during the period between October 17/90 and December 6/90

(approximately 7 week period)

* c) irradiated with varying amounts (0 - 21,600 J/M^2) of short wave

ultra violet light using a Stratagene Linker

d) partially buried for up to 10 days in soil taken from an alley in

Vancouver

* - The amount of UV irradiation was converted to an equivalent period of exposure to sunlight.

The conversion factor used for this calculation was obtained by using a UV light meter (UVX-25

Sensor) to measure the quantity of short wave (254 nm) UV irradiation from the sun. Readings

were made at 12:00 noon on three consecutive sunny days in mid January 1991 and the highest

value was used in the calculation.

Isolation of DNA From Bloodstains For VNTR Analysis (Protocol D) :

Deoxyribonucleic acid was extracted from each blood stain as described in a protocol used

by the FBI. The blood stain was cut into approximately 2 mm square pieces, placed into a 1.5 ml

microcentrifuge tube containing 400 ul stain extraction buffer and 10 ul of Proteinase K (20 mg/ml) and incubated overnight in a 56^oC water bath. The liquid from the tube was then removed and saved in another tube. Using a syringe needle head, a hole was made in the bottom of the original tube which, at this point, only contained the damp pieces of stain. The tube with the hole and pieces of stain were placed on top of a new tube and spun for 2 minutes at 13,000 xg to draw out the remaining fluid from the damp pieces of stain. The fluid collected in the new tube was pooled with the fluid saved previously. This fluid was then extracted once with 500 ul of phenol/chloroform/isoamyl alcohol (25:24:1) and once with 500 ul of chloroform/isoamyl alcohol (24:1). The DNA in the fluid was precipitated with 1.0 ml of cold absolute ethanol at -20^oC for at least 30 minutes. The precipitated DNA was collected by centrifugation at 13,000 xg for 15 minutes at -20^oC and washed with 70% ethanol. The resulting DNA pellet was either air dried or dried in a vacuum desiccator and resolubilized overnight in 40 ul TE^{-4} in a 56^oC water bath. The entire sample was used for one VNTR typing analysis.

Purification of DNA Extracted From Stains For VNTR Analysis :

The resolubilized DNA was placed on a Millipore disc filter (0.05 uM) and dialysed for two hours in LTE.

Isolation of DNA From Liquid Blood or Bloodstains With Chelex 100 For PCR Analysis (Protocol E) :

Deoxyribonucleic acid was extracted according to a protocol used by the FBI. Briefly, in a 1.5 ml microcentrifuge tube containing 1.0 ml of dH_2O, either a volume (50-100 ul) of liquid blood or a cut up piece of cotton blood stain (50-100 ul blood) was added. The tube with its contents was incubated at room temperature for 30 minutes and then spun in a microcentrifuge for 2-3 minutes at 13,000 xg. Without disturbing the pellet, as much as possible of the supernatant was discarded and the pellet (including the pieces of fabric in the case of the bloodstain) was made up to a final volume of 400 ul with 5% (w/v) Chelex 100 (Biorad). This mixture was incubated for 30

minutes in a 56°C water bath (Blue M) and then vortexed at high speed for 10 seconds before

boiling in a water bath for 8 minutes. The boiled mixture was once again vortexed at high speed

for 10 seconds and the supernatant (containing the DNA) was then collected by centrifugation at

13,000 xg for 2-3 minutes.

Purification of Chelex 100 Extracted DNA :

The supernatant containing the extracted DNA was purified with GENECLEAN (Bio 101

Inc.) according to the kit protocol. Approximately 2.5 volumes of NaI solution was added to the

DNA solution (supernatant) and mixed before the addition of 8 ul of glass milk (suspension of

silica matrix in water that binds single and double stranded DNA without binding DNA

contaminants). This mixture was incubated for 5 minutes in an ice bath to promote the binding of

the DNA molecules to the silica matrix. Following this incubation, the mixture was spun for 5

seconds at 13,000 xg and the supernatant containing the unwanted contaminants, was discarded.

The silica matrix pellet (with bound DNA) was washed three times with ice cold NEW WASH

buffer (a mixture of Tris-acid and Tris-base with a pH of 7.0-8.5), the reagents of which were

provided with the GENECLEAN kit. After the third wash, the DNA was eluted from the silica

matrix with dH_2O and incubated for 3 minutes at 50°C. The suspension was spun for 30 seconds

at 13,000 xg and the supernatant containing released DNA (80% of total) was collected. The

pellet was eluted once more to recover residual DNA (10-20%). The final total volume of the

extracted stain DNA was 300 ul.

2. PCR AMPLIFICATION OF DNA EXTRACTED FROM BLOODSTAINS

PCR For the APRT, PALB, and ADA Loci From DNA Extracted From Bloodstains :

Fifteen microliters (1/20) of Geneclean purified Chelex 100 extracted DNA from one bloodstain

was used for each amplification reaction. The polymerase chain reaction was carried out as

described previously for each loci except for changes in the $MgCl_2$ concentrations and the number

of cycles completed. The MgCl$_2$ concentrations were increased for each loci so that it was 2.0 mM for the APRT locus, 2.5 mM for the PALB locus, and 3.3 mM for the ADA locus. Each PCR was allowed to continue for 33 cycles.

Control For PCR Amplification :

As with the data base experiments a positive control (stock DNA sample with reliable amplification) was included in each PCR run. The absence of negative controls does not affect the recognition of amplification activity, the basic purpose of these experiments. However, negative controls should be included in actual forensic case analysis.

3. VNTR ANALYSIS OF DNA EXTRACTED FROM BLOODSTAINS

Refer to chapter 1 - methods section (VNTR analysis)

RESULTS AND DISCUSSION

Technical Conditions of DNA Extraction For PCR Analysis of Forensic

Samples :

The ability to amplify DNA isolated from forensic samples is a vital factor in assessing the

feasibility of employing this dimorphic RES system for forensic identification. At the onset of this

project, PCR amplification of DNA from forensic samples was at its infancy. Therefore, despite

the success of other laboratories, it was important to demonstrate that amplification of forensic

specimens was also easily achievable and reproducible in this study with these specifically chosen

dimorphic loci.

Initially, in the absence of other more suitable methods known to this laboratory, DNA was

extracted from stains using protocol D and used in the polymerase chain reaction (PCR).

Unfortunately, amplification of these DNA samples was not successful, possibly due to the presence

of contaminants which inhibit the activity of the Taq DNA polymerase. The contaminant suspected

of causing this inhibition was sodium dodecyl sulfate (SDS) used in the extraction procedure. The

presence of low levels (0.01%) of SDS has been reported to be inhibitory for Taq Polymerase I and

the addition of 0.1% of Tween 20 and 0.1% of NP40 completely reverses this inhibition [173].

However, the addition of Tween 20/NP40 still did not result in amplification of the stain extracted

DNA. The strategy was then to switch the extraction method to one which produces purified

DNA.

The next DNA extraction procedure attempt was that of the BRL DNA Capture Reagent

protocol. The DNA Capture Reagent (modified Sepharose beads) immobilizes DNA from

buffered DNA solutions or complex biological samples leading to its purification from undesired

materials by filtration or centrifugation. In product developmental studies, 80-100% of input DNA

(in various buffer solutions) was immobilized by the Capture Reagent and 90% of this bound DNA

was subsequently recovered [174]. PCR amplification of stain DNA extracted with this BRL

protocol was not successful. Amplification was also not seen with a DNA solution recovered from

the DNA capture of 300 ng of DNA (in LTE). Since this "spiked" DNA (extracted from whole blood using protocol A) had previously been shown to be amplifiable, it appears that in this situation, the DNA Capture Reagent protocol is not as efficient as previously reported.

Amplification of stain DNA was finally achieved with DNA extracted using Protocol E obtained from the FBI and which has since been published by Walsh et al. [169]. This protocol does not require detergents but revolves around the use of high temperatures (100^oC) in the presence of an alkaline chelating ion exchange resin, Chelex 100 [175]. The alkalinity of Chelex 100 (pH 10.8 at room temperature) not only facilitates the lysis of the cells but also assures complete DNA denaturation. Furthermore, this resin has a high affinity for polyvalent metal ions and its presence during boiling of stain suspension is thought to prevent DNA degradation by chelating metals ions that serve as catalysts for such degradation.

In order to simulate possible environmental conditions under which forensic specimens exist, bloodstains were exposed to the external elements for various time intervals and attempts were made to amplify DNA extracted from such samples with three of the five dimorphic RES primer sets. Only 1/20 of the purified DNA extracts were used because larger proportions resulted in lower amounts of amplified products. Low concentrations of red blood cell components such as haemoglobin, hematin, protoporphyrin IX, $FeCl_{13}$, and $FeSO_4$ have been demonstrated to inhibit PCR [176]. Inhibition by hematin occurs at a level lower (0.8 uM) than the other components (20 uM), which is significant since this component is found in high levels in aged bloodstains and co-purifies with DNA. It is suggested that hematin is a competitive inhibitor of Taq DNA polymerase.

Technical Conditions of DNA Extraction For VNTR Analysis of Forensic Samples : Initially, protocol C (p27) was used for the isolation of DNA from bloodstains, blood swabs, and cells attached to hair shafts. The change to protocol D coincided with the change from protocol A to B for DNA extraction from whole blood (chapter 1). Protocols B and D were obtained together from the FBI laboratory and it became apparent that these protocols were more suitable for forensic VNTR analysis. For example, with protocol C, forensic specimens are initially incubated overnight

at 37°C and in protocol D, the temperature of incubation is 56°C. Cellular nucleases released during cell lysis or external nucleases (contaminants of the specimen) may still be active at 37°C and not at 56°C. Hence, incubation at the higher temperature would inhibit possible DNA degradation. To allow for comparisons between the two profiling systems, simulated forensic bloodstains identical to those analyzed by PCR were also typed by probing of 4 VNTR loci. In order to save time, two probe combinations (i.e. 3'HVR α-globin/YNH24 and phins 310/CMM101), were used per hybridization reaction to probe simultaneously at two loci.

Effects of Aging On DNA Genetic Typing of Bloodstains :

Deoxyribonucleic acids extracted from bloodstains stored in a laboratory drawer for up to 11 months were successfully amplified by the polymerase chain reaction (Figure 39). VNTR (blot-hybridization) band patterns were also obtained from such stains up to 8 months old (Figure 40). (Older stains for VNTR analysis were unavailable.) The degree of amplification or intensity of DNA bands was decreased with increased age of bloodstains. This suggests that the amount of DNA fragments smaller than 1 kb increases with increased age of sample. A value of 1 kb is suggested since that is the approximate length of the target fragment amplified with these dimorphic RES loci and the entire fragment must be intact if amplification is to occur.

However, DNA degradation may not be a sole cause of decreased levels of amplification or band intensity (in the case of VNTR blot-hybridization system) in aged stains. Although Pääbo et.al. found, in ancient tissue samples, an inverse relationship between age and amplification efficiency [163], the amount of DNA degradation did not correlate with age of specimen [164]. High levels of modified thymine in ancient samples, suggest that significant amounts of oxidative damage had occurred [164]. Hence, DNA in dehydrated tissues may be protected from hydrolytic damage and remain relatively intact, but it is still susceptible to oxidative damage.

Effects of Environmental Factors Particular to Vancouver, on DNA Genetic Typing of Bloodstains

:   Possible causes for DNA degradation of forensic samples exposed to the external

Figure 39 : Amplification of DNA From Aged Stains

DNA extracted from human bloodstains kept at room temperature without light for various time periods were amplified at the ADA locus.

Lane m = molecular length standards

1 = lab stock DNA from blood sample

2 = DNA from 2 day old stain

3 = DNA from 3 month old stain

4 = DNA from 5 month old stain

5 = DNA from 11 month old stain

6 = negative control (no DNA)
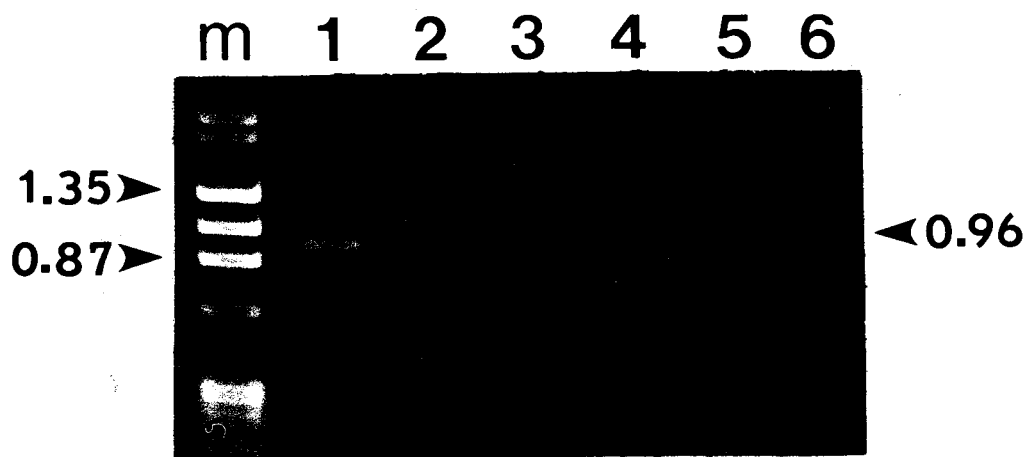
Figure 40 : VNTR Analysis of DNA from Aged Stains

DNA extracted from human bloodstains kept at room temperature without light for various

time periods were simultaneously probed with 3'HVR α-globin and YNH24 to detect alleles

at the D16S85 and D2S44 loci.

Lane   M = molecular length standards

   1 = DNA from 2 day old stain

   2 = DNA from 5 month old stain

   3 = DNA from 8 month old stain

Long arrows indicate position of alleles for the two loci.
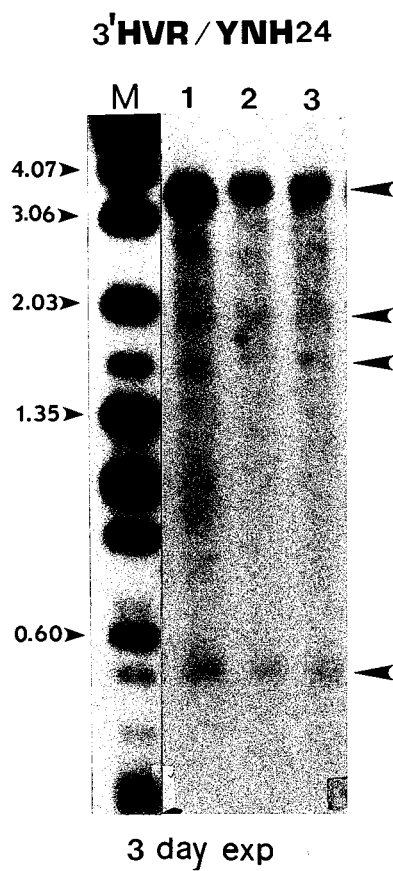

3'HVR / YNH24

3 day exp

environment would be : temperature, humidity, air pollutants, ultraviolet light, and soil. Hence bloodstains were subjected to these various conditions prior to analysis by either the VNTR (blot-hybridization) or dimorphic RES (PCR) system.

During the period between October 17 (1990) and December 6 (1990), the amount of precipitation (as measured at the Vancouver International Airport), was slightly higher than normal for this time of the year [177]. Daily temperatures were also slightly higher than normal with a mean temperature of not greater than 9.9^{o}C for any of the 3 months. However the total hours of sunshine during October to December, was 57 hours (25%) less than normal. Humidity was high with relative humidity of not less than 87% during the entire 7 week period.

Deoxyribonucleic acids were extracted from bloodstains hung for up to 7 weeks (during October to December) on the external side of an east facing window (at 10th and Heather St.) in Vancouver. All samples were successfully amplified by the polymerase chain reaction. Undigested amplified products of APRT, PALB, and ADA from blood DNA (left panel), 4 hour old stain (middle panel) and 7 week old stain (right panel), are shown in Figure 41. Digestion products of PALB amplified samples are shown in Figure 42. Not surprisingly, the degree of amplification decreased with age of stain (Figure 42). This conclusion appears to be in contradiction with the results seen for DNA from a bloodstain only 4 hour old (Figure 42 - lane 1). Previous amplifications of DNA aliquots from the same sample demonstrated amplification levels comparable to that of DNA from stains 1 day old. The low level observed in this one reaction is the result of inconsistent amplification. This explanation is confirmed by variable levels of amplification of DNA aliquots from identical samples (Figure 43 - lanes 1/A, 2/B, 3/C and 4/D). Amplification inconsistencies may be a consequence of minute differences in the amount of starting DNA which leads to great differences as a result of exponential accumulation of DNA. Another possibility may be the failure of amplification during the first few of cycles such that amplification does not actually begin until cycle 2 or 3, resulting in a 2 or 4 fold decrease in the amount of product.

Figure 41 : Amplification of DNA From Bloodstain Exposed to External
Environment (APRT/PALB/ADA)

DNA extracted from liquid blood (Left panel), 4 hour old bloodstain middle panel) and

7 week old external bloodstain (right panel) were amplified at the APRT, PALB and ADA

loci.

Lane   m = molecular length standards

         1 = APRT locus (0.73 kb)

         2 = PALB locus (1.01 kb)

         3 = ADA locus (0.96 kb)

Figure 42 : Amplification of DNA From Bloodstain Exposed to
External Environment (PALB)

DNA extracted from human bloodstains either kept at room temperature without light

(internal) or exposed to the external environment for various time periods, were

amplified at the PALB loci. The amplification products were digested with Fnu 4H1

which resulted in digestion products of 0.65 and 0.36 (represented by long arrows).

Lane m = molecular length standards

L = lab stock DNA from blood

B = DNA from blood extracted as per stains

1 = DNA from 4 hour old stain (internal)

2 = DNA from 1 day old stain (internal)
3 = " " " " " "   (external - environment)

4 = DNA from 1 week old stain (internal)
5 = " " " " " "   (external)

6 = DNA from 2 week old stain (internal)
7 = " " " " " "   (external)

8 = DNA from 3 week old stain (internal)
9 = " " " " " "   (external)

10 = DNA from 4 week old stain (internal)
11 = " " " " " "   (external)

12 = DNA from 5 week old stain (internal)
13 = " " " " " "   (external)

14 = DNA from 6 week old stain (internal)
15 = " " " " " "   (external)

16 = DNA from 7 week old stain (internal)
17 = " " " " " "   (external)

Figure 42 : Amplification After Exposure to External Environment (PALB)
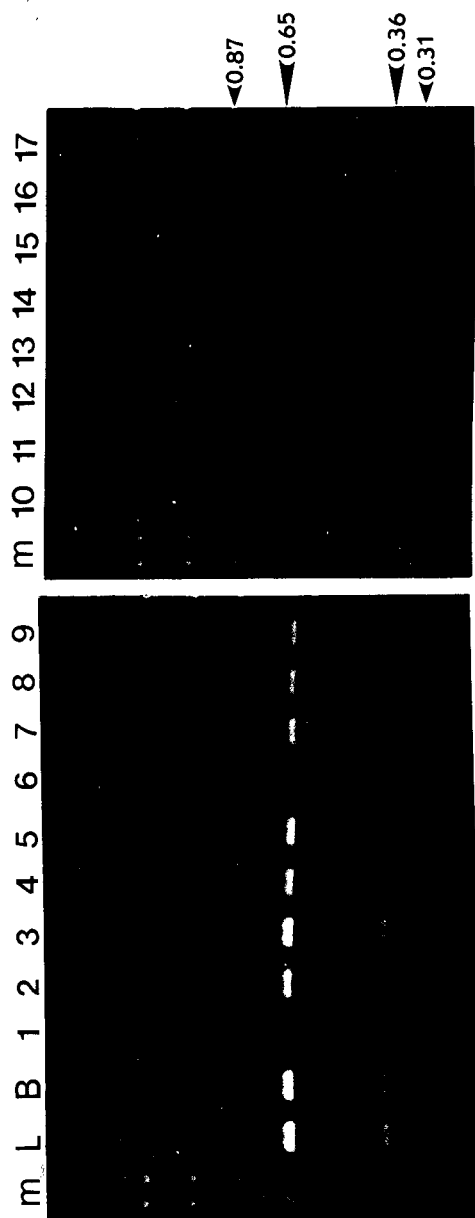
Figure 43 : Inconsistent Amplification

Identical samples of DNA were amplified at the PALB locus.
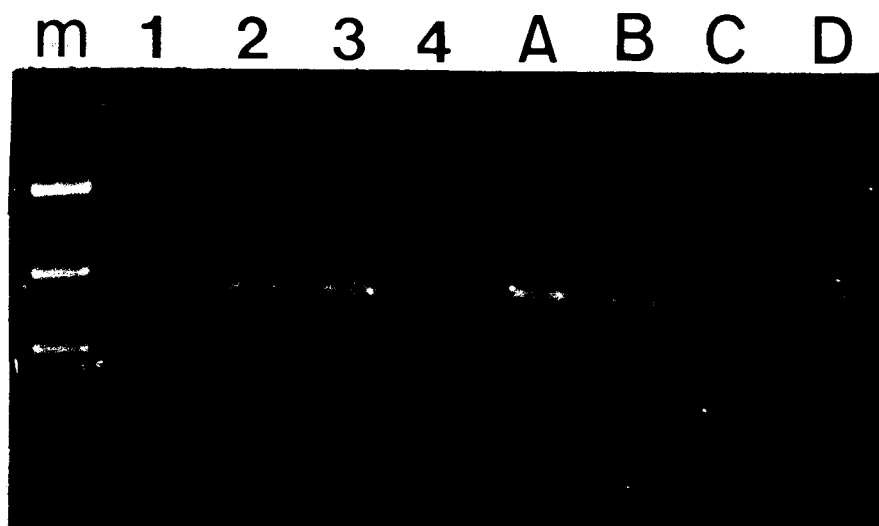
Lane   m = molecular length standards

Lane   1 and A = lab stock DNA extracted from blood

       2 and B = blood DNA extracted per stain

       3 and C = 4 hour old stains (internal)

       4 and D = 1 day old stains (internal)

Comparisons of PCR products of DNA from stains kept in a laboratory drawer (control sample) and stains exposed to the external environment (test sample) for a corresponding period of time, indicated no consistent difference in the level of amplification (Figure 42). This suggests that the polymerase chain reaction is not affected by the environmental factors prevalent during that time period, but are influenced by the effects of aging.

In contrast, observable differences were found with VNTR analysis of these two types of stains. A seven week old stain stored in a laboratory drawer (Figure 44 - lane 5) produced one band with 3'HVR α-globin, no bands with YNH24 or phins 310 and two bands with CMM101 (results with phins310/CMM101 not shown). A bloodstain exposed to environmental conditions for 7 weeks (Figure 44 - lane 6), failed to produce a pattern with any of the four probes even after prolonged exposure of the autoradiograph. The 7 week old control and test stains (Figure 44 - lanes 5 and 6) compare poorly with the more fresh stain seen in Figure 44 - lane 1. This may be due to insufficient quantities of high molecular length DNA molecules. This was concluded from the observation of only a faint low molecular length DNA smear in each lane (especially in the case of the test sample - lane 6), after agarose gel electrophoresis of the restriction nuclease digested samples (data not shown). As the degree of DNA degradation increases, the intensity of high molecular length bands decreases relative to the lower molecular length fragments [178]. DNA samples in the adjacent lanes which were successfully typed demonstrate a DNA smear through out the length of the gel after digestion with Hae III. These results suggest that both age of sample and environmental conditions influence the quality and/or quantity of extractable DNA.

This observation that environmental factors affect VNTR analysis confirms the results of a FBI validation study. In that study, VNTR analysis of 8 weeks old bloodstains (equivalent volume of blood unknown) exposed to ambient outdoor temperatures (average daily high of 21°C) during March through May, was also unsuccessful due to DNA degradation [179]. However, DNA extracted from control bloodstains kept in the dark, were of sufficient size and quantity to produce a VNTR banding pattern. The difference in the degrees of success of VNTR analysis observed

Figure 44 : VNTR Analysis of DNA Extracted From
External/UV/Soil Bloodstains

DNA extracted from bloodstains exposed to the external environment, or ultraviolet

radiation, or soil contaminants, were simultaneously probed with 3'HVR α-globin and

YNH24 to detect alleles at the D16S85 and D2S44 loci.

Lane   M = molecular length standards

    1 = DNA from 4 hour old stain (internal)

    2 = DNA from stain exposed to 3600 J/M^2 of UV

    3 = "   "   "   "   " 9000   " " "

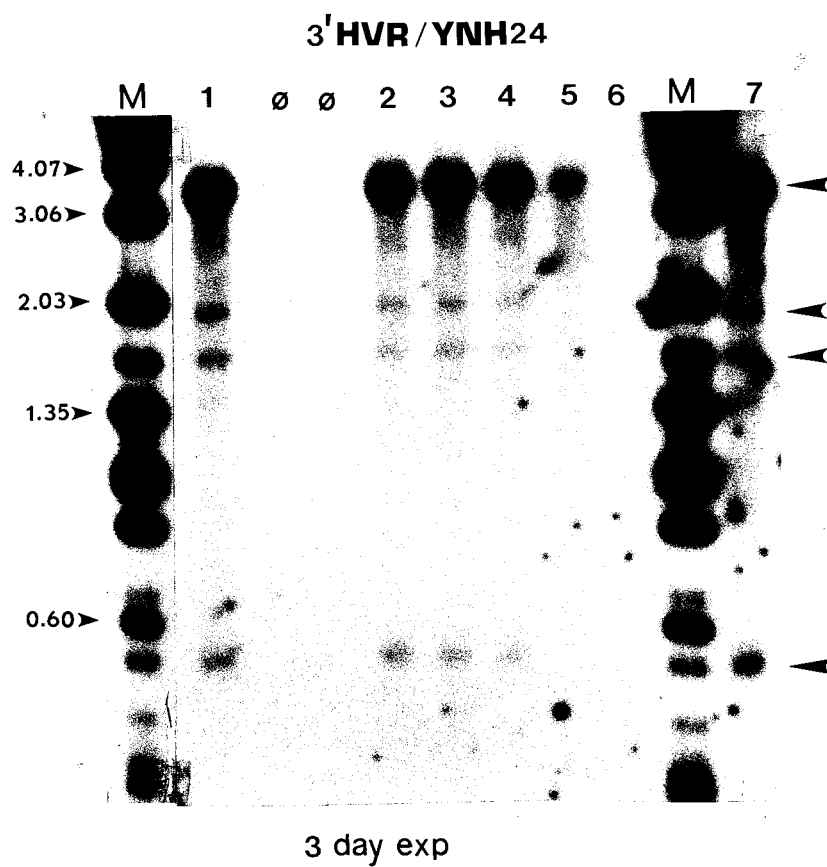   4 = "   "   "   "   " 14400 " " "

    5 = DNA from 7 week old stain (internal)

    6 = "   " " " "   " (external)

    7 = DNA from 10 day old bloodstain buried in soil

Figure 44 : VNTR Analysis of DNA From External/U.V./Soil Bloodstains

3'HVR/YNH24

3 day exp

between the control stains (no external exposure) of this study and that of the FBI, may be a result of non-comparable volumes of blood in stains. Additionally, the partial VNTR band pattern obtained with the 7 week (less than 2 months) old non-exposed bloodstain in this experiment is not consistent with the greater success achieved with 5 and 8 months old non-exposed stains discussed previously (page 194 : Figure 40). The only difference between the stains used in the two experiments is the environment of the stains during their drying period. The 5 and 8 month old stains were dried during the month of February (1990) and the 7 week old stain was dried during October. Differences in humidity levels and temperature may have played a role in the variable ability to type aged stains. As mentioned previously, the humidity was high during October and this may have prolonged the drying period of the stain.

Since there are indications of an environmental role in the quality and/or quantity of DNA extracted from forensic samples, it was necessary to attempt to identify the environmental factors involved.

A) Weather -

RFLP (VNTR) patterns consistent with those obtained from liquid blood samples of the same origin, were obtained from bloodstains incubated at $37^{\circ}C$ for up to 5 days [170], and 20 weeks [179]. Hence prolonged high humidity, rather than temperatures levels, was a more likely cause of DNA degradation in the test sample exposed to the external environment. The large amount of precipitation could also have influenced the amount of DNA available for extraction as some of the blood in the stain may have been washed away.

B) Air Pollution -

The contribution of air pollutants such as nitrogen oxides (nitric oxide and nitrogen dioxide) and sulfur dioxide to DNA damage or degradation was also questioned. Nitrous acid (HNO_2) formed from the reduction of nitrogen dioxide (NO_2), reacts with primary amino groups to oxidatively deaminate adenine to hypoxanthine, cytosine to uracil and guanine to xanthine [180].

Aqueous sulfur dioxide (SO_2), which is the equivalent of bisulfite (HSO_3), also catalyses the deamination of cytosine to uracil [181]. Such changes in DNA bases could affect DNA amplification and restriction endonuclease digestion but would probably not promote DNA degradation.

Nitrogen oxides (as with sulfur dioxide) are produced by the combustion of fossil fuels [182]. Levels of nitrogen oxides are higher than for sulfur dioxide as NO_x are discharged from steam boilers, building heating systems and internal combustion engines. Although the mean hourly levels of NO_x (0.027 ppm) [183] during the period of stain exposure did not exceed the maximum acceptable air quality levels (0.210 ppm) [182], DNA damage could have occurred without any apparent affect on DNA amplification.

Sulfur dioxide (SO_2) is formed by the combustion of fossil fuels containing sulfur [182]. Except for oil refineries and cement plants, natural gas rather than oil is used in almost all heat requiring installations in the Greater Vancouver Regional District, thus the level of sulfur dioxide is low relative to most other cities. The mean hourly concentration of sulfur dioxide during October to December was 0.006 ppm [183] which is well within the maximum acceptable air quality level of 0.340 ppm [182]. Due to such low levels of SO_2, it appears that any possible DNA damage attributable to sulfur dioxide would be negligible.

C) Ultraviolet Irradiation -

Another possible factor affecting the state of DNA molecules is ultraviolet irradiation. DNA damage can occur as a result of UV-induced pyrimidine dimers, interstrand DNA crosslinks and DNA-protein crosslinks. McNally et al. found that bloodstains exposed to ultraviolet irradiation (from a UV light box), for periods up to 5 days, produced consistent RFLP (VNTR) patterns but the intensity of the radiolabeled DNA bands decreased with increased UV exposure time [170]. Stains exposed to sunlight for 10 days failed to produce a VNTR (blot system) band pattern while identical stains stored in darkness at the same temperature provided a banding pattern [179]. RFLP analysis using VNTR probes produced DNA profiles from bloodstains

irradiated with up to 14,400 J/M^2 of short wave (254 nm) ultraviolet light (Figure 44 - lane 4).

This amount of UV light is equivalent to at least 4 hours of direct sunlight. (UV crosslinking of

DNA to membranes only require 1,200 J/M^2 or the equivalent of 20 minutes of direct sunlight).

Irradiated stain DNA produced consistent VNTR bands (Figure 44 - lanes 2,3,4) but were

somewhat less intense than bands from the untreated 4 hour old stain (Figure 44 - Lane 1).

Although band intensities decreased slightly with increased doses of UV irradiation, there does not

appear to be a simple direct relationship. Bloodstains irradiated with UV light at quantities which

differ by a factor of 4, demonstrated only minor differences in band intensities (Figure 44 - lane 2

and 4). Since forensic samples are often not in direct sunlight for prolonged periods i.e. folded

and/or partially covered or shaded, the amounts of UV irradiation per time interval received by

such samples, are probably much less than the values suggested here. Hence in terms of actual

forensic casework stains, the amount of UV irradiation received from short term (weeks as

opposed to months) exposure to sunlight, should not be a major cause of DNA damage and/or

degradation.

As previously discussed (p202) amplification of stains exposed to the external environment

did not appear to be influenced by environment in this situation, but were affected by age. Despite

this observation, it was necessary to determine whether high levels of UV irradiation would affect

amplification. Bloodstains irradiated with various doses of ultraviolet light were all successfully

amplified. However, the level of amplification was much less than that obtained from untreated

samples and did not correlate with the dosage of UV irradiation received (Figure 45 - panel A and

B). Amplification at the APRT locus was possible only with stains irradiated with up to 14,400

J/M^2 (Figure 45 - panel B) whereas, amplification was successful at levels up to 21,600 J/M^2 at the

ADA and PALB loci (Figure 46 - panel A and B, lane 4). Furthermore, amplification of all 3 loci

was not constant. Amplification occurred in one experiment and not in another and there was no

apparent pattern to this inconsistency which occurred more frequently than in all other non-

irradiated samples in the entire project.

Figure 45 : Amplification of DNA Extracted From
UV Irradiated Bloodstains

DNA extracted from bloodstains exposed to ultraviolet radiation were amplified at the

PALB (Panel A) and APRT (Panel B) loci.

Panel A:

Lane   m = molecular length standards

1 = lab stock DNA from blood

2 = DNA from 4 hour old stain (no treatment)

3 = DNA from stain irradiated with 600 J/M^2 UV

4 = "   "   "   "   "   1200   "   "

5 = "   "   "   "   "   2400   "   "

6 = "   "   "   "   "   3600   "   "

7 = "   "   "   "   "   9000   "   "

8 = "   "   "   "   "   14400   "   "

Panel B:

Lane   m = molecular length standards

1 = lab stock DNA from blood

2 = DNA from 4 hour old stain (no treatment)

3 = DNA from stain irradiated with 2400 J/M^2 UV

4 = "   "   "   "   "   14400   "   "

Figure 45 : Amplification of DNA From U.V. Irradiated Bloodstains
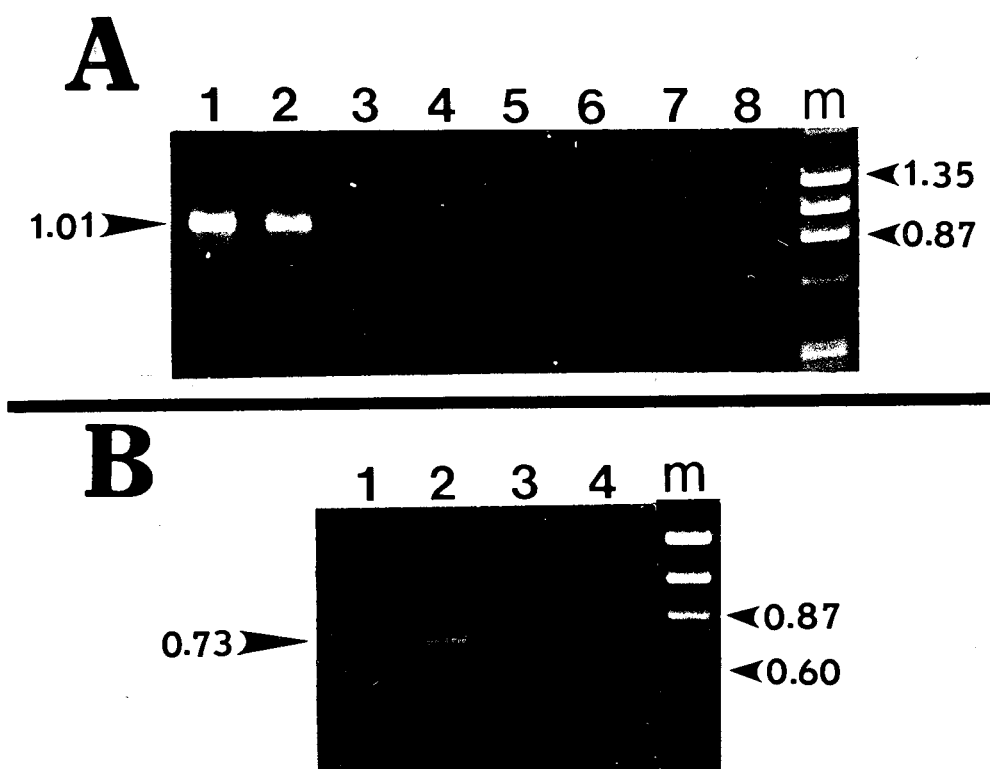
Figure 46 : Amplification of DNA Extracted From
UV Irradiated and Soil Bloodstains

DNA extracted from bloodstains exposed to ultraviolet radiation or soil contaminants

were amplified at the ADA (Panel A) and PALB (Panel B) loci.
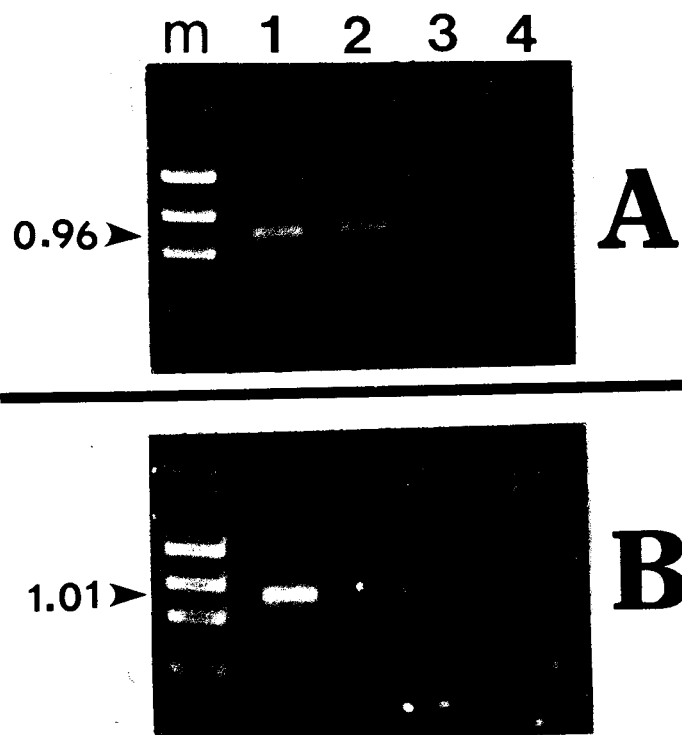
Lane   m = molecular length standards

1 = lab stock DNA from blood

2 = DNA from 4 hour old stain (no treatment)

3 = DNA from 10 day old bloodstain buried in soil

4 = DNA from stain irradiated with 21600 J/M^2 UV

An example of such variation can be seen in Figure 45 (panel A - lane 8), where DNA

from a stain irradiated with 14,400 J/M^2 of UV light was not amplified with PALB in this PCR

experiment but amplification was possible with a dosage of 21,600 J/M^2 in another experiment

shown in Figure 46: panel B-lane 4.

The results observed in this set of experiments do not correlate with the results of the

stains exposed to the external environment. A stain irradiated with an equivalent of only 10

minutes of sunlight (600 J/M^2) exhibited a noticeable decrease in the level of amplification. The

conclusion reached earlier that the absence of an influence exerted by the environment on

amplification may be erroneous. The affects of environment may have been observed but were

wrongly attributed to only the age of the sample. Since, degradation was occurring with age, UV

damaged was not be separately distinguished as the effects of the two are not apparently additive.

Buoncristiani et al. found that short wave UV exposure of naked DNA in solution of more

than 200 J/M^2 reduced amplification and at levels exceeding 2000 J/M^2, no amplification was

observed [172]. It was further determined that short wave UV radiation causes pyrimidine dimers

(which blocks Taq polymerase) and a good distribution of such dimers was evident at doses of 120

J/M^2. However, amplification of liquid whole blood, liquid semen and bloodstains exposed to high

levels of short wave UV radiation, was not significantly affected. The bloodstains results of

Buoncristiani et al. are not in agreement with the results of the present study. This discrepancy

may be partially explained by differences in base sequence of the amplified DNA.

Studies on the origin of ultraviolet damage in DNA by Becker and Wang had shown that

there are sequence and structural requirements for ultraviolet photoproduct formation in DNA

[184]. Becker and Wang found that the sequences surrounding purines or pyrimidines and the

flexibility of DNA at that site will influence the ability of any base to form ultraviolet photoproducts

(dimers). For double-stranded DNA, UV photoreactivity (predominantly dimerizations) occurs

between adjacent pyrimidines and very infrequently in non-adjacent pyrimidine residues. In

contrast, purine photoreactivity will only occur readily if the purines are flanked on the 5'side by at

least two contiguous pyrimidines which transfer the neccessary activation energy. The flexibility of DNA refers to presence or absence of torsional constraints which may inhibit the adoption of a geometrical configuration neccessary for the photoreaction.

Furthermore, effects UV irradiation on amplification of bloodstains by the first group was not observed possibly because of the size of their target fragments (110 bp and 242 bp). The size of the smallest target fragments in the present study is 730 bp and hence there is a greater chance for the photoreactivity (i.e. DNA dimerization and strand breaks) within the target fragment.

D) Soil -

Forensic specimens are often exposed to soil and its contaminants as a result of the physical location of the materials under investigation. RFLP (blot-hybridization) analysis of soiled stains (case samples) by McNally et al. indicated that partial degradation of DNA in soiled stains does occur suggesting that soil has a negative effect on DNA integrity [171]. This result is substantiated by two other studies in which RFLP analysis (blot-hybridization) was unsuccessful with four and 5 day old blood-soil samples [170,179]. However, there was bacterial contamination of some of the blood-soil samples which was detected with bacterial probe pAC267 [170].

In contrast with the three previous studies, this laboratory obtained an intense RFLP (blot-hybridization) band pattern with DNA extracted from bloodstains partially buried in soil for 10 days (Figure 44 - lane 7). Two of these bands are slightly distorted and the background is higher than the untreated control sample (lane 1). However, DNA from this soiled stain does not appear to have degraded as a result of aging or the effects of soil and/or soil contaminants. This conclusion can be drawn since the intensity of its DNA bands are at least comparable to, if not more than, that of the 4 hour old untreated control sample.

Amplification of the 10 day old soiled bloodstain was also successful at the ADA and PALB loci (Figure 46 - lane 3) but was not possible at the APRT locus. Of the five loci, the ADA and PALB loci are the most efficiently amplified. Hence, the inability to amplify the APRT may eventually be resolved with further development of this system. The specificity of these primers for

human DNA eliminates the problem of bacterial contamination found in soil samples [179]. In retrospect, a bacterial probe could have been used to probe the amplified products of the soiled bloodstain to ensure that the products were of human origin. However, this writer is confident that the amplified products are of human origin because the size of the amplified DNA agrees with that of theoretical calculations, and the specificity experiments discussed earlier in chaper 2 (p161) are conclusive.

CONCLUSIONS

The application of the VNTR (blot-hybridization) and the dimorphic RES (polymerase chain reaction) systems, to forensic specimens i.e. bloodstains, has met with varying degrees of success. Protocols have been chosen and altered to provide the most accurate and reliable results with cotton bloodstains. The success of restriction fragment length polymorphism (RFLP) with either system is largely dependent on the amount of high molecular length DNA and/or undamaged DNA extracted from such forensic specimens. The environmental exposure studies in this thesis were not rigorously performed because the main purpose of these environmental exposure studies was to determine the general effects of environment on DNA and precise quantitative and qualitative aspects of extracted DNA were not analysed. Since actual forensic specimens may be of variable quantity and subjected to an infinite variety of conditions, it is also valid to determine environmental assaults under partially defined conditions that would allow DNA typing analysis.

Factors in this study which may be detrimental to the VNTR system include aging and humidity. These same factors, with the addition of ultraviolet radiation, may also be influential factors for the dimorphic RES system. Despite the affects of these elements on both systems, varying degrees of positive results have been obtained with simulated forensic bloodstains. Hence, one does not know for certain whether a particular sample will provide a DNA profile. The value of testing either real or simulated forensic specimens is to provide guidelines as to what may be expected and to indicate direction for the further development of test system(s).

REFERENCES

1. Stockdale R.E.
    Science Against Crime
    Publ. Marshall Cavendish Books Ltd., London
    p30-31 (1982)

2. Gaensslen R.E., Desio P.J., and Lee H.C.
    Forensic Science : Genetic marker systems - Individualization of
    Blood and Body fluids
    Ed. G. Davies
    p209-240 (1986)

3. Oepen I.
    Identification of characteristics in blood and semen
    stains - a review
    Foren. Sci Int. 36, p183-191 (1988)

4. Kuo M.C.
    Linking a bloodstain to a missing person by genetic inheritance
    J. Foren. Sci. 27:2, p438-444 (1982)

5. Denault G.C., Takimoto H.H., Kwan Q.Y., and Pallos A.
    Detectability of selected genetic markers in dried blood on aging
    J. Foren. Sci. 25:3, p479-498 (1980)

6. Jenkins G.C., Brown J., Lincoln P.J., and Dodd B.E.
    The problem of the acquired B antigen in forensic serology
    J. Foren. Sci. Soc. 12, p597-603 (1972)

7. Sensabaugh G.F., Blake E.T., and Northey D.H.
    Genetic markers in semen. III: Alteration of phosphoglucomutase isozyme patterns in
    semen contaminated with saliva
    J. Foren. Sci. 25:3 p470-478 (1980)

8. Gaudette B.D.
    DNA Typing : A new service to Canadian police
    RCMP Gazette 52:4, p1-7 (1990)

9. Jeffreys A.J., Wilson V., Thein S.L.
    Hypervariable "minisatellite" regions in human DNA
    Nature 314:7, p67-73 (1985)

10. Jeffreys A.J., Wilson V., and Thein S.L.
    Individual-specific "fingerprints" of human DNA
    Nature 316:4, p76-79 (1985)

11. Gill P.G., Jeffreys A.J., and Werrett D.J.
    Forensic application of DNA "fingerprints"
    Nature 318:12, p577-579 (1985)

12.   Kanter W., Baird M., Shaler R., and Balazs I.
        Analysis of restriction fragment length polymorphisms in deoxyribonucleic acid (DNA)
        recovered from dried bloodstains
        J. Foren. Sci. 31:2 p403-408 (1986)

13.   Guisti A., Baird M., Pasquale S., et al.
        Application of deoxyribonucleic acid (DNA) polymorphisms to the analysis of DNA
        recovered from sperm
        J. Foren. Sci. 31:2, p409-417 (1986)

14.   Wainscoat J.S., Pilkington S., Peto T.E.A., Bell J.I.,
        and Higgs D.R.
        Allele-specific DNA identity patterns
        Hum. Genet. 75, p384-387 (1987)

15.   Gill P.
        A new method for sex determination of the donor of forensic samples using a recombinant
        DNA probe
        Electrophoresis 8, p35-38 (1987)

16.   Gill P., Lygo J.E., Fowler S.J., and Werrett S.J.
        An evalution of DNA fingerprinting for forensic purposes
        Electrophoresis 8, p38-44 (1987)

17.   Fowler S.J., Gill P., Werrett D.J., and Higgs D.R
        Individual specific DNA fingerprints from a hypervariable region probe: alpha-globin
        3'HVR
        Hum. Genet. (1989)

18.   Haglund W.D., Reay S.T., and Tepper S.L.
        Identification of decomposed human remains by deoxyribonucleic acid (DNA) profiling
        J. Foren. Sci. 35:3 p724-729 (1990)

19.   Gusella J.F.
        DNA polymorphism and human disease
        Ann. Rev. Biochem. 55, p831-854 (1986)

20.   Fowler J.C.S., Burgoyne L.A., Scott A.C., and Harding H.W.J.
        Repetitive deoxyribonucleic acid (DNA) and human genome
        variation - A concise review relevant to forensic biology
        J. Foren. Sci. 33:5, p1111-1126 (1988)

21.   Nei M.
        Molecular Evolutionary Genetics
        Columbia University Press, N.Y.
        Chapter 6, p143 (1987)

22.   Alberts B., Bray D., Lewis J., et. al.
        Molecular Biology of Cell
        Garland Publishing Company, New York,
        p214-215 (1983)
        [As reported by Fowler et al., JFS 33:5, p1111 (1988)]

23. Harris H., and Hopkinson D.A.
Average heterozygosity per locus in man: an estimate based on the incidence of enzyme polymorphism
Ann. Hum. Genet. 36, p9-20 (1972)

24. Botstein D., White R.L., Skolnick M., and Davis R.W.
Construction of a genetic linkage map in man using restriction fragment length polymorphisms
Am. J. Hum. Genet. 32, p314-331 (1980)

25. Cooper D.N., Smith B.A., Cooke H.J., et al.
An estimate of unique DNA sequence heterozygosity in the human genome
Hum. Genet. 69, p201-205 (1985)

26. Jeffreys A.J.
DNA sequence variants in the ^{G}gamma-, $^\alpha$gamma-, delta- and beta-globin genes of man
Cell 18, p1-10 (1979)

27. Li W.H. and Graur D.
Fundamentals of Molecular Evolution
p67-77
Sinauer, Sutherland, Mass. (1991)

28. Levitan M.
Textbook of Human Genetics - 3rd Edition
Oxford University Press (1988)
p148

29. Sambrook J., Fritsch E.F., and Maniatis T.
Molecular Cloning: A Laboratory Manual (2nd Edition)
p5.3 (1989)
Cold Spring Harbor

30. Southern E.M.
Detection of specific sequences among DNA fragments separated by gel electrophoresis
J. Mol. Biol. 98, p503-517 (1975)

31. Watkins P.C.
Restriction fragment length polymorphism (RFLP): applications in human chromosome mapping and genetic disease research
Biotechniques 6:4, p310-320 (1988)

32. Gusella J.F., Wexler N.S., Conneally P.M., et al.
A polymorphic DNA marker genetically linked to Huntington's disease
Nature 306:17, p234-238 (1983)

33. St. George-Hyslop P.H., Tanzi R.E., Polinsky R.J., et al.
The genetic defect causing familial Alzheimer's Disease maps on chromosome 21
Science 235, p885-890 (1987)

34. Ray P.N., Belfall B., Duff C., et al.
Cloning of the breakpoint of an X;21 translocation associated with Duchenne muscular dystrophy
Nature 318:19, p672-675 (1985)

35. Tsui L.C., Barker D., Braman J.C., et al.
Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker
Science 230, p1054-1057 (1985)

36. Kan Y.W., and Dozy A.M.
Antenatal diagnosis of sickle-cell anaemia by D.N.A. analysis of amniotic-fluid cells
The Lancet ii, p910-911 (1978)

37. Chang J.C., and Kan Y.W.
A sensitive new prenatal test for sickle-cell anemia
New Engl. J. Med. 307:1, p30-32 (1982)

38. Woo S.L.C., Lidsky A.S., Guttler F., et al.
Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria
Nature 306:10, p151-155 (1983)

39. Darras B.T., Harper J.F., and Francke U.
Prenatal diagnosis and detection of carriers with DNA probes in Duchenne's muscular dystrophy
New Engl. J. Med. 316:16, p985-992 (1987)

40. Oberle I., Camerino G., Heilig R., et al.
Genetic screening for hemophilia A (Classic hemophilia) with a polymorphic DNA probe
New Engl. J. Med. 312:11, p682-686 (1985)

41. Wyman A.R., and White R.
A highly polymorphic locus in human DNA
Proc. Natl. Acad. Sci. USA 77:11, p6754-6758 (1980)

42. Bell G.I., Karam J.H., and Rutter W.J
Polymorphic DNA region adjacent to the 5' end of the human insulin gene
Proc. Natl. Acad. Sci. USA 78:9, p5759-5763 (1981)

43. Bell G.I., Selby M.J., and Rutter W.J.
The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences
Nature 295:7, p31-35 (1982)

44. Higgs D.R., Goodbourn S.E.Y., Wainscoat J.S., et al.
Highly variable regions of DNA flank the human α-globin genes
Nucleic Acids Res. 9:17, p4213-4224 (1981)

45. Higgs D.R., Wainscoat J.S., Flint J., et al.
Analysis of the human α-globin gene cluster reveals a highly informative genetic locus
Proc. Natl. Acad. Sci. 83, p5165-5169 (1986)

46.   Kobilinsky L. and Levine L.
        Recent application of DNA analysis to issues of paternity
        J. Foren. Sci. 33, p1107-1108 (1989)

47.   Waye J.S., Fourney R.M., Bowen J.H, et al.
        Forensic analysis of restriction fragment length polymorphism: theoretical and practical
        considerations for design and implementation
        Proceedings for The International Symposium on Human Identification
        1989-Promega   p117-152 (1990)

48.   Pignatti P.F. and Marigo M.
        Discussion of "Recent application to DNA analysis to issues of paternity"
        J. Foren. Sci. 35, p517-518 (1990)

49.   Jeffreys A.J., Brookfield J.F.Y., and Semeonoff R.
        Positive identification of an immigration test-case using human DNA fingerprints
        Nature 317:31, p818-819 (1985)

50.   Gill P., and Werrett D.J.
        Exclusion of a man charged with murder by DNA fingerprinting
        Foren. Sci. Internat. 35, p145-148 (1987)

51.   Wambaugh J.
        The Blooding
        Perigod Press/Bamtam Books (1989)

52.   Newmark P.
        DNA fingerprinting to be used for British immigrants?
        Nature 331:18, p556 (1988)

53.   McGourty C.
        Genetic tests made official by U.K. courts
        Nature 339:8 p408 (1989)

54.   Wong Z., Wilson V., Jeffreys A.J., and Thein S.L.
        Cloning a selectedfragment from a human DNA "fingerprint": isolation of an extremely
        polymorphic minisatellite
        Nucleic Acids Res. 14:11, p4605-4616 (1986)

55.   Wong Z., Wilson V., Patel I., Povey S., and Jeffreys A.J.
        Characterization of a panel of highly variable minisatellites cloned from human DNA
        Ann. Hum. Genet. 51, p269-288 (1987)

56.   Jeffreys A.J., Royle N.J., Wilson V., and Wong Z.
        Spontaneous mutation rates to new length alleles at tandem-repetitve hypervariable loci in
        human DNA
        Nature 332:17, p278-281 (1988)

57.   Budowle B., Deadman H.A., Murch R.S., and Baechtel F.S.
        An introduction to the methods of DNA analysis under investigation in the FBI laboratory
        Crime Lab. Digest 15:1 p8-22 (1988)

58. Wetton J.H., Carter R.E., Parkin D.T., and Walters D.
Demographic study of a wild house sparrow population by DNA fingerprinting
Nature 327:14, p147-149 (1987)

59. Burke T., and Bruford M.W.
DNA fingerprinting in birds
Nature 327:14, p149-152 (1987)

60. Jeffreys A.J., Wilson V., Kelly R., et al.
Mouse DNA "fingerprints": analysis of chromosome localization and germ-line stability of
hypervariable loci in recombinant inbred strains
Nucleic Acids Res. 15:7, p2823-2836 (1987)

61. Hill W.G.
DNA fingerprinting applied to animal and bird populations
Nature 327, p98-99 (1987)

62. Wells R.A., Green P., and Reeders S.T.
Simultaneous genetic mapping of multiple human minisatellite sequences using DNA
fingerprinting
Genomics 5, p761-772 (1989)

63. Waye J.S., and Fourney R.M.
Identification of complex DNA polymorphisms based on variable number of tandem repeats
(VNTR) and restriction site polymorphism
Hum. Genet. 84, p223-227 (1990)

64. Lewis R.
Witness for the prosecution : DNA fingerprints
Discover p44-52 (June 1988)

65. Lifecodes Corporation
Background information for DNA-Print identification test
Technical bulletin

66. Cellmark Diagnostics
Court cases in the United States
Technical bulletin

67. Castonguay R.T.
Message
Crime Lab. Digest 15:1, p1 (1988)

68. Shutler G.G.
A perspective on DNA fingerprinting
RCMP Gazette 50:1, p22-27 (1988)

69. Knoll P.J.
An auspicious beginning for DNA genotyping in Canada
Paper presented at the Western Canada Crown Seminar
Banff, April 11, 1989

70. Thompson W.C., and Ford S.
Is DNA fingerprinting ready for the courts?
New Scientists 31, p38-43 (1990)

71. Lander E.
DNA fingerprinting on trial
Nature 339:15, p501-505 (1989)

72. Neufeld P.J., and Colman N.
When Science takes the witness stand
Scientific American 262:5, p46-53 (1990)

73. Lander E.
Expert's Report on People vs. Castro
Court report

74. Ford S., and Thompson W.C.
A question of identity: some reasonable doubts about DNA "fingerprints"
Sciences p37-43 (1989)

75. Witkowski J.
Fingerprint of the future?
NEB Transcript 2:1, p2-7 (1989)

76. Budowle B., Baechtel S., Giusti A.M., and Monson K.L.
Applying highly polymorphic variable number of tandem repeats loci genetic markers to identity testing
Clin. Biochem. 23, p287-293 (1990)

77. Mudd J.L., and Presley L.A.
Quality control DNA typing: a proposed protocol
Crime Lab. Dig. 15:4, p109-113 (1988)

78. Ferrara P.B.
The forensic laboratory manager's responsibility for quality assurance in DNA analysis
Crime Lab. Dig. 17:2, P40-42 (1990)

79. Statement of the working group on statistical standards for DNA analysis
Crime Lab. Dig. 17:3, p53-58 (1990)

80. Kearney J.J., and Mudd J.L.
Guidelines for a proficiency testing program for DNA restriction fragment length polymorphism analysis
Crime Lab. Dig. 17:3, p59-64 (1990)

81. Erickson D.
Do DNA fingerprints protect the innocent?
Sci. Amer., p18 (August 1991)

82. Moss D.C.
DNA - The new fingerprints
Am. Bar Asso. J., May 1 , p66-70 (1988)

83. Wallace B.
Basic Population Genetics
Columbia University Press
New York (1981) Chapter 5, and 8

84. Ayala F.J.
Population and Evolutionary Genetics: A Primer
Benjamin/Cummings Company, Inc.
Menlo Park, California 94025 (1982) p61

85. Keeton W.T.
Biological Science 3rd Edition
W.W. Norton & Company, Inc.
New York (1980) p755-757

86. Neufeld P.J., and Scheck B.C.
Factors affecting the fallibility of DNA profiling: Is there less than meets the eye?
Expert Evidence 1:4, 93-97 (1989)

87. Baird M., Balazs I., Giusti A., et al.
Allele frequency distribution of two highly polymorphic DNA sequences in three ethnic groups and its application to the determination of paternity
Am. J. Hum. Genet. 39, p489-501 (1986)

88. Balazs I., Baird M., Clyne M., and Meade E.
Human population genetic studies of five hypervariable DNA loci
Am. J. Hum. Genet. 44, p182-190 (1989)

89. Cohen J.E.
DNA fingerprinting for forensic identification: Potential effects on data interpretation of subpopulation heterogeneity and band number variability
Am. J. Hum. Genet. 46, p358-368 (1990)

90. Mullis K.B., Faloona F., Scharf S., et al.
Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction
CSH Quantitative Biol. 51, p263-273 (1986)

91. Mullis K.B., and Faloona F.
Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction
Meth. Enzymol. 155, p335-350 (1987)

92. Saiki R.K., Scharf S., Faloona F., et al.
Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia
Science 230, p1350-1354 (1985)

93. Budowle B. and Baechtel F.S.
Modifications to improve the effectiveness of restriction fragment length polymorphism typing
Appl. Theor. Electrophoresis 1, p181-187 (1990)

94.   Sambrook J., Fritsch E.F., and Maniatis T.
        Molecular Cloning: A Laboratory Manual (2nd Edition)
        pE.10-E.13   (1989)
        Cold Spring Habor

95.   Schaffer H.E., and Sederoff R.R.
        Improved estimation of DNA fragment lengths from agarose gels
        Analyt. Biochem. 115, p113-122 (1981)
        [Programing modifications by Jeffrey Bonnycastle]

96.   Southern E.M.
        Measurement of DNA length by gel electrophoresis
        Analt. Biochem. 100, p319-323 (1979)

97.   Budowle B., Waye J.S. Shutler G.G., et al.
        Hae III-A suitable restriction endonuclease for restriction fragment length polymorphism
        analysis of biological evidence samples
        J. Foren. Sci. 35:3, p530-536 (1990)

98.   Nakamura Y., Leppert M., O'Connell P., et al.
        Variable number of tandem repeat (VNTR) markers for human gene mapping
        Science 235, p1616-235 (1987)

99.   Elder J.K., and Southern E.M.
        Measurement of DNA length by gel electrophoresis II: comparison of methods for relating
        mobility to fragment length
        Analyt. Biochem. 128, p227-231 (1983)

100.   Zar J.H.
        Biostatistical Analysis - 2nd Edition
        Publ. - Prentice-Hall, Inc., Englewood Cliffs, New Jersey
        Chapter 11, p162-184 (1984)

101.   Johnson P.H., Miller M.J., and Grossman L.I.
        Electrophoresis of DNA in agarose gels - II. Effects of loading mass and electroendosmosis
        on electrophoretic mobilities
        Analt. Biochem. 102, p159-162 (1980)

102.   Zar J.H.
        Biostatistical Analysis - 2nd Edition
        Publ. - Prentice-Hall, Inc., Englewood Cliffs, New Jersey
        Chapter 2, p32 (1984)

103.   Devlin B., Risch N., and Roeder K.
        Estimation of allele frequencies for VNTR loci
        Am. J. Hum. Genet. 48, p662-676 (1991)

104.   Budowle B. and Monson K.L.
        A statistical approach for VNTR analysis
        Proceedings of the International Symposium on the Forensic Aspects of DNA Analysis
        1989 - FBI Academy, Quantico, Virginia
        p121-135 (1989)

105. Promega Corporation GenePrint data
DNA tying products technical bulletin
No. 3 (1990-91)

106. Budowle B., Giusti A.M., Waye J.S. et. al.
Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from
VNTR loci, for use in forensic comparisons
Am. J. Hum. Genet. 48, p841-855 (1991)

107. Budowle B., Baechtel S., Giusti A.M., et al.
Data for forensic matching criteria for VNTR profiles
Proceedings for The International Symposium on Human Identification 1989 - Promega
p103-115 (1990)

108. Lebo R.V., Chakravarti A., Buetow K.H., et al.
Recombination within and between the human insulin and B-globin gene loci
Proc. Natl. Acad. Sci. USA 80, p4808-4812 (1983)

109. Promega Corporation GenePrint data
DNA tying products technical bulletin,
No. 10 (1990-91)

110. Budowle B. and Stafford J.
Response to expert report by D.C.Hartl
Crime Lab. Dig. 18:3, p101-108 (1991)

111. van Eede P.H., Henke L., Fimmers R., et al.
Size calculation of restriction enzyme HaeIII-generated fragments detected by probe
YNH24 by comparsion of data from two laboratories: the generation of fragment-size
frequencies
Foren. Sci. Int. 49, p21-31 (1991)

112. Statistics Canada : 1981 census of Canada, Vancouver, British Columbia
Ottawa, Ont. : Minister of Supply and Services (1983)

113. Fourney R.M., Shutler G.G., Monteith N., et al.
DNA typing in the Royal Canadian Mounted Police
Proceedings of the International Symposium on the Forensic Aspects of DNA Analysis
1989 - FBI Academy, Quantico, Virginia
p137-146 (1989)

114. Action R.T., Harman L., Rodney M.S., et al.
Comparison of VNTR allele frequencies in White and Black populations
Proceedings for The International Symposium on Human Identification
1989 - Promega p5-20 (1990)

115. Brenner C. and Morris J.W.
Paternity index calculations in single locus hypervariable DNA probes: validation and other
studies
Proceedings for The International Symposium on Human Identification
1989 - Promega p21-53 (1990)

116. Devlin B., Risch N. and Roeder K.
No excess of homozygosity at loci used for DNA fingerprinting
Sci. 249, p1416-1449 (1991)

117. Technical Comments
Forensic DNA tests and Hardy-Weinberg Equilibrium
Sci. 253, p1037-1041 (1991)

118. Stoney D.
Statistics applicable to the inference of a victim's blood type from familial testing
J. Foren. Sci. 24:1, p9-22 (1984)

119. Tyler M.G., Kirby L.T., Wood S., et al.
Human blood stain identification and sex determination in dried blood stains using recombinant DNA techniques
Foren. Sci. Int. 31, p267-272 (1986)

120. Yokoi T. and Sagisaka K.
Sex determination of blood stains with a recombinant DNA probe: comparison with radioactive and non-radioactive labeling methods
Foren. Sci. Int. 41, p117-124 (1989)

121. Lee H.C., Pagliaro E.M., Berka K.M., et al.
Genetic markers in human bone: I. Deoxyribonucleic acid (DNA) analysis
J. Foren. Sci. 36:2, p320-330 (1991)

122. Pääbo S.
Molecular cloning of Ancient Egyptian mummy DNA
Nature 314:18, p644-645 (1985)

123. Perry W.L., Bass W.M., Riggsby W.S., and Sirotkin K.
The autodegradation of Deoxyribonucleic acid (DNA) in human rib bone and its relationship to the time interval since death
J. Foren. Sci. 33:1, p144-153 (1988)

124. Bär W., Kratzer A., Machler M., and Schmid W.
Postmortem stability of DNA
Foren. Sci. Int. 39, p59-70 (1989)

125. Saiki R.K., Gelfand D.H., Stoffel S., et al.
Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase
Science 239, p487-491 (1988)

126. Embury S.H., Scharf S.J., Saiki R.K.
Rapid prenatal diagnosis of sickle cell anemia by a new method of DNA analysis
New Engl. J. Med. 316:11, p656-661 (1987)

127. Kogan S.C., Doherty M., and Gitschier J.
An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences
N. Engl. J. Med. 317, p985-990 (1987)

128. Claustres M., Williams C., and Williamson R.
Amplification of DNA for detection of cystic fibrosis-linked polymorphisms
J. Ped. 115:5, p749-752 (1989)

129. Rosa P.A., and Schwan T.G.
A specific and sensitive assay for the Lyme disease spirochete
Borrelia burgdorferi using the polymerase chain reaction
J. Infect. Dis. 160:6, p1018-1029 (1989)

130. Demmler G.J., Buffone F.J., Schimbor C.M., and May R.A.
Detection of Cytomegalovirus in urine from newborns by using polymerase chain reaction
DNA amplification
J. Infect. Dis. 158:6 p1177-1184 (1988)

131. Kasai K.K., Nakamura Y., and White R.
Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the
polymerase chain reaction (PCR) and its application to forensic science
J. Foren. Sci. 35:5, p1196-1200 (1990)

132. Horn G.T., Richards B. and Klinger K.W.
Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction
Nucleic Acids Res. 17:5, p2140 (1989)

133. Boerwinkle E., Xiong W., Fourest E., and Chan L.
Rapid typing of tndemly repeated hypervariable loci by the polymerase chain reaction:
application to the apoliprotein B 3'hypervariable region
Proc. Natl. Acad. Sci. USA 86, p212-216 (1989)

134. Broderick T.P., Schaff D>A., Bertino A.M., et al.
Comparative anatomy of the human APRT gene and enzyme: nucleotide sequence
divergence and conservation of a nonrandom CpG dinucleotide arrangement
Proc. Natl. Acad. Sci. USA 84, p3349-3353 (1987)

135. Sasaki H., Yoshioka N., Takage Y., and Sakaki Y.
Structure of the chromosomal gene for human serum prealbumin
Gene 37, p191-197 (1985)

136. Wiginton D.A., Kaplan D.J., States J.C., et al.
Complete sequence and structure of the gene for human adenosine deaminase
Biochem. 25, p8234-8244 (1986)

137. Venta P.J.
Personal communication 1990 (unpublished)

138. Oka K., Tkalcevic G.T., Stocks J., Galton D.J., et al.
Nucleotide sequence of Pvu II polymorphic site at the human lipoprotein lipase gene locus
Nucleic Acids Res. 17:16, p6752 (1989)

139. Sensabaugh G.F.
Consequences of nucleotide misincorporation during the polymerase chain reaction
Proceedings of the International Symposium on the Forensic Aspects of DNA Analysis
1989 - FBI Academy, Quantico, Virginia

p263-264 (1989)

140.   Stambrook P.J., Dush M.K., Trill J.J., and Tischfield J.A.
        Cloning of a functional human adenine phosphoribosyl-transferase (APRT) gene:
        identification of a restriction fragment length polymorphism and preliminary analysis of
        DNAs from APRT-deficient families and cell mutants
        Som. Cell and Mol. Genet. 10, p359-367 (1984)

141.   Sensabaugh G.
        Biochemical markers of individuality
        Forensic Science Handbook
        Ed. R. Saferstein (Prentice-Hall)
        p338-415 (chapter 8) (1982)

142.   Li S., Oka K., Galton D., and Stocks J.
        Pvu-II RFLP at the human lipoprotein lipase (LPL) gene locus
        Nucleic Acids Res. 16:5, p2358 (1988)

143.   Yoshioka k., Sasali H., Yoshioka N., et al.
        Structure of the mutant prealbumin gene responsible for familial amyloidotic
        polyneuropathy
        Mol. Biol. Med. 3, p319-328 (1986)

144.   Tzall S., Ellenbogen A., Eng F., and Hirschhorn R.
        Identification and characterization of nine RFLPs at the adenosine deaminase (ADA) locus
        Am. J. Hum. Genet. 44, p864-875 (1989)

145.   Lee B.L., Venta P.J., and Tashian R.E.
        DNA polymorphism in the 5'flanking region of the human carbonic anhydrase II gene on
        chromosome 8
        Hum. Genet. 69, p337-339 (1985)

146.   Jones D.A.
        Blood samples : probability of discrimination
        J.Foren. Sci. Soc. 12, p355-359 (1972)

147.   Kimura M.
        The Neutral Theory of Molecular Evolution
        Cambridge University Press, Cambridge
        p238 (1983)

148.   Keeton W.T.
        Biological Science 3rd Edition
        W.W. Norton & Company, Inc.
        New York (1980) p613-617

149.   Nei M.
        Molecular Evolutionary Genetics
        Columbia University Press, N.Y.
        Chapter 7 p176-175 (1987)

150. Weir B.S.
Inferences about linkage disequilibrium
Biometrics 35, p235-254 (1979)

151. Thompson E.A., Deeb S., Walker D., and Motulsky A.G.
The detection of linkage disequilibrium between closely linked markers: RFLPs at the
AI-CIII apolipoprotein genes
Am. J. Hum. Genet. 42, p113-124 (1988)

152. Chamberlain J., Gibbs R.A., Ranier J.E., et al.
Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA
amplification
Nucleic Acids Res. 16:23, p11141-11156 (1988)

153. Gibbs R.A., Chamberlain J.S., and Caskey C.T.
Diagnosis of new mutation diseases using the polymerase chain reaction
In PCR Technology : Principles and Applications for DNA Amplification
p171-191 (1989)
Ed. H.A. Erlich
Stockton Press, New York, N.Y.

154. Jeffreys A.J., Wilson V., Wong A., et al.
Multilocus and single locus minisatellite DNA probes in forensic medicine
An International Symposium on the Forensic Aspects of DNA Analysis
FBI Academy, Quantico, Virginia
p77-85 (June 1989)

155. Budowle B., Chakraborty R., Giusti A.M., et al.
Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE
Am. J. Hum. Genet. 48, p137-144 (1991)

156. Cheung L.L.C., McNeil B.K., Ferris J.A.J. and Autor A.P.
Determination of twin zygosity by DNA genetic typing
Abstract
Canadian Society of Forensic Sciences Meeting
Toronto, Ont., Canada
Oct. 18-21, 1988

157. Chimera J.A., Harris C.R. and Litt M.
Population genetics of the highly polymorphic locus D16S7 and its use in
paternity evaluation
Am. J. Hum. Genet. 45, p926-931 (1989)

158. Akane A., Matsubara K. et. al.
Paternity testing: blood group systems and DNA analysis by variable number of tandem
repeat markers
J. Foren. Sci. 35, p1217-1225 (1990)

159. Bartlett S.E. and Davidson W.S.
Forensically informative nucleotide sequences (FINS) identify different species of tuna
Poster presentation
Canadian Society of Forensic Science

Ottawa, Ont., Canada
Oct. 1-5, 1990

160. Harris A.S., Bentzen P. and Wright J.M.
Cloning of hypervariable loci from fish DNA
Poster presentation
Canadian Society of Forensic Science
Ottawa, Ont., Canada
October 1-5, 1990

161. Fey M.F., Wells R.A., Wainscoat J.S. and Thein S.L.
Assesment of clonality of gastrointestinal cancer by DNA fingerprinting
J. Clin. Invest. 82:5, p1532-1537 (1988)

162. Pääbo S.
Amplifying ancient DNA
PCR Protocols - A Guide to Methods and Applications
Ed.- M.A.Innis, D.H.Gelfand, J.J.Sninsky, and T.J. White
Academic Press, Inc., San Diego, California
Chapter 20, p159-166 (1990)

163. Pääbo S., Gifford J.A., and Wilson A.C.
Mitochondrial DNA sequences from a 7000-year old brain
Nucleic Acids Res. 16:20, p9775-9787 (1988)

164 Pääbo S.
Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification
PNAS 86, p1939-1943 (1989)

165. Reynolds R., von Beroldingen C., and Sensabaugh G.F.
Effects of DNA degradation on amplification by the polymerase chain reaction
An International Symposium on the Forensic Aspects of DNA Analysis
FBI Academy, Quantico, Virginia, p257 (June 1989)

166. Comey C.T.
The use of DNA amplification in the analysis of forensic evidence
Crime Lab. Dig. 15:4, p99-103 (1988)

167. VonBeroldingen C., Walsh S., Roby R., and Sensabaugh G.
DNA in hair
An International Symposium on the Forensic Aspects of DNA Analysis
FBI Academy, Quantico, Virginia
p265-266 (June 1989)

168. Blake E.T., Pääbo S., and Stolorow M.D.
DNA amplification and typing from aged biological evidence
An International Symposium on the Forensic Aspects of DNA Analysis
FBI Academy, Quantico, Virginia
p267-268 (June 1989)

169. Walsh P.S., Metzger D.A., and Higuchi R.

Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material
Biotechniques 10:4, p506-513 (1991)

170.   McNally L., Shaler R.C., Baird M., et al.
Evaluation of deoxyribonucleic acid (DNA) isolated from human bloodstains exposed to ultraviolet light, heat, humidity, and soil   contamination
J. Foren. Sci. 34:5, p1059-1069 (1989)

171.   McNally L., Shaler R.C., Baird M., et al.
The effects of environment and substrata on deoxyribonucleic acid (DNA): the use of casework samples from New York city
J. Foren. Sci. 34:5, p1070-1077 (1989)

172.   Buoncristiani M., von Beroldingen C., and Sensabaugh G.
Effects of UV damage on DNA amplification by the polymerase chain reaction
An International Symposium on the Forensic Aspects of DNA Analysis FBI Academy, Quantico, Virginia
p259 (June 1989)

173.   Gelfand D.H.
PCR Technology : Taq DNA polymerase-Chapter 2
Ed. H.A. Erlich
Publ. Stockton Press 1989
p20

174.   Harding J.D., Gebeyehu G., Bebee R. et al.
Rapid isolation of DNA from complex biological samples using a novel capture reagent-methidium-spermine-sepharose
Nucleic Acids Res. 17, p6947-6958 (1989)

175.   Singer-Sam J., Tanguay R.L., and Riggs A.D.
Use of Chelex to improve the PCR signal from a small number of cells
Amplifications : A forum for PCR users
Issue 3, p11 (September 1989)

176.   Walsh S. and Higuchi R.
PCR inhibition and bloodstains
An International Symposium on the Forensic Aspects of DNA Analysis
FBI Academy, Quantico, Virginia
p281-282 (June 1989)

177.   Environment Canada
Monthly meteorological summarys
Taken at Vancouver International Airport, B.C.
October 1990 - December 1990

178.   Konzak K.C., Reynolds R., von Beroldingen C., et al.
Effects of DNA damage and degradation of RFLP analysis
An International Symposium on the Forensic Aspects of DNA Analysis
FBI Academy, Quantico, Virginia
p255 (June 1989)

179.   Adams D.E.
       Validation of the FBI procedure for DNA analysis: a summary
       Crime Lab Dig. 15:4, p106-108 (1988)

180.   Stryer L.
       Biochemistry
       W.H. Freeman and Company, San Francisco
       Second edition, p636 (1981)

181.   Blackburn G.M. and Gait M.J.
       Nucleic Acids in Chemistry and Biology
       IRL Press, Oxford University Press, New York
       p265 (1990)

182.   Greater Vancouver Regional District
       Ambient Air Quality - Annual Report
       1989

183.   Greater Vancouver Regional District
       Ambient Air Quality - Quarterly Report
       1990

184.   Becker M.M. and Wang Z.
       Origin of ultraviolet damage in DNA
       J. Mol. Biol. 210, p429-438 (1989)