Aspects of Human Relationship Identification using Short Tandem Repeats

By

Chang En Pu, MS

Centre for Forensic Science
The Department of
Pure & Applied Chemistry
University of Strathclyde
U.K.

Supervisor,
Dr. Adrian M.T. Linacre

2009

Statement of authenticity and author's right

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has lead to the award of a degree.

Signed:  Date: April 12, 2009

Aspects of Human Relationship Identification using Short Tandem Repeats

By

Chang En Pu, MS

A report submitted to the Department of Pure and Applied Chemistry, University of Strathclyde, in part fulfilment of the regulations for the degree of PhD in Chemistry.

ACKNOWLEDGEMENTS

The author is indebted to the following people whose contributions made this work possible:

Dr. Adrian Linacre, an internationally well-known forensic biologist, my supervisors, for his brilliant instructions, valuable advice, and smart guidance in the whole process of studying, although it was a long way, he did not lose his consideration, hospitality and patience. Especially he spent a lot of energy and times helping me publish 3 papers of population study, and that forms the basis of the dissertation.

Dr. Will Goodwin and Dr. Nigel Watson for their inspiring questions, integrated comments and advanced suggestions. Additional thanks to Dr. Goodwin for being the second time learning from him, the first chance was when he worked for University of Glasgow, I spent about 3 months following his forensic DNA steps, and in that time he has shown his highly sufficient knowledge as a forensic genetics. Dr. Kathleen A. Savage, for her hospitality and consideration for organizing the viva, I felt comfortable and friendly.

All members of the Forensic Science Unit, University of Strathclyde, they has been so kind and patient to a foreign student who have not shown up regularly, and they didn’t blame me for the lagging in PhD programme, and colleagues of the Forensic Science Department and the Forensic DNA Laboratory, the Investigation Bureau of Ministry of Justice, Taiwan ROC, for their hard working of processing routine cases and diversified researching topics, discussion, challenging and even questioning, anyway that helped me for getting out the piles of data.

My parents, my wife I-Fong, and my son Shuai-Chen, for their familial support and pressure, encouragement, and patience, especially the time when I got twice surgeries on my two eyes.

Statement of Publications as a Result of this Research

During the studying period of 2002-2008 in the University, the author and supervisor Dr. Adrian Linacre have had 3 publications published, 2 further paper under review and 8 conference presentations at the time of writing. These studies are as follows:

Publications:

I. Pu CE, Linacre A. A study on the false avuncular inclusion rates in human identification. (Under revision by Dr. Linacre, planning to submit to Int. J. of Leg Med.)
II. Pu CE, Linacre A. Systematic evaluation of minimum CPI requirement and its effects on specificity and sensitivity for paternity duo testing. (under final revision by Dr. Linacre, planning to submit to Transfusion)
III. Pu CE, Linacre A. Increasing the confidence in half-sibship determination based upon 15 STR loci. J Forensic Leg Med. 2008; 15(6): 373-7.
IV. Pu CE, Linacre A. Systematic evaluation of sensitivity and specificity of sibship determination by using 15 STR loci. J Forensic Leg Med. 2008; 15(5): 329-34.
V. Pu CE, Linacre A. CPI distribution and cutoff values for duo kinship testing. Chin J Physiol. 2007; 50(5): 232-9.

Conference presentations:

Local, In Chinese:

I. Pu CE. Study on false inclusion versus CPI cutoffs in paternity test. Proceedings of Taiwan Academy of Forensic Sciences, Nov. 21, 2003.
II. Pu CE. An EXCEL program for calculating allele frequencies and forensic DNA parameters. Proceedings of Taiwan Academy of Forensic Sciences, Nov. 19, 2004.
III. Pu CE. An EXCEL program for automatic STR matching for disasters. Proceedings of Taiwan Academy of Forensic Sciences, Oct. 14, 2005.

International, in English:

I. Pu CE. et al. Experiences of Human Bodies Identified by DNA typing in Singapore Airlines SA-006 Crash in Taipei. 54th Annual meeting of American Academy of Forensic Sciences, Feb. 18-23, Atlanta, MA., 2002.

II. Pu, CE. et al. Evaluation the identification power of CODIS 13 STR for the identification of unidentified bodies in Taiwan, 55th annual meeting, American Academy of Forensic Sciences, Feb. 17-22, Chicago, IL., USA, 2003.

III. Pu, CE. et al. Evaluating the false parentage rate and CPI cutoff of CODIS 13 STR for seven populations, 56th annual meeting, American Academy of Forensic Sciences, Feb. 16-21, Dallas, TX., USA, 2004.

IV. Pu, CE. and Linacre A. A Computer Program for Calculating Forensic Population Study Parameters of STR loci, 57th annual meeting, American Academy of Forensic Sciences, Feb. 21-26, New Orleans, UA., USA, 2005.

V. Pu CE. et al. CPI Distribution and Cutoff Value for Duo Paternity Building－International Forensic Science Symposium, Nov. 7-9, Taipei, Taiwan, 2005.

VI. Pu CE. and Linacre A. A Work-Sheet Based Computer Program for Forensic Population Study of STR Loci, International Forensic Science Symposium, Nov. 7-9, Taipei, Taiwan, 2005.

VII. Pu CE. et al. Study on the DNA STR Profiling for Artificially Degraded Bones. International Forensic Science Symposium, Nov. 7-9, Taipei, Taiwan, 2005.

VIII. Pu CE. and Linacre A. Study on Automatic Forensic DNA STR Blood-Relative Comparing for Un-Identified Body Database Processing－International Forensic Science Symposium, Nov. 7-9, Taipei, Taiwan, 2005.

IX. Pu, CE. et al. CPI distribution and cutoff value for duo paternity building, 58th annual meeting, American Academy of Forensic Sciences, Feb. 20-25, Seattle, WA., USA, 2006.

X. Pu, CE. et al. Convenient database and kinship matching for unidentified human remains. 91st Annual meeting, International Association for Identification, Jun. 2-7, Boston, MA., USA, 2006.

XI. Pu CE, Meng LM, Chao CM, Wu FC. Study on the DNA STR Profiling for Artificially Degraded Bones. International Forensic Science Symposium, Nov. 7-9, Taipei, Taiwan, 2005.

XII. Pu CE and Linacre A. Study on Automatic Forensic DNA STR Blood-Relative Comparing for Un-Identified Body Database Processing－International Forensic Science Symposium, Nov. 7-9, Taipei, Taiwan, 2005.

Contains of publications and conference presentations are attached as appendix.

Signed: Date: April 12, 2009

Abstract

Human short tandem repeat (STR) loci identification can be for linking material from a crime scene to an individual or linking two individuals. Paternity testing is one form of linkage where typically samples are tested between the mother, child and putative father (trio) and a probability (CPI) of paternity (or maternity) can be calculated. If only one parent is available for a paternity test (duo), the probability of paternity will be reduced. Besides, the study also focuses on only two in putative siblings' and only two putative half-siblings' relationship determination, also strength of probabilities are expressed by Combined Sibling Indices (CSI) and Combined Half-sibling Indices (CHSI).

Three real populations were used to generate visual offspring and random pairs to check if the existence of Coincidental Matched Pairs (CMPs) where unrelated individuals share alleles for all loci become false parent and child (nonexcluded inclusion). When real duos also observed for the distribution of CPI, then under different CPI cutoffs, the specificity and sensitivity of the STR systems can be determined. Siblingship and Half-siblingship were also examined the same way for observing different index cutoffs and their corresponding specificity and sensitivity, besides allele sharing situation, Two-Allele-Sharing-Locus (TASL) for sibling and All-Shared-Alleles (ASA) for half-sibling were added up for enhancing the specificity and sensitivity. A combination of these two data sets as index and allele-sharing increased the confidence in an determination of inclusion, especially for the low CSI and low CHSI cases.

Recommendations are made in this thesis for flexible cutoff values to assist in determination of human relationship in normal and non-optimal situation.

Table of Contents

Table of Tables

Table of Figures

* The numbering of the Figures reflected the Chapter number.

Table of Boxes

Table of Abbreviations

Ch:	Chinese
Ca:	Caucasian
AA:	African Americans
RD:	Real duo
CMP:	Coincidental Matched Pair
TASL:	Two-Alleles-Shared-Locus
ASA:	All-Shared-Alleles
SPE:	Specificity
SEN:	Sensitivity
sib:	Sibling
half-sib:	Half-sibling
rp:	Random pairs

Chapter 1: A review of the problem and the aim of the study

1.1 Development of Forensic DNA Analysis

Since the first use of DNA in the criminal justice system the technology behind forensic DNA profiling has gradually evolved. A major event in the development of DNA profiling was the advent of the polymerase chain reaction (PCR) (1). PCR has had a revolutionary affect on forensic DNA development allowing increased sensitivity and speed of analysis.

DNA Typing Prior to the Advent of PCR

Genetic polymorphisms were identified and characterised by using restriction fragment length polymorphism (RFLP) techniques between the 1970s and 1990s. RFLP required the digestion of genomic DNA with restriction endonucleases and then hybridization to specific probes to the size fragments (1, 2). The application of RFLP to detect and characterise genetic polymorphisms in a forensic context was first reported by Dr. Alec Jeffreys (3). Dr Jeffreys (now Professor Sir Alec Jeffreys) reported that highly repetitive sections of minisatellites DNA can differentiate individuals based upon the length of the restriction fragment spanning a minisatellite. The length of the restriction fragment was detected by using a radioactive probe that would bind to the

central core sequence of the minisatellite. The original choice of the probe sequence meant that it would bind to a number of other restriction fragments and hence bind to multiple minisatellites resulting in a large number of positions where the probe bound. These positions were detected using an autoradiograph resulting in multiple bands corresponding to a lane on an agarose gel. Multi-locus probes (MLP) as they were called resulted in bar-code like bands that were termed a DNA fingerprint. This MLP method was firstly applied to resolve an immigration case (4) and shortly thereafter to solve a double homicide case in UK (5).

MLP suffered from a number of drawbacks. It required around 1 μg of DNA as the starting template; electrophoretic motilities affected the matching of bands; inter-plate and inter-laboratory comparison was not possible; and the technique was found to be labour intensive. Due to these reasons MLP was superseded by single locus probes (SLP).

SLP typing varied from MLP in that the probe was designed to be specific to the minisatellite locus. This would therefore result in recognizing a maximum of two bands for each minisatellite locus, and a serial or a cocktail of probes could be used to detect composite banding patterns, if the patterns are

distinctive then a high power of discrimination can be reported. The results could be recorded digitally according to the co-run ladders making inter-lab comparisons feasible, and the clear banding patterns made the developing of computerized databank of DNA profiles possible. Although the storing and comparison of data could be handled by computer, the automation of the DNA profiling was seen to be very difficult, especially for processing large amounts of samples in a short time. Fortunately the application of PCR technology to the forensic process increased greatly the sensitivity and speed of the process (1).

PCR stage

PCR not only revolutionised medical genetics but also forensic science. The PCR technique enabled the scientists to replicate (amplify) DNA templates from trace amounts of DNA (6, 7). Further, the process works most effectively on small size templates allowing trace amounts of degraded DNA to be processed (8, 9).

One of the first applications of PCR in a forensic context was the development of amplified length polymorphisms (AMFLP) profiling systems. D1S80 and Apo B (10) are smaller versions of minisatellites. These two AMFLP loci were generally recognized as simple and robust, and were adopted by

laboratories in Europe and other areas. The two loci attracted interest from commercial suppliers and kits from the commercial company Applied Biosystems were developed. As there were only two loci the power of discrimination was low and the two could not be amplified in the same reaction.

At the same time commercial companies had started to produce kits for some marker systems using PCR techniques. The first well-known kit for the detection of a sequence polymorphism within HLA DQa locus was launched in 1990 (11). Although this kit had relatively low discrimination powers it alerted the forensic community to the use of PCR methods.

In 1991, Dr. Jeffreys found differences between the repeating sequences within a single minisatellite and a novel method called minisatellite variant repeat PCR (MVR-PCR) was developed which applying PCR and SLP techniques to detect differences in the internal variation of a minisatellites (12). The technology was also termed digital DNA typing owing to having the potential to provide results for developing database. This process, although highly discriminating, had a problem with mixed samples.

In 1992, the National Research Council of the United States and the European DNA Profiling group (EDNAP) recognized that the future of forensic

DNA was the use of new form of polymorphism generated from short tandem repeats (STRs) (13, 14), also known as microsatellites. These are a type of VNTR which consisted small repeating sequence units, usually three, four or five bases, also called simple sequence repeats. The maximum length of the loci was around 300 base pairs (9), which is an ideal size to be amplified by PCR.

The Forensic Science Service (FSS) in the UK had been working on a multiplex of four STR loci for routine use as well as the start for developing databases, and then expanded the number of loci to six autosomal STR loci and another locus for sex determination; this increased the power of discrimination of the test. The six loci STR multiplex was called the second generation multiplex (SGM) (15, 16). The multiplex systems for STR profiling allowed for simultaneous amplification of a number of STR loci in a single reaction, and the fluorescent dye labelled primer sets for different loci could be detected by automated DNA sequencers (17).

Seeing the success of the STR multiplex of FSS, commercial companies had been encouraged to make the STR kits available as off-the-shelf products. There are several multiplexes available now on the market covering between 9 to 15 autosomal STR systems as well as the sex-specific amelogenin locus (see

table 1.1).

Table 1.1: Main commercially available STR multiplex kits profiled by capillary electrophoresis

	Identifiler	SGM Plus	Profiler Plus	COfiler	Profiler	Minifiler	SEfiler	Poweplex 16	Poweplex ES
	15	10	9	7	9	8	11	15	8
D8S1179	*	*	*				*	*	*
D21S11	*	*	*			*	*	*	*
D7S820	*		*	*	*	*		*	
CSF1PO	*			*	*	*		*	
D3S1358	*	*	*	*	*		*	*	*
TH01	*	*		*	*		*	*	*
D13S317	*		*			*		*	
D16S539	*	*		*	*	*	*	*	
D2S1338	*	*			*	*	*		
D19S433	*	*					*		
VWA	*	*	*		*		*	*	*
TPOX	*			*	*			*	
D18S51	*	*	*			*	*	*	*
Amelogenin	*	*	*	*	*	*	*	*	*
D5S818	*							*	
FGA	*	*	*		*	*	*	*	*
SE33							*		*
Penta D								*	
Penta E								*	

Identifiler, SGM Plus, Profiler Plus, COfiler, Profiler, Minifiler, SEfiler: STR Kits from Applied Biosystems (Foster City, CA), Homepage (18)

PowerPlex 16, PowerPlex ES : STR Kits from Promega Corporation (Madison, WI), Homepage (19)

** Commercial STR Kits are popularly used by forensic DNA laboratories around the world for their ease of handling and the automated profiling.

In 1995, the UK NDNAD (National DNA Database) was launched (20), shortly after DNA profiling using six STR loci (the SGM – second generation multiplex) was available for criminal casework. The necessary legislation was the Criminal Justice and Police Order Act 1994, which came into force on 10th April 1995. Subsequent legislation has expanded the scope of samples that may be collected and retained on the NDNAD.

STR profiles are in digital form can be digitalised very easily and this has allowed for the effective processing of the DNA information by digitalisation of the data. That makes the searching and comparison of DNA profiles much easier.

The match probability of SGM was 1 in 10^8 and, with the population of the UK being approximately 58 million, this power of discrimination was acceptable when the database was originally established. In order to exclude adventitious hits in part, in 1999 the six-locus SGM test was changed to the test of ten-locus kit-AmpF*I*STR® SGM Plus®. The probability that two STR profiles from unrelated people will share all alleles at ten loci is estimated to be more than 1 billionth. To date no two unrelated individuals have been found to

have exactly the same STR profiles at all ten loci.

The effective application of the DNA database, in particular in the UK, has become a leading force for the establishment and expansion of DNA database in other countries, including the USA and many European and Asian countries that now have databases with hundreds of thousands of DNA profiles. In an average year the UK NDNAD produces around 40,000 crime scene to individual matches. With such a large number of DNA profiles held on the NDNAD there is currently a 45% chance that a DNA profile obtained from an incident will match a DNA profile on NDNAD (21).

In 1998, Interpol established a register of sexual offenders that would contain DNA information, and THO1, VWA, FGA and D21S11 (15) were selected; these became known as European standard set of loci (ESS). Three more loci of D3S1358, D8S1179 and D18S51 were added one year later leading to a standardisation in the loci that should form part of a multiplex (16).

These multiplex kits have been initially developed to meet the requirements of the FSS and latterly the Federal Bureau of Investigation (FBI) in the USA. In the USA the Combined DNA Index System (CODIS) was authorized by a Federal DNA Identification Act of 1994. Originally a panel of

13 STRs and a sex-specific locus were proposed by FBI in 1997 (22) and these loci can be found in Profiler Plus and Cofiler kits supplied by Applied Biosystems. The procedures for obtaining and managing DNA data have become one of the standard procedures for forensic DNA analysis. The success of any DNA profiling is can only be measured by the number of true inclusions resulting in the identifying of a perpetrator of a crime, the exoneration of falsely accused persons, and the reduction in crime.

STR typing has superseded all previous forms of DNA typing and is generally accepted for human identification and relationship testing (13). Examples of the use of STR typing are shown in table 1.2.

Table 1.2: Application of STR profiling in human identity testing

Type of application	Type of comparison
Forensic DNA Casework	Comparison of two or more genotypes.
DNA Databases for convicted or suspect	Comparison of a genotype to all those stored on a database.
Mass fatalities from disasters	Comparison of a genotype from a reference sample to family members or reference ante mortem samples.
Missing Persons	Comparison of alleles to family members.
Human relationship testing	Parentage test: half allele sharing matching. Kinship testing: siblingship, half-siblingship.
Monitoring of bone marrow engrafting	Full allele sharing matching.

(23)

Despite the large number of commercial kits as shown in table 1, there are core loci used in Northern America, European, and Asia etc. which allows the exchange of the STR data. This is further illustrated in table 1.3.

Table 1.3: STR loci used in some countries and INTERPOL

	Identifiler®	U.S. Core loci	UK/European Core Loci	German	Interpol Standard Set of Loci	Australia	Taiwan
	15	13	10	8	7	9	15
D8S1179	*	*	*	*	*	*	*
D21S11	*	*	*	*	*	*	*
D7S820	*	*				*	*
CSF1PO	*	*			*		*
D3S1358	*	*	*	*	*	*	*
TH01	*	*	*	*			*
D13S317	*	*				*	*
D16S539	*	*	*				*
D2S1338	*		*				*
D19S433	*		*				*
VWA	*	*	*	*	*	*	*
TPOX	*	*					*
D18S51	*	*	*	*	*	*	*
Amelogenin	*	*	*	*	*	*	*
D5S818	*	*					*
FGA	*	*	*	*	*	*	*
SE33				*			

*Identifiler® is the name of the 15 STR kit produced by Applied Biosystems, Foster City, CA USA.

Fifteen STR core loci system, including the 13 STR loci of CODIS, D2S1338, D19S433 and a sex test loci, is currently the one of the largest number of loci that is amplified in one reaction as a commercial kit. Owing to the convenience of a single reaction to amplify 15 STR loci, these loci have become one of the major standard DNA tests for human identity testing and human relationship testing in many later-developed DNA testing laboratories around the world.

The major events in the evolvement of forensic DNA are summarized in Table 1.4.

Table 1.4: Evolution of forensic DNA.

Year	Events
1980	Arlene Wyman and Ray White described first polymorphic RFLP marker (24).
1983	PCR techniques first described by Kary Mullis (1).
1984	Alec Jeffreys developed RFLP methods and termed the pattern a “DNA fingerprinting” (2).
1985	Police in UK first use forensic DNA profiling (25).
1986	DNA testing became public in US (25).
1987	In UK forensic investigators use DNA to help solve the “Black Pad” murders and to identify the killer. The first case in which DNA evidence is used to determine the identity of a murderer, the first case in which a suspect was exonerated due to DNA (5). VNTR loci described and a series VNTR probes for use in RFLP analysis developed (13). DNA was first introduced as evidence in U.S. court system (14, 26).

continues

Table 1.5 (continued)

1988	Jeffreys showed that PCR can be used to faithfully amplify entire VNTR loci, genetic testing to be performed could be with far less DNA than previous techniques (12). First commercial kit using PCR techniques became available (14). EDNAP had its first meeting in London UK, collaboration and 'harmonization' were agreed (27).
1989	Promega first introduces probes for isotopic detection of VNTR loci (19). The Technical Working Group on DNA Analysis Methods (TWGDAM) in US proposed combing forensic DNA with computer advancements to aid in resolving violent crimes (20). First DNA exoneration took place, in US; Gary Dotson had a conviction overturned on the basis of DNA evidence (25). US *People v. Castro* raised important issues concerning reliability and quality of forensic DNA testing (25, 26). FBI US proposed a national DNA database for North America that would contain profiles from SLP analysis (27). EDNAP began a series of experiments aimed at achieving common standards (28).
1990	HLA DQA1 kit using the PCR technique was introduced (26). CODIS pilot program started (29)
1991	STR markers became recognized as an effective tool for human identity testing (16). Fluorescent STR markers first described (14). DNA extracted by Chelex was developed (14).
1992	Promega first introduces probes for non-isotopic detection of VNTR loci (13). In US, National Research Council found that DNA testing was a reliable method for criminal identification (26) Working group within ISFG concluded that RFLP would be replaced by PCR based analysis (27). STR polymorphisms are discovered on the Y chromosome (29). Innocence Project was founded by Barry C. Scheck and Peter J. Neufeld to assist prisoners who could be proven innocent through DNA testing (31).

continues

Table 1.6 (continued)

1993	First STR kit became available and sex-specific typing (amelogenin) method developed (16, 26) DNA Commission of the International Society of Forensic Haemogenetics (ISFH) recommended nomenclature for STR systems which is commonly used today (27). First disaster case using PCR-based DNA typing method for identification of victims (32), the techniques proved to be reliable and robust.
1994	NDIS and DAB (DNA Advisory Board created) were established. (26) The first FSS multiplex applied to casework published (THO1, VWA, FES/FPS and F13A1) (16). The DNA Identification Act was launched in US. Federal DNA Identification Act was enacted, authorized the FBI to establish a National DNA index for law enforcement (29) The EDNAP lead collaborative exercises demonstrated that simple STR systems were suitable candidate for standardization run by different laboratories in European (33)
1995	UK DNA database established (SGM loci: THO1, VWA, FGA, D8S1179, and D21S11), the world's first national DNA database (20). The ABI Prism 310 Genetic Analyzer with multicolour fluorescence detection capability launched (34).
1996	FBI US started mtDNA testing (13) The US National Missing Persons DNA Database Program initially authorized (29). The International Commission on Missing Persons (ICMP) was created at G-7 Summit (35). Fluorescence detection methods produced a clearer banding than that from silver staining (36)
1997	Short Tandem Repeat DNA Internet Database launched (23). CODIS 13 core loci defined (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, VWA, TPOX, D18S51, D5S818, FGA) (20, 29).

continues

Table 1.7 (continued)

1998	Austria, Germany and The Netherlands introduced national databases (13). Interpol uses THO1, VWA, FGA and D21S11, became known as the European standard set (ESS) (16. Federal Bureau of Investigation announced quality assurance standards for forensic DNA testing laboratories (23). FBI launched national Combined DNA Index System (26).
1999	Interpol set of STR was expanded to include three more loci (D3S1358, D8S1179 and D18S51) (16). SGM plus kit became available (16). Commercial multiplex STR kits were validated (26). In US, 13 assaults linked through DNA analysis and perpetrator identified (29). In Switzerland, a crime analysis system integrates forensic case data with operational police intelligence developed (37).
2000	PowerPlex 16 kit enables first single amplification of CODIS STRs (Amelogenin, Penta D, and Penta E included in the kit) (13). FBI approved revisions to the minimum quality assurance standards required for DNA laboratories to join CODIS system (23). FBI in US stops processing RFLP and converted to multiplex of STRs (26).
2001	Identifiler STR kit with 5 dyes released (including the CODIS 13, amelogenin, D2S1338 and D19S433) (16). Multi-capillary ABI 3100 analyzer with increased level of throughput available (34). CODIS STR Loci data from 41 Sample Population announced, provided a solid foundation for STR profile frequency estimate (38).
2002	FBI in US released mtDNA population database (39) Microchip capillary array electrophoresis analyzers with rapid high-throughput of forensic DNA analysis was described (40).

continues

Table 1.8 (continued)

2003	Human Genome Project completed (16) 50th anniversary of Watson-Crick DNA discovery Function of NDNAD maximized by identifying a suspect through his close familial members (41). SGM plus population study of 24 European populations was published for the foundation of frequency calculation (42). Y STRs kit available (43).
2004	The effectiveness and success of DNA Expansion Program in UK was described (37). Portable DNA analysis systems described (40). AABB annual report described that 98.34% of relationship report using STR analysis technologies (44).
2006	The UK and much of the European laboratories uses SGM plus kit or the equivalent (45).
2008	A large comprehensive YSTR reference database with more than 13,000 haplotypes available online at: wwww.usystrdatabase.org (46)

1.2 Parentage testing

Civil law or immigration related paternity (or maternity) cases are performed routinely around the world when the identity of the parent of a child is in dispute. The cases typically involve the mother, the child, and at least one alleged father. The children will inherit a combination of paternal and maternal alleles from the STR loci. In standard paternity testing it is relatively straightforward to examine alleles shared between the biological mother and child to identify any alleles not in common. These alleles must have been contributed by the biological father; the alleles that are identified as coming

from the biological father are termed paternal obligatory alleles.

STRs loci are the main operating method for paternity testing laboratories (44, 47, 50). Their speed, relatively low cost, robustness, and reproducibility have long been recognized (49, 51). In order to obtain high levels of discrimination, typically more than 10 STRs systems are employed. The initial determination of parentage is made based upon whether or not paternal obligatory alleles were found between the child and the alleged father at the STR markers examined (47). If no allele is shared between the putative father and child at more than two STR loci, the putative man may be excluded as being the biological father of the child (49). This assumes that it is exceedingly that two mutational events will have occurred, although this could be factored into a likelihood ratio.

In 1938 the theoretical procedures for calculating the strength of being a biological father, the paternity index (PI) was first developed in the publication of Essen-Moller (49). Since then it has become common to include a likelihood ratio into a paternity report that quantifies the information from DNA evidence under two competing hypothesizes if an inclusion is found. The role of the calculation is to aid in understanding the strength of the parentage relationship (47). The

paternity index (PI) is the ratio of two conditional probabilities, represented in the formula X/Y, where the numerator (X) assumes that the putative father is the biological father, and the denominator (Y) that assumes that a random man from the same ethnicity is the biological father. Event Y is a chance event dependent upon the commonality of the shared allele, so the value of Y is dependent on the frequency of the allele in the population. The value of X is dependent upon the genotypes of the three persons tested but may typically take a value of 1, if the observed locus of the alleged father (AF) is homozygous, 0.5 is assigned if the AF is heterozygous, or 0.25 if the mother and child share are heterozygous and share the same alleles. These events are shown in box 1.

Box 1: Punnett square of the STR allele inheritance and relation to paternity index.

Locus: D3S1358.

Allele type and frequency:

allele	frequency	allele	frequency	allele	frequency
12	0.0013	13	0.0014	14	0.0391
15	0.3471	16	0.3106	17	0.2347
18	0.0593	19	0.0057	20	0.0007

Event 1:

		Allele from father A	
		14	15
Allele from mother B	12	12, 14	12, 15
	13	13, 14	13, 15

1. Mother B is genotype 12, 13 and child is genotype 13, 14, if A (14, 15) is the father, then mother must pass on allele 13, and the father must pass on allele 14.

If he is not the father then the mother still has to pass on allele 13 but some other man from random population should pass on allele 14.

2. If A is the biological father then 13, 14 can happen one in four ways, equals a probability of 0.25.
3. If A is not the biological father, then B still has to pass on allele 13 with probability of 0.5 and the chance that a random male other than A is the father is dependent on the frequency of allele 14 (f) in the population. This gives a likelihood ratio of PI= 0.25/ (0.5* 0.0391) = 1/ (2* 0.0391) = 12.78.

Event 2:

		Allele from father A	
		15	15
Allele from mother B	12	12, 15	12, 15
	13	13, 15	13, 15

1. If A is the biological father and since A is homozygous, the assortments of children can only be two kinds, 13, 15 with a probability of 0.5.
2. If A is not the father the mother must pass on 13 with probability of 0.5 and the probability that a male other than A can be a father depends the frequency of 15.
3. The likelihood ration of PI= 0.5/(0.5* 0.3471) = 2.88

Event 3:

		Allele from father A	
		12	15
Allele from mother B	12	12, 12	12, 15
	13	12, 13	13, 15

1. Since the child is 12, 13, the mother must pass on 13, if A is the father, the assortments of children STR appears in four ways, the 12, 13 with a probability of 0.25.
2. If A is not the biological father, then either 12 or 13 in 12, 13 can be from mother (0.5 of 12 and 0.5 of 13), meaning the chance for 12 and 13 coming from a random man other than A is 0.5 * frequency of 12 + 0.5* frequency of 13.
3. This gives a likelihood ratio of PI= 0.25/ (0.5* 0.0013+ 0.5* 0.0014) = 1/ 2(0.0013+ 0.0014) = 185.18.

Calculation process and description of events based on the text in (52), and other allele combination events and equivalent formula can be accessed also.

A population database containing frequency distributions of various alleles for the tested STR markers, such as that described in appendix 1, is used to calculate the probability of a randomly selected man passing the obligatory alleles to the child (47, 49).

As all the STR loci are assumed to be inherited independently, the individual PI values are multiplied together to obtain the combined paternity index (CPI). Some laboratories use a threshold value, termed a cut off value, to assume genetic linkage if the CPI is over preset value. There are various cutoff values for different laboratories, ranging from "0" to "10,000" (44). A generally accepted minimum standard for an inclusion of paternity is a CPI of 100 or greater; this was originally suggested by Coleman and Swenson in 2000 (49). With *a priori* probability of 0.5, a CPI of 100 indicates that the putative father is 100 times more likely to be the biological father of a child compared to a random man. The strength of the evidence can be transformed into percentage by using Formula (1)

$$W = CPI/(CPI + 1) \quad \text{---------- (1)}$$

to convert the likelihood ratio to a percentage probability of paternity (47). In this case it would be reported as a 99.0% probability of paternity. See box 2 for a

worked example.

Box 2: A standard worksheet for a standard Chinese family.

	Mother		Child		Father		formula	p=	q=	PI=
STR D3S1358	17	17	17	17	16	17	1/(2.q)	0.0000	0.2347	2.13
STR vWA	16	17	16	17	17	19	1/(2.p+2.q)	0.1575	0.2467	1.24
STR FGA	23	24	23	24	23	24	1/(p+q)	0.1738	0.2151	2.57
STR TH01	8	9	9	9	8	9	1/(2.q)	0.0000	0.4748	1.05
STR TPOX	8	11	8	11	8	8	1/ (p+q)	0.2869	0.5546	1.19
STR CSF1PO	11	12	12	12	12	13	1/(2.q)	0.0000	0.3647	1.37
STR D5S818	10	10	10	12	11	12	1/(2.q)	0.0000	0.2145	2.33
STR D13S317	8	9	9	11	9	11	1/(2.q)	0.0000	0.2399	2.08
STR D7S820	9	13	9	13	9	10	1/(2.p+2.q)	0.0360	0.0668	4.86
STR D8S1179	11	12	12	13	13	15	1/(2.q)	0.0000	0.2080	2.40
STR D21S11	28	33	28	33	28	31	1/(2.p+2.q)	0.0520	0.0485	4.98
STR D18S51	15	15	12	15	12	15	1/(2.q)	0.0000	0.0386	12.95
STR D16S539	11	12	9	11	9	14	1/(2.q)	0.0000	0.2657	1.88
								CPI=	80,087.22	

p = 0.0000 means the frequency of p is not necessary.

In a number of instances a complete parentage trio is not always available. Sometimes a sample from either of the parents is not available. This may be because one of them refused to provide a sample, one of them is missing or one of them is deceased. A mother-not-tested (MNT) case, also termed parentage duo cases, will lead to less statistical certainty (44, 50). A relationship between a child and putative father can still be assessed to a reasonable degree of confidence if the frequencies of shared alleles are rare. In such instances the addition of a DNA profile from a close relative can provide further confirmation. When testing a duo

case, the degree of confidence (or specificity) of the CPI still needs to be evaluated. This is particularly the case when the CPI is smaller than the preset cutoff values. There are few reports of such cases (53).

Genetic linkage of putative relatives is necessary in missing person cases, mass fatality victim identification events, and mass graves. The specimens and reference samples for examination may be the same as a standard trio met in a parentage testing. If this is not the case then a sample from only a single parent (child) or sibling is often compared (35, 56), making the determination of the relationship more challenging and in line with a MNT instance. If the samples are old or 'environmentally affected' they may produce degraded DNA, leading to reduce of DNA information, which would further reduce the value of the DNA test. A floating CPI requirement following different number of loci used for testing, and their corresponding specificities can be identified under such scenarios.

1.3 Other kinship testing: Siblingship and half-siblingship testing

In certain Disaster Victim Identification (DVI) instances neither of the biological parents of the two individuals is available for testing (56). Comparison of two DNA profiles from putative siblings (siblingship) or half siblings where

only one parent is in common (half-siblingship) can be undertaken. It would be expected that two siblings have a 0.25 chance of sharing both alleles a 0.5 chance of sharing one allele and a 0.25 chance of sharing no alleles. A probability of siblingship can be determined based upon the frequency of the matching alleles in the two participants. The resulting siblingship index (SI) should increase if genetic loci with higher powers of discrimination are examined. An SI can therefore be used to determine the probability of a sibling genetic relationship between two individuals. The SI is generated from a likelihood ratio to evaluate the hypothesis that the evidence profile is from two siblings compared to the hypothesis from the two are not siblings (57).

Two half-siblings are defined as two individuals sharing one parent only. On occasion it may be necessary to determine whether two DNA samples have come from two half-siblings. The confidence of such a DNA test, if there are matching alleles at the loci tested, will also increase with the number of loci examined. The confidence will also increase if there are multiple family members that can be examined. The probability indicating half-siblingship is based on the allele frequencies for any alleles that are shared between the two tested individuals is framed under the two hypotheses that either they are half-siblings or they are not

half-siblings. From this likelihood ratio calculation a combined half-sibship indices (CHSI) can be reported (57). There remains greater uncertainty for half-sibling relationships than that for full siblingship due to the chance that two half-siblings can share no alleles in common at a locus with double chance (0.5 compared to 0.25).

1.4 Mass kinship matching in mass fatalities identification

Recently there have been many natural and man-made disasters, such as earthquakes, tsunami, aviation crashes, maritime disasters etc. that resulted in hundreds or thousands of fatalities. In many cases mass fatality incidents (MFI) lead to severe human remains fragmentation and a corresponding large number of reference samples for comparison (32, 58, 59, 60, 56, 61, 62, 63, 64). Some of the prominent disaster events in recent 10 years are shown in Table 1.5. In a number of these incidents DNA tests and comparison were made (65) to identify some of the human remains.

Table 1.9: Different kinds of disasters with top 3 fatalities in 10 years.

Fatalities	Events	Place	Year
Earthquake			
283,100	Earthquake/tsunami	Indian Ocean	2004
86,100	Kashmir earthquake	Pakistan	2005

Continues

Table 1.5 (continued)

69,181	Sichuan earthquake	Eastern Sichuan, China	2008
Aviation			
2,998	September 11 attacks	New York, Pennsylvania and Washington D.C.	2001
276	Air Force of the Army of the Guardians of the Islamic Revolution Ilyushin Il-76	Sirach Mountains	2003
265	American Airlines Flight 587	New York City	2001
Maritime			
1,863	MV *Joola*	Senegal	2002
1,018	MV *al-Salam Boccaccio 98*	Red Sea	2006
800	*MV Princess of Stars* capsized by Typhoon Fengshen off Sibuyan Island	Philippines	2008
Explosives			
1,000	Ammunition dump fire	Lagos, Nigeria	2002
234	PetroChina Chuandongbei natural gas field explosion	Guoqiao, Kai, Chongqing, China	2003
Industrial accidents			
234	PetroChina Chuandongbei natural gas field explosion	Guoqiao, Kai, Chongqing, China	2003
167	Piper Alpha oil rig disaster	North Sea	1988
Coal mine			
214	coal mine	Sunjiawan, Fuxing, Liaoning, China	2005
181	coal mine with flooding	Huayuan, Xintai, Shandong, China	2007
166	coal mine	Chenjiashan, Tongchuan, Shanxi, China	2004
*Data were obtained from reference (66).			

In order to identity human remains from the mass fatalities, ante mortem data may be used first. In some cases this may not be reliable and hence there is a

need for DNA testing. Comparison of DNA profiles to reference samples taken from personal items is undertaken routinely. If these are not available, or are questioned, then linkage to a genetic relative is necessary. This form of indirect matching lacks the level of confidence of matching a crime scene sample to a suspect, but has been used effectively in the past (56).

The establishment DNA databases and kinship matching mechanisms for missing persons identification has been such a successful development (67) that has led to increasing numbers of successful identification of missing people (68). National guidelines (69) and international (70, 71) guidelines have been developed to allow for uniformity in the DNA process.

The majority of the DNA based kinship matching of potential first degree genetic relatives are “blind hits”, when an allele is shared at all the loci tested. The chance that the two samples originated from first degree relatives can be determined compared to coming from two unrelated individuals (72, 73). This calculation can be incorporated into the DNA report and combined with any ante mortem records prior to the putative identification of the human remains (73).

When the number of missing people is in thousands then the number of family reference samples is also such a large number. In many cases the linkage of

the samples to a family member is the only means of identification (63, 73, 74, 75, 76). In such cases, coincidental matches even when using 15 STR loci are possible.

To minimise the number of false inclusions, multiple family members have to be tested for multiple kinship matching. Alternatively extended DNA testing including additional autosomal STR loci, mitochondrial DNA typing or Y chromosome STR typing may be needed. Multiple family members of first-degree consanguinity are not always available and multiple DNA profiling techniques may not be applicable. For this reason one aim of the thesis is to establish cutoff values for use in kinship testing. In the first instance first degree parent child combinations will be considered followed by two potential siblings and then half-siblings to mimic real case scenarios (78). Some recommendation to help determine these types of genetic relatives are proposed.

The aims of the thesis are:

I. Achieve a better understanding of the specificity and sensitivity related to the indexes calculated from the standard kinship testing formulae.

II. Improve the method to determine the siblingship relationship when traditional index measuring methods are not effective in low index cases.

III. Increase the specificity of the DNA test from non-optimal DNA samples by developing a new triplex.

Chapter 2: Paternity duo test

2.1 Introduction

In a survey conducted by American Association of Blood Banks (AABB), an apparent increase in the number of duo cases has been found (44). A duo case occurs when a DNA sample can be obtained from only one parent for comparison to one DNA profile obtained from one offspring. Based upon the data obtained a Combined Paternity Index (CPI) can be determined to report on the probability of a first degree genetic relationship. For many of the laboratories involved in paternity testing a minimal CPI requirement is preset for the purposes of determining a first degree genetic linkage.

In the same report by the AABB (44), the minimum CPI value that was required for determining a duo case varied from 100 (25 out of the 39 laboratories) to 10,000 (1 out of the 39 laboratories). CPI cutoff values varied from "whatever is obtained" (17 out of the 35 laboratories) to 1,000 (1 out of the 35 laboratories) for use in family reconstruction cases. When a CPI value greater than the cutoff figure is obtained there is a higher degree of confidence that the samples tested are from first degree relatives than being unrelated. There is still the chance that two unrelated people will share one allele at all the 15 loci tested, resulting in a possible false

inclusion (75, 77). The use of fewer loci increases the opportunity for more false inclusions; this has been reported previously (78, 79, 80, 81, 82). Only a limited study on the evaluation of specificity of some CPI cutoff values has been reported previously (82). This chapter will examine the specificity and sensitivity of the current CPI cutoff values and will propose CPI cutoff values resulting 99% to 99.99% specificity of the duo paternity test.

2.2 Materials and Methods

DNA profiles from 450 members of the Chinese population were obtained using the ABI AmpF*l*STR Identifiler® PCR Amplification Kit (Applied Biosystems, Foster City, CA, USA). The STR products were analysed with an ABI Prism 3100 XL Genetic Analyzer. All the 450 copies of the 15 STR profiles were added to the Microsoft Excel data sheets and processed by a built-in Visual Basic programme written by authors of this study (programme codes are described in Table 2.1). Every individual of the population was paired with every other individual to form pairs, e.g. for the Chinese population in this study (450 x 449)/2 equalling 101,025 pairs were made. Every pair was processed by the same programme to have two children to be used for counting two duos CPI following a

standard formula (48). Totally 202,050 CPIs were obtained.

Specific loci of the 15 STR core kit were removed to simulate the STR profiling data obtained from kits using fewer STR loci, e.g. D2S1338 and D19S433 were omitted to mimic the CODIS 13 system (the loci in Profiler Plus kit add Cofiler kit). Additional loci were removed to mimic commercial kits with fewer loci (different kits with different number of loci is described Table 1).

To simulate DNA degradation, the loci creating the larger PCR products were taken off one by one, e.g. starting from the loci CSF1PO in the Identifiler kit, and then D2S1338, D18S51, FGA, D7S820 , D16S539 …to ultimately obtain only 8 loci.

From the 101,025 pairs generated from 450 members of the Chinese population picked at random, CPIs values were obtained for those pairs with coincidental matches where at least one allele was shared at all loci. The data were treated in the same manner for full profiles as well as in the reduced loci and degradation studies.

DNA profiles from the 15 STR loci from a Caucasian (n = 301) and American African (n = 256) population were obtained from the Short Tandem Repeat DNA Internet Database (23). These profiles were also processed by the

same programme to generate real duos and to determine the number of coincidental matching pairs (CMPs). There were 90,300 real duos generated for CPI counting and 45,150 random pairs analysed for CMPs for the Caucasian population. For the African American population 65,280 real duos were made and 32,640 pairs used to generate the number of CMPs. Inevitably for computer based populations no account of substructure is made and mating occurs randomly.

The allele frequency table used for the Chinese population was from previous studies (83, 84, 85). The Caucasian and African Americans allele frequencies were downloaded from the same origin as their DNA profiles (23). All the frequencies of alleles were adjusted by using the 5/2N rule (49).

The specificity of the duo test was calculated as 1 – the % of false inclusions and sensitivity of the CPI was based upon 1 – the % of false negatives (86). The rate of false positives, seen as a false inclusion in a paternity test, was determined by the percentage of pairs sharing STR loci coincidentally and when their CPI values were greater than any preset minimum CPI requirements. The percentage of false negatives (seen as an exclusion of paternity) was the proportion of real duos that might not be recognized as real duos based upon their CPIs lower than the preset cutoff values.

Table 2.1: Main programme (top modules) for generating real duos and processing coincidental matched pairs

(the complete programme was attached as appendix 4).

```
Sub duo_cpi_main_Ch( )
'function: make every pair of STR profiles from random population to
'generate two children STR profiles, calculate the CPI and
'record them into the worksheets-duo-cpis
'loading the frequency table from worksheets
Dim freqtab(54, 15)
Worksheets("freq tab").Select
Range("A4").Select
For i = 0 To 54
    For j = 0 To 15
        freqtab(i, j) = ActiveCell.Offset(i, j)
    Next j
Next i
'loading all the random population
Dim randompopulation(999, 30)
id_columns = 31
Worksheets("Ch").Select
individuals = Range("C1")
Range("A3").Select
For i = 0 To individuals - 1
    For j = 0 To id_columns - 1
        randompopulation(i, j) = ActiveCell.Offset(i, j)
    Next j
Next i
'generate all couples
Dim cpi_base(203000)
cpis = 0
Dim parent(1, 30)
For i = 0 To individuals - 2
    For j = i + 1 To individuals - 1
        For k = 0 To id_columns - 1
```

```
                parent(0, k) = randompopulation(i, k)
                parent(1, k) = randompopulation(j, k)
            Next k
            Call get2children_Ch(freqtab, parent, cpi_base, cpis)
            cpis = cpis + 1
        Next j
Next i
Worksheets("real duo cpis").Select
Range("A2").Select
Lines = 0
nextcolumns = 0
For i = 0 To cpis * 2 - 1
        ActiveCell.Offset(Lines, nextcolumns) = cpi_base(i)
        Lines = Lines + 1
        If Lines Mod 60000 = 0 Then
            Lines = 0
            nextcolumns = nextcolumns + 1
        End If
Next i
Range("A1").Select
End Sub
```

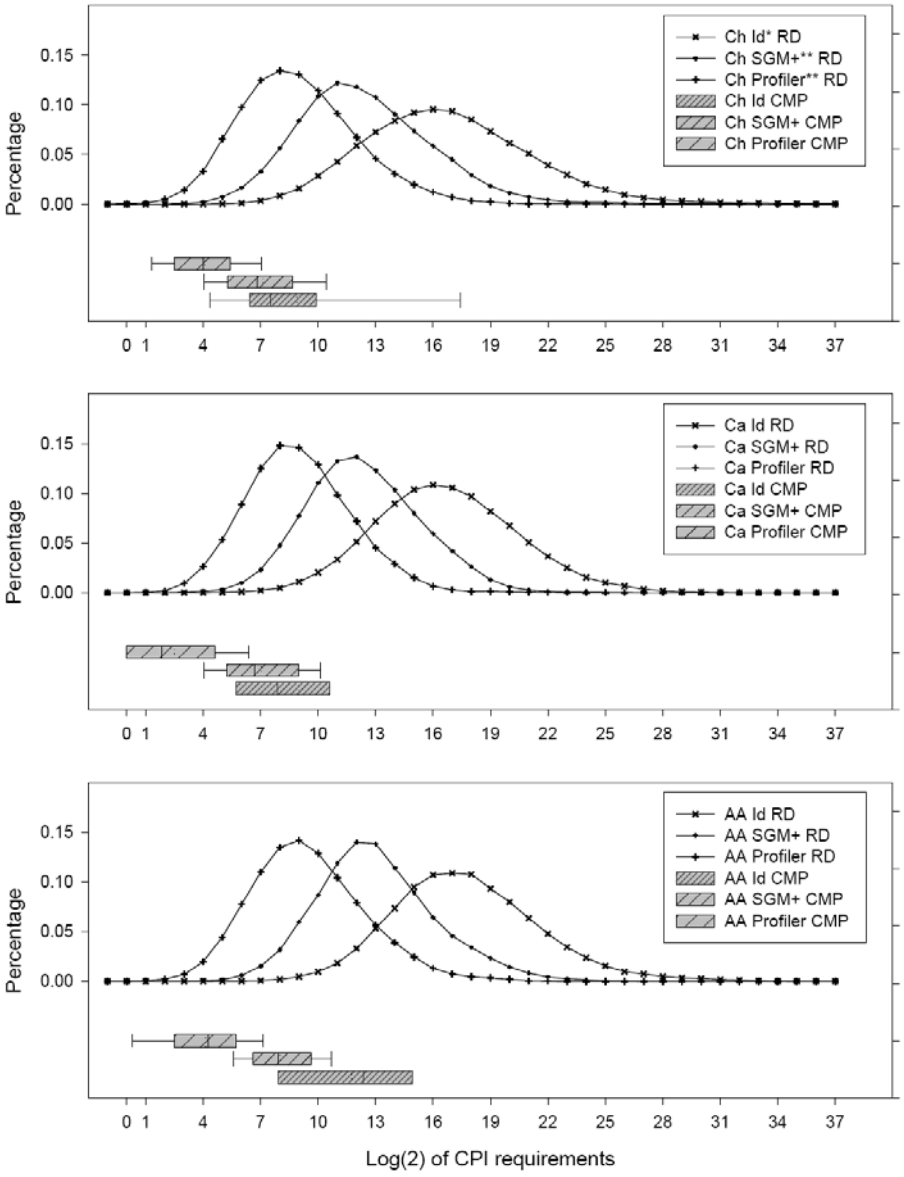
2.3 Results and Discussion

Distribution of CPI of Real Duos (RD) and Coincidental Matched Pairs (CMP)

The ratio distribution of CPI values of 202,025 Chinese Real Duos (RDs), 90,300 Caucasian RDs and 65,280 RDs for the STR loci in the Identifiler®, SGM plus® and Profiler® kits are shown in figure 2. The box-plots for the distribution of CPIs of CMPs are also illustrated. When there

is an increase in the number of STR loci used, larger CPIs distributions are obtained. These results are as expected. There were no clear boundaries to differentiate the distribution of RDs and CMPs when smaller CPI cutoff values are set.

Figure 3.1: Ratio distribution and box plot of CPI for 3 systems and 3 populations (Ch: Chinese, Ca: Caucasian, AA: African Americans, *:Identifiler, **:SGM plus and ***: Profiler are PCR Amplification kits made by Applied Biosystem, Foster City, LA, USA).



RD: real duo

CMP: Coincidental Mathed Pair

* there are two boxplots without tails owing to less samples

For many relationship testing laboratories, a minimum CPI value requirement or cutoff point is used for determining the existence of a first degree genetic relationship. Due to the RDs with low CPI values and CMPs with high CPI values observed in this study, false exclusions could occur by setting the cutoff higher than required. Equally a false inclusion could occur if these values are set too low. Further evaluation of specificity and sensitivity of different CPI cutoff values under different situations was undertaken.

Number of Real Duos and Coincidental Matched Pairs that Meet the Different CPI Requirements

Twenty five CMPs were observed in the 101,025 random pairs from the 450 members of the Chinese population taken at random using 15 STR loci. Eight CMPs for the Caucasian population (from 45,150 pairs) and 4 CMPs (from 32,540 pairs) for the African American population were obtained, as shown in Table 2.2. When using STR kits which amplify fewer loci than the 15 used Identifiler, more CMPs were found. The nine loci in the Profiler® kit resulted in 1,474 CMPs for the Chinese population, 522 CMPs for the Caucasian population and 315 CMPs for African American population.

Specificity and Sensitivity of different CPI Requirements

In Table the specificity and sensitivity for duos using varying numbers of STR loci are illustrated. As the CPI cutoff value increased, there was a subsequent decrease in sensitivity

and an increase in specificity. Using 15 STR loci the CPI cutoff value varying from 0 (stated as "whatever is obtained") to 10,000, then the sensitivity varied from 100% to 74.48% and the specificity from 99.98% to 100.00% for the Chinese population. The change was more pronounced with fewer STR loci confirming the findings of a previous study (82).

If a minimum CPI value of 100 was chosen using 15 STR loci, the specificity versus sensitivity for the Chinese population was 99.98% versus 99.72%. When the same test was preformed on the Caucasian population it was 99.99% versus 99.76% and for the African American population it was 99.99% versus 99.96%. Again the percentages are reduced if fewer STR loci are used. If CPI = 100 for 7 loci (shown as the column 15-8 in Table 2.3), the specificity versus sensitivity for the Chinese population was 99.82% versus 42.83%. For the Caucasian population the same values were 99.85% versus 45.36% and for the African American population they were 99.79% vs. 56.01%. The conclusion from these data is that pre-setting a minimum CPI requirement may not be appropriate for all scenarios and may result in an increase in false inclusions and false exclusions.

Table 2.1: Number of RDs and CMPs that meet the minimum CPI requirements for 3 populations on 5 STR systems.

System name	Identifiler		CODIS 13		SGM+		Profiler+*		Profiler	
Minimum CPI requirement	RD	CMP	RD	CMP	RD	CMP	RD	CMP	RD	CMP
Population	Chinese									
What ever is obtained	0	25	0	98	0	162	0	434	0	1,474
10	2	25	105	90	160	153	712	392	5,585	919
100	571	17	4,564	47	8,864	87	21,629	138	59,456	176
101	576	17	4,620	47	8,982	87	21,894	137	59,810	175
150	1,172	14	7,424	37	13,743	75	31,661	103	74,918	133
200	1,743	12	10,104	30	18,248	67	39,780	82	86,242	95
300	3,015	11	15,182	28	26,104	51	53,589	56	102,081	63
500	5,552	9	23,358	17	38,745	36	72,804	35	121,301	31
1,000	11,161	5	38,076	10	60,505	22	100,041	22	144,491	9
1,001	11,172	5	38,101	10	60,552	22	100,074	22	144,521	9
10,000	51,557	3	104,456	2	137,984	2	168,727	4	188,703	1
number of RD	202,050									
number of CMP comparing	101,025									

RD: real duo

CMP: Coincidental Matched Pair

continues

Table 2.1 (continued)

System name	Identifiler		CODIS 13		SGM+		Profiler+*		Profiler	
Minimum CPI requirement	RD	CMP	RD	CMP	RD	CMP	RD	CMP	RD	CMP
Population	Caucasian									
What ever is obtained	0	8	0	31	0	67	0	159	0	522
10	0	8	11	30	46	64	118	151	1,657	357
100	218	4	1,135	20	2,425	34	7,724	73	23,007	72
101	220	4	1,148	20	2,467	34	7,809	73	23,202	72
150	389	4	2,007	15	4,087	28	11,640	53	30,178	55
200	584	4	2,793	10	5,714	27	15,586	39	35,813	41
300	1,005	4	4,472	10	8,875	21	21,273	24	43,731	29
500	1,748	4	7,267	8	14,407	16	30,441	14	53,622	15
1,000	3,477	2	12,980	4	24,232	8	43,631	4	65,215	7
1,001	3,485	2	12,991	4	24,244	8	43,647	4	65,237	7
10,000	20,183	1	44,910	0	63,181	1	78,648	0	86,155	0
number of RD	90,300									
number of CMP comparing	45,150									
System name	Identifiler		CODIS 13		SGM+		Profiler+*		Profiler	
Minimum CPI requirement	RD	CMP	RD	CMP	RD	CMP	RD	CMP	RD	CMP

RD: real duo

CMP: Coincidental Matched Pair

continues

Table 2.1 (continued)

System name	Identifiler		CODIS 13		SGM+		Profiler+*		Profiler	
Minimum CPI requirement	RD	CMP	RD	CMP	RD	CMP	RD	CMP	RD	CMP
Population	African Americans									
What ever is obtained	0	4	0	21	0	30	0	110	0	315
10	0	4	6	21	128	30	189	103	975	224
100	27	4	503	19	5,258	21	6,259	45	14,495	52
101	28	4	509	19	5,323	21	6,345	44	14,617	52
150	71	3	869	16	7,671	19	9,204	38	19,172	31
200	105	3	1,256	15	9,965	18	11,647	29	22,867	23
300	228	3	2,066	10	13,491	12	15,798	19	28,154	12
500	437	3	3,530	9	18,864	9	21,790	12	34,840	7
1,000	1,046	3	6,709	7	26,633	5	30,719	6	43,169	4
1,001	1,050	3	6,717	7	26,639	5	30,734	6	43,176	4
10,000	8,987	2	28,036	1	50,208	0	54,751	2	60,144	1
number of RD	65,280									
number of CMP comparing	32,640									

Profiler+*: abbreviation of Profiler Plus AmpliSTR kit from Applied Biosystem, USA.

RD: real duo

CMP: Coincidental Matched Pair

Table 2.2: Sensitivity and specificity versus minimum CPI requirements for 3 populations on 5 STR systems.

System name	Identifiler		CODIS 13		SGM+		Profiler+		Profiler	
Minimum CPI requiremer	spe*	sen**	spe	sen	spe	sen	spe	sen	spe	sen
Population	Chinese									
What ever is obtained	99.98%	100.00%	99.90%	100.00%	99.84%	100.00%	99.57%	100.00%	98.54%	100.00%
10	99.98%	100.00%	99.91%	99.95%	99.85%	99.92%	99.61%	99.65%	99.09%	97.24%
100	99.98%	99.72%	99.95%	97.74%	99.91%	95.61%	99.86%	89.30%	99.83%	70.57%
101	99.98%	99.71%	99.95%	97.71%	99.91%	95.55%	99.86%	89.16%	99.83%	70.40%
150	99.99%	99.42%	99.96%	96.33%	99.93%	93.20%	99.90%	84.33%	99.87%	62.92%
200	99.99%	99.14%	99.97%	95.00%	99.93%	90.97%	99.92%	80.31%	99.91%	57.32%
300	99.99%	98.51%	99.97%	92.49%	99.95%	87.08%	99.94%	73.48%	99.94%	49.48%
500	99.99%	97.25%	99.98%	88.44%	99.96%	80.82%	99.97%	63.97%	99.97%	39.96%
1,000	100.00%	94.48%	99.99%	81.16%	99.98%	70.05%	99.98%	50.49%	99.99%	28.49%
1,001	100.00%	94.47%	99.99%	81.14%	99.98%	70.03%	99.98%	50.47%	99.99%	28.47%
10,000	100.00%	74.48%	100.00%	48.30%	100.00%	31.71%	100.00%	16.49%	100.00%	6.61%
Population	Caucasian									
What ever is obtained	99.98%	100.00%	99.93%	100.00%	99.85%	100.00%	99.65%	100.00%	98.84%	100.00%
10	99.98%	100.00%	99.93%	99.99%	99.86%	99.95%	99.67%	99.87%	99.21%	98.17%
100	99.99%	99.76%	99.96%	98.74%	99.92%	97.31%	99.84%	91.45%	99.84%	74.52%
101	99.99%	99.76%	99.96%	98.73%	99.92%	97.27%	99.84%	91.35%	99.84%	74.31%
150	99.99%	99.57%	99.97%	97.78%	99.94%	95.47%	99.88%	87.11%	99.88%	66.58%
200	99.99%	99.35%	99.98%	96.91%	99.94%	93.67%	99.91%	82.74%	99.91%	60.34%
300	99.99%	98.89%	99.98%	95.05%	99.95%	90.17%	99.95%	76.44%	99.94%	51.57%
500	99.99%	98.06%	99.98%	91.95%	99.96%	84.05%	99.97%	66.29%	99.97%	40.62%
1,000	100.00%	96.15%	99.99%	85.63%	99.98%	73.17%	99.99%	51.68%	99.98%	27.78%
1,001	100.00%	96.14%	99.99%	85.61%	99.98%	73.15%	99.99%	51.66%	99.98%	27.76%
10,000	100.00%	77.65%	100.00%	50.27%	100.00%	30.03%	100.00%	12.90%	100.00%	4.59%
Population	African Americans									
What ever is obtained	99.99%	100.00%	99.94%	100.00%	99.91%	100.00%	99.66%	100.00%	99.03%	100.00%
10	99.99%	100.00%	99.94%	99.99%	99.91%	99.80%	99.68%	99.71%	99.31%	98.51%
100	99.99%	99.96%	99.94%	99.23%	99.94%	91.95%	99.86%	90.41%	99.84%	77.80%
101	99.99%	99.96%	99.94%	99.22%	99.94%	91.85%	99.87%	90.28%	99.84%	77.61%
150	99.99%	99.89%	99.95%	98.67%	99.94%	88.25%	99.88%	85.90%	99.91%	70.63%
200	99.99%	99.84%	99.95%	98.08%	99.94%	84.73%	99.91%	82.16%	99.93%	64.97%
300	99.99%	99.65%	99.97%	96.84%	99.96%	79.33%	99.94%	75.80%	99.96%	56.87%
500	99.99%	99.33%	99.97%	94.59%	99.97%	71.10%	99.96%	66.62%	99.98%	46.63%
1,000	99.99%	98.40%	99.98%	89.72%	99.98%	59.20%	99.98%	52.94%	99.99%	33.87%
1,001	99.99%	98.39%	99.98%	89.71%	99.98%	59.19%	99.98%	52.92%	99.99%	33.86%
10,000	99.99%	86.23%	100.00%	57.05%	100.00%	23.09%	99.99%	16.13%	100.00%	7.87%

spe*: specificity.

sen**: sensitivity.

Table 2.3: Sensitivity and specificity versus minimum CPI requirements mimicking degradation from long loci for 3 populations on 8 degradation situation.

Degradation situation*	15-0		15-1		15-2		15-3		15-4		15-5		15-6		15-7		15-8	
Minimum CPI requirement	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen
Population	Chinese																	
What ever is obtained	99.98%	100.00%	99.97%	100.00%	99.92%	100.00%	99.84%	100.00%	99.62%	100.00%	99.37%	100.00%	98.97%	100.00%	97.91%	100.00%	96.38%	100.00%
10	99.98%	100.00%	99.97%	100.00%	99.93%	99.96%	99.86%	99.87%	99.69%	99.35%	99.52%	98.01%	99.27%	96.46%	98.79%	93.18%	98.31%	87.63%
100	99.98%	99.72%	99.98%	99.46%	99.96%	98.02%	99.93%	95.56%	99.88%	89.23%	99.83%	80.44%	99.81%	72.20%	99.79%	56.66%	99.82%	42.83%
101	99.98%	99.71%	99.98%	99.45%	99.96%	97.98%	99.93%	95.52%	99.88%	89.16%	99.84%	80.31%	99.81%	72.03%	99.80%	56.49%	99.83%	42.56%
150	99.99%	99.42%	99.98%	99.00%	99.96%	96.86%	99.94%	93.34%	99.91%	85.23%	99.88%	74.70%	99.86%	65.27%	99.88%	48.37%	99.91%	34.59%
200	99.99%	99.14%	99.99%	98.51%	99.97%	95.74%	99.95%	91.34%	99.92%	81.78%	99.90%	70.37%	99.90%	60.22%	99.92%	42.72%	99.94%	29.26%
300	99.99%	98.51%	99.99%	97.59%	99.98%	93.70%	99.96%	87.85%	99.95%	76.46%	99.93%	63.71%	99.94%	52.72%	99.95%	35.01%	99.96%	22.63%
500	99.99%	97.25%	99.99%	95.79%	99.98%	90.28%	99.97%	82.61%	99.96%	68.87%	99.97%	55.01%	99.97%	43.76%	99.98%	26.32%	99.99%	15.59%
1,000	100.00%	94.48%	99.99%	92.08%	99.99%	84.05%	99.99%	73.80%	99.98%	57.80%	99.98%	43.29%	99.98%	32.27%	99.99%	16.74%	99.99%	9.14%
1,001	100.00%	94.47%	99.99%	92.07%	99.99%	84.04%	99.99%	73.79%	99.98%	57.78%	99.98%	43.27%	99.98%	32.25%	99.99%	16.73%	99.99%	9.13%
10,000	100.00%	74.48%	100.00%	68.44%	100.00%	53.44%	100.00%	38.35%	100.00%	22.93%	100.00%	13.52%	100.00%	7.97%	100.00%	3.05%	100.00%	1.43%
Population	Caucasian																	
What ever is obtained	99.98%	100.00%	99.98%	100.00%	99.93%	100.00%	99.84%	100.00%	99.65%	100.00%	99.40%	100.00%	99.03%	100.00%	98.14%	100.00%	96.52%	100.00%
10	99.98%	100.00%	99.98%	99.99%	99.93%	99.99%	99.84%	99.91%	99.65%	99.64%	99.40%	99.01%	99.03%	98.32%	98.14%	96.04%	98.43%	91.49%
100	99.99%	99.76%	99.98%	99.67%	99.93%	98.88%	99.84%	96.57%	99.65%	92.24%	99.40%	86.14%	99.03%	78.37%	98.14%	62.89%	99.85%	45.36%
101	99.99%	99.76%	99.98%	99.67%	99.93%	98.86%	99.84%	96.53%	99.65%	92.15%	99.40%	86.01%	99.03%	78.18%	98.14%	62.67%	99.86%	45.09%
150	99.99%	99.57%	99.98%	99.39%	99.93%	98.08%	99.85%	94.69%	99.71%	88.67%	99.56%	80.86%	99.34%	71.44%	98.93%	53.95%	99.90%	36.32%
200	99.99%	99.35%	99.99%	99.10%	99.96%	97.37%	99.92%	93.00%	99.89%	85.69%	99.87%	76.62%	99.84%	66.13%	99.85%	47.60%	99.94%	30.66%
300	99.99%	98.89%	99.99%	98.56%	99.96%	95.87%	99.92%	89.99%	99.89%	80.50%	99.88%	69.74%	99.84%	58.00%	99.85%	38.72%	99.97%	23.39%
500	99.99%	98.06%	99.99%	97.46%	99.97%	93.25%	99.94%	84.78%	99.92%	72.61%	99.90%	60.26%	99.89%	47.74%	99.91%	28.63%	99.99%	16.16%
1,000	100.00%	96.15%	99.99%	94.88%	99.97%	87.79%	99.95%	75.42%	99.94%	60.51%	99.93%	46.59%	99.91%	34.61%	99.94%	17.83%	100.00%	9.02%
1,001	100.00%	96.14%	99.99%	94.87%	99.98%	87.78%	99.96%	75.41%	99.96%	60.49%	99.94%	46.56%	99.95%	34.60%	99.96%	17.80%	100.00%	9.02%
10,000	100.00%	77.65%	99.99%	72.21%	99.99%	54.31%	99.97%	36.02%	99.97%	21.27%	99.97%	12.59%	99.97%	7.10%	99.98%	1.80%	100.00%	0.86%
Population	African Americans																	
What ever is obtained	99.99%	100.00%	99.98%	100.00%	99.95%	100.00%	99.87%	100.00%	99.75%	100.00%	99.53%	100.00%	99.14%	100.00%	98.27%	100.00%	97.58%	100.00%
10	99.99%	100.00%	99.98%	100.00%	99.95%	99.99%	99.88%	99.95%	99.77%	99.73%	99.61%	99.51%	99.36%	99.00%	98.97%	96.77%	98.65%	95.34%
100	99.99%	99.96%	99.98%	99.90%	99.97%	99.32%	99.93%	97.93%	99.87%	94.06%	99.86%	89.88%	99.79%	83.04%	99.77%	67.31%	99.79%	56.01%
101	99.99%	99.96%	99.98%	99.89%	99.97%	99.31%	99.93%	97.90%	99.87%	93.97%	99.86%	89.79%	99.79%	82.96%	99.77%	66.98%	99.79%	55.78%
150	99.99%	99.89%	99.98%	99.73%	99.97%	98.79%	99.94%	96.69%	99.90%	91.07%	99.89%	85.65%	99.84%	76.97%	99.84%	58.77%	99.88%	45.96%
200	99.99%	99.84%	99.98%	99.63%	99.97%	98.31%	99.95%	95.35%	99.92%	88.46%	99.91%	81.91%	99.88%	72.03%	99.89%	52.36%	99.92%	39.17%
300	99.99%	99.65%	99.98%	99.24%	99.98%	97.34%	99.95%	93.11%	99.94%	83.90%	99.94%	76.10%	99.92%	64.58%	99.92%	43.99%	99.96%	30.37%
500	99.99%	99.33%	99.98%	98.58%	99.98%	95.59%	99.97%	89.11%	99.95%	76.90%	99.97%	67.51%	99.96%	54.62%	99.95%	34.32%	99.97%	21.38%
1,000	99.99%	98.40%	99.99%	96.81%	99.99%	91.63%	99.97%	81.58%	99.98%	65.71%	99.98%	54.81%	99.98%	41.16%	99.98%	22.80%	99.99%	12.29%
1,001	99.99%	98.39%	99.99%	96.81%	99.99%	91.61%	99.97%	81.56%	99.98%	65.69%	99.98%	54.78%	99.98%	41.15%	99.98%	22.79%	99.99%	12.27%
10,000	99.99%	86.23%	99.99%	79.14%	100.00%	62.70%	100.00%	43.53%	100.00%	27.01%	100.00%	18.41%	100.00%	10.62%	100.00%	3.37%	100.00%	1.52%

Degradation situation*: degradation of loci following the sequence of CSF1PO, D2S1338, D18S51, FGA, D7S820, D16S539, D21S11, D13S317 in the Identifiler kit.

Evaluation of CPI cutoff based on Different Specificity in different situation

By setting the specificity at 99% (equivalent to index of 100) , 99.9% and 99.99%, the correlated CPI cutoff could be calculated for varying number of STR loci (shown in Table and Table). A specificity set at 99.99% using the 15 STR loci was obtained with a CPI of 139 for the Chinese population, 94 for the Caucasian population and "whatever is obtained" for the African American population. When using the CODIS 13 STRs with the same specificity of 99.99% the CPI cutoff for the Chinese population was 628, 565 for the Caucasian population and 2,024 for the African American population. The sensitivity would be reduced to 86.29%, 91.06%, and 82.03% for the three populations respectively.

By using the optimal cutoff selection method proposed by Zou (111), the optimal CPI cutoff that maximizes the specificity and sensitivity for all the situations was "CPI = 1" (data not shown). This is in line with the findings of similar low CPI cutoff values reported by a previous study using only a few hundred real duos (82). For automatic database searching, the efficiency of the test depends on how many candidate pairs that are picked by the computer programme that need further confirmation. Ultimately it is the role of the decision-maker to set the cutoff when balancing the specificity and sensitivity as improving the capability of one will result in the decrease in the capability of the other. No CPI minimum value can be set for paternity duo tests that will never result in false inclusions or false exclusions. The laboratories that joined the AABB survey strongly supported that testing without a mother should be

processed only when the mother is unavailable or if she is deceased. Otherwise, every effort should be made to include the mother in the test (44).

Table 2.4: Predicted minimum CPI requirements of specificity of 99.90% and 99.99% for 3 populations and 5 STR systems.

System name	Identifiler		CODIS 13		SGM+		Profiler+		Profiler	
Minimum CPI requirement	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen
Population	Chinese									
What ever is obtained	99.98%	100.00%	99.90%	100.00%	99.84%	100.00%	99.57%	100.00%	-	-
7.7	-	-	-	-	-	-	-	-	99.00%	98.12%
62	-	-	-	-	99.90%	97.56%	-	-	-	-
139	99.99%	99.49%	-	-	-	-	-	-	-	-
145	-	-	-	-	-	-	99.90%	84.76%	-	-
182	-	-	-	-	-	-	-	-	99.90%	59.09%
628	-	-	99.99%	86.29%	-	-	-	-	-	-
680	-	-	-	-	-	-	-	-	99.99%	34.61%
1,381	-	-	-	-	99.99%	64.42%	-	-	-	-
1,433	-	-	-	-	-	-	99.99%	43.74%	-	-
Population	Caucasian									
What ever is obtained	99.98%	100.00%	99.93%	100.00%	99.85%	100.00%	99.65%	100.00%	-	-
4.2	-	-	-	-	-	-	-	-	99.00%	99.70%
44	-	-	-	-	99.90%	99.18%	-	-	-	-
94	99.99%	99.79%	-	-	-	-	-	-	-	-
166	-	-	-	-	-	-	99.90%	85.53%	-	-
171	-	-	-	-	-	-	-	-	99.90%	63.86%
565	-	-	99.99%	91.06%	-	-	-	-	-	-
759	-	-	-	-	-	-	99.99%	57.62%	-	-
1,010	-	-	-	-	-	-	-	-	99.99%	27.61%
1,109	-	-	-	-	99.99%	71.35%	-	-	-	-
Population	African Americans									
What ever is obtained	99.99%	100.00%	99.94%	100.00%	99.91%	100.00%	99.66%	100.00%	99.03%	100.00%
140	-	-	-	-	-	-	-	-	99.90%	71.96%
168	-	-	-	-	-	-	-	-	-	-
703	-	-	-	-	-	-	-	-	99.99%	40.01%
1,088	-	-	-	-	99.99%	57.85%	-	-	-	-
2,024	-	-	99.99%	82.03%	-	-	-	-	-	-
2,236	-	-	-	-	-	-	99.99%	37.53%	-	-

SPE: specificity, SEN: sensitivity

Table 2.5: Predicted minimum CPI requirements of specificity of 99.90% and 99.99% for 3 populations and 8 different degradation situation.

Degradation situation	15-0		15-1		15-2		15-3		15-4		15-5		15-6		15-7	
Minimum CPI requirement	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen
Population	Chinese															
What ever is obtained	99.98%	100.00%	99.97%	100.00%	99.92%	100.00%	99.84%	100.00%	99.62%	100.00%	99.37%	100.00%	-	-	-	-
1	-	-	-	-	-	-	-	-	-	-	-	-	99.00%	99.90%	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.00%	90.01%
20.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
37	-	-	-	-	-	-	99.90%	98.70%	-	-	-	-	-	-	-	-
124	-	-	-	-	-	-	-	-	99.90%	87.22%	-	-	-	-	-	-
139	99.99%	99.49%	-	-	-	-	-	-	-	-	-	-	-	-	-	-
135	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
186	-	-	99.99%	98.66%	-	-	-	-	-	-	99.90%	71.54%	-	-	-	-
170	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	45.98%
198	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	60.39%	-	-
707	-	-	-	-	99.99%	87.40%	-	-	-	-	-	-	-	-	-	-
945	-	-	-	-	-	-	99.99%	74.58%	-	-	-	-	-	-	-	-
749	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	20.40%
470	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,171	-	-	-	-	-	-	-	-	99.99%	55.25%	-	-	-	-	-	-
1,080	-	-	-	-	-	-	-	-	-	-	99.99%	42.05%	-	-	-	-
1,240	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	29.01%	-	-

continues

Table 2.5 (continued)

Degradation situation	15-0		15-1		15-2		15-3		15-4		15-5		15-6		15-7	
Minimum CPI requirement	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen
Population	Caucasian															
What ever is obtained	99.98%	100.00%	99.98%	100.00%	99.93%	100.00%	99.84%	100.00%	99.65%	100.00%	99.40%	100.00%	99.03%	100.00%	-	-
12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.00%	94.92%
19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
41	-	-	-	-	-	-	99.90%	98.87%	-	-	-	-	-	-	-	-
68	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
94	99.99%	99.79%	-	-	-	-	-	-	-	-	-	-	-	-	-	-
99	-	-	99.99%	99.68%	-	-	-	-	-	-	-	-	-	-	-	-
104	-	-	-	-	-	-	-	-	99.90%	91.91%	-	-	-	-	-	-
133	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	56.67%
137	-	-	-	-	-	-	-	-	-	-	99.90%	82.15%	-	-	-	-
142	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
157	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	70.76%	-	-
394	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
359	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
722	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	22.54%
793	-	-	-	-	99.99%	89.81%	-	-	-	-	-	-	99.99%	38.80%	-	-
915	-	-	-	-	-	-	-	-	-	-	99.99%	48.29%	-	-	-	-
1,181	-	-	-	-	-	-	-	-	99.99%	57.41%	-	-	-	-	-	-
1,246	-	-	-	-	-	-	99.99%	72.19%	-	-	-	-	-	-	-	-

continues

Table 2.5 (continued)

Degradation situation	15-0		15-1		15-2		15-3		15-4		15-5		15-6		15-7	
Minimum CPI requirement	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen
Population	African Americans															
What ever is obtained	99.99%	100.00%	99.98%	100.00%	99.95%	100.00%	99.87%	100.00%	99.75%	100.00%	99.53%	100.00%	99.14%	100.00%	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.00%	96.57%
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	-	-	-	-	99.90%	99.75%	-	-	-	-	-	-	-	-
139	-	-	-	-	-	-	-	-	99.90%	91.71%	-	-	-	-	-	-
156	-	-	-	-	-	-	-	-	-	-	99.90%	85.15%	-	-	-	-
161	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
222	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	70.18%	-	-
230	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	49.49%
724	-	-	99.99%	97.80%	-	-	-	-	-	-	-	-	-	-	-	-
881	-	-	-	-			-	-	-	-	-	-	-	-	-	-
982	-	-	-	-	99.99%	91.75%	-	-	-	-	-	-	-	-	-	-
1425	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	34.77%	-	-
1,438	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	17.66%
1,495	-	-	-	-	-	-	-	-	-	-	99.99%	47.47%	-	-	-	-
1,708	-	-	-	-	-	-	-	-	99.99%	56.26%	-	-	-	-	-	-
1,983	-	-	-	-	-	-	99.99%	71.70%	-	-	-	-	-	-	-	-

2.4 Conclusion

DNA laboratories using STR typing for relationship determination are suggested to apply suitable CPI cutoff values based upon the loci number of STR panel tested. Specificity values as shown in Table 2.4 had better be used in cases of low CPI values from non-optimal DNA samples. When establishing database for mass comparison such as for unidentified human remains or mass disasters, it is better to reference the data of this study to set the CPI policy for screening the potential matches.

Chapter 3: Siblingship test

3.1 Introduction

Siblingship determination is encountered in instances such as linking human remains to a relative (putative brother or sister) and when neither of the biological parents of the two individuals is available for testing. Theoretically at any one locus there is a 0.5 chance that two siblings will share one allele, a 0.25 chance that they will share neither alleles and a 0.25 chance that they will share both alleles, this sharing of alleles is termed two-allele-sharing-locus (TASL) (113). The chance that two unrelated individuals share either one or both alleles at any one locus is dependent upon the frequency of the alleles (112, 113). A probability of siblingship can be determined based upon the frequency of the matching alleles in the population and will increase when many loci are examined or loci with high powers of discrimination (113). There is an increase in uncertainty using DNA testing to resolve siblingship if the parents are heterozygous rather than homozygous (111, 117). There is reduced evidence of siblingship if within the loci tested there are no loci where both alleles are shared between the individuals, on the contrary confidence of siblingship is increased if the number of TASL increase (112).

Based upon the degree of sharing of alleles between two DNA profiles it is possible to determine a combined siblingship indices (CSI) (111). A CSI index less than 1 supports the premise that the two individuals tested are not siblings. If the index is over 1 then the data supports the existence of a sibling relationship. Other cutoff points have been recommended such as CSI $\geq$ 3 (112). These figures are guides as that were found in this study that 1.6% of random pairs of DNA

profiles had CSI greater than 1 when using 15 STR loci. In a previous study using 16 STR loci 0.1% of unrelated pairs of DNA profiles were found to have CSI > 100 and 0.01% for CSI > 1,000 (114). Using the 15 STR loci used in the Identifiler® kit, none of the non-sibling pairs were found with CSI ≥ 1, while all sibling pairs have CSI > 10 (115). In a different study using the same 15 STR loci, 6.06% of sibling pairs exhibited CSI < 1 and 9.1% of random unrelated pairs had CSI > 1 (116). These differences are best explained by the fact that different allele frequencies were used in this published study and the study in this thesis.

This study extended the scope of the experiments by using the 15 STR core loci to study three populations using 357,630 full sibling pairs and 178,815 non-sibling pairs generated from DNA profiles of random members of the three populations. Using this high number of sample pairs, it is possible to evaluate the specificity and sensitivity of the 15 STR loci for discriminating between full and non-siblings. Combining a determined CSI value with TASL counts further provides confidence in the specificity of the result.

3.2 Materials and Methods

The same 450 STR profiles of random members of the Taiwanese Chinese population were also processed by a Microsoft Excel Macros coded by Visual Basic programme (programme codes described in Table 3.1). Every member of the population was paired with every other member to form random pairs, e.g. for Chinese population in this study (450 x 449)/2, resulting in 101,025 pairs. Every pair was set to have two children, resulting in 202,050 sibling pairs being

generated. DNA profiles from the 15 STR loci from the Caucasian (n = 301) and the African American (n = 256) populations were processed in the same way as those of the Taiwanese population to generate sibling pairs (90,300 of Caucasians, 65,280 of African Americans) and random pairs (45,150 of Caucasians, 32,640 of African Americans). The Combined Siblingship Indices (CSI) was calculated for each simulated sibling pairs or random pairs by using standard formulae (111). By using built-in function provided by the software EXCEL, the formulae were transformed into cells of worksheet (Table 3.2). STR profiles from pairs were pasted onto the assigned area to produce the CSI value. The allele frequency tables used for calculation of CSI were adjusted by using 5/2N rule.

The rate of false negatives equalled the percentage of real siblingship testing cases (in this study the simulated sibling pairs) that would be excluded based upon any given cutoff point of CSI or TASL. The rate of false positives equalled the percentage of random pairs of DNA profiles where their CSI or TASL was greater than any recommended cutoff value. The sensitivity of the test is based upon 1 – the % of false negatives, the specificity of the test is based upon 1 – the % of false positives.

Table 3.1: Programme for generating real siblings and processing random pairs.

```
Sub sibling_generating_CSI_calculation_main( )
'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
'loading of the profiles of random population
'for Chinese=450
'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Dim Individual_str(449, 30)
Worksheets("Ch450").Select
    ' loading of the number of couples
    individuals = Range("C1")
    'columns of data= 31 for Identifiler
    infocells = 31
    Range("A3").Select
    'loading of array randompopulation
    For i = 0 To individuals - 1
        For J = 0 To infocells - 1
            Individual_str(i, J) = ActiveCell.Offset(i, J)
        Next J
    Next i
'call subroutine
Call calculate_si(Individual_str, individuals, infocells)
End Sub
```

Table 3.2: Calculation worksheet of CSI.

Description of Function:

- ✧ EXCEL worksheet is a table containing multiple cells, in each cell data or formula can be put. In this study, author designed a functional worksheet to calculate the CSI between two individuals automatically after posting the STR profiles.
- ✧ In this sheet the original STR profiles can be posted on the upper part of the worksheet, and the lower part was filled with built-in EXCEL functions and commands for choosing the right CSI index formula, and calculating the index.

This is the upper part of the worksheet:

(post the profiles into columns to get the CSI)

CSI= 7.85E+03

	individual 1		individual 2	
STR D8S1179	11	15	15	15
STR D21S11	29	32.2	30	32.2
STR D7S820	8	12	12	12
STR CSF1PO	9	10	11	13
STR D3S1358	15	17	15	18
STR TH01	9	9	9	9
STR D13S317	8	10	8	9
STR D16S539	9	12	9	12
STR D2S1338	20	22	22	23
STR D19S433	13	15	13	14
STR vWA	14	17	17	18
STR TPOX	8	14	8	14
STR D18S51	16	17	16	16
STR Sex	X	X	X	Y
STR D5S818	10	12	11	12
STR FGA	19	26	19	22

continues

Table 3.2 (continued)

STR D8S1179							k0		k1	k2	
Sibling Index =	10.0156	B$19			vlook col=	11		0.25	0.25	0.25	
		shared with ind 2 position		Hor ck	freq	a value	formula		0n/0ff		
check child homo	1st allele of C	0	0	0	0.0000	0.0128	(K2+2K1*a+k0*a*a)/a*a		0	0.0000	
hetero	2nd allele of C	17	17	17	0.0128	0.0000	(K1+k0*a)/a		0	0.0000	
check father homo	Ver ck	17	17	1		b value	K0		0	0.0000	
hetero			2	1	<=no. sha allele	0	(K2+k1*a+k1*b+k0*2*a*b)/2*a*b		0	0.0000	
						check for b	(k1+2*k0*a)/2a		1	10.0156	
						0		sum		10.0156	

The cell shows the shared allele

Frequency of allele obtained

Index formula chosen and index calculated

continues

Table 3.2 (continued)

=A$5				
Sibling Index =	=IF(T9,T9,1)	B$19		
		shared with ind 2 position		Hor ck
check child homo	1st allele of ind1	=IF(B$5=D5,B$5,0)	=IF(B$5=E$5,B$5,0)	=IF(M4,M4,N4)
=IF(B$5=C$5, "homo", "hetero")		=IF(C$5=D5,C$5,0)	=IF(C$5=E$5,C$5,0)	=IF(M5,M5,N5)
check father homo	Ver ck	=IF(M4,M4,M5)	=IF(N4,N4,N5)	=COUNTIF(O4:O5,">0")
=IF(D$5=E$5, "homo", "hetero")			=COUNTIF(M6:N6,">0")	=IF(O6<N7,O6,N7)

An "if" check is written to check heterozygous or not

			k0
vlook col=		11	0.25
freq		a value	formula
=IF(O4>0,VLOOKUP(O4,'freq tab'!A$1:P$67,Q2,FALSE),0)		=IF(Q9,Q5,MAX(P4:P5))	(K2+2K1+k0*a*a)/a*a
=IF(O5>0,VLOOKUP(O5,'freq tab'!A$1:P$67,Q2,FALSE),0)		=IF(Q9,MAX(P4:P5),0)	(K1+k0*a)/a
		b value	K0
<=no. sha allele		=IF(Q9=1,MIN(P4:P5),0)	(K2+k1*a+k1*b+k0*2*a*b)/2*a*b
		check for b	(k1+2*k0*a)/2a
		=IF(K5="hetero",1)*IF(K7="hetero",1)*IF(O7=2,1)	

Use VLOOKUP function to obtain the allele frequency

continues

Table 3.2 (continued)

k1	k2
0.25	0.25
0n/0ff	
=IF(K5="homo",1)*IF(K7="homo",1)*IF(O6=2,1)	=IF(S4=1,(T2+2*S2+R2*Q4*Q4)/(Q4*Q4),0)
=(IF(K5="hetero",1)*IF(K7="homo",1)+IF(K5="homo",1)*IF(K7="hetero",1))*(IF(O7=1,1))	=IF(S5=1,(S2+R2*Q4)/Q4,0)
=IF(O7=0,1,0)	=IF(S6=1, R2,0)
=IF(K5="hetero",1)*IF(K7="hetero",1)*IF(O7=2,1)	=IF(S7=1,(T2+S2*Q4+S2*Q7+R2*2*Q4*Q7)/(2*Q4*Q7),0)
=IF(K5="hetero",1)*IF(K7="hetero",1)*IF(O7=1,1)	=IF(S8=1, (S2+2*R2*Q4)/(2*Q4),0)
sum	=SUM(T4:T8)

Five "if" checks here to choose the right Index

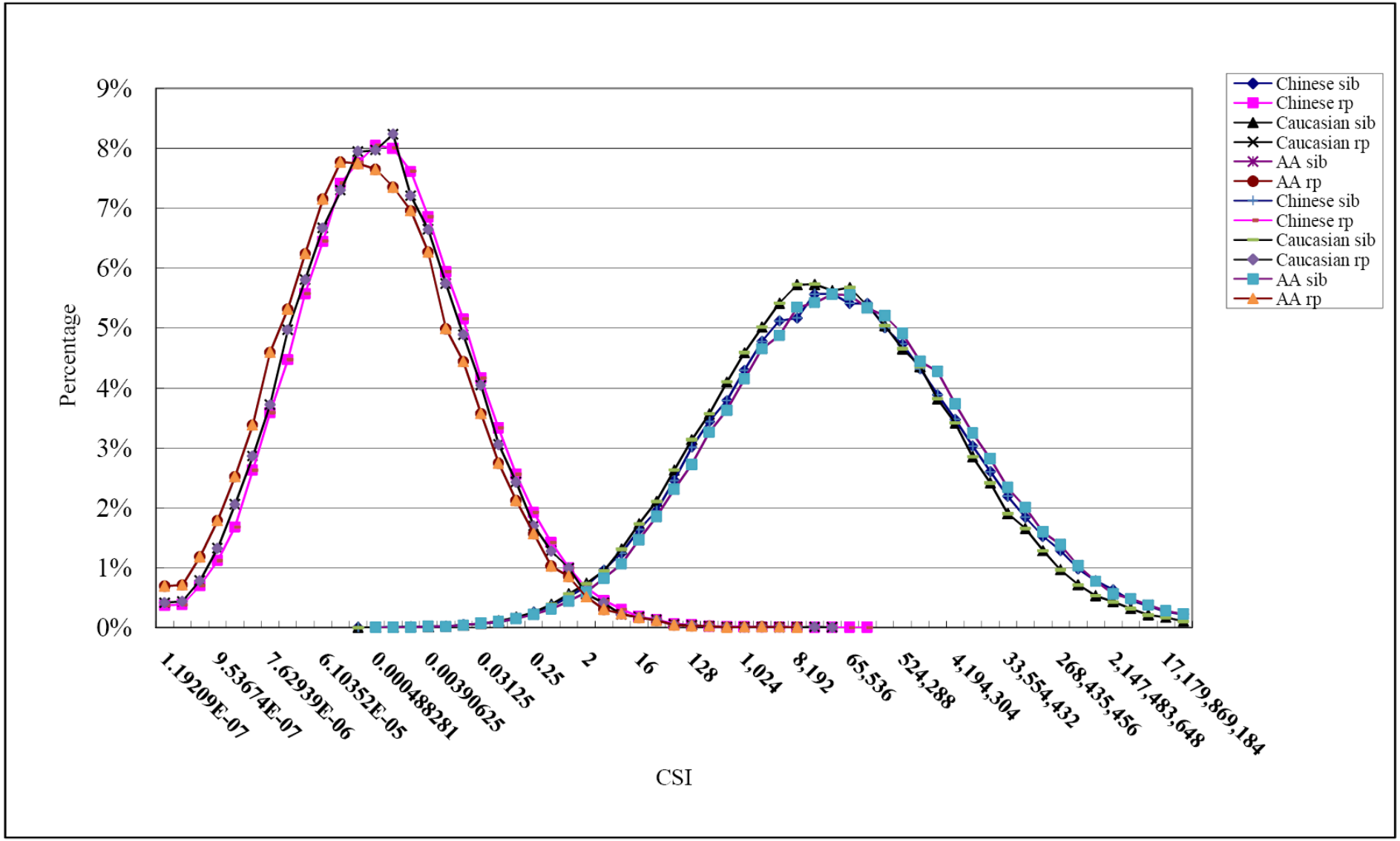
Five index calculation processes

3.3 Results and Discussion

Ratio Distribution of CSI for Three Populations

The CSI ratio distribution of simulated sibling pairs and random pairs are depicted in Figure 3.1. Bipolar models with a widespread ratio distribution were found for all three populations. From the data obtained CSI values of less than 1 were found for 1.500%, 1.608% and 1.340% of simulated sibling pairs for the Chinese, Caucasian and African American populations respectively. In contrast when examining random pairs of DNA profiles resulted in 1.875%, 1.590% and 1.415% with a CSI larger than 1 (Table 3.3). Simulated siblings pairs with very low CSI values (1.68E-05 for Chinese population) and random pairs with very high CSI values (6.24E+08 for African American population) were observed in this study. These data indicate the wide possible CSI values that can be obtained from real siblingship data with a high degree of uncertainty in any report. It has been proposed that when two DNA profiles produce CSI values between 0.067 and 10 further loci should be analysed (116). In this present study the lowest CSI value for a sibling pair was significantly less than 0.067 and CSI for random pairs was much larger than 10.3. Our data indicate that simply applying a minimum CSI requirement may result in false exclusions if set too high or false inclusions if too low.

Figure 3.1: Ratio distribution of CSI for 3 populations.



AA: African Americans

Sensitivity and Specificity under Different CSI Cutoffs

The ability of the DNA test to correctly classify kinship testing results into two categories (sibling or not sibling) is assessed by specificity and sensitivity (86). In Table 3.3 the sensitivity and specificity using a range of CSI cutoff values is illustrated. CSI cutoff values at 0.067, 3 and 10.3 were added following previous recommendations (112, 116). As the CSI cutoff values increased there was a corresponding decrease in sensitivity and increase in specificity. When adopting a simple CSI value of 1, below which indicates non-sibling and above supports a sibling pair, then for the three populations the sensitivity was 98.500% and the specificity was 98.125% for the Chinese population, for Caucasian population the sensitivity was 98.392% and the specificity was 98.410% and for African American population the sensitivity was 98.660% and the specificity was 98.585%.

Ratio Distribution, Sensitivity and Specificity of TASL

The TASL ratio of simulated sibling pairs and random pairs is depicted in Figure 3. Bipolar models were found for all of the three populations.

The sensitivity and specificity for 1 to 15 of TASL are illustrated in Table . As the number of cutoff loci increased there was a corresponding decrease in sensitivity and increase in specificity. At the TASL cutoff ≥ 5, the specificity was greater than 99% for all of the populations, but the sensitivity was only around 75%. Based upon these data, TASL alone is not an optimal screening method for the siblingship determination.

Table 3.3: Evaluation of sensitivity and specificity for siblinship determination using. 15 STR systems.

population	Chinese		Caucasians		African Americans	
CSI	**SPE	*SEN	SPE	SEN	SPE	SEN
0.03125	87.905%	99.849%	88.977%	99.868%	90.291%	99.859%
0.067	91.538%	99.747%	92.330%	99.750%	93.269%	99.755%
0.125	93.795%	99.615%	94.454%	99.590%	95.150%	99.619%
0.25	95.713%	99.364%	96.146%	99.338%	96.713%	99.403%
0.5	97.131%	99.012%	97.420%	98.952%	97.736%	99.096%
1	98.125%	98.500%	98.410%	98.392%	98.585%	98.660%
2	98.788%	97.816%	98.966%	97.659%	99.093%	98.068%
3	99.050%	97.298%	99.207%	97.174%	99.277%	97.603%
10	99.613%	95.151%	99.670%	94.876%	99.691%	95.760%
10.3	99.624%	95.074%	99.672%	94.791%	99.694%	95.695%
33	99.852%	91.953%	99.880%	91.466%	99.893%	92.785%
100	99.935%	87.735%	99.949%	87.009%	99.942%	88.919%
150	99.953%	85.825%	99.969%	85.012%	99.966%	87.109%
200	99.958%	84.429%	99.978%	83.627%	99.972%	85.846%
300	99.969%	82.331%	99.980%	81.349%	99.975%	83.791%
330	99.971%	81.818%	99.984%	80.788%	99.975%	83.346%
500	99.976%	79.492%	99.987%	78.303%	99.975%	81.150%
1,000	99.984%	75.218%	99.989%	73.754%	99.982%	77.004%
minimum	87.905%	75.218%	88.977%	73.754%	90.291%	77.004%
maximum	99.984%	99.849%	99.989%	99.868%	99.982%	99.859%

*.SEN : sensitivity; **.SPE : specificity

Figure 1.2: Ratio distribution of TASL for three populations (*sib: siblings; **rp: random pairs; ***AA: African Americans.

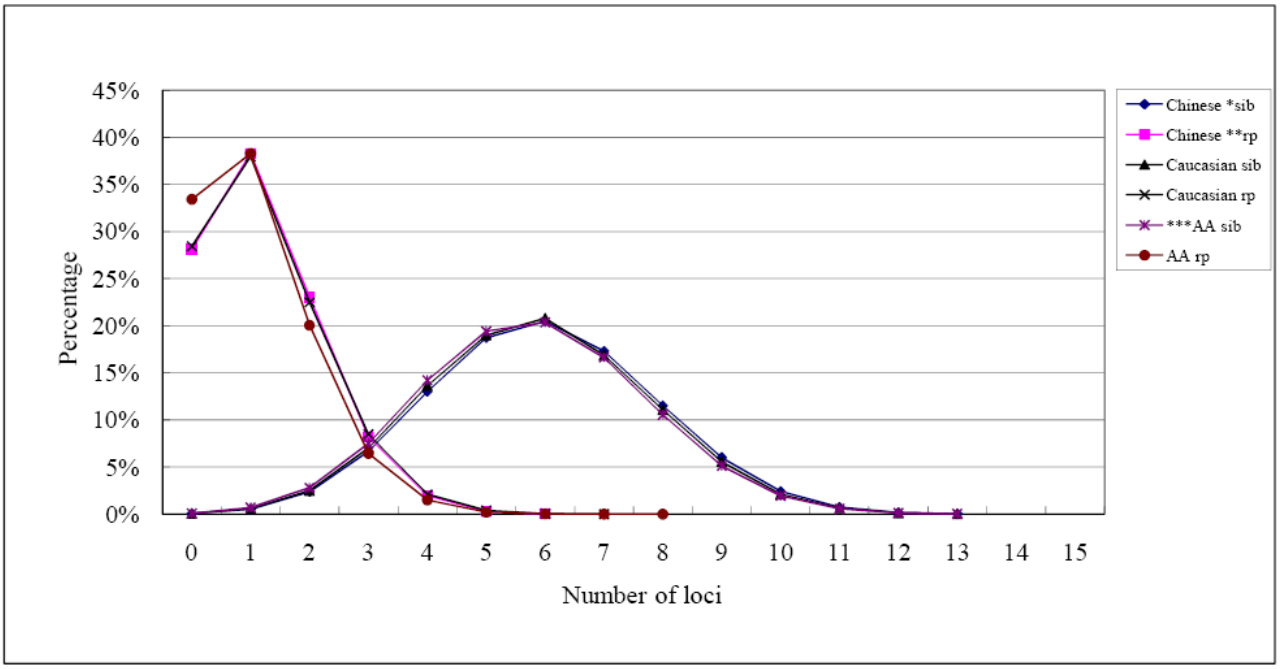


Table 3.4: Sensitivity and specificity versus two allele sharing cutoff points.

population	Chinese		Caucasians		African Americans	
Loci	SPE	SEN	SPE	SEN	SPE	SEN
1	28.147%	99.952%	28.430%	99.956%	33.447%	99.917%
2	66.459%	99.445%	66.412%	99.419%	71.752%	99.226%
3	89.484%	97.075%	88.913%	96.904%	91.801%	96.437%
4	97.689%	90.414%	97.415%	89.889%	98.254%	88.932%
5	99.662%	77.376%	99.526%	76.230%	99.752%	74.710%
6	99.957%	58.641%	99.940%	57.224%	99.963%	55.277%
7	99.997%	38.132%	99.996%	36.412%	99.994%	34.910%
8	100.000%	20.828%	100.000%	19.542%	99.997%	18.284%
9	100.000%	9.295%	100.000%	8.475%	100.000%	7.736%
10	100.000%	3.284%	100.000%	2.907%	100.000%	2.627%
11	100.000%	0.879%	100.000%	0.788%	100.000%	0.700%
12	100.000%	0.169%	100.000%	0.155%	100.000%	0.146%
13	100.000%	0.025%	100.000%	0.025%	100.000%	0.023%
14	100.000%	0.003%	100.000%	0.002%	100.000%	0.000%
15	100.000%	0.000%	100.000%	0.000%	100.000%	0.000%

Evaluation of Synergy Effects of Two Criteria for the Siblingship Determination

The sensitivity and specificity of both the CSI and TASL values were combined. Values of CSI varying from 0.125 to 100 were analysed against the values of 1 to 9 for TASL (Table 3.5). A combination of these two data sets increased the confidence of an exclusion or inclusion. A CSI of 0.125 and TASL of 5 resulted in a specificity of 99.727% for Chinese population, 99.626% for the Caucasian population and 99.804% for African American population. Instances of medium CSI values reflected the true specificity, e.g. when CSI = 3 and TASL = 5, resulted in the specificity increasing to 99.885% instead of 99.050% (Table 3.5), when based on CSI=3 only), 99.825% instead of 99.207% for Caucasian population and 99.905% instead of 99.277% for African American population.

The data obtained in this study indicates that values of CSI and TASL to be adopted as the cutoffs may vary from population to population. The values used may also vary if the DNA test is used for identification of human remains or in criminal cases if the burden of proof required varies. The ideal situation to resolve siblingship cases with greater confidence would be to use more autosomal STR loci if possible or to use mitochondrial DNA analysis or STR loci on either of the sex determining chromosomes.

3.4 Conclusion

The use of CSI and TASL together resolved the difficulty in siblingship determination for cases with low CSI value. When medium index value was obtained then the confidence could be raised if an increasing number of TASLs.

Table 3.5: Synergy effects of TASL and CSI for siblingship determination.

	CSI=0.1		CSI=0.5		CSI=1		CSI=2		CSI=3		CSI=10		CSI=33		CSI=100	
TASL	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE
	Chinese															
1	99.590%	93.943%	98.997%	97.173%	98.488%	98.150%	97.807%	98.797%	97.290%	99.054%	95.147%	99.614%	91.952%	99.852%	87.734%	99.935%
2	99.180%	94.862%	98.659%	97.443%	98.191%	98.293%	97.559%	98.883%	97.073%	99.113%	95.004%	99.632%	91.872%	99.854%	87.686%	99.936%
3	96.949%	96.883%	96.633%	98.238%	96.319%	98.767%	95.862%	99.145%	95.489%	99.307%	93.801%	99.685%	91.028%	99.872%	87.114%	99.941%
4	90.389%	98.777%	90.292%	99.205%	90.156%	99.401%	89.940%	99.559%	89.764%	99.632%	88.810%	99.817%	86.944%	99.912%	83.996%	99.957%
5	77.372%	99.727%	77.358%	99.789%	77.331%	99.824%	77.281%	99.864%	77.229%	99.885%	76.909%	99.925%	76.153%	99.963%	74.670%	99.978%
6	58.641%	99.959%	58.640%	99.964%	58.638%	99.966%	58.630%	99.970%	58.624%	99.975%	58.567%	99.981%	58.403%	99.991%	58.009%	99.994%
7	38.132%	99.997%	38.132%	99.997%	38.132%	99.997%	38.131%	99.998%	38.131%	99.998%	38.127%	99.998%	38.113%	99.998%	38.057%	99.999%
8	20.828%	100.000	20.828%	100.000	20.828%	100.000	20.828%	100.000	20.828%	100.000	20.828%	100.000	20.827%	100.000	20.823%	100.000
	Caucasia															
1	99.568%	94.534%	98.937%	97.429%	98.382%	98.412%	97.652%	98.968%	97.167%	99.209%	94.876%	99.670%	91.467%	99.880%	87.010%	99.949%
2	99.124%	95.178%	98.585%	97.635%	98.083%	98.481%	97.408%	99.008%	96.947%	99.231%	94.735%	99.670%	91.396%	99.880%	86.966%	99.949%
3	96.776%	96.746%	96.444%	98.202%	96.136%	98.802%	95.636%	99.181%	95.267%	99.342%	93.483%	99.697%	90.493%	99.887%	86.390%	99.949%
4	89.859%	98.587%	89.736%	99.092%	89.607%	99.333%	89.377%	99.508%	89.200%	99.579%	88.208%	99.772%	86.274%	99.903%	83.208%	99.953%
5	76.226%	99.626%	76.208%	99.743%	76.184%	99.783%	76.143%	99.812%	76.105%	99.825%	75.790%	99.900%	75.008%	99.951%	73.553%	99.969%
6	57.223%	99.942%	57.220%	99.945%	57.220%	99.949%	57.215%	99.956%	57.207%	99.958%	57.168%	99.971%	56.980%	99.984%	56.564%	99.984%
7	36.411%	99.996%	36.411%	99.996%	36.411%	99.996%	36.411%	99.996%	36.411%	99.996%	36.410%	99.996%	36.394%	99.996%	36.347%	99.996%
8	19.542%	100.000	19.542%	100.000	19.542%	100.000	19.542%	100.000	19.542%	100.000	19.542%	100.000	19.542%	100.000	19.538%	100.000
	African Americans															
1	99.557%	95.245%	99.047%	97.770%	98.617%	98.603%	98.032%	99.102%	97.575%	99.283%	95.749%	99.691%	92.783%	99.893%	88.919%	99.942%
2	98.984%	95.938%	98.572%	97.953%	98.197%	98.695%	97.670%	99.145%	97.255%	99.305%	95.533%	99.697%	92.644%	99.893%	88.831%	99.942%
3	96.345%	97.549%	96.132%	98.594%	95.905%	99.047%	95.538%	99.354%	95.250%	99.452%	93.940%	99.740%	91.481%	99.905%	88.047%	99.945%
4	88.912%	99.059%	88.848%	99.369%	88.787%	99.550%	88.623%	99.672%	88.494%	99.709%	87.819%	99.850%	86.345%	99.942%	84.000%	99.963%
5	74.710%	99.804%	74.701%	99.828%	74.698%	99.856%	74.660%	99.893%	74.635%	99.905%	74.429%	99.942%	73.938%	99.966%	72.895%	99.975%
6	55.277%	99.966%	55.276%	99.966%	55.276%	99.966%	55.274%	99.969%	55.271%	99.969%	55.242%	99.979%	55.155%	99.982%	54.914%	99.985%
7	34.910%	99.994%	34.910%	99.994%	34.910%	99.994%	34.910%	99.994%	34.910%	99.994%	34.907%	99.994%	34.903%	99.994%	34.874%	99.994%
8	18.284%	99.997%	18.284%	99.997%	18.284%	99.997%	18.284%	99.997%	18.284%	99.997%	18.284%	99.997%	18.284%	99.997%	18.284%	99.997%

Chapter 4: Half-Siblingship Test

4.1 Introduction

A claim of half-sibling relationship may be made in cases of inheritance, immigration (118) and genetic counselling (119). A genetic link as a half-sibling may assist in the identification of human remains if no closer genetic relative is available for testing. A half-sibling is defined as two individuals sharing one parent, this may be a mother when it is termed uterine, or one father when it is called agnate or consanguine. Theoretically there is a 0.5 chance that two half-siblings will share 1 allele at any one locus and a 0.5 chance that they will share no alleles (112). The confidence that two individuals are half-siblings, or not, will increase if more loci are examined, or if there are more potential half-siblings for comparison. When determining the probability whether or not two individuals are half-siblings it is necessary to consider the allele frequencies of any alleles that are shared. From this calculation a combined half-siblingship indices (CHSI) (111) is reported. If CHSI is less than 1 then the evidence supports the two individuals as not being related as half-siblings, otherwise the data supports the existence of a half-sibling relationship. There remains much uncertainty after this calculation (111, 112, 118). Uncertainty remains with the evaluation of the resulting LR figure in terms of its reliability in determining whether two samples have come from half-siblings.

Three populations containing 355,620 simulated half-sibling pairs and 178,815 random pairs generated from 15 STR profiles were created within the populations. Using this large number of sample pairs, it is possible to evaluate the sensitivity and specificity for the 15 STR core set in

discriminating between half-siblings and random pairs. The discriminating power combining the CHSI and all shared alleles (ASA) is reported to assist a laboratory with an interpretation of test results.

4.2 Materials and Methods

STR DNA profiles from 450 individuals were processed by a Microsoft Excel Macros written by built in Visual Basic, with minor modification of the code stated in Table 3.1 for sibling study. Every individual of the population was paired with every other individual to form random pairs, e.g. for Taiwan population in this study (450 x 449)/2, equalling 101,025 pairs were made. For generating half-siblings, every individual in the population was paired with two other individuals, e.g., (449x448)/2, equalling 100,576 triples (one person with two mates). One offspring was generated from each of the two pairs within a triple, resulting in 201,152 half-sibling pairs. DNA profiles from the 15 STR loci from the Caucasian (n= 301) and the African American (n= 256) populations were also processed in the same way as those for the Taiwanese population to generate half-sibling pairs and random pairs. For the Caucasian population 89,700 half-sibling pairs were generated and 45,150 random pairs were obtained and for African American population 64,770 half-sibling pairs and 32,640 random pairs were generated. The CHSI values were calculated according to standard formulae (111). By using built-in function provided by the software EXCEL, the formulae were transformed into cells of worksheet (Table 4.1). The CHSI values were obtained when STR profiles of the pairs were

pasted into the assigned area of the worksheet. The allele frequency tables used for Caucasian and African American populations were downloaded from the same source as the DNA profiles (35) and the frequency table for the Chinese population was from previous studies (83, 84). All the frequencies of alleles were adjusted by using 5/2N rule for this study.

The rate of false negatives was calculated as the percentage of real half-siblingship that could be excluded based upon any given cutoff point of CHSI. The rate of false positives equalled the percentage of random pairs of DNA profiles where their CHSI was greater than any chosen CHSI threshold value. The sensitivity of the test is based upon 1 – the % of false negatives and the specificity is based upon 1 – the % of false positives (86).

All-shared-alleles (ASA) were determined by counting all the alleles shared by paired profiles. In the case of both alleles at any one locus being shared a score of 2 is registered. If 1 allele is in common a score of 1 is registered.

The above formulae were also applied to four situations to evaluate the specificity and sensitivity of the test. These four situations included low CHSI with low ASA, low CHSI with high ASA, high CHSI with low ASA and high CHSI with high ASA.

Table 4.1: Calculation worksheet of CHSI.

Description of Function:

- ✧ EXCEL worksheet is a table containing multiple cells, in each cell data or formula can be put. In this study, author designed a functional worksheet to calculate the CSI between two individuals automatically after posting the STR profiles.
- ✧ In this sheet the original STR profiles can be posted on the upper part of the worksheet, and the lower part was filled with built-in EXCEL functions and commands for choosing the right CHSI index formula, and calculating the index.

This is the upper part of the worksheet:

(post the profiles into columns to get the CHSI)

CHSI= 4.40E+00

	individual 1		individual 2	
STR D8S1179	11	15	15	15
STR D21S11	30	31	30	31
STR D7S820	10	11	12	10
STR CSF1PO	11	11	12	12
STR D3S1358	15	16	15	15
STR TH01	9	7	9	9
STR D13S317	11	10	11	9
STR D16S539	12	13	12	13
STR D2S1338	20	26	23	20
STR D19S433	13	15	13	15.2
STR vWA	15	18	17	18
STR TPOX	11	11	11	11
STR D18S51	12	15	12	12
STR Sex	X	X	X	X
STR D5S818	12	11	12	10
STR FGA	19	24	19	20

continues

Table 4.1 (continued)

This is the lower part of the worksheet, there are multiple cells with allele-sharing-check, formula picking and calculation function, “description box” in dotted line are also presented.

STR D8S1179								k0			k1	k2	
Sibling Index =		2.0042 B$9			vlook col=		11				0.5	0.25	0.5
				Hor ck	freq		a value	formula			0n/0ff		
check child homo	1st allele of ind1	0	0	0		0	0.1662	(2K1+k0*a)/a				0	0.0000
hetero	2nd allele of ind1	15	15	15		0.1662	0	(K1+k0*a)/a				1	2.0042
check father homo	Ver ck	15	15	1			b value	K0				0	0.0000
homo			2	1	<=no. sha allele		0	(k1*a+k1*b+k0*2*a*b)/2*a*b				0	0.0000
							check for b	(k1+2*k0*a)/2a				0	0.0000
							0				sum		2.0042

The cell shows the shared allele

Frequency of allele obtained

Index formula chosen and index calculated

continues

Table 4.1 (continued)

The formula developed by the author that residing at the multiple cells, the formula is not seen bu users when using the calculation table, details are described in “description box” (in dotted line).

=A$5			
Sibling Index =	=IF(O9,O9,1)	B$9	
check child homo	1st allele of ind1	=IF(B$5=D5,B$5,0)	=IF(B$5=E$5,B$5,0)
=IF(B$5=C$5, "homo", "hetero")	2nd allele of ind1	=IF(C$5=D5,C$5,0)	=IF(C$5=E$5,C$5,0)
check father homo	Ver ck	=IF(H4,H4,H5)	=IF(I4,I4,I5)
=IF(D$5=E$5, "homo", "hetero")			=COUNTIF(H6:I6,">0")

An “if” check is written to check heterozygous or not

	vlook col=	11	
Hor ck	freq	a value	
=IF(H4,H4,I4)	=IF(J4>0,VLOOKUP(J4,'freq tab'!A$1:P$67,L2,FALSE),0)	=IF(L9,L5,MAX(K4:K5))	
=IF(H5,H5,I5)	=IF(J5>0,VLOOKUP(J5,'freq tab'!A$1:P$67,L2,FALSE),0)	=IF(L9,MAX(K4:K5),0)	
=COUNTIF(J4:J5,">0")		b value	
=IF(J6<I7,J6,I7)	<=no. sha allele	=IF(L9=1,MIN(K4:K5),0)	
		check for b	
		=IF(F5="hetero",1)*IF(F7="hetero",1)*IF(J7=2,1)	

Use VLOOKUP function to obtain the allele frequency

continues

Table 4.1 continued

k0	k1
0.5	0.25
formula	0n/0ff
(2K1+k0*a)/a	=IF(F5="homo",1)*IF(F7="homo",1)*IF(J6=2,1)
(K1+k0*a)/a	=(IF(F5="hetero",1)*IF(F7="homo",1)+IF(F5="homo",1)*IF(F7="hetero",1))*(IF(J7=1,1))
K0	=IF(J7=0,1,0)
(k1*a+k1*b+k0*2*a*b)/2*a*b	=IF(F5="hetero",1)*IF(F7="hetero",1)*IF(J7=2,1)
(k1+2*k0*a)/2a	=IF(F5="hetero",1)*IF(F7="hetero",1)*IF(J7=1,1)
	sum

Five "if" checks here to choose the right Index

k2
0.5
=IF(N4=1,(2*N2+M2*L4)/L4,0)
=IF(N5=1,(N2+M2*L4)/L4,0)
=IF(N6=1, M2,0)
=IF(N7=1,(N2*L4+N2*L7+M2*2*L4*L7)/(2*L4*L7),0)
=IF(N8=1, (N2+2*M2*L4)/(2*L4),0)
=SUM(O4:O8)

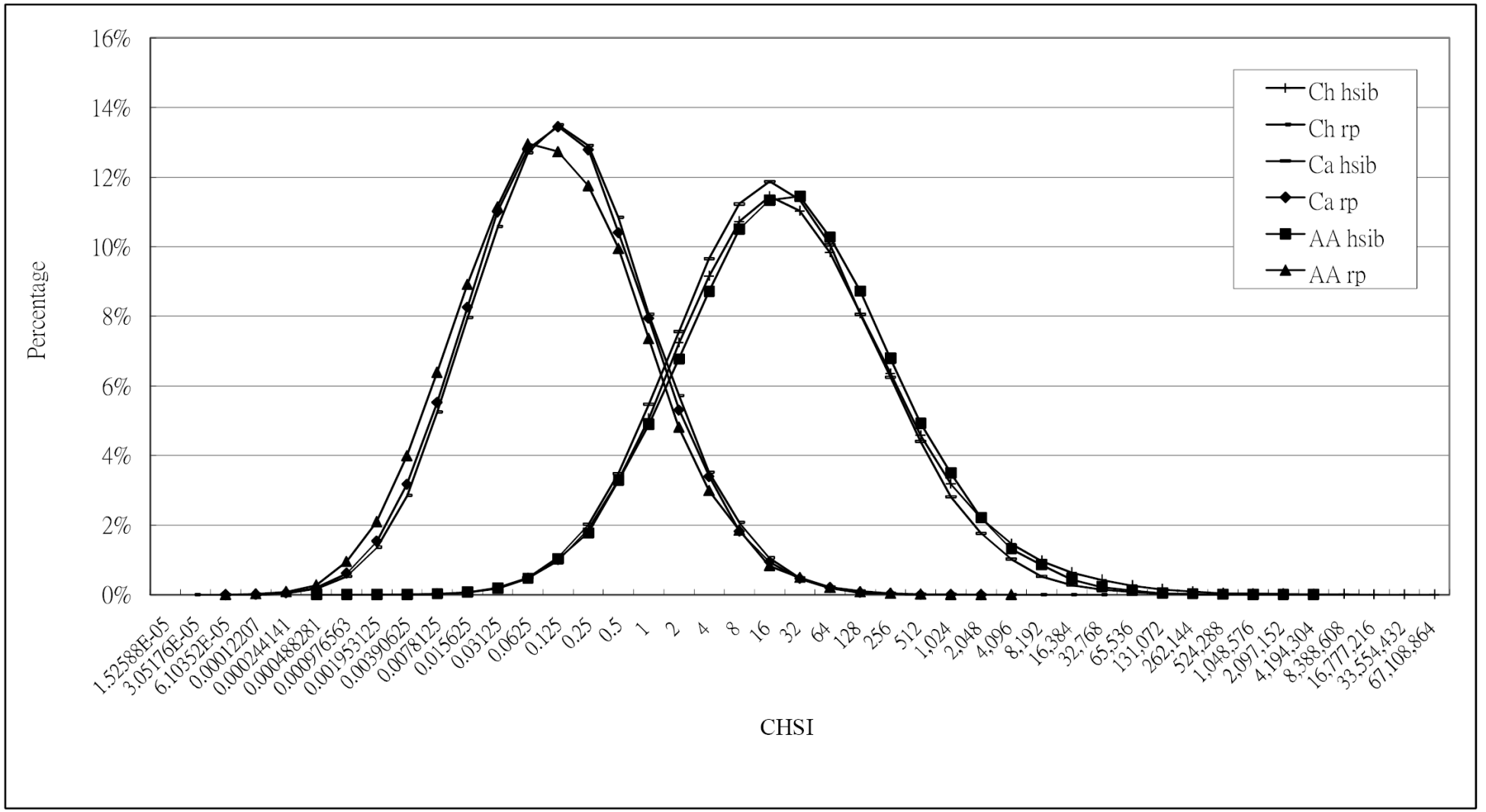
Five index calculation processes

4.3 Results and Discussion

Ratio Distribution of CHSI for Three Populations by using 15 STR core set

The CHSI ratio distribution of simulated half-sibling pairs and random pairs for three populations is shown in figure 4.1. Bipolar models with a widespread ratio distribution were found for all of the three populations. Simulated siblings pairs with very low CHSI values (1.40E-03 for the Chinese population) and random pairs with very high CHSI values (2.29E+05 for African American population) were observed in this study. From the data obtained for the 15 STR loci (Table 4.2), 12.02%, 12.86% and 11.73% of CHSI values from simulated half-sibling pairs were found to be less than 1 for Chinese, Caucasians and African Americans populations respectively. Random pairs that produced CHSI values greater than 1 for these populations were 13.24%, 12.22% and 11.33% for these three populations.

Figure 2.1: CHSI ratio distribution for 3 populations (*hsib: half-siblings, **rp: random pairs).



Ch: Chinese Ca:Caucasian AA: African Americans

rp: random pair hsib: half-sibling

Table 4.2: Evaluation of sensitivity, specificity and CHSI threshold value for half-siblingship determination using 15 STR systems.

population	Chinese		Caucasians		African Americans	
CHSI threshold value	*SPE	**SEN	SPE	SEN	SPE	SEN
0.03125	28.76%	99.70%	30.37%	99.69%	33.88%	99.71%
0.0625	41.46%	99.22%	43.18%	99.21%	46.85%	99.24%
0.125	54.96%	98.23%	56.63%	98.12%	59.59%	98.21%
0.25	67.87%	96.34%	69.42%	96.10%	71.35%	96.44%
0.5	78.70%	93.04%	79.83%	92.62%	81.31%	93.16%
1	86.76%	87.98%	87.78%	87.14%	88.67%	88.27%
2	92.48%	80.74%	93.09%	79.58%	93.48%	81.50%
3	94.80%	75.63%	95.30%	74.20%	95.39%	76.62%
10	98.49%	57.19%	98.67%	54.92%	98.69%	58.83%
33	99.63%	37.93%	99.71%	34.97%	99.68%	39.02%
100	99.92%	23.14%	99.96%	19.99%	99.91%	23.37%
150	99.95%	18.83%	99.98%	15.70%	99.94%	18.74%
200	99.97%	16.19%	99.99%	13.05%	99.96%	15.86%
300	99.98%	12.95%	100.00%	9.90%	99.98%	12.33%
330	99.98%	12.25%	100.00%	9.23%	99.98%	11.59%
500	99.99%	9.68%	100.00%	6.79%	99.99%	8.89%
1,000	100.00%	6.46%	100.00%	3.95%	99.99%	5.37%

*SPE: Specificity
**SEN: Sensitivity

Sensitivity and Specificity under Different CHSI Cutoffs for 15 STR systems

The ability of the DNA test to correctly classify kinship testing results into two categories is assessed by their specificity and sensitivity (114). In Table 4.2 the sensitivity and specificity are illustrated using a range of CHSI values from 0.03125 to 1,000. The CHSI requirements as cutoff values followed previous recommendations (117). As the CHSI cutoff values increased there was a corresponding decrease in sensitivity and increase in specificity. According to Table 4.2, when adopting CHSI value of 1 as the cutoff value, and using the 15 STR loci for the three populations, the sensitivity was 87.98% and the specificity was 86.76% for the Chinese population, for Caucasian population the sensitivity was 87.14% and the

specificity was 87.78% and for African American population the sensitivity was 88.27% and the specificity was 88.67%. Using a CHSI of 1 would falsely exclude more than 10% of real half-sibling, indicating the potential problems with this type of DNA testing.

Ratio Distribution of Allele sharing for three populations

The ratio distribution of allele-sharing for simulated half-sibling pairs and random pairs is shown in Figure 4.2. Three instances of allele sharing between two DNA profiles across the loci tested are illustrated. The three graphs illustrate the percentage of loci where there are two alleles shared (A), one alleles shared (B) and all the alleles shared (C). With more alleles shared there is a corresponding separation of real and false half-sibling pairs. Graph C illustrates the ASA situation where there is greatest separation between half-sibling pairs and random pairs (with smaller percentage of overlapping area in the bipolar model). These data indicate the potential value of ASA in the determination of half-siblingship.

Evaluation of Synergy Effects of Two Criteria for the Half-siblingship Determination

The CHSI versus ASA ratio distribution was calculated for all the simulated half-siblings and random pairs, and the sensitivity and specificity for each CHSI and ASA combination was evaluated. Values of CHSI varying from 0.125 to 10 were analysed against the values of 7 to 18 for ASA using the 15 STR loci (Table 4.3). A combination of these two data sets increased the specificity thus reducing the chance of a false inclusion. A CHSI of

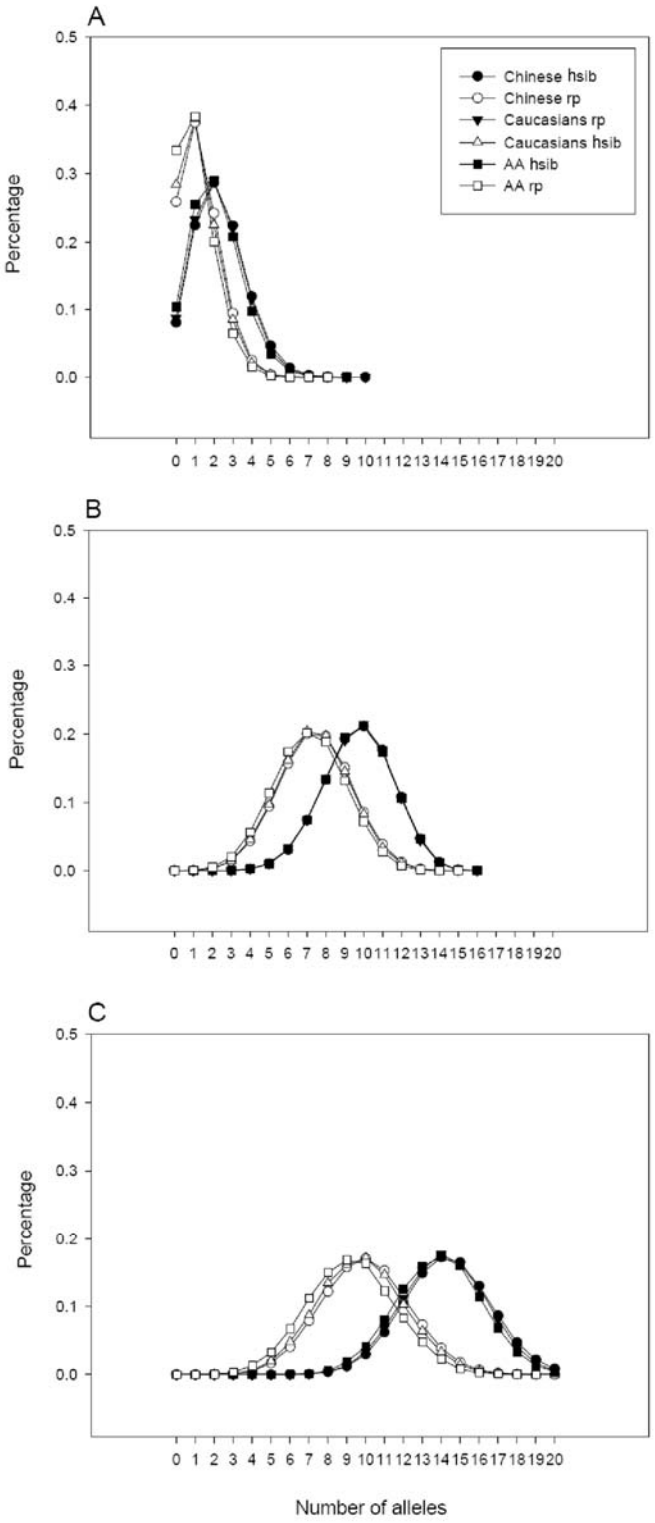
0.125 and ASA of 13 resulted in a specificity of 86.14%, whereas if only considering the CHSI specificity was only 54.96% (Table 4.2) for the Taiwanese population. When CHSI = 1 and ASA = 13, the specificity increased to 91.22% instead of 86.76% for the Taiwanese population (see Table 4.2 when based on CHSI = 1 only), 92.27% instead of 87.78% for Caucasian population and 93.97% instead of 88.67% for African American population. Combining the two criterion increases the specificity for half-sibling determination especially for the cases of "low CHSI with high ASA". If these two factors are not considered in tandem there is a higher chance that two real half siblings will be falsely excluded.

An alternative approach to resolve half-siblingship cases with greater confidence would be by using more autosomal STR loci, using mitochondrial DNA analysis if appropriate or by using STR loci on either of the sex determining chromosomes.

4.4 Conclusion

The combination of CHSI and ASA can be used to resolve complex cases of alleged half-siblingship. This may be where there are low CHSI values obtained. By combining the number of alleles where all the alleles are shared the specificity can be increased thus reducing the false exclusions.

Figure 4.2: Ratio distribution of allele sharing across the loci. A for the number of 2-allele sharing loci, B for the number of loci with one allele shared and C when all shared alleles are counted.



hsib: siblings
rp: random pairs
AA: African Americans

Table 4.3: Synergy effects of ASA and CHSI for half-siblingship determination by using 15 STR.

	0.125		0.25		0.5		1		1.25		1.5		2		3		10	
ASA	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE
Chinese																		
7	98.23%	54.96%	96.34%	67.87%	93.04%	78.71%	87.98%	86.76%	85.87%	88.89%	84.02%	90.39%	80.74%	92.48%	75.63%	94.80%	57.19%	98.49%
8	98.22%	55.03%	96.33%	67.89%	93.04%	78.71%	87.98%	86.76%	85.87%	88.89%	84.02%	90.39%	80.74%	92.48%	75.63%	94.80%	57.19%	98.49%
9	98.11%	55.57%	96.27%	68.04%	93.00%	78.75%	87.96%	86.78%	85.85%	88.90%	84.00%	90.40%	80.73%	92.49%	75.62%	94.81%	57.19%	98.49%
10	97.42%	58.16%	95.82%	68.98%	92.74%	79.07%	87.80%	86.87%	85.72%	88.98%	83.89%	90.46%	80.64%	92.52%	75.56%	94.82%	57.17%	98.49%
11	94.96%	65.15%	93.92%	72.24%	91.45%	80.27%	86.98%	87.30%	85.02%	89.28%	83.28%	90.68%	80.13%	92.66%	75.18%	94.89%	57.03%	98.50%
12	89.00%	75.86%	88.57%	78.72%	87.15%	83.38%	83.86%	88.58%	82.26%	90.19%	80.79%	91.39%	78.07%	93.14%	73.62%	95.16%	56.39%	98.54%
13	78.46%	86.14%	78.37%	86.78%	77.84%	88.49%	76.13%	91.22%	75.15%	92.23%	74.19%	93.03%	72.29%	94.23%	68.93%	95.75%	54.22%	98.64%
14	63.57%	93.34%	63.56%	93.41%	63.46%	93.73%	62.91%	94.60%	62.53%	95.00%	62.11%	95.35%	61.15%	95.96%	59.21%	96.84%	48.89%	98.84%
15	46.36%	97.24%	46.36%	97.24%	46.35%	97.29%	46.24%	97.43%	46.14%	97.52%	46.01%	97.62%	45.71%	97.81%	44.96%	98.16%	39.51%	99.17%
16	29.86%	99.02%	29.86%	99.02%	29.86%	99.02%	29.85%	99.03%	29.84%	99.04%	29.82%	99.06%	29.76%	99.10%	29.58%	99.17%	27.53%	99.52%
17	16.86%	99.70%	16.86%	99.70%	16.86%	99.70%	16.86%	99.71%	16.86%	99.71%	16.86%	99.71%	16.85%	99.71%	16.83%	99.72%	16.29%	99.81%
18	8.18%	99.93%	8.18%	99.93%	8.18%	99.93%	8.18%	99.93%	8.18%	99.93%	8.18%	99.93%	8.18%	99.93%	8.18%	99.94%	8.10%	99.94%
Caucasian																		
7	97.47%	56.63%	95.46%	69.42%	92.00%	79.83%	86.57%	87.78%	84.34%	89.69%	82.46%	91.15%	79.05%	93.09%	73.71%	95.30%	54.55%	98.67%
8	97.46%	56.70%	95.45%	69.44%	92.00%	79.84%	86.56%	87.78%	84.34%	89.69%	82.46%	91.15%	79.05%	93.09%	73.71%	95.30%	54.55%	98.67%
9	97.31%	57.26%	95.37%	69.57%	91.95%	79.85%	86.54%	87.78%	84.32%	89.69%	82.44%	91.15%	79.03%	93.09%	73.70%	95.30%	54.55%	98.67%
10	96.57%	60.16%	94.91%	70.55%	91.69%	80.12%	86.40%	87.85%	84.21%	89.73%	82.35%	91.17%	78.96%	93.10%	73.66%	95.30%	54.54%	98.67%
11	93.78%	67.91%	92.75%	74.22%	90.21%	81.54%	85.52%	88.33%	83.46%	90.07%	81.71%	91.40%	78.45%	93.25%	73.30%	95.37%	54.44%	98.68%
12	87.29%	78.64%	86.88%	81.01%	85.46%	84.94%	82.15%	89.70%	80.52%	91.06%	79.08%	92.14%	76.29%	93.70%	71.67%	95.60%	53.85%	98.71%
13	76.17%	88.10%	76.09%	88.61%	75.59%	89.96%	73.98%	92.27%	73.03%	93.09%	72.12%	93.80%	70.22%	94.86%	66.80%	96.25%	51.75%	98.83%
14	60.90%	94.41%	60.89%	94.46%	60.81%	94.73%	60.31%	95.44%	59.94%	95.77%	59.53%	96.02%	58.64%	96.54%	56.73%	97.27%	46.35%	99.01%
15	43.31%	97.76%	43.31%	97.76%	43.30%	97.79%	43.22%	97.90%	43.15%	97.96%	43.05%	98.01%	42.78%	98.15%	42.09%	98.42%	36.85%	99.29%
16	27.20%	99.21%	27.20%	99.21%	27.19%	99.21%	27.18%	99.22%	27.17%	99.23%	27.16%	99.24%	27.11%	99.26%	26.95%	99.34%	25.16%	99.63%
17	14.52%	99.75%	14.52%	99.75%	14.52%	99.75%	14.52%	99.76%	14.52%	99.76%	14.52%	99.76%	14.51%	99.76%	14.50%	99.78%	14.11%	99.84%
18	6.49%	99.94%	6.49%	99.94%	6.49%	99.94%	6.49%	99.94%	6.49%	99.94%	6.49%	99.94%	6.49%	99.94%	6.49%	99.94%	6.43%	99.95%
African Americans																		
7	97.35%	59.60%	95.53%	71.35%	92.10%	81.31%	87.05%	88.67%	85.05%	90.66%	83.26%	91.75%	80.08%	93.48%	75.12%	95.39%	57.27%	98.69%
8	97.32%	59.74%	95.52%	71.38%	92.09%	81.31%	87.05%	88.67%	85.04%	90.66%	83.26%	91.75%	80.08%	93.48%	75.12%	95.39%	57.27%	98.69%
9	97.13%	60.71%	95.40%	71.66%	92.02%	81.39%	87.02%	88.68%	85.02%	90.67%	83.24%	91.76%	80.07%	93.49%	75.11%	95.39%	57.27%	98.69%
10	95.82%	64.72%	94.43%	73.27%	91.42%	81.90%	86.66%	88.81%	84.73%	90.74%	82.97%	91.80%	79.86%	93.52%	74.97%	95.40%	57.23%	98.69%
11	92.09%	73.69%	91.33%	77.86%	89.15%	83.76%	85.14%	89.52%	83.42%	91.16%	81.80%	92.11%	78.92%	93.69%	74.27%	95.49%	56.98%	98.69%
12	84.38%	83.73%	84.14%	85.04%	83.05%	87.63%	80.56%	91.20%	79.28%	92.41%	78.03%	93.13%	75.73%	94.36%	71.79%	95.89%	55.96%	98.73%
13	71.83%	91.69%	71.79%	91.89%	71.48%	92.54%	70.45%	93.97%	69.80%	94.57%	69.07%	94.97%	67.74%	95.66%	65.13%	96.68%	52.71%	98.86%
14	56.13%	96.48%	56.13%	96.49%	56.07%	96.55%	55.78%	96.86%	55.56%	97.06%	55.30%	97.18%	54.72%	97.44%	53.47%	97.88%	45.72%	99.09%
15	38.95%	98.72%	38.95%	98.72%	38.95%	98.72%	38.90%	98.74%	38.84%	98.76%	38.79%	98.78%	38.63%	98.84%	38.26%	98.99%	34.76%	99.45%
16	23.19%	99.55%	23.19%	99.55%	23.19%	99.55%	23.18%	99.55%	23.18%	99.55%	23.16%	99.56%	23.14%	99.56%	23.07%	99.59%	22.05%	99.70%
17	11.95%	99.87%	11.95%	99.87%	11.95%	99.87%	11.95%	99.87%	11.95%	99.87%	11.95%	99.87%	11.95%	99.87%	11.94%	99.87%	11.73%	99.88%
18	5.18%	99.97%	5.18%	99.97%	5.18%	99.97%	5.18%	99.97%	5.18%	99.97%	5.18%	99.97%	5.18%	99.97%	5.18%	99.97%	5.16%	99.97%

Chapter 5: Develop a non-CODIS triplex

5.1 Introduction

STR typing using commercially available kits may amplify up to 15 STR loci in one reaction. When a DNA sample is subject to microbial degradation, or if sub-optimal amounts of DNA are present, the larger alleles in the multiplex are most likely to no longer generate a product. It is possible to compensate for this by using shortened PCR products than those in commercially available kits (120). Further, when a sample containing degraded DNA is being linked to a family member, such as a sibling or a mother, lower powers of association will be obtained. To combat this problem either novel loci are required or the use of new primers designed to DNA sequences closer to the STR is needed. This chapter outlines the design of novel STR primer sets to amplify three tetra repeat loci; namely D10S674, D14S608 and D15S659.

5.2 Materials and Methods

The three STR loci (D10S674, D14S608 and D15S659) were selected from the Taiwan Polymorphic Marker Database (TPMD) (121). The size of each amplicon was reduced by designing primers as close to the STR sequence as possible. This led to all the alleles being shorter than 250 bp. The primers were designed by using the software Primer3 (122, 123), which is freely available. All the newly designed primers were entered into the software BLAST in NCBI (124) to identify any mispriming sites beyond the target region. Then all the sequences of these 6 primers were checked by the software AutoDimer, available on the internet (23), before being combined into a triplex. The primers sequences and tandem repeats are described in Table

5.1.

Table 5.1: The fluorescent dye and primer sequences of the triplex for the three selected loci.

No.	Locus name	Primers
1	D10S674	FAM —5' AGGATGTGAACTGGAAATTCAT 5' GTGTCTTTGCATGTGCAGACAGACAGA
2	D14S608	VIC —5' CGTGGTACAGGTAGATAAATGGA 5' GTGTCTTGGATCTCCTTCTTTTTATGGATG
3	D15S659	NED—5' CCCAACATAACATATTGCTTAAACT 5' GTGTCTTTGGATAGACACATGACAGATAGGT

The triplex was analysed using a Chinese population living in Taiwan to examine allele frequencies and various forensic parameters (236 optimal profiles of the triplex were obtained). Total DNA was extracted from saliva using the Qiagen blood kit (Qiagen, Hidden, Germany) or Genomic DNA Extraction kit (Real Genomic's ^{TM,} Taiwan).

The triplex was amplified in a 15 μL reaction including 10X buffer (Gold ST*R, Promega) 1.5 μL, primers 2 μL [primers were made into the concentration of 10 pmol/μL, and mixed in the ratio of 4(D10): 6(D14): 3(D15)], DNA template 2 μL (conc. 0.5 ng/μL), polymerase 0.3 μL (Ampli Taq Gold ^{TM} 250 unit 5units/μL, ABI) and ddH_2O 9.2 μL. Amplification was performed using a GeneAMP PCR system 9700 (Appiled Biosystem, Foster City, USA). The cycling

reaction consisted of denaturation at 95 °C for 10 min, and then 30 cycles at 95 °C for 45 sec, 59 °C for 45 sec, and 72 °C for 45 sec, followed by final incubation at 60 °C for 45 min. Aliquotes of 1 μL of the PCR products were mixed with 12 μL deionized formamide containing an internal size standard (Genescan-500, LIZ, ABI), denatured at 95 °C for 3 min, cooled on ice a few seconds, and subjected to laser-induced foluorescence capillary electrophoresis on an ABI Prism 3130 Genetic Analyzer using POP 7, 36-cm capillaty arrays, and default instrument settings.

Homozygote alleles of all the 3 markers were amplified using unlabelled forward and reverse primers in singleplex reactions with a cycler protocol the same as that for labelled primers. Two samples of the PCR products of each locus were sequenced by the T7-cloning method to confirm the length of the products and the assortments of the tandem repeats.

The allele frequencies, forensic DNA parameters, and Hardy-Weinberg equilibrium were analysed by using a Visual Basic Programme written by the author published at the local annual meeting of Taiwan Association of Forensic Science (125), the code stated in Table 5.2.

Table 5.2: Main programme to process STR profiles and generate forensic DNA population study parameters (complete programme was posted in appendix 6).

```
Sub main_popu_study()

'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

' ' export data into text mode

' import the data using Excel software

' change the worksheet name to "Identifiler STR"
```

```
' profiling of 15 loci can be processed at once

'xxxxxxxxxxxxxxxxxxxxxxxxx

'array to load all the locus name

Dim locusnames(15)

'array to load all the identifiler profiles for all the individuals

Dim all_alleles(5999, 30)

'array to record all the possible alleles

Dim allele_cat(79)

'array to record all the allele frequencies

Dim freq_tab(79, 15)

'get the basic information from STR profiles

Worksheets("Basic Info").Select

individuals = Range("A2")

locitot = Range("A4")

mintot = Range("C4")

maxtot = Range("C5")

allele_no = 0

'load the locus name

Worksheets("STR profiles").Select

Range("B1").Select
```

```
locus = 0

For i = 1 To locitot

    locus = (i - 1)

    locusnames(locus) = ActiveCell.Offset(0, (i - 1) * 2)

Next i

'call the sunroutine to pick all the possible alleles

Call pick_allele_category(individuals, mintot, maxtot, allele_cat, all_alleles, allele_no)

'assign the unique alleles to frequency table

For i = 0 To allele_no - 1

    freq_tab(i, 0) = allele_cat(i)

Next i

'call the subroutine to calculate the allele frequencies

Call count_allele_freq(individuals, locitot, allele_no, freq_tab, all_alleles)

'call output frequency table

Call freq_tab_output(allele_no, locusnames, freq_tab, locitot)

'check the fintness and calculate all the forensic DNA parmeters

Call g_test_and_parameter(individuals, locitot, allele_no, freq_tab, all_alleles)

'call output parameters

Call freq_tab_output(allele_no, locusnames, freq_tab, locitot)

Call parameter_output(freq_tab, locitot)
```

```
'Call 5/2Ncorrection()

…..
```

5.3 Results and Discussion

The design of the new primers reduced the amplicon size for the three loci. The new allele size ranges for the three loci are shown in Table 5.3. Homozygous alleles were sequenced to determine the true nucleotide length and sequence within the tandem repeats. The sequencing was performed in both directions to confirm the DNA sequence. Electrophoresis peaks are depicted in Figure 5.1 and the peaks from the triplex reaction are depicted in figure 5.2. All three loci were found to be highly polymorphic and no deviation from Hardy-Weinberg equilibrium was observed (p>0.05).

The allele frequencies for the triplex, homozygosity, heterozygosity, power of discrimination (PD), power of matching (PM), mean exclusion probability (MEP), polymorphism information content (PIC), typical paternity index (PI), and probability value (p value of G-test) of exact tests for Hardy-Weinberg equilibrium are given in Table 5.4. The most frequent allele types for each locus were: D10S674 allele 12, D14S608 allele 10, D15S659 allele 12. The most common allele frequencies range from 0.2076 (D15) to 0.2945 (D10).

The allele distribution of D14S608 varied from 5 to 13, which is in line with the 7-11 allelic range of a previous study (126). The allelic size range of D15S659 was between 11 and 19, which is different from that of 6-17 in the previous study (127). This difference may be reflection

of the different populations sampled in these two studies.

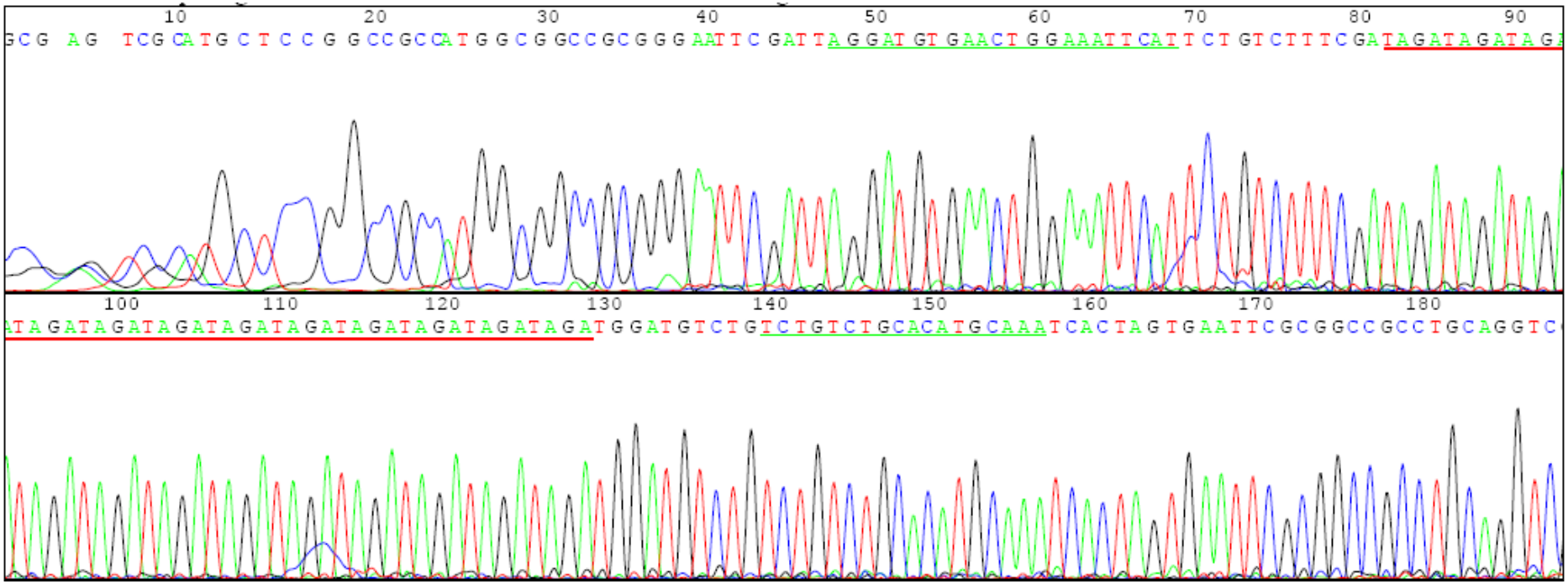
The PD for the 3 three markers were 0.9083 for D10S674, 0.9355 for D14S608, and 0.9472 for D15S659. These figures are comparable to the average PD of routinely used forensic STR systems, adding specificity to the determination of human relationship test is anticipated.

Table 5.3: Information on the genetic loci used in the novel triplex.

Locus name	Original length range in GeneBank	Shortened length range	Sequence	tandem repeats	Length of the sequencing
D10S674	218-254	92-120	GATGCCATGGCGGCCGCGGGAATTCGATTAGGATGTGAACTGGAAATTCATTCTGTCTTTCGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATGGATGTCTGTCTGTCTGCACATGCAAATCACTAGTGAATTCGCGGCCGCCTGCAG	TAGA* 12	108 bp
D14S608	188-224	155-178	CCGCCATGGCGGCCGCGGGAATTCGATTCGTGGTACAGGTAGATAAATGGATGATAGATAGATAATGGAGATAGATGATAGACAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAGTATATATATAATGGACTATATAATATGTGACATAGCATATATAATGGAATATTATTCATCCATAAAAAGAAGGAGATCAATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCT	GATA * 11	182 bp
D15S659	174-206	109-137	CGGTCGCATGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTTGGATAGACACATGACAGATAGGTATGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATACAGAGAACTTAGTTTAAGCAATATGTTATGTTGGGAATCACTAGTGAATTCGCGGCCGCCTGCAGGTCG	GATA * 13	113 bp

Figure 5.1: Electrophoresis peaks of sequencing of the triplex.

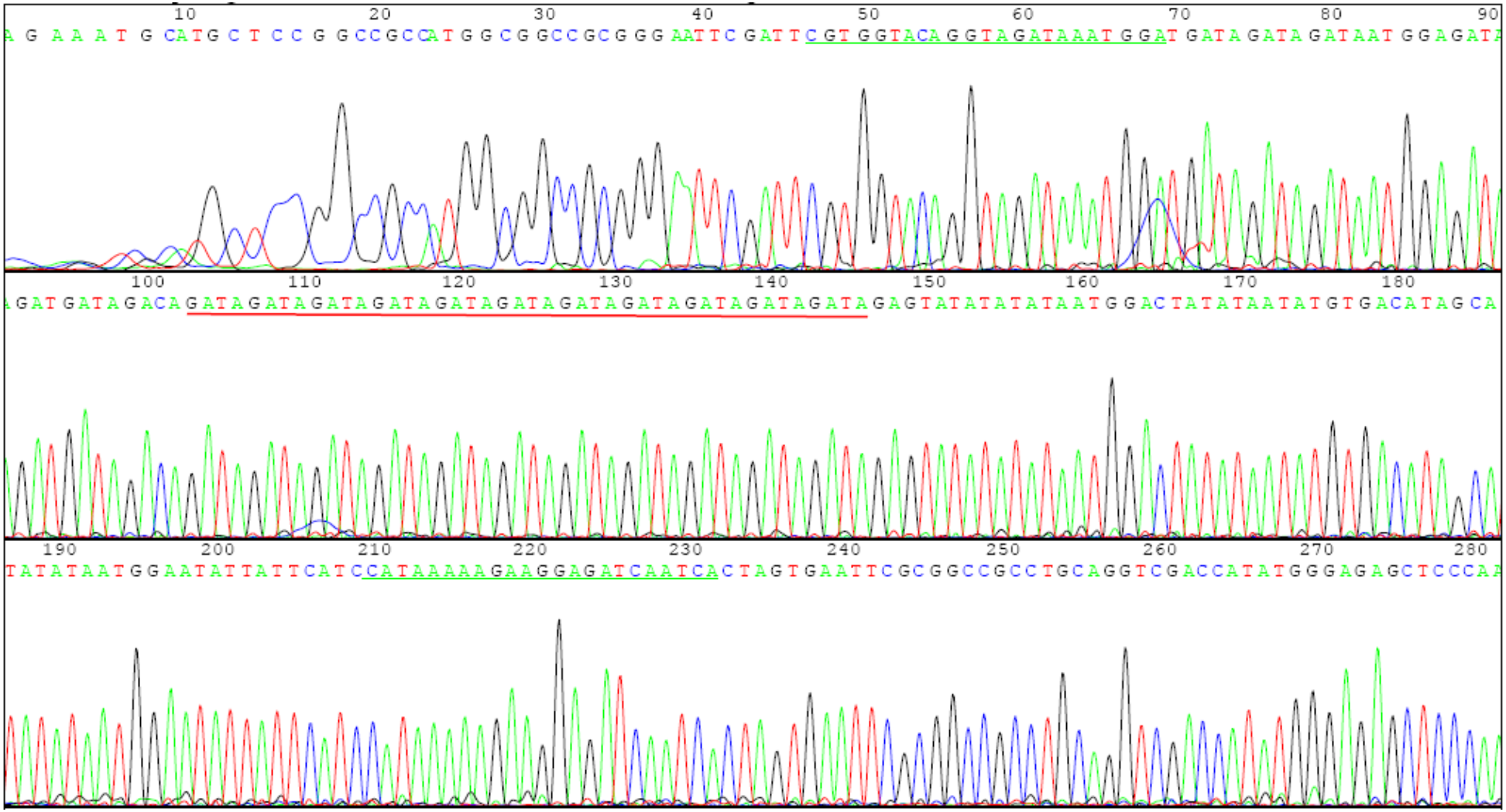
D10S674



continues

Figure 5.1 (continued)

D14S608



continues

Figure 5.1 (continued)

D15S659

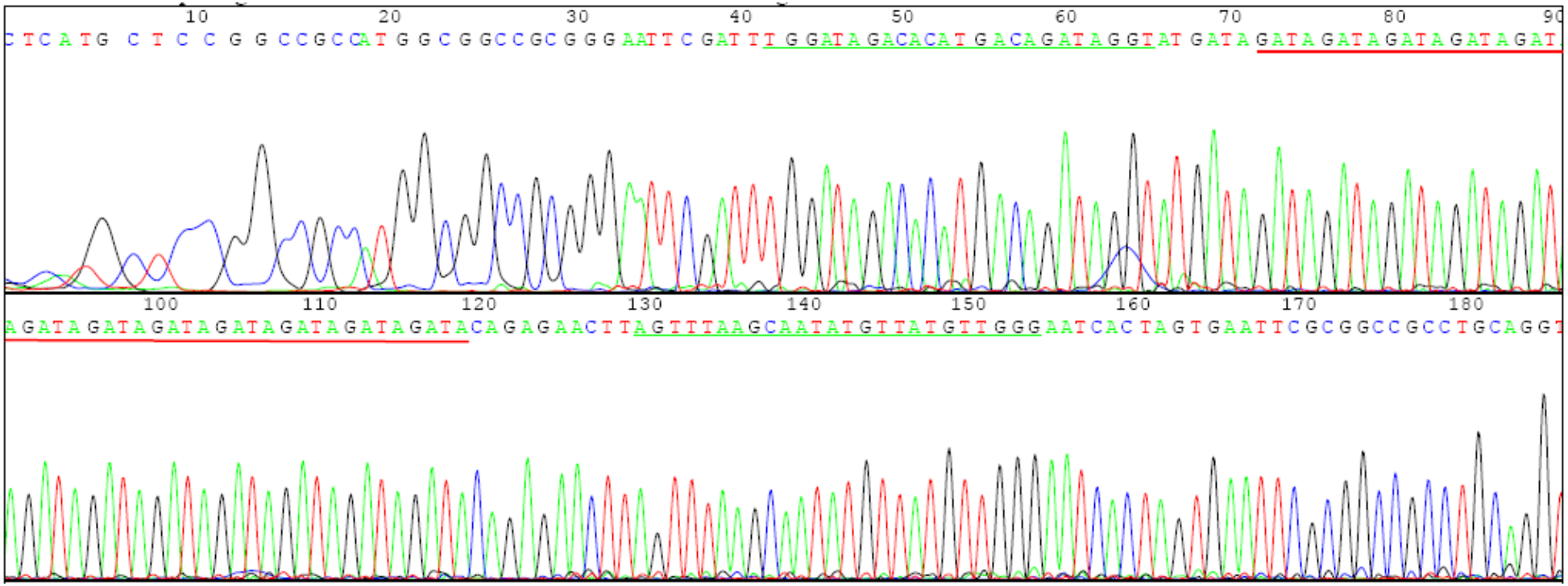
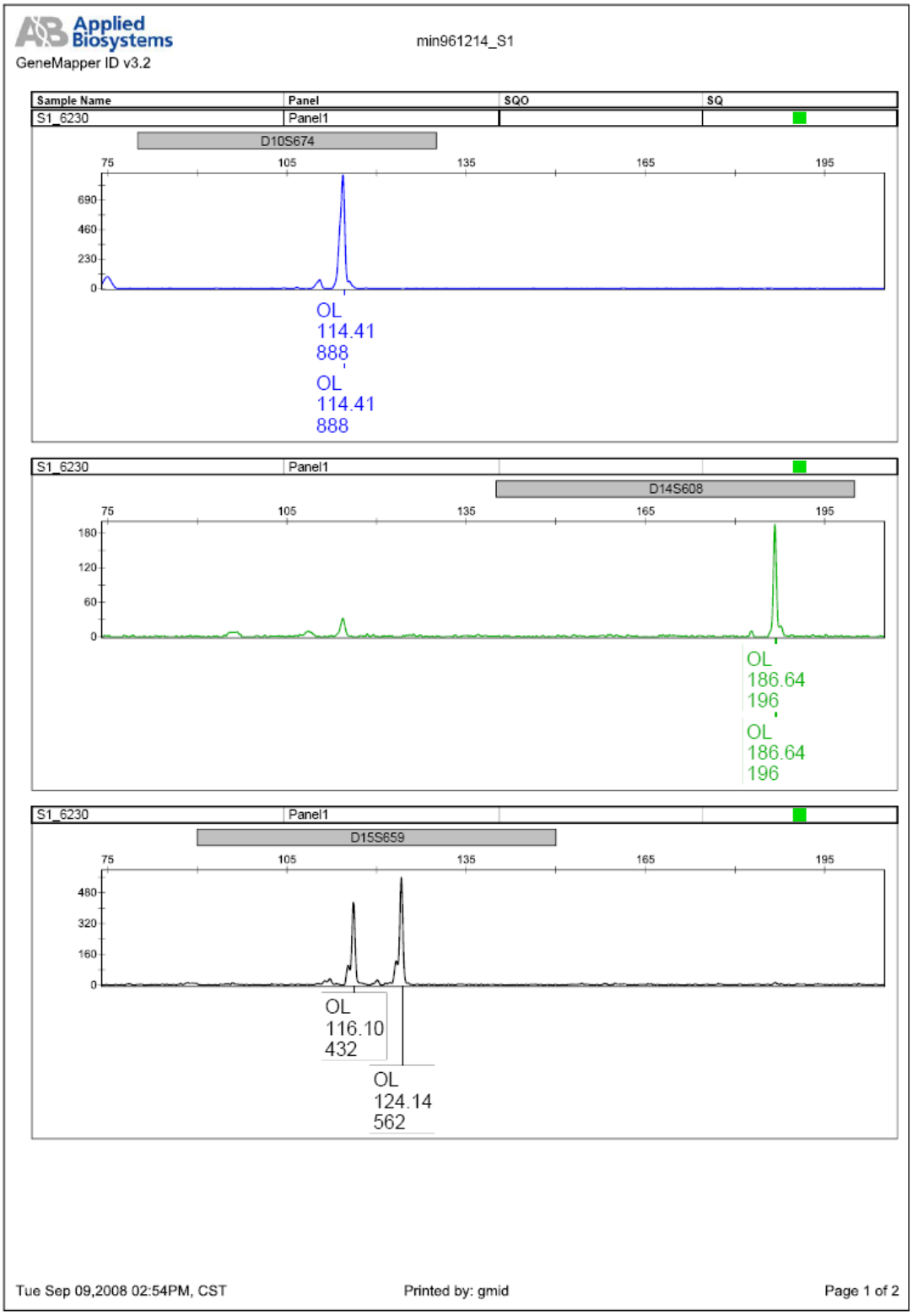


Figure 5.2: The peaks of the triplex.



continues

Figure 5.2 (continued)

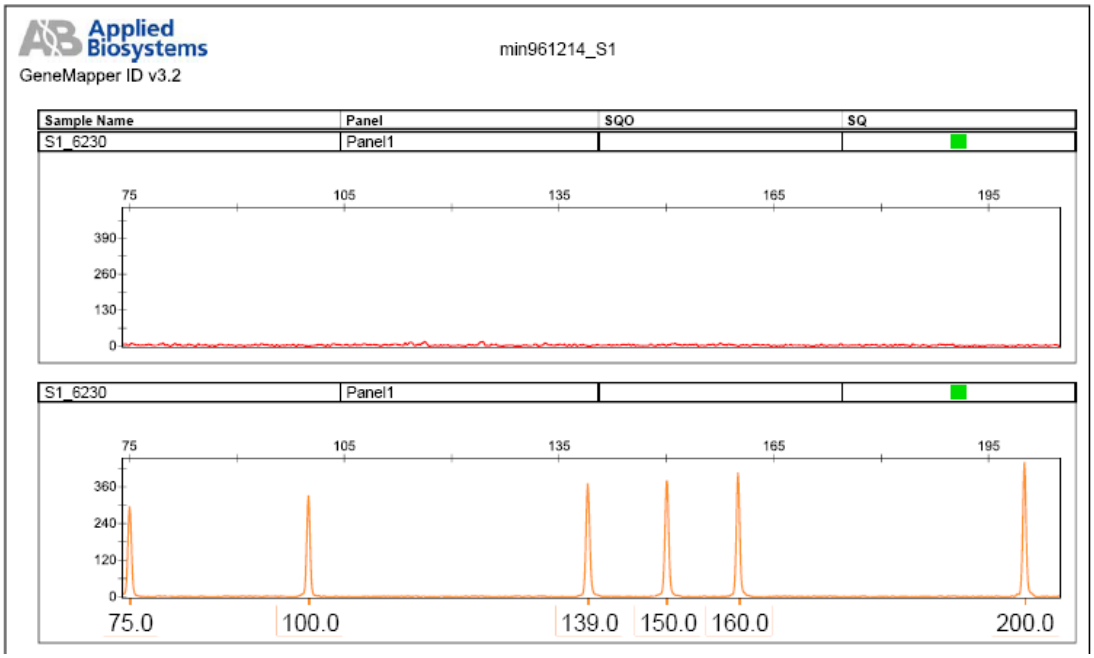


Table 5.4: Allele frequencies and forensic parameters for the triplex using a Taiwanese population.

Allele	D10S674	D14S608	D15S659
5	-	0.0021	-
5.3	-	0.0042	-
6	-	0.0975	-
6.3	-	0.0021	-
7	-	0.2076	-
8	0.0297	0.0318	-
9	0.1017	0.1292	-
10	0.0424	0.2564	-
11	0.1144	0.1631	0.1186
11.3	-	0.0042	-
12	0.2945	0.0784	-
12.3	-	0.0064	-
13	0.2669	0.0169	0.2013
13.3	0.0021	-	-
14	0.1144	-	0.1356
15	0.0339	-	0.0508
16	-	-	0.1462
17	-	-	0.2076
18	-	-	0.1208
19	-	-	0.0191
Homozygosity	0.1992	0.1568	0.1653
Heterozygosity	0.8008	0.8432	0.8347
PD	0.9083	0.9355	0.9472
PM	0.0917	0.0645	0.0528
PIC	0.7754	0.8095	0.8257
PE	0.5594	0.6318	0.6166
PI	2.5106	3.1892	3.0256
p value	0.2786	0.6497	0.7473

PD: power of discrimination
PM: power of matching
PIC: polymorphism information content
PE: power of exclusion
PI: typical paternity index
P: probability value of G tests for Hardy-Weinberg disequilibrium test.

5.4 Conclusion

The addition of these three STR loci can only be of benefit when analysing poor quality DNA samples. This non-CODIS triplex may be applied when increased powers of association are required in resolving relationship disputes.

Chapter 6: General Discussion

6.1 Unique Characteristics of this Study

Development of the methods of studying CPI cutoff values

Duo cases requiring the comparison of two DNA profiles believed to have come from two related persons is not a rare event (107). The starting process is to obtain STR profiles from the samples of interest. In the case of a paternity test (duo) case there should be an allele in common at all the loci tested, unless there is a mutational event. If there are matching alleles then a CPI can be calculated. If the CPI meets any preset minimum requirement (CPI cutoff values) then there is confidence that the two individuals tested are first-degree relatives. Unfortunately there is no universal CPI cutoff value, and though some cutoff values are accepted by many testing laboratories, these values may not supported by empirical data. The lack of an agreed CPI cutoff value may be due to the variation in the number of STR loci used in relationship testing laboratories. The effect of the CPI cutoff values based upon the STR loci used has not been reported previously. The duo paternity testing studies reported in this thesis highlight the need to consider kinship identification based upon the STR loci used, the CPI values obtained, specificity and sensitivity of the system

and their interaction with each other. Only then is there confidence in assigning an optimal preset CPI cutoff value.

The sensitivity (1 –rate of false exclusions) and specificity (1- rate of false positives) of the loci used need to be determined prior to setting CPI cutoff values. For evaluating sensitivity, sufficient numbers of paternity pairs (duos) with a proven genetic relationship (e.g. biological father and son) is the most important requirement. Unfortunately there is no easy way to access these data. This is a pioneering report by pairing 15 STR profiles from a random population to generate "real" off-springs. When using 15 STR profiles it is possible to obtain sufficient and reliable duo combinations, e.g. for the Chinese population 202,050 real duos were obtained. Based on access to this large amount of samples, the normal distribution of CPI can be seen, and the percentage of real duos excluded by different CPI cutoff values could be surveyed. Using these data it is possible to evaluate the sensitivity of the test.

In common with the sensitivity studies to examine the specificity of the DNA test under different CPI cutoff values, a sufficient amount of random pairs is also needed to check coincidental matches, typical of those observed in a paternity duo case. Using these data it is then possible to observe the effect of a

wide range of CPI values relative to false inclusions. If this study was based upon real data it would be prohibitively time consuming and may be the reason why there have been few such studies.

Development of the method of combining the index and allele-sharing number to determine the specificity of siblingship

The main method to assist with the determination of a potential sibling relationship between two people relies on the strength of the calculated DNA sibling index. Previous reports suggested that in addition to the indices obtained that by surveying the number of allele-shared at each locus further confirmation of a sibling relationship may be obtained (81). Thus allele-sharing could be an indicator to support a sibling linkage. Chapters 3 and 4 in this thesis are the first attempts to consider both the sibling index and the number of alleles that are shared. Using the combination of data real cases where a low DNA index is obtained may gain further support if TASL is considered. An example of this situation is when the CSI value is lower than "1", the result is inconclusive owing to the low index, but based on the data in Table 3.5 and given a TASL of 4, the specificity would be raised to 99.40% for Chinese.

Use of EXCEL Software to aid in the determination of various values

EXCEL™ is a part of OFFICE software provided by Microsoft Company. It is used widely in analytical laboratories for data analysis. The software is capable of manipulating STR data as part of the output by the DNA Sequencer. During the course of these studies STR data was imported into the worksheets of EXCEL. The output of the software was to produce a Visual Basic programme and Macros (a process to run a series of commands) to process the pairing of STR profiles, the generation of offspring DNA data, checking the allele-sharing of locus, and calculating of index. This software is widely available and could be used by almost all of those involved in the identification of human remains and linkage of DNA samples.

6.2 Future Studies

STR profiles from three populations were analysed. A similar method of analysis could be conducted on many more or smaller populations.

STR profiles can be searched rapidly against those held on criminal justice databases. Familial searching is possible if there are no matches but there is a possibility of a genetic relative being on the DNA database. The cutoff parameters recommended in these studies may be applied to identify various genetic relatives.

6.3 Fulfilling the goal of the study

One aim of this study was to better understand the specificity and sensitivity related to the indexes calculated from the standard kinship testing formulae. Based upon the recommendations in this thesis it is hoped that there can be greater confidence when determining siblingship when traditional index measuring methods result in low indexes.

Paternity duo determination

In this study 15 STR loci were examined from a total of 358,080 real duos, and 179,040 random pairs from 3 populations of Chinese in Taiwan, American Caucasians and African Americans. Due to the large sample size and resulting number of sample pairs meaningful data can be obtained that would not be possible by observations of non-virtual populations. The examination of a variable number of STR loci, from 7 to 15 loci, with the CPI cutoff values from 0 to 10,000. The CPI cutoff value applied affected the increase or decrease in specificity and sensitivity of the test. In short, under the same cutoff values, if more STR systems were applied, both specificity and sensitivity would be increased at the same time. If fewer STR loci were used the specificity could be maintained but sensitivity would be sacrificed with the potential increase in the number of false exclusions. For example, if the 9 loci that comprise STR Profiler kit is applied to the Chinese population in Taiwan

and the CPI cutoff value are set at 100, the specificity is 99.83% and sensitivity is 70.57%. However if 15 STR loci were used in this same population then specificity is increased to 99.98% and at the same time the sensitivity is increased to 99.72%. A similar situation was found for the other two populations. Based upon these data it is clear that at least 15 STR loci should be used if possible to maximise confidence in any DNA test as there should be an increase in both the specificity and sensitivity of the test.

Besides surveying the specificity of CPI under cutoff values, the probability of paternity could be obtained by using the Formula (1).

With a CPI value of 1 then the probability would be 1/ (1+1) = 50%. With a CPI value of 1,000 then the probability would be calculated as W=1,000/ (1,000+1) = 99.9%.

$$W = CPI/(CPI+1) \quad \text{---------- (1)}$$

When cases are examined using the 10 loci in SGM+, with a lower CPI value than that of Identifiler, a CPI value as low as 10 can be obtained from a duo case using two members of the Chinese population in Taiwan. The probability of paternity is only about 91% with a resulting low confidence that the two individuals are first degree relatives. It should be noted that according to Table 2.3 in this study, a

specificity of 99.85% can be reached. A similarly high specificity was found when using data from the populations of American Caucasians and African Americans. Apparently, when dealing with low CPI cases, this study provides a method for reporting the specificity based upon real data rather than calculating a probability of paternity in a duo case.

Siblingship determination

When comparing two DNA profiles to determine if they may be from siblings, traditionally determination of siblingship relies upon the calculation of the CSI. Most commonly this is performed using the formula issued by Wenk (57). The index can be transformed into probability of siblingship by using Formula (2)..

$$W = CSI/(CSI+1) \quad (2)$$

For example, if CSI > 1 then the probability of siblingship is larger than 50%. With the larger CSI indices then the probability of siblingship will also increase. Unfortunately random pairs of DNA profiles with CSI higher than 1 were found commonly within all three population. Even if the CSI value is increased to 10 there is no significant increase in the specificity, although there is no reduction in the sensitivity. Previous reports have suggested that the TASL (two-alleles-sharing-locus) is a crucial indicator to the determination of siblingship (112); this is also supported by the data presented in this thesis.

A software programme termed Visual Basic was specifically written to generate 15 STR loci data to mimic the DNA profiles of real siblings and of random pairs of individuals. This was conducted for all 3 populations generating a total of 357,630 real sibling pairs and 178,815 random pairs. Due to the large number of combinations generated it was possible to analyse the affect of CSI cutoff values with regard to the specificity and sensitivity.

The affect of various commonly used CSI cutoff values, ranging from 0.03125 to 1,000, including the affect on the rise and fall of specificity and sensitivity, is tabulated (Table 3.3). While this is the first table that describes such an extensive range of cutoff values and their affects to specificity and sensitivity, there remains an issue with putative sibling cases where there is a low CSI. A survey of routine cases by the author found that some comparisons with low CSI values had a high number of TASL. In this current study CSI cutoff values were combined with TASL to analyse their specificity and sensitivity. Eight variations of CSI (0.125-100) versus eight TASL situations (from 1 to 8), resulting in a total of 64 combinations showing the variation of these two parameters (Table 3.4) was presented. Instances of low CSI cases may be aided by the addition of TASL (from 1 to 8) thus altering the corresponding specificity and sensitivity.

$$W = CSI/(CSI + 1) \quad \text{---------- (2)}$$

For example, if the CSI is only 0.5, by using the traditional Formula (2) the two putative siblings might be excluded. However when applying a TASL = 5, from data in Table 3.4, a specificity of 99.79 % for the Chinese population in Taiwan can be obtained. This high specificity indicates a high confidence for determining siblingship. The same combination for the Caucasians and African American populations reaches a specificity of 99.74 %. This is the first study to illustrate the synergistic effects of CSI and TASL for the resolution of the siblingship determination, especially for the low index situation that could result in inconclusive cases.

Half-siblingship determination

When there is only one of the parents that are in common for the two individuals this is termed as half-siblingship. The uncertainty is higher in such a test compared to true siblings. The formula developed by Wenk (57) can still be used to obtain the CHSI for an evaluation. It has been predicted that the ratio of real half-siblingship with CHSI less than 1, and the ratio of random pairs with CHSI larger than 1, is higher than that for siblingship study. This study used the self-written programme to process the same populations in which there were a total of 355,622 real half-sibling pairs and 178,815 random pairs. Owing to the large number of sample pairs, all the

variations of CHSI cutoffs and their corresponding specificity and sensitivity could be analysed reliably.

Varying CHSI cutoff values ranging from 0.03125 to 1,000 were tabulated (Table 4.3) along with their relevant specificity and sensitivity. It was found that TASL could also differentiate real half-sibling group from random pairs group, but the differentiating capability was less than ASA. So ASA and CHSI index were combined to analyse the two groups.

Low CHSI index (0.125-10, 9 values) when paired to 12 various ASA situation (7-18, 12 values) resulted in 108 (9 x 12) combinations. Using these data the effects on both of specificity and sensitivity of CHSI cutoffs values could be examined

If the resultant CHSI was 0.5, by using the index-evaluation method described, half-siblingship could be excluded with confidence. Using the data in Table 4.3 if the ASA was 15 then a specificity of 97.29% could be referenced for members of the Chinese in Taiwan. For American Caucasians and African Americans, the specificity could be greater than 97.79 %. Combining the index and ASA for evaluating half-siblingship from STR profiles results in increased specificity and sensitivity in this initial and pioneering study.

The consistency of CPI cutoffs, CSI cutoffs, CHSI cutoffs and their

corresponding affects on specificity and sensitivity was found in all the three populations used in this study. The data presented indicates that these methods and programme are robust and reliable.

Conclusions

From this study 6 points could be concluded:

1. DNA laboratories using STR typing for kinship relationship testing are recommended to use suitable and appropriate CPI cutoff values based upon the number of STR loci tested.

2. Specificity values as shown in Table 2.4 should be used in cases of low CPI values from non-optimal DNA samples.

3. DNA laboratories are recommended to set a suitable CPI cutoff value to maximise the specificity and sensitivity. The data in Table 2.5 illustrates that with increasing STR loci only low CPI cutoff values are needed to reach a specificity of 99.99% while at the same time the sensitivity is also high. For example when using 15 STR and a CPI cutoff value of 94 the specificity was 99.99% and the sensitivity was 99.79%. When using the CODIS 13 and a CPI cutoff value of 565 the specificity was 99.99% but the sensitivity was only 91.06%.

4. It is better for a DNA laboratory to enhance the capability for evaluating low index cases in sibling determination by combining CSI with TASL. To enhance the capability for evaluating low index cases in half-sibling determination by

combining CHSI with ASA, for avoiding losing of cases.

5. For processing the siblingship and half-siblingship cases, it is suggested to use more STR loci, e.g. the 15 Core STR system in this study, to maintain the sufficient specificity and sensitivity simultaneously.

6. The cutoff index values combined with allele sharing number in this study can be applied to victim identification in mass disasters and familial search of DNA database as a means of screening for potential matches as first degree relatives or second degree of relatives.

References

1. Mullis KB and Faloona FA. Specific synthesis of DNA in vitro via a polymerase-atalyzed chain-reaction. Methods in Enzymology. 1987; 155: 335-350.
2. Jeffreys AJ, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. Nature 1985; 316(6023): 76-9.
3. Jeffreys AJ, Brookfield JF, Semeonoff R. Positive identification of an immigration test-case using human DNA fingerprints. Nature 1985; 317(6040): 818-9.
4. Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. Nature 1985; 314(6006): 67-73.
5. Jeffreys AJ. The forensic role of DNA profiling in Scotland. Royal Society of Edinburgh DNA symposium, JFSS 1993; 33(4): 204-211.
6. Easteal S, McLeod N and Reed K. The Future, New Approaches & Future Development. In: DNA Profiling Principles, Pitfalls and Potential, Chur, Switzerland: Harwood Academic Publishers GmbH, 1991. pp.129-182.
7. Edwards AI, Civitello A, Hammond HA and Caskey CT. DNA Typing and Genetic Mapping with Trimeric and Tetrameric Tandem Repeats. Am J Hum Genet 1991; 49:746-756.
8. Evison MP, Smillie DM, Chamberlain AT. Extraction of single-copy nuclear DNA from forensic specimens with a variety of postmortem histories. J Forensic Sci 1997; 42(6):1032-8.
9. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 Short Tandem Repeat Loci for Use in Personal Identification Application. Am J Hum Genet 1994; 55:175-189.
10. Latorra D, Schanfield MS. Analysis of human specificity in AFLP system APOB, PAH, and D1S80. Forensic Sic Int 1996; 83(1): 15-25..
11. Helmuth R, Fildes N, Blake E, Luce MC, Chimera J, Madej R. HLA-DQa allele and genotype frequencies in various human populations, determined by using enzymatic amplification and oligonucleotide probes. Am J Hum Genet, 1990; 47: 515-523.
12. Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG. Minisatellite repeat coding as a digital approach to DNA typing. Nature. 1991; 354(6350): 204-9.
13. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhans S, Finis CJ, Tomsey CS, Zachetti JM, Masibay A, Rabbach DR, Amlott EA, Sprecher CJ. Validation of a

16 locus fluorescent multiplex system. J Forensic Science, 2002; 47(4): 773-85.
14. Varsha. DNA fingerprinting in the criminal justice system: an overview, DNA Cell Biol. 2006; 25(3): 181-8. Review.
15. Kimpton CP, Oldroyd NJ, Watson SK, Frazer RR, Johnnson PE, Millcan ES, Urguhrt A, Sparks BL, Gill P. Validation of highly discriminating multiplex short tandem repeat amplification system for individual identification. Electrophoresis 1996; 17(8): 1283-93.
16. Martin PD, Schmitter H, Schneider PM. A brief history of the formation of DNA databases in forensic science within Europe. Forensic Sci Int. 2001; 119(2): 225-31.
17. Oldroyd NJ, Urquhart AJ, Kimpton CP, Millican ES, Watson SK, Downes T, Gill P. A highly discriminative octaplex short tandem repeat polymerase chain reaction system suitable for human individual identification. Electrophoresis 1995; 6: 334-337.
18. Applied Biosystems, Homepage, http://www.appliedbiosystems.com.tw/ (accessed May 1 2008) .
19. Promega, Homepage, http://www.promega.com/ (accessed May 1 2008).
20. Werret DJ. The national DNA database. Forensic Sci Int 1997; 88: 33-42.
21. Goodwin W, Linacre A, Hadi S. Databases of DNA profiles. In: An Introduction to Forensic Genetics. John Wiley & Sons, Ltd. West Sussex, UK, 2007. pp97-104.
22. Budowle B, Moretti TR, Niezgoda SJ, and Brown BL. CODIS and PCR-Based Short Tandem Repeat Loci: Law Enforcement Tools. Genetic Identity Conference Proceedings, Second European Symposium on Human Identification, 1998. http://www.promega.com/geneticidproc/eusymp2proc/17.pdf (accessed May 1 2008)
23. Short Tandom Repeat DNA Internet Database, http://www.cstl.nist.gov/div831/strbase/ (accessed May 1 2008)
24. Wyman AR and White R. A highly polymorphic locus in human DNA. 1980; Proc Natl Acad Sci USA. 1980; 77(11): 6754-8.
25. Key Dates in the History of DNA Profiling, http://www.crimtrac.gov.au/systems_projects/KeyDatesintheHistoryofDNAProfiling.html (accessed April 1 2008)
26. NRC I—National Research Council Committee on DNA Technology in Forensic Science, Board on Biology, Commission on Life Sciences, DNA Technology in Forensic Science. 1992; National Academy Press: Washington,

DC.
27. Gill P, brinkmann B, d'Aloja F, Anderson J, Bar W, Carracedo A, Dupuy B, Eriksen B, Jangblad M, Johnson V, Kloosterman AD, Lincoln P, Morling N, Rand S, Sabatler M, Scheithauer R, Schneider P, Vide MC. Consideration from the European DNA profiling group (EDNAP) concerning STR nomenclature. Forensic Sci Int. 1997; 87(3): 185-92.
28. Carracedo A, Barros F, Pestoni C, Lareu MV. Standardization of Forensic DNA Analysis in Europe Forensic Science International 1997; 86: 87-102.
29. Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Mahamians, Jamaicans, and Trinidadians. J Forensic Sci. 1999;44(6): 1277-86.
30. Roewer L, Arnemann J, Spurr NK, Grzeschik KH, Zpplen. Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum Genet 1992; 89: 389–94.
31. INNOCENCE PRSOJECT Homepage, http://www.innocenceproject.org/ (accessed October 1 2008).
32. Whitaker JP, Clayton TM, Urquhart AJ, Millican ES, Downes TJ, Kimpton CP, Gill P. Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. Biotechniques. 1995; 18(4): 670-7
33. Gill P, Kimpton C, D'Aloja E, Andersen JF, Bar W, Brinkmann B, Holgersson S, Johnsson V, Kloosterman AD, Lareu MV. Report of the European DNA profiling group (EDNAP)--towards standardization of short tandem repeat (STR) loci. Forensic Sci Int 1994; 65(1): 51-9.
34. Butler JM, Buel E, Crivellente F, McCord BR. Forensic DNA typing capability electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis 2004; 25: 1937-1412.
35. Huffine E, Crews J, Kennedy B, Bomberger K, Zinbo Asta. Mass Identification of Persons Missing from the Break-up of the Former Yugoslavia: Structure, Function, and Role of the identification Commission on Missing Persons. Croatian Medical Journal 2001; 42(3); 271-275.
36. Lins AM, Sprecher CJ, Puers C, Schumm JW. Multiplex sets for the amplification of polymorphic short tandem repeat loci--silver stain and fluorescence detection. Biotechniques 1996; 20(5): 882-9.
37. Walsh S. Recent advances in forensic genetics. Expert Rev Mol Diagn 2004; 4(1): 31-40.

38. Budowle B, Shea B, Niezgoda S, Chakraborty. CODIS STR Loci Data from 41 Sample Populations. J Forensic Sci 2001; 46(3): 453-489.
39. Butler JM. Overview and history of DNA typing. In: Forensic DNA Typing, 2nd, Elsevier Academic Press, MA, USA, 2005, pp 1-13.
40. Greenspoon SA, Yeung SHI, Ban JD, Mathies RA. Microchip Capillary Electrophoresis: Progress Toward an Integrated Forensic Analysis System. Profiles in DNA. 2007; 10(2):16-18.
41. Mennell J, Shaw I. The future of forensic and crime scene science. Part I. A UK forensic science user and provider perspective, Forensic Sci Int. 2006; 14;157 Suppl 1: S7-12.
42. Gill P, Forreman L, Buckleton HS, Triggs CM, Allen H. A comparison of adjustment methods to test the robustness of an STR DNA database compared of 24 European populations. Forensic Sci Int 2003; 131: 184-196.
43. Knijff PD. Son, Give Up Your Gun: Presenting Y-STR Results in Court. Profiles in DNA 2003; 6(1):3-5, http://www.promega.com/profiles/602/profilesin DNA_602_03.pdf (Accessed September 2008).
44. Relationship Testing Program Unit. ANNUAL REPORT SUMMARY FOR TESTING IN 2004, http: //www.aabb.org/ Documents/ Accreditation/ Parentage_Testing_Accreditation_Program/ rtannrpt04.pdf (accessed May 1 2008)
45. Gill P, Fereday L, Morling N, Schneider PM. The evolution of DNA databases—Recommendation for new European STR Loci. 2006; Forensic Sci Int 156: 242-244.
46. Fatolitis L, Ballantyne J. The US Y-STR Database. Profiles in DNA 2008; 11(1): 13-14.
47. Tracey M. Short Tandem Repeat-based Identification of Individuals and Parents. Croat Med J. 2001; 42(3): 233-238.
48. Alford RL, Hammond HA, Coto I, Caskey CT. Rapid and efficient resolution of parentage by amplification of short tandem repeats. Am J Hum Genet. 1994; 55(1): 190-5.
49. Butler JM. Kinship and parentage testing. In: Forensic DNA Typing, Elsevier Academic Press, MA, USA, 2005, pp 529-537.
50. Lu D, Lu H, Guo J. Exclusion of 15 STRs in single parent paternity testing. Experiences of management of aviation disasters in Taiwan. 17th the Meeting of the International Association of Forensic Science Hong Kong 21-26 August 2005. pp345.
51. Wenk RE, Chiafari FA, Gorlin J, Polesky HF. Better tools are needed for

parentage and kinship studies. Transfusion. 2003; 43(7): 979-81.
52. Goodwin W, Linacre A, Hadi S. Kinship testing. In: An Introduction to Forensic Genetics. John Wiley & Sons, Ltd. West Sussex, UK, 2007. pp105-123.
53. Pu CE, Linacre A. CPI distribution and cutoff values for duo kinship testing. Chinese J. Physiology 50(5):232-239, 2007.
54. Brinkmann B, Pfeiffer H, Schurenkamp M, Hohoff C. The evidential value of STRs. An analysis of exclusion cases. Int J Legal Med. 2001; 114(3):173-7.
55. AABB, Guidance for standards for relationship testing laboratories, MD, USA, American association of blood banks, 2006, pp 139.
56. Hsu CM, Huang NE, Tsai LC, Kao LG, Chao CH, Linacre A, Lee JC. Identification of victims of the 1998 Taoyuan Airbus crash accident using DNA analysis. Int J Legal Med. 1999; 113(1): 43-6.
57. Wenk RE, Traver, M. and Chiafari, FA. Determination of sibship in any two persons. Transfusion 1996, 36:259-262.
58. Clayton TM, Whitaker JP, Maguire CN. Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci. Forensic Sci Int. 1995; 76(1):7-15.
59. Corach D, Sala A, Penacino G, Sotelo A. Mass disasters: rapid molecular screening of human remains by means of short tandem repeats typing. Electrophoresis. 1995; 16(9): 1617-23.
60. Primorac D, Andelinovic S, Definis-Gojanovic M, Drmic I, Rezic B, Baden MM, Kennedy MA, Schanfield MS, Skakel SB, Lee HC. Identification of war victims from mass graves in Croatia, Bosnia, and Herzegovina by use of standard forensic methods and DNA typing. J Forensic Sci 1996; 41(5): 891-4.
61. Goodwin W, Linacre A, Vanezis P. The use of mitochondrial DNA and short tandem repeat typing in the identification of air crash victims. Electrophoresis 1999; 20: 1707-1711.
62. Lorente JA, Entrala C, Alvarez JC, Lorente M, Arce B, Heinrich B, Carrasco F, Budowle B, Villanueva E. Social benefits of non-criminal genetic databases: missing persons and human remains identification. Int J Legal Med. 2002; 116 (3):187-90.
63. Leclair B, Fregeau CJ, Bowen KL, Fourney RM. Enhanced kinship analysis and STR-based DNA typing for human identification in mass fatality incidents: the Swissair flight 111 disaster. J Forensic Sci. 2004; 49(5): 939-53.
64. Deng YJ, Li YZ, Yu XG, Li L, Wu DY, Zhou J, Man TY, Yang G, Yan JW, Cai DQ, Wang J, Yang HM, Li SB, Yu J. Preliminary DNA identification for the tsunami victims in Thailand. Geno Prot Bioinfo 2005; 3(3): 143-157.

65. Alonso A, Martin P, Albarrán C, Garcia P, Fernandez de Simon L, Jesús Iturralde M, Fernández-Rodriguez A, Atienza I, Capilla J, García-Hirschfeld J, Martinez P, Vallejo G, García O, García E, Real P, Alvarez D, León A, Sancho M. Challenges of DNA profiling in mass disaster investigations. Croat Med J. 2005; 46(4): 540-8.
66. List of accidents and disasters by death toll, http://en.wikipedia.org/wiki/List_of_accidents_and_disasters_by_death_toll. (Accessed on September 15 2008).
67. International Commission on Missing Persons (ICMP) Homepage, http://www.ic-mp.org/home.php . (Accessed on September 15 2008).
68. Andelinović S, Sutlović D, Erceg Ivkosić I, Skaro V, Ivkosić A, Paić F, Rezić B, Definis-Gojanović M, Primorac D. Twelve-year experience in identification of skeletal remains from mass graves. Croat Med J. 2005; 46 (4): 530-9.
69. Mass Fatality Incidents: A Guide for Human Forensic Identification. June 2005, http://www.ojp.usdoj.gov/nij/pubs-sum/199758.htm.(Accesed on September 15 2008)
70. Disaster Victim Identification Homepage, http://www.interpol.int/Public/DisasterVictim/Default.asp (Accesed on September 15 2008)
71. INTERPOL. INTERPOL GUIDE-DISASTER VICTIM IDENTIFICATION, 1997, http://www.interpol.int/Public/DisasterVictim/guide/guide.pdf
72. Ballantyne J. mass disaster genetics. Nature genetics 1997;15: 329331.
73. Brenner CH, Weir BS. Issues and strategies in the DNA identification of World Trade Center victims. Theor Popul Biol. 2003; 63 (3):173-8.
74. Konjhodzic R, Omerovic E, Davoren J, Vanek D, Crews J and Huffine E. Performance Comparison of the Penta D and Penta E Loci With the D2S1338 and D19S433 Loci in the Massive DNA Identification Effort in Former Yugoslavia, In: Proceedings of the 55th Annual Meeting of the American Academy of Forensic Sciences, Chicago, IL, USA, 2003, pp 37.
75. Rizvic A and Huffine E. Coordinating the Identification Efforts of the Missing from Former Yugoslavia, In: Proceedings of the 55th Annual Meeting of the American Academy of Forensic Sciences, Chicago, IL, USA, 2003, pp 64-65.
76. Leclair B, Shaler R, Carmody GR, Eliason K, Hendrickson BC, Judkins T, Norton MJ, Sears C, Scholl T. Bioinformatics and human identification in mass fatality incidents: the world trade center disaster. J Forensic Sci. 2007; 52(4): 806-19.
77. Birus I, Marcikic M, Lauc D, Dzijan S and Lauc G. How high should paternity

index be for reliable identification of war victims by DNA typing? Croat. Med. J. 44: 322-6, 2003.
78. Maguire C and Woodward M. Kinship Analysis: DNA-based Kinship Analysis. Profiles in DNA 2008; 11(1): 3-6.
79. De Ungria MC, Frani AM, Magno MM, Tabbada KA, Calacal GC, Delfin FC and Halos SC. Evaluating DNA tests of motherless cases using a Philippine genetic database. Transfusion 42:954-957, 2002.
80. Gornik I, Marcikic M, Kubat M, Primorac D, and Lauc G. The identification of war victims by reverse paternity is associated with significant risks of false inclusion. Int. J. Legal. Med. 116:255-7, 2002.
81. Presciuttini S, Ciampini F, Alu M, Cerri N, Dobosz M, and Domenici R Allele sharing in first-degree and unrelated pairs of individuals in the Ge.F.I. AmpFlSTR® Profiler Plus™ database, Forensic Sci. Int. 2003; 131: 85-89.
82. Pu, CE and Linacre A. CPI distribution and cutoff value for duo paternity building. In: Proceedings of the 58th Annual Meeting of the American Academy of Forensic Sciences, Seattle, WA, USA, 2006, pp 172-173.
83. Pu CE, Wu FC, Cheng CL, Wu KC, Chao CH, and Li JM. DNA STR profiling of Chinese population in Taiwan determined by using an automated sequencer, Forensic Sci. Int. 1998; 97:47-51.
84. Pu, CE. et al. DNA profiling of Chinese population in Taiwan on the STR loci D8S1179, D16S539, D18S51 and D21S11 52th annual meeting, American Academy of Forensic Sciences, Reno, NV, USA, 2000.
85. Lee JC, Chang YY, Su CW, Tzeng CH, Pu CE, Han TH, Chang CW, Liu DC, Lin FR, Lin YC, Chen MH. The Screening of STR and YSTR Loci in Taiwanese Han Population, Proceedings of Taiwan Academy of Forensic Sciences, 2006, 27-38. (in Chinese)
86. Kirkwood BR and Sterne JAC. Measurement error: assessment and implications. In: Medical statistics, MA, USA, Blackwell publishing, pp. 429-446, 2003.
87. Andjelinovic S, Sutlovic D, Erceg DI, Skaro V, Ivkosic A, Definis-Gojanovic BRM, Primorac D. Identification of Skeletal Remains From Mass Graves: Ten-Year Experence of Our Work. In: Proceedings of the 55th Annual Meeting of the American Academy of Forensic Sciences, Chicago, IL, USA, 2003, pp 70.
88. Huffine E, Crews JD, Davoren J, Vanek D, Konihodzic R. A DNA-Led Identification Program for the Former Yugoslavia. In: Proceedings of the 55th Annual Meeting of the American Academy of Forensic Sciences, Chicago, IL, USA, 2003, pp 50.

89. Brenner CH. A note on paternity computation in cases lacking a mother. Transfusion. 1993; 33(1): 51-4.
90. Brenner CH. Some mathematical problems in the DNA identification of victims in the 2004 tsunami and similar mass fatalities. Forensic Sci Int. 2006; 157(2-3):172-80.
91. Calafell F. The probability distribution of the number of loci indicating exclusion in a core set of STR markers. Int J Legal Med. 2000; 114(1-2):61-5.
92. Chakraborty R, Stivers DN. Paternity exclusion by DNA markers: effects of paternal mutations. J of Forensic Sci 1996; 41(4):671-677.
93. Forensic Sci Int. 2000 Nov 13; 114(2):57-65. Comment in: Forensic Sci Int. 2002; 125(2-3):284. Motherless case in paternity testing. Lee HS, Lee JW, Han GR, Hwang JJ.
94. Gjertson DW, Morris JW. Assessing probability of paternity and the product rule in DNA systems. Genetica. 1995; 96(1-2): 89-98.
95. Hallenberg C, Morling N.A report of the 2000 and 2001 paternity testing workshops of the English speaking working group of the international society for forensic genetics. Forensic Sci Int. 2002; 129(1): 43-50.
96. Int J Legal Med. 1996; 108(4):229. Comment on: Int J Legal Med. 1994; 107(3): 127-31.
97. Karlsson AO, Holmlund G, Egeland T, Mostad P. DNA-testing for immigration cases: the risk of erroneous conclusions.Forensic Sci Int. 2007; 172(2-3):144-9.
98. Lee CL, Lebeck L, Pothiawala M. Exclusion of paternity without testing the mother. Am J Clin Pathol. 1980; 74(6): 809-12.
99. Luque JA, Valverde JL. Paternity evaluation in cases lacking a mother and nondetectable alleles. J Forensic Sci. 1996; 41(4): 671-7.
100. Morling N, Allen RW, Carracedo A, Geada H, Guidet F, Hallenberg C, Martin W, Mayr WR, Olaisen B, Pascali VL, Schneider PM; Paternity Testing Commission of the International Society of Forensic Genetics. Paternity Testing Commission of the International Society of Forensic Genetics: recommendations on genetic investigations in paternity cases. Forensic Sci Int. 2002; 129(3):148-57.
101. Pena SD. Am J Hum Genet. 1995; 56(6):1503-6. Comment on: Am J Hum Genet. 1994; 55(1):190-5. Pitfalls of paternity testing based solely on PCR typing of minisatellites and microsatellites.
102. Putsch M, Lucked C, Repenning A, Fischer L, Malyusz V, Simeoni E, Lignitz E, Oehmichen M, von Wurmb-Schwark N.The problem of single parent/child paternity analysis--practical results involving 336 children and 348 unrelated

men.Forensic Sci Int. 2006 ;159(2-3): 98-103.
103. Presciuttini S, Toni C, Tempestini E, Verdiani S, Casarino L, Spinetti I, De Stefano F, Domenici R, Bailey-Wilson JE. Inferring relationships between pairs of individuals from locus heterozygosities. BMC Genet. 2002; 3:23.
104. Pretty IA, Hildebrand DP.The forensic and investigative significance of reverse paternity testing with absent maternal sample. Am J Forensic Med Pathol. 2005; 26(4):340-2.
105. Bieber FR, Brenner CH, Lazer D.Human genetics. Finding criminals through DNA of their relatives.Science. 2006; 312(5778):1315-6.
106. Tai JJ, Chen MH. A non-excluded set method for the calculation of paternity probability. Proc Natl Sci Counc Repub China B. 1994; 18(1):36-43.
107. Thompson WC, Taroni F, Aitken CG. How the probability of a false positive affects the value of DNA evidence.J Forensic Sci. 2003; 48(1): 47-54.
108. Wenk RE, Gjertson DW, Chiafari FA, Houtz T. The specific power of parentage exclusion in a child's blood relatives. Transfusion. 2005; 45(3): 440-4.
109. Wenk RE, Houtz T, Chiafari FA.Maternal typing and test sufficiency in parentage analyses.Transfusion. 2006; 46(2): 199-203.
110. Wurmb-Schwark N, Malyusz V, Simeoni E, Lignitz E, Poetsch M.Possible pitfalls in motherless paternity analysis with related putative father. Forensic Sci Int 2006; 159: 92-97.
111. Zou KH, Wells III WM, Kikinis R and Warfield SK. Three validation metrics for automated probabilistic image segmentation of brain tumors, Statist. Med. 23:1259-1282, 2004.
112. Tzeng CH, Lyou JY, Chen YR, Hu HY, Lin JS, Wang SY, Lee JC. Determination of sibship by PCR-amplified short tandem repeat analysis in Taiwan. Transfusion. 2000, 40: 840-845.
113. Wenk RE and Chiafari FA. Distinguishing full siblings from half-siblings in limited pedigrees. Transfusion 2000, 40: 1148-9.
114. Henke J and Henke L. Which short tandem repeat polymorphisms are required for identification? Lessons from complicated kinship cases. Croat Med J 2005, 46(4): 593-597.
115. Thomson JA, Ayres KL, Pilotti V, Barrett MN, Walker JIH. and Debenham PG. Analysis of disputed single-parent/child and sibling relationships using 16 STR loci. Int J Legal Med 2001, 115: 128-134.
116. Reid TM, Wolf CA, Kraemer CM, Lee SC, Baird ML and Lee RF. Specificity of sibship determination using the ABI Identifiler multiplex system. J Forensic Sci 2004, 49: 1-3.

117. Giroti RI, Verma S, Singh K, Malik R and Talwar I. A gray zone approach for evaluation of 15 short tandem repeat loci in sibship analysis: A pilot study in Indian subjects. 2006 J. Clin. Forensic Med. doi: 10.1016/j.jcfm.2006.08.002.
118. Gaytmenn R, Hildebrand DP, Sweet D and Pretty IA. Determination of the sensitivity and specificity of sibship calculations using AmpFISTR Profiler Plus. Int J Legal Med 2002, 116:161-164.
119. Allen RW, Fu J, Reid TM, Baird M. Considerations for the interpretation of STR results in cases of questioned half-sibship. Transfusion. 2007 47:515-519. http://www.cbsnews.com/stories/2006/03/17/60minutes/main1414965.shtml (accessed June 1 2007)
120. Minifiler™ kit, http://marketing.appliedbiosystem.com/mk/get/MINIFILER_LANDIMG?isource=fr_E_Alias_minifiler_prodRegMinifiler_20061114 (accessed June 1 2007)
121. TPMD, http://tpmd.nhri.org/php-bin/index.php (accessed March 1 2007)
122. Rozen S, Skaletsky H, Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 2000; 132: 365-86.
123. http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi (accessed April 1 2007)
124. NCBI, http://www.ncbi.nlm.nih.gov/ (accessed May 1 2007)
125. Pu CE, Chen MY, Meng LM, Cheng SM, Wu FC and Linacre AM. An EXCEL based program for forensic DNA population study, Proceedings of Taiwan Academy of Forensic Sciences, 2004, 189-196. (Abstracts in Chinese)
126. Grubwieser P, Zimmermann B, Niederstatter H, Pavlic M, Steinlechner M, Parson W. Evaluation of an extended set of 15 candidate STR loci for paternity and kinship analysis in an Austrian population sample. Int J Legal Med. 2007, 121(2): 85-9.
127. Shi M, Yu X, Bai R, Shu X, Zhu G, Lv J, Tu Y. Genetic polymorphism of 14 non-CODIS STR loci for forensic use in southeast China population. Forensic Sci Int. 2008, 15; 174(1): 77-80.

Appendix

Catalog of appendix

Appendix 1: Table of Alllele frequencies of a population taken from the Chinese population in Taiwan

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
5						0.0025	0.0025							0.0025	
6						0.1130	0.0025								
7			0.0025	0.0030		0.2845	0.0015							0.0260	
8			0.1395	0.0030		0.0635	0.2810	0.0025				0.5795		0.0040	
9	0.0025		0.0730	0.0525		0.4575	0.1240	0.2875		0.0007		0.0995		0.0810	
9.3						0.0240									
10	0.1395		0.1590	0.2635		0.0560	0.1555	0.1265		0.0007		0.0250	0.0025	0.1910	
10.2										0.0007					
11	0.1030		0.3560	0.2510		0.0025	0.2460	0.2455		0.0021		0.2770	0.0030	0.3335	
11.2										0.0009					
12	0.1195		0.2290	0.3340	0.0025		0.1515	0.2230		0.0357		0.0190	0.0410	0.2205	
12.2										0.0054					
13	0.2095		0.0345	0.0800	0.0025		0.0325	0.0960		0.3032	0.0025		0.1560	0.1350	
13.2										0.0347					
14	0.1940		0.0075	0.0100	0.0430		0.0065	0.0165	0.0007	0.2318	0.2640		0.2155	0.0080	0.0025
14.2										0.1186					
15	0.1450			0.0025	0.3560			0.0025		0.0687	0.0355		0.1930	0.0025	
15.2										0.1503					
16	0.0700				0.3060				0.0158	0.0119	0.1580		0.1435		
16.2										0.0329					
17	0.0155			0.0025	0.2210				0.0590	0.0007	0.2430		0.0865		0.0025
17.2										0.0026					
18	0.0025				0.0640				0.0992	0.0007	0.1845		0.0440		0.0225
19					0.0085				0.1900		0.0945		0.0370		0.0590
20	0.0025								0.1115		0.0160		0.0280		0.0625
21	0.0025								0.0339		0.0030		0.0230		0.1155
21.2															0.0030
22									0.0484				0.0195		0.1795
22.2															0.0040
23									0.1973				0.0060		0.2185
23.2															0.0095
24									0.1676				0.0025		0.1800

Continues

Appendix 1 (continued)

24.2				0.0055
25		0.0643	0.0025	0.0805
25.2				0.0025
26		0.0100		0.0355
26.2				0.0025
27	0.0025	0.0021	0.0025	0.0145
27.2				0.0025
28	0.0485	0.0007		0.0025
28.2	0.0025			
28.3	0.0025			
29	0.2680			
29.2	0.0025			
30	0.2640			0.0025
30.2	0.0080			
30.3	0.0025			
31	0.1120			
31.2	0.0645			
32	0.0330			
32.2	0.1375			
33	0.0050			
33.2	0.0520			
34.2	0.0025			
35.2	0.0025			

*The frequency table generated by using the author written programme on 450 random Chinese STR profilers.

Appendix 2: Table of allele frequency of a Caucasian population.

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
5						0.0083						0.0083			
6						0.2318						0.0083			
7			0.01821			0.1904								0.00828	
8	0.01159		0.15066	0.0083		0.0844	0.11258	0.01821				0.5348		0.00828	
8.1			0.00828												
9	0.00828		0.17715	0.0116		0.1142	0.0745	0.11258				0.1192		0.04967	
9.3						0.3676									
10	0.10099		0.24338	0.2169		0.0083	0.05132	0.05629		0.00828		0.0563	0.00828	0.05132	
10.3															
11	0.08278		0.20695	0.3013	0.00828	0.0083	0.3394	0.32119		0.00828		0.2434	0.01656	0.36093	
12	0.18543		0.16556	0.3609			0.24834	0.32616		0.08113		0.0414	0.12748	0.38411	
12.2										0.00828					
13	0.30464		0.03477	0.0960			0.12417	0.1457		0.25331	0.0083	0.0083	0.13245	0.14073	
13.2										0.00828					
14	0.16556		0.00828	0.0083	0.10265		0.04801	0.01987		0.36921	0.0944		0.13742	0.00828	
14.2										0.01821			0.00828		
15	0.11424				0.26159		0.00828		0.00828	0.15232	0.1109		0.15894	0.00828	
15.2										0.03477					
16	0.03146				0.25331				0.03311	0.04967	0.2003		0.13907		
16.2										0.0149					
17					0.21523				0.18212	0.00828	0.2815		0.12583		
17.2										0.00828					
18					0.15232				0.07947		0.2003		0.07616		0.0265
18.2										0.00828					
19					0.01159				0.11424		0.1043		0.03808		0.0530
19.2															
20					0.00828				0.1457		0.0083		0.02152		0.1275
21									0.04139		0.0083		0.00828		0.1854
21.2															0.0083
22									0.03808				0.00828		0.2185
22.2															0.0116
22.3															
23									0.11755						0.1341

continues

Appendix 2 (continued)

23		0.1176	0.1341
23.2			0.0083
24		0.1225	0.1358
24.2			0.0083
25		0.0927	0.0712
25.2	0.0083		
26		0.0298	0.0232
27	0.0265	0.0083	0.0083
28	0.1589		
29	0.1954		
29.2	0.0083		
30	0.2782		
30.2	0.0282		
31	0.0828		
31.2	0.0993		
32	0.0083		
32.2	0.0844		
33	0.0083		
33.1			
33.2	0.0265		
34			
34.2	0.0083		
35	0.0083		

*Allele frequency table adopted from (23).

Appendix 3: Table of allele frequency of a population of African Americans

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
5						0.0097									
6			0.0097			0.1240						0.1008			
7			0.0155	0.0525		0.4205						0.0174			
8	0.0097		0.2364	0.0603		0.1938	0.0330	0.0388				0.3721		0.0485	
8.1															
9	0.0097		0.1085	0.0370		0.1512	0.0330	0.1957				0.1783	0.0097	0.0388	
9.3			0.0097			0.1047									
10	0.0291		0.3314	0.2568		0.0097	0.0233	0.1163		0.0097		0.0892	0.0097	0.0698	
10.3															
11	0.0446		0.2035	0.2490			0.3062	0.3178		0.0620		0.2190	0.0097	0.2326	
12	0.1415		0.0872	0.2977			0.4244	0.1957		0.1143	0.0097	0.0213	0.0778	0.3527	
12.2										0.0349					
13	0.2171		0.0136	0.0370	0.0097		0.1454	0.1182		0.2461	0.0097	0.0097	0.0525	0.2384	
13.2										0.0523			0.0097		
14	0.3004			0.0097	0.0892		0.0349	0.0174		0.2229	0.0775		0.0720	0.0155	
14.2										0.0795					
15	0.1841				0.3023					0.0775	0.1861		0.1615	0.0097	
15.2					0.0097					0.0601			0.0097		
16	0.0698				0.3353				0.0584	0.0097	0.2481		0.1576		
16.2										0.0271					0.0097
17	0.0097				0.2054				0.0992		0.2423		0.1518		
17.2										0.0097					
18	0.0097				0.0601				0.0389		0.1550		0.1226		0.0097
18.2										0.0097					0.0116
19					0.0097				0.1479		0.0620		0.0992		0.0620
19.2															0.0097
20									0.1031		0.0155		0.0642		0.0562
21									0.1440		0.0097		0.0097		0.1163
21.2													0.0097		
22									0.1304				0.0097		0.1957
22.2															0.0097
22.3															0.0097
23									0.1109				0.0097		0.1705

continues

Appendix 3 (continued)

23.2				0.0097
24		0.07977	0.00969	0.1221
24.2				
25		0.07198		0.1240
25.2				
26	0.00969	0.01167		0.0814
27	0.07752	0.00969		0.0233
28	0.25775			0.0116
29	0.19767			0.0097
29.2				
30	0.17442			0.0097
30.2	0.00969			0.0097
31	0.0814			
31.2	0.04651			0.0097
32	0.00969			
32.2	0.05814			
33	0.00969			
33.1	0.00969			
33.2	0.03488			
34	0.00969			
34.2				
35	0.02326			
36	0.00969			
37	0.00969			
38	0.00969			
39	0.00969			

*Allele frequency table adopted from (23).

Appendix 4: Complete Programme for generating real duos and processing coincidental matched pairs (continuation of Table 2.1)

```
Sub duo_cpi_main_Ch()
'function: make every pair of STR profiles from random population to
'generate two children STR profiles, calculate the CPI and
'record them into the worksheets-duo-cpis
'loading the frequency table from worksheets
Dim freqtab(54, 15)
Worksheets("freq tab").Select
Range("A4").Select
For i = 0 To 54
    For j = 0 To 15
        freqtab(i, j) = ActiveCell.Offset(i, j)
    Next j
Next i
'loading all the random population
Dim randompopulation(999, 30)
id_columns = 31
Worksheets("Ch").Select
individuals = Range("C1")
Range("A3").Select
For i = 0 To individuals - 1
    For j = 0 To id_columns - 1
        randompopulation(i, j) = ActiveCell.Offset(i, j)
    Next j
Next i
'generate all couples
```

```
Dim cpi_base(203000)
cpis = 0
Dim parent(1, 30)
For i = 0 To individuals - 2
    For j = i + 1 To individuals - 1
        For k = 0 To id_columns - 1
            parent(0, k) = randompopulation(i, k)
            parent(1, k) = randompopulation(j, k)
        Next k
        Call get2children_Ch(freqtab, parent, cpi_base, cpis)
        cpis = cpis + 1
    Next j
Next i
Worksheets("real duo cpis").Select
Range("A2").Select
Lines = 0
nextcolumns = 0
For i = 0 To cpis * 2 - 1
    ActiveCell.Offset(Lines, nextcolumns) = cpi_base(i)
    Lines = Lines + 1
    If Lines Mod 60000 = 0 Then
        Lines = 0
        nextcolumns = nextcolumns + 1
    End If
Next i
Range("A1").Select
```

```
End Sub
Sub duo_cpi_main_Ca_AA_HA()
'function: make every pair of STR profiles from random population to
'generate two children STR profiles, calculate the CPI and
'record them into the worksheets-duo-cpis
'loading the frequency table from worksheets
Dim freqtab(58, 15)
Worksheets("freq 3p").Select
Range("A4").Select
Caoffset = 0
AAoffset = 1
HAoffset = 2
For i = 0 To 58
        freqtab(i, 0) = ActiveCell.Offset(i)
Next i
Range("B4").Select
For i = 0 To 58
        For j = 1 To 15
                freqtab(i, j) = ActiveCell.Offset(i, (j - 1) * 3 + AAoffset)
        Next j
Next i
'Worksheets("temp").Select
'Range("A1").Select
'For i = 0 To 58
'      For j = 0 To 15
'               ActiveCell.Offset(i, j) = freqtab(i, j)
```

```
'       Next j
'Next i
'x = 1
'loading all the random population
Dim randompopulation(999, 30)
id_columns = 31
Worksheets("AA").Select
individuals = Range("C1")
Range("A3").Select
For i = 0 To individuals - 1
        For j = 0 To id_columns - 1
                randompopulation(i, j) = ActiveCell.Offset(i, j)
        Next j
Next i
'Worksheets("real duo cpis").Select
'Range("A1") = individuals
'Stop
'generate all couples
Dim cpi_base(203000)
cpis = 0
Dim parent(1, 30)
For i = 0 To individuals - 2
        For j = i + 1 To individuals - 1
                For k = 0 To id_columns - 1
                        parent(0, k) = randompopulation(i, k)
                        parent(1, k) = randompopulation(j, k)
```

```
            Next k
            Call get2children_Ca_AA(freqtab, parent, cpi_base, cpis)
            cpis = cpis + 1
        Next j
Next i
Worksheets("real duo cpis").Select
Range("A2").Select
Lines = 0
nextcolumns = 0
For i = 0 To cpis * 2 - 1
        ActiveCell.Offset(Lines, nextcolumns) = cpi_base(i)
        Lines = Lines + 1
        If Lines Mod 60000 = 0 Then
            Lines = 0
            nextcolumns = nextcolumns + 1
        End If
Next i
End Sub
Sub get2children_Ch(freqtab, parent, cpi_base, cpis)
Dim children(1, 30)
'generating 2 children
For i = 0 To 1
        Rnd1f = Rnd()
        Rnd1m = Rnd()


        Rnd2f = Rnd()
```

```
Rnd2m = Rnd()


Rnd3f = Rnd()
Rnd3m = Rnd()


Rnd4f = Rnd()
Rnd4m = Rnd()


Rnd5f = Rnd()
Rnd5m = Rnd()


Rnd6f = Rnd()
Rnd6m = Rnd()


Rnd7f = Rnd()
Rnd7m = Rnd()


Rnd8f = Rnd()
Rnd8m = Rnd()


Rnd9f = Rnd()
Rnd9m = Rnd()


Rnd10f = Rnd()
Rnd10m = Rnd()
```

```
        Rnd11f = Rnd()
        Rnd11m = Rnd()

        Rnd12f = Rnd()
        Rnd12m = Rnd()

        Rnd13f = Rnd()
        Rnd13m = Rnd()

        Rnd14f = Rnd()
        Rnd14m = Rnd()

        Rnd15f = Rnd()
        Rnd15m = Rnd()

children(i, 0) = parent(0, 0) & "+" & parent(1, 0) & "-" & i
'loci 1
If Rnd1f > 0.5 Then
        children(i, 1) = parent(0, 1)
        Else
        children(i, 1) = parent(0, 2)
End If
If Rnd1m > 0.5 Then
        children(i, 2) = parent(1, 1)
        Else
        children(i, 2) = parent(1, 2)
```

```
End If
'loci 2
If Rnd2f > 0.5 Then
    children(i, 3) = parent(0, 3)
    Else
    children(i, 3) = parent(0, 4)
End If
If Rnd2m > 0.5 Then
    children(i, 4) = parent(1, 3)
    Else
    children(i, 4) = parent(1, 4)
End If
'loci 3
If Rnd3f > 0.5 Then
    children(i, 5) = parent(0, 5)
    Else
    children(i, 5) = parent(0, 6)
End If
If Rnd3m > 0.5 Then
    children(i, 6) = parent(1, 5)
    Else
    children(i, 6) = parent(1, 6)
End If
'loci 4
If Rnd4f > 0.5 Then
    children(i, 7) = parent(0, 7)
```

```
        Else
        children(i, 7) = parent(0, 8)
End If
If Rnd4m > 0.5 Then
        children(i, 8) = parent(1, 7)
        Else
        children(i, 8) = parent(1, 8)
End If
'loci 5
If Rnd5f > 0.5 Then
        children(i, 9) = parent(0, 9)
        Else
        children(i, 9) = parent(0, 10)
End If
If Rnd5m > 0.5 Then
        children(i, 10) = parent(1, 9)
        Else
        children(i, 10) = parent(1, 10)
End If
'loci 6
If Rnd6f > 0.5 Then
        children(i, 11) = parent(0, 11)
        Else
        children(i, 11) = parent(0, 12)
End If
If Rnd6m > 0.5 Then
```

```
        children(i, 12) = parent(1, 11)
        Else
        children(i, 12) = parent(1, 12)
End If
'loci 7
If Rnd7f > 0.5 Then
        children(i, 13) = parent(0, 13)
        Else
        children(i, 13) = parent(0, 14)
End If
If Rnd7m > 0.5 Then
        children(i, 14) = parent(1, 13)
        Else
        children(i, 14) = parent(1, 14)
End If
'loci 8
If Rnd8f > 0.5 Then
        children(i, 15) = parent(0, 15)
        Else
        children(i, 15) = parent(0, 16)
End If
If Rnd8m > 0.5 Then
        children(i, 16) = parent(1, 15)
        Else
        children(i, 16) = parent(1, 16)
End If
```

```
'loci 9
If Rnd9f > 0.5 Then
        children(i, 17) = parent(0, 17)
        Else
        children(i, 17) = parent(0, 18)
End If
If Rnd9m > 0.5 Then
        children(i, 18) = parent(1, 17)
        Else
        children(i, 18) = parent(1, 18)
End If

'loci 10
If Rnd10f > 0.5 Then
        children(i, 19) = parent(0, 19)
        Else
        children(i, 19) = parent(0, 20)
End If
If Rnd10m > 0.5 Then
        children(i, 20) = parent(1, 19)
        Else
        children(i, 20) = parent(1, 20)
End If
'loci 11
If Rnd11f > 0.5 Then
        children(i, 21) = parent(0, 21)
```

```
        Else
        children(i, 21) = parent(0, 22)
End If
If Rnd11m > 0.5 Then
        children(i, 22) = parent(1, 21)
        Else
        children(i, 22) = parent(1, 22)
End If
'loci 12
If Rnd12f > 0.5 Then
        children(i, 23) = parent(0, 23)
        Else
        children(i, 23) = parent(0, 24)
End If
If Rnd12m > 0.5 Then
        children(i, 24) = parent(1, 23)
        Else
        children(i, 24) = parent(1, 24)
End If

'loci 13
If Rnd13f > 0.5 Then
        children(i, 25) = parent(0, 25)
        Else
        children(i, 25) = parent(0, 26)
End If
```

```
If Rnd13m > 0.5 Then
        children(i, 26) = parent(1, 25)
        Else
        children(i, 26) = parent(1, 26)
End If
'loci 14
If Rnd14f > 0.5 Then
        children(i, 27) = parent(0, 27)
        Else
        children(i, 27) = parent(0, 28)
End If
If Rnd14m > 0.5 Then
        children(i, 28) = parent(1, 27)
        Else
        children(i, 28) = parent(1, 28)
End If

'loci 15
If Rnd15f > 0.5 Then
        children(i, 29) = parent(0, 29)
        Else
        children(i, 29) = parent(0, 30)
End If
If Rnd15m > 0.5 Then
        children(i, 30) = parent(1, 29)
        Else
```

```
        children(i, 30) = parent(1, 30)
End If
Next i

Call pi_calculator_Ch(freqtab, children, parent, cpi_base, cpis)
End Sub
Sub get2children_Ca_AA(freqtab, parent, cpi_base, cpis)
Dim children(1, 30)
'generating 2 children
For i = 0 To 1
        Rnd1f = Rnd()
        Rnd1m = Rnd()
        Rnd2f = Rnd()
        Rnd2m = Rnd()
        Rnd3f = Rnd()
        Rnd3m = Rnd()

        Rnd4f = Rnd()
        Rnd4m = Rnd()

        Rnd5f = Rnd()
        Rnd5m = Rnd()

        Rnd6f = Rnd()
        Rnd6m = Rnd()
```

```
Rnd7f = Rnd()
Rnd7m = Rnd()

Rnd8f = Rnd()
Rnd8m = Rnd()

Rnd9f = Rnd()
Rnd9m = Rnd()

Rnd10f = Rnd()
Rnd10m = Rnd()

Rnd11f = Rnd()
Rnd11m = Rnd()

Rnd12f = Rnd()
Rnd12m = Rnd()

Rnd13f = Rnd()
Rnd13m = Rnd()

Rnd14f = Rnd()
Rnd14m = Rnd()

Rnd15f = Rnd()
Rnd15m = Rnd()
```

```
  children(i, 0) = parent(0, 0) & "+" & parent(1, 0) & "-" & i
'loci 1
If Rnd1f > 0.5 Then
        children(i, 1) = parent(0, 1)
        Else
        children(i, 1) = parent(0, 2)
End If
If Rnd1m > 0.5 Then
        children(i, 2) = parent(1, 1)
        Else
        children(i, 2) = parent(1, 2)
End If
'loci 2
If Rnd2f > 0.5 Then
        children(i, 3) = parent(0, 3)
        Else
        children(i, 3) = parent(0, 4)
End If
If Rnd2m > 0.5 Then
        children(i, 4) = parent(1, 3)
        Else
        children(i, 4) = parent(1, 4)
End If
'loci 3
```

```
If Rnd3f > 0.5 Then
        children(i, 5) = parent(0, 5)
        Else
        children(i, 5) = parent(0, 6)
End If
If Rnd3m > 0.5 Then
        children(i, 6) = parent(1, 5)
        Else
        children(i, 6) = parent(1, 6)
End If
'loci 4
If Rnd4f > 0.5 Then
        children(i, 7) = parent(0, 7)
        Else
        children(i, 7) = parent(0, 8)
End If
If Rnd4m > 0.5 Then
        children(i, 8) = parent(1, 7)
        Else
        children(i, 8) = parent(1, 8)
End If
'loci 5
If Rnd5f > 0.5 Then
        children(i, 9) = parent(0, 9)
        Else
        children(i, 9) = parent(0, 10)
```

```
End If
If Rnd5m > 0.5 Then
        children(i, 10) = parent(1, 9)
        Else
        children(i, 10) = parent(1, 10)
End If
'loci 6
If Rnd6f > 0.5 Then
        children(i, 11) = parent(0, 11)
        Else
        children(i, 11) = parent(0, 12)
End If
If Rnd6m > 0.5 Then
        children(i, 12) = parent(1, 11)
        Else
        children(i, 12) = parent(1, 12)
End If
'loci 7
If Rnd7f > 0.5 Then
        children(i, 13) = parent(0, 13)
        Else
        children(i, 13) = parent(0, 14)
End If
If Rnd7m > 0.5 Then
        children(i, 14) = parent(1, 13)
        Else
```

```
        children(i, 14) = parent(1, 14)
End If
'loci 8
If Rnd8f > 0.5 Then
        children(i, 15) = parent(0, 15)
        Else
        children(i, 15) = parent(0, 16)
End If
If Rnd8m > 0.5 Then
        children(i, 16) = parent(1, 15)
        Else
        children(i, 16) = parent(1, 16)
End If
'loci 9
If Rnd9f > 0.5 Then
        children(i, 17) = parent(0, 17)
        Else
        children(i, 17) = parent(0, 18)
End If
If Rnd9m > 0.5 Then
        children(i, 18) = parent(1, 17)
        Else
        children(i, 18) = parent(1, 18)
End If

'loci 10
```

```
If Rnd10f > 0.5 Then
        children(i, 19) = parent(0, 19)
        Else
        children(i, 19) = parent(0, 20)
End If
If Rnd10m > 0.5 Then
        children(i, 20) = parent(1, 19)
        Else
        children(i, 20) = parent(1, 20)
End If
'loci 11
If Rnd11f > 0.5 Then
        children(i, 21) = parent(0, 21)
        Else
        children(i, 21) = parent(0, 22)
End If
If Rnd11m > 0.5 Then
        children(i, 22) = parent(1, 21)
        Else
        children(i, 22) = parent(1, 22)
End If
'loci 12
If Rnd12f > 0.5 Then
        children(i, 23) = parent(0, 23)
        Else
        children(i, 23) = parent(0, 24)
```

```
End If
If Rnd12m > 0.5 Then
        children(i, 24) = parent(1, 23)
        Else
        children(i, 24) = parent(1, 24)
End If

'loci 13
If Rnd13f > 0.5 Then
        children(i, 25) = parent(0, 25)
        Else
        children(i, 25) = parent(0, 26)
End If
If Rnd13m > 0.5 Then
        children(i, 26) = parent(1, 25)
        Else
        children(i, 26) = parent(1, 26)
End If
'loci 14
If Rnd14f > 0.5 Then
        children(i, 27) = parent(0, 27)
        Else
        children(i, 27) = parent(0, 28)
End If
If Rnd14m > 0.5 Then
        children(i, 28) = parent(1, 27)
```

```
        Else
        children(i, 28) = parent(1, 28)
End If

'loci 15
If Rnd15f > 0.5 Then
        children(i, 29) = parent(0, 29)
        Else
        children(i, 29) = parent(0, 30)
End If
If Rnd15m > 0.5 Then
        children(i, 30) = parent(1, 29)
        Else
        children(i, 30) = parent(1, 30)
End If
Next i
Call pi_calculator_Ca_AA(freqtab, children, parent, cpi_base, cpis)
End Sub
Sub pi_calculator_Ch(freqtab, children, parent, cpi_base, cpis)
Dim pi15(14)
cpi = 1
systemname = 0
identifiler_loci = 15
'cpi calculated--first child with first parent
        cpi = 1
        systemname = 0
```

```
        For j = 1 To 30 Step 2
        child1 = children(0, j)
        child2 = children(0, j + 1)
        father1 = parent(0, j)
        father2 = parent(0, j + 1)
        systemname = systemname + 1
        pi15(systemname - 1) = pi(systemname, child1, child2, father1, father2,
freqtab)
        Next j
        'combine all the pi
        'control block for degradation, mutation, different kit
        'degradation--from long loci, CSF-5, D2-13, D18-11, FGA-2, D7-8, D16-12,
D21-10, D13-7
        'D3
        'pi15(0) = 1
        'vWA
        'pi15(1) = 1
        'FGA
        'pi15(2) = 1
        'THO1
        'pi15(3) = 1
        'TPOX
        'pi15(4) = 1
        'CSF1PO
        'pi15(5) = 1
        'D5
```

```
        'pi15(6) = 1
        'D13
        'pi15(7) = 1
        'D7
        'pi15(8) = 1
        'D8
        pi15(9) = 1
        'D21
        pi15(10) = 1
        'D18
        pi15(11) = 1
        'D16
        pi15(12) = 1
        'D2
        pi15(13) = 1
        'D19
        pi15(14) = 1
        For k = 0 To identifiler_loci - 1 - codis_minus2
            cpi = cpi * pi15(k)
        Next k
        cpi_base(cpis * 2) = cpi
'cpi calculated--second child with second parent
        cpi = 1
        systemname = 0
        For j = 1 To 30 Step 2
        child1 = children(1, j)
```

```
        child2 = children(1, j + 1)
        father1 = parent(1, j)
        father2 = parent(1, j + 1)
        systemname = systemname + 1
        pi15(systemname - 1) = pi(systemname, child1, child2, father1, father2,
freqtab)
        Next j
        'combine all the pi
        'control block for degradation, mutation, different kit
        'degradation--from long loci, CSF-5, D2-13, D18-11, FGA-2, D7-8, D16-12,
D21-10, D13-7
        'D3
        'pi15(0) = 1
        'vWA
        'pi15(1) = 1
        'FGA
        'pi15(2) = 1
        'THO1
        'pi15(3) = 1
        'TPOX
        'pi15(4) = 1
        'CSF1PO
        'pi15(5) = 1
        'D5
        'pi15(6) = 1
        'D13
```

```
        'pi15(7) = 1
        'D7
        'pi15(8) = 1
        'D8
        pi15(9) = 1
        'D21
        pi15(10) = 1
        'D18
        pi15(11) = 1
        'D16
        pi15(12) = 1
        'D2
        pi15(13) = 1
        'D19
        pi15(14) = 1
        For k = 0 To identifiler_loci - 1 - codis_minus2
            cpi = cpi * pi15(k)
        Next k
        cpi_base(cpis * 2 + 1) = cpi
End Sub
Sub pi_calculator_Ca_AA(freqtab, children, parent, cpi_base, cpis)
Dim pi15(14)
cpi = 1
systemname = 0
identifiler_loci = 15
'cpi calculated--first child with first parent
```

```
        cpi = 1
        systemname = 0
        For j = 1 To 30 Step 2
        child1 = children(0, j)
        child2 = children(0, j + 1)
        father1 = parent(0, j)
        father2 = parent(0, j + 1)
        systemname = systemname + 1
        pi15(systemname - 1) = pi(systemname, child1, child2, father1, father2,
freqtab)
        Next j
        'combine all the pi
        'control block for degradation, mutation, different kit
        'degradation--from long loci, CSF-0, D2-13, D18-11, FGA-1, D7-7, D16-10,
D21-12, D13-9
        'CSF1PO
        pi15(0) = 1
        'FGA
        pi15(1) = 1
        'THO1
        'pi15(2) = 1
        'TPOX
        'pi15(3) = 1
        'vWA
        'pi15(4) = 1
        'D3
```

```
        'pi15(5) = 1
        'D5
        'pi15(6) = 1
        'D7
        'pi15(7) = 1
        'D8
        'pi15(8) = 1
        'D13
        'pi15(9) = 1
        'D16
        'pi15(10) = 1
        'D18
        pi15(11) = 1
        'D21
        'pi15(12) = 1
        'D2
        pi15(13) = 1
        'D19
        'pi15(14) = 1
        For k = 0 To identifiler_loci - 1
                cpi = cpi * pi15(k)
        Next k
        cpi_base(cpis * 2) = cpi
'cpi calculated--second child with second parent
        cpi = 1
        systemname = 0
```

```
        For j = 1 To 30 Step 2
        child1 = children(1, j)
        child2 = children(1, j + 1)
        father1 = parent(1, j)
        father2 = parent(1, j + 1)
        systemname = systemname + 1
        pi15(systemname - 1) = pi(systemname, child1, child2, father1, father2,
freqtab)
        Next j
        'combine all the pi
        'control block for degradation, mutation, different kit
        'degradation--from long loci, CSF-0, D2-13, D18-11, FGA-1, D7-7, D16-10,
D21-12, D13-9
        'CSF1PO
        pi15(0) = 1
        'FGA
        pi15(1) = 1
        'THO1
        'pi15(2) = 1
        'TPOX
        'pi15(3) = 1
        'vWA
        'pi15(4) = 1
        'D3
        'pi15(5) = 1
        'D5
```

```
        'pi15(6) = 1
        'D7
        'pi15(7) = 1
        'D8
        'pi15(8) = 1
        'D13
        'pi15(9) = 1
        'D16
        'pi15(10) = 1
        'D18
        pi15(11) = 1
        'D21
        'pi15(12) = 1
        'D2
        pi15(13) = 1
        'D19
        'pi15(14) = 1
        For k = 0 To identifiler_loci - 1
                cpi = cpi * pi15(k)
        Next k
        cpi_base(cpis * 2 + 1) = cpi
End Sub
Function sharingcheck(c1, c2, f1, f2)
If c1 = f1 Or c1 = f2 Or c2 = f1 Or c2 = f2 Then
        sharingcheck = True
        Range("D2") = sharingcheck
```

```
Else
    sharingcheck = False
    Range("D2") = sharingcheck
End If
End Function
Function pi(system, c1, c2, f1, f2, freqtab)
If c1 > c2 Then
    temp = c1
    c1 = c2
    c2 = temp
End If
If f1 > f2 Then
    temp = f1
    f1 = f2
    f2 = temp
End If
childhomo = False
fatherhomo = False
If c1 = c2 Then
childhomo = True
End If
If f1 = f2 Then
fatherhomo = True
End If
allele1 = 0
allele2 = 0
```

```
sharingtot = 0
If c1 = f1 Then
        allele1 = 1
End If
If c1 = f2 Then
        allele2 = 1
End If
If c2 = f1 Then
        allele1 = 1
End If
If c2 = f2 Then
        allele2 = 1
End If
sharingtot = allele1 + allele2
'Range("D7") = sharingtot
'formula 1. q q 1/q
If childhomo = True And fatherhomo = True Then
    q = c1
    pi = 1 / allele_freq(system, q, freqtab)
    'Range("F2") = "Formula1"
End If
'formula 2. pq q 1/2q
If childhomo = False And fatherhomo = True Then
    q = f1
    pi = 1 / (2 * allele_freq(system, q, freqtab))
    'Range("F2") = "Formula2"
```

```
End If
'formula 3. q qr 1/2q
If childhomo = True And fatherhomo = False Then
    q = c1
    pi = 1 / (2 * allele_freq(system, q, freqtab))
    'Range("F2") = "Formula3"
End If
'formual 4. pq pq p+q/4pq
If childhomo = False And fatherhomo = False And sharingtot = 2 Then
    q = f1
    p = f2
    q1 = allele_freq(system, q, freqtab)
    p1 = allele_freq(system, p, freqtab)
    pi = (p1 + q1) / (4 * p1 * q1)
    'Range("F2") = "Formula4"
End If
'formula 5. pq qr 1/4q
If childhomo = False And fatherhomo = False And sharingtot = 1 And allele1 = 1
Then
    q = f1
    pi = 1 / (4 * allele_freq(system, q, freqtab))
    'Range("F2") = "Formula5"
End If
If childhomo = False And fatherhomo = False And sharingtot = 1 And allele2 = 1
Then
    q = f2
```

```
        pi = 1 / (4 * allele_freq(system, q, freqtab))
        'Range("F2") = "Formula5"
End If
End Function
Function allele_freq(system, allele, freqtab)
found = False
For i = 0 To 54
        If freqtab(i, 0) = allele Then
        allele_freq = freqtab(i, system)
        If allele_freq = 0 Then
                allele_freq = 0.0025
        End If
        found = True
        End If
Next i
If found = False Then
        allele_freq = 0.0025
End If
End Function
```

Appendix 5: Complete Programme for generating real siblings, half-siblings and processing random pairs (continues of Table 3.1)

```
Sub sibling_generating_CSI_calculation_main()
'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
'loading of the profiles of random population
'for Chinese=450
'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Dim Individual_str(449, 30)
Worksheets("Ch450").Select
    ' loading of the number of couples
    individuals = Range("C1")
    'columns of data= 31 for Identifiler
    infocells = 31
    Range("A3").Select
    'loading of array randompopulation
    For i = 0 To individuals - 1
        For J = 0 To infocells - 1
            Individual_str(i, J) = ActiveCell.Offset(i, J)
        Next J
    Next i
'call subroutine
Call calculate_si(Individual_str, individuals, infocells)
End Sub
Sub calculate_si(couples_str, couples, infocells)
'one couple generate 2 pairs of si
Dim si_hsi_base(250000, 2)
```

```
Dim least_sib_pairs1(3, 34)
Dim least_sib_pairs2(3, 34)
Dim largest_sib_pairs1(3, 34)
Dim largest_sib_pairs2(3, 34)
' set the cutoffs to obtain the least & largest CSI
least_si = 0.004
largest_si = 10000000000#
'paste the profiles to CSI claculation worksheets to get the index
Worksheets("Ch formula").Select
Range("A1").Select
Lines = 0
For i = 0 To couples - 2
    Range("A1").Select
    For m = 0 To infocells - 1
        ActiveCell.Offset(0, m) = couples_str(i, m)
    Next m
    For J = i + 1 To couples - 1
        Range("A1").Select
        For m = 0 To infocells - 1
            ActiveCell.Offset(1, m) = couples_str(J, m)
        Next m
        si_hsi_base(Lines, 0) = Range("AH4")
        si_hsi_base(Lines, 1) = Range("AI4")
        si_hsi_base(Lines, 2) = Range("AR4")
        'find out the least si
        If Range("AH4") < least_si Then
```

```
        For m = 0 To infocells - 1
            least_sib_pairs1(0, m) = couples_str(i, m)
            least_sib_pairs1(1, m) = couples_str(J, m)
        Next m
        least_sib_pairs1(0, 32) = si_hsi_base(Lines, 0)
        least_sib_pairs1(0, 33) = si_hsi_base(Lines, 1)
        least_sib_pairs1(0, 34) = si_hsi_base(Lines, 2)
        Range("A4").Select
        For m = 0 To infocells - 1
            least_sib_pairs1(2, m) = ActiveCell.Offset(0, m)
            least_sib_pairs1(3, m) = ActiveCell.Offset(1, m)
        Next m
        least_si = Range("AH4")
        least_sib_pairs1(0, 32) = least_si
    End If
    'find out the largest
    If Range("AH4") > largest_si Then
        For m = 0 To infocells - 1
            largest_sib_pairs1(0, m) = couples_str(i, m)
            largest_sib_pairs1(1, m) = couples_str(J, m)
        Next m
        largest_sib_pairs1(0, 32) = si_hsi_base(Lines, 0)
        largest_sib_pairs1(0, 33) = si_hsi_base(Lines, 1)
        largest_sib_pairs1(0, 34) = si_hsi_base(Lines, 2)
        Range("A4").Select
        For m = 0 To infocells - 1
```

```
                largest_sib_pairs1(2, m) = ActiveCell.Offset(0, m)
                largest_sib_pairs1(3, m) = ActiveCell.Offset(1, m)
        Next m
        largest_si = Range("AH4")
        largest_sib_pairs1(0, 32) = largest_si
End If
Lines = Lines + 1
'activate another rand() to get another pair of sibling
Range("A6") = 1
si_hsi_base(Lines, 0) = Range("AH4")
si_hsi_base(Lines, 1) = Range("AI4")
si_hsi_base(Lines, 2) = Range("AR4")
If Range("AH4") < least_si Then
        For m = 0 To infocells - 1
                least_sib_pairs2(0, m) = couples_str(i, m)
                least_sib_pairs2(1, m) = couples_str(J, m)
        Next m
        least_sib_pairs2(0, 32) = si_hsi_base(Lines, 0)
        least_sib_pairs2(0, 33) = si_hsi_base(Lines, 1)
        least_sib_pairs2(0, 34) = si_hsi_base(Lines, 2)
        Range("A4").Select
        For m = 0 To infocells - 1
                least_sib_pairs2(2, m) = ActiveCell.Offset(0, m)
                least_sib_pairs2(3, m) = ActiveCell.Offset(1, m)
        Next m
        least_si = Range("AH4")
```

```
                least_sib_pairs2(0, 32) = least_si


            End If
            If Range("AH4") > largest_si Then
                For m = 0 To infocells - 1
                    largest_sib_pairs2(0, m) = couples_str(i, m)
                    largest_sib_pairs2(1, m) = couples_str(J, m)
                Next m
                largest_sib_pairs2(0, 32) = si_hsi_base(Lines, 0)
                largest_sib_pairs2(0, 33) = si_hsi_base(Lines, 1)
                largest_sib_pairs2(0, 34) = si_hsi_base(Lines, 2)
                Range("A4").Select
                For m = 0 To infocells - 1
                    largest_sib_pairs2(2, m) = ActiveCell.Offset(0, m)
                    largest_sib_pairs2(3, m) = ActiveCell.Offset(1, m)
                Next m
                largest_si = Range("AH4")
                largest_sib_pairs1(0, 32) = largest_si
            End If
            Lines = Lines + 1
            Next J
Next i
si_hsi_ct = Lines - 1
Worksheets("Ch si vs hsi").Select
Range("A3:EA60000").ClearContents
Lines = 0
```

```
nextcolumns = 0
Range("A3").Select
For i = 0 To si_hsi_ct
    ActiveCell.Offset(Lines, 0 + nextcolumns) = si_hsi_base(i, 0)
    ActiveCell.Offset(Lines, 1 + nextcolumns) = si_hsi_base(i, 1)
    ActiveCell.Offset(Lines, 2 + nextcolumns) = si_hsi_base(i, 2)
    Lines = Lines + 1
    If Lines Mod 60000 = 0 Then
        Lines = 0
        nextcolumns = nextcolumns + 3
    End If
Next i
Range("AA1").Select
For n = 0 To 3
    For m = 0 To 34
        ActiveCell.Offset(n, m) = least_sib_pairs1(n, m)
    Next m
Next n
Range("AA5").Select
For n = 0 To 3
    For m = 0 To 34
        ActiveCell.Offset(n, m) = least_sib_pairs2(n, m)
    Next m
Next n
Range("AA9").Select
For n = 0 To 3
```

```
        For m = 0 To 34
            ActiveCell.Offset(n, m) = largest_sib_pairs1(n, m)
        Next m
Next n
Range("AA13").Select
For n = 0 To 3
        For m = 0 To 34
            ActiveCell.Offset(n, m) = largest_sib_pairs2(n, m)
        Next m
Next n
Range("E1") = least_si
Range("E2") = largest_si
Range("A1").Select
End Sub
Sub rp_csi_calculation_main()

'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
'main program of random pair csi calculation
'loading of the profiles
'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Dim Individual_str(449, 30)
Worksheets("Ch450").Select
        ' loading of the number of couples
        individuals = Range("C1")
        'columns of data= 31 for Identifiler
        infocells = 31
```

```
        Range("A3").Select
        'loading of array randompopulation
        For i = 0 To individuals - 1
            For J = 0 To infocells - 1
                Individual_str(i, J) = ActiveCell.Offset(i, J)
            Next J
        Next i
Call rp_calculate_si(Individual_str, individuals, infocells)
End Sub
Sub rp_calculate_si(couples_str, couples, infocells)
'one couple generate 2 pairs of si and hsi
Dim si_hsi_base(250000, 2)
Dim least_sib_pairs1(3, 34)
Dim least_sib_pairs2(3, 34)
Dim largest_sib_pairs1(3, 34)
Dim largest_sib_pairs2(3, 34)
least_si = 0.004
largest_si = 1000#
'paste the profiles to get the index
Worksheets("Ch rp formula").Select
Range("A1").Select
Lines = 0
For i = 0 To couples - 2
        Range("A1").Select
        For m = 0 To infocells - 1
            ActiveCell.Offset(0, m) = couples_str(i, m)
```

```
Next m
For J = i + 1 To couples - 1
    Range("A1").Select
    For m = 0 To infocells - 1
        ActiveCell.Offset(1, m) = couples_str(J, m)
    Next m
    si_hsi_base(Lines, 0) = Range("AH4")
    si_hsi_base(Lines, 1) = Range("AI4")
    si_hsi_base(Lines, 2) = Range("AR4")
    'find out the least si
    If Range("AH4") < least_si Then
        For m = 0 To infocells - 1
            least_sib_pairs1(0, m) = couples_str(i, m)
            least_sib_pairs1(1, m) = couples_str(J, m)
        Next m
        least_sib_pairs1(0, 32) = si_hsi_base(Lines, 0)
        least_sib_pairs1(0, 33) = si_hsi_base(Lines, 1)
        least_sib_pairs1(0, 34) = si_hsi_base(Lines, 2)
        Range("A4").Select
        For m = 0 To infocells - 1
            least_sib_pairs1(2, m) = ActiveCell.Offset(0, m)
            least_sib_pairs1(3, m) = ActiveCell.Offset(1, m)
        Next m
        least_si = Range("AH4")
        least_sib_pairs1(0, 32) = least_si
    End If
```

```
            'find out the largest
            If Range("AH4") > largest_si Then
                For m = 0 To infocells - 1
                    largest_sib_pairs1(0, m) = couples_str(i, m)
                    largest_sib_pairs1(1, m) = couples_str(J, m)
                Next m
                largest_sib_pairs1(0, 32) = si_hsi_base(Lines, 0)
                largest_sib_pairs1(0, 33) = si_hsi_base(Lines, 1)
                largest_sib_pairs1(0, 34) = si_hsi_base(Lines, 2)
                Range("A4").Select
                For m = 0 To infocells - 1
                    largest_sib_pairs1(2, m) = ActiveCell.Offset(0, m)
                    largest_sib_pairs1(3, m) = ActiveCell.Offset(1, m)
                Next m
                largest_si = Range("AH4")
                largest_sib_pairs1(0, 32) = largest_si
            End If
            Lines = Lines + 1
        Next J
Next i
si_hsi_ct = Lines - 1
Worksheets("Ch rp si").Select
Range("A3:EA60000").ClearContents
Lines = 0
nextcolumns = 0
Range("A3").Select
```

```
For i = 0 To si_hsi_ct
    ActiveCell.Offset(Lines, 0 + nextcolumns) = si_hsi_base(i, 0)
    ActiveCell.Offset(Lines, 1 + nextcolumns) = si_hsi_base(i, 1)
    ActiveCell.Offset(Lines, 2 + nextcolumns) = si_hsi_base(i, 2)
    Lines = Lines + 1
    If Lines Mod 60000 = 0 Then
        Lines = 0
        nextcolumns = nextcolumns + 3
    End If
Next i
Range("AA1").Select
For n = 0 To 3
    For m = 0 To 34
        ActiveCell.Offset(n, m) = least_sib_pairs1(n, m)
    Next m
Next n
Range("AA5").Select
For n = 0 To 3
    For m = 0 To 34
        ActiveCell.Offset(n, m) = least_sib_pairs2(n, m)
    Next m
Next n
Range("AA9").Select
For n = 0 To 3
    For m = 0 To 34
        ActiveCell.Offset(n, m) = largest_sib_pairs1(n, m)
```

```
        Next m
Next n
Range("AA13").Select
For n = 0 To 3
        For m = 0 To 34
                ActiveCell.Offset(n, m) = largest_sib_pairs2(n, m)
        Next m
Next n
Range("E1") = least_si
Range("E2") = largest_si
Range("A1").Select
End Sub
Sub count_si_1_ratio_main()
'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
'main program
'discriminating the pairs with si<1 by number of two allele sharing loci
'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Dim two_allele_pairs(15, 1)
Worksheets("Ch si vs hsi").Select
si_hsi_ct = 202050
Lines = 0
nextcolumns = 0
margin_csi = 2
Range("A3").Select
For i = 0 To si_hsi_ct - 1
        si = ActiveCell.Offset(Lines, 0 + nextcolumns)
```

```
        loci = ActiveCell.Offset(Lines, 1 + nextcolumns)
        If si < margin_csi Then
            two_allele_pairs(loci, 0) = two_allele_pairs(loci, 0) + 1
        Else
            two_allele_pairs(loci, 1) = two_allele_pairs(loci, 1) + 1
        End If
        Lines = Lines + 1
        If Lines Mod 60000 = 0 Then
            Lines = 0
            nextcolumns = nextcolumns + 3
        End If
Next i
Worksheets("si>=<1 ct").Select
Range("b104").Select
For i = 0 To 15
        ActiveCell.Offset(i, 0) = two_allele_pairs(i, 0)
        ActiveCell.Offset(i, 1) = two_allele_pairs(i, 1)
Next i
Range("b122") = si_hsi_ct
End Sub
```

Appendix 6: Complete programme to process STR profiles and generate forensic DNA population study parameters (continues of Table 5.2)

```
Sub main_popu_study()
'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
' a program to process the Identifiler data from ABI sequencer
' export data into text mode
' import the data using Excel software
' change the worksheet name to "Identifiler STR"
' profiling of 15 loci can be processed at once
'xxxxxxxxxxxxxxxxxxxxxxxx
'array to load all the locus name
Dim locusnames(15)
'array to load all the identifiler profiles for all the individuals
Dim all_alleles(5999, 30)
'array to record all the possible alleles
Dim allele_cat(79)
'array to record all the allele frequencies
Dim freq_tab(79, 15)
'get the basic information from STR profiles
Worksheets("Basic Info").Select
individuals = Range("A2")
locitot = Range("A4")
mintot = Range("C4")
maxtot = Range("C5")
allele_no = 0
```

```
'load the locus name
Worksheets("STR profiles").Select
Range("B1").Select
locus = 0
For i = 1 To locitot
    locus = (i - 1)
    locusnames(locus) = ActiveCell.Offset(0, (i - 1) * 2)
Next i
'load all the profiles and check empty or non-digital cell
Dim emptycell(100, 1)
Dim textcell(100, 1)
empty_cell = 0
text_cell = 0
Range("B2").Select
For i = 0 To individuals - 1
    For j = 0 To locitot * 2 - 1
        all_alleles(i, j) = ActiveCell.Offset(i, j)
        If all_alleles(i, j) = "" Then
            emptycell(empty_cell, 0) = i
            emptycell(empty_cell, 1) = j
            empty_cell = empty_cell + 1
        End If
        If TypeName(all_alleles(i, j)) = "String" Then
            textcell(text_cell, 0) = i
            textcell(text_cell, 1) = j
            text_cell = text_cell + 1
```

```
            End If
        Next j
Next i
'if empty cells or string cells found, output a table describing them
Worksheets("Check rep").Select
Range("A4:D103").Select
Selection.ClearContents
If empty_cell > 0 Or text_cell > 0 Then
        Range("A4").Select
        For i = 0 To empty_cell - 1
            ActiveCell.Offset(i, 0) = emptycell(i, 0) + 1
            ActiveCell.Offset(i, 1) = emptycell(i, 1) + 1
        Next i
        Range("C4").Select
        For i = 0 To text_cell - 1
            ActiveCell.Offset(i, 0) = textcell(i, 0) + 1
            ActiveCell.Offset(i, 1) = textcell(i, 1) + 1
        Next i
        Range("A1").Select
        MsgBox "empty cell or string cell found, proceed or not?"
        answer = InputBox("Y / N ?")
        If answer = "N" Then
            MsgBox "Please make sure that there is no empty or string cell"
            GoTo Label_end
        End If
End If
```

```
'call the sunroutine to pick all the possible alleles
Call pick_allele_category(individuals, mintot, maxtot, allele_cat, all_alleles, allele_no)
'assign the unique alleles to frequency table
For i = 0 To allele_no - 1
    freq_tab(i, 0) = allele_cat(i)
Next i
'call the subroutine to calculate the allele frequencies
Call count_allele_freq(individuals, locitot, allele_no, freq_tab, all_alleles)
'call output frequency table
Call freq_tab_output(allele_no, locusnames, freq_tab, locitot)
'check the fintness and calculate all the forensic DNA parmeters
Call g_test_and_parameter(individuals, locitot, allele_no, freq_tab, all_alleles)
'call output parameters
Call freq_tab_output(allele_no, locusnames, freq_tab, locitot)
Call parameter_output(freq_tab, locitot)
'Call 5/2Ncorrection()
MsgBox "Do you want 5/2N correction for the minimum allele percentage ?"
answer = InputBox("Y / N ?")
If answer = "Y" Then
    Call correction5_2N(freq_tab, locitot)
    MsgBox "minimum allele frequency = 5/2N processed"
End If
'show the frequency table
Worksheets("freq tab").Select
Range("A1").Select
Label_end:
```

```
End Sub
Sub pick_allele_category(individuals, mintot, maxtot, allele_cat, all_alleles, allele_no)
'xxxxxxxxxxxxxxxxxxxxxxxx
Dim all_possible_allele(159, 1)
allele_increament = 0
i = 0
'list all the possible alleles into array
'make the allele template for allele-chosing
While mintot <= maxtot
    all_possible_allele(i, 0) = mintot + allele_increament
    'ActiveCell.Offset(i, 0) = alleles(i, 0)
    i = i + 1
    allele_increament = allele_increament + 0.1
    If allele_increament >= 0.4 Then
        allele_increament = 0
        mintot = mintot + 1
    End If
Wend
'compare all the alleles to the allele template
'if there is a match, make the template with "1"
i = 0
While all_possible_allele(i, 0) > 0
    j = 0
    Do While j < individuals
        k = 0
        Do While k < 31
```

```
                If all_possible_allele(i, 0) = all_alleles(j, k) Then
                    all_possible_allele(i, 1) = 1
                    j = individuals
                    Exit Do
                End If
                k = k + 1
            Loop
            j = j + 1
        Loop
        i = i + 1
Wend
'screen the real allele cateagory out
j = 0
For i = 0 To 159
        If all_possible_allele(i, 1) = 1 Then
            allele_cat(j) = all_possible_allele(i, 0)
            j = j + 1
        End If
Next i
'paste the allele number to worksheets
allele_no = j
'Worksheets("Basic Info").Select
'Range("A6") = allele_no
End Sub
Sub count_allele_freq(individuals, locitot, allele_no, freq_tab, all_alleles)
'xxxxxxxxxxxxxxxxxxxx
```

```
Dim locusminmax(15, 1)
Worksheets("Basic Info").Select
Range("C2").Select
For loci = 1 To locitot
    locusminmax(loci, 0) = ActiveCell.Offset(0, (loci - 1) * 2)
    locusminmax(loci, 1) = ActiveCell.Offset(1, (loci - 1) * 2)
Next loci
For loci = 1 To locitot
    locusmin = locusminmax(loci, 0)
    locusmax = locusminmax(loci, 1)
    For i = 0 To allele_no - 1
        If freq_tab(i, 0) = locusmin Then
            min_position = i
        End If
        If freq_tab(i, 0) = locusmax Then
            max_position = i
        End If
    Next i
    For i = 0 To individuals - 1
        For j = min_position To max_position
            If freq_tab(j, 0) = all_alleles(i, (loci - 1) * 2) Then
                freq_tab(j, loci) = freq_tab(j, loci) + 1
            End If
            If freq_tab(j, 0) = all_alleles(i, (loci - 1) * 2 + 1) Then
                freq_tab(j, loci) = freq_tab(j, loci) + 1
            End If
```

```
            Next j
        Next i
Next loci
'sum the counting of every locus
For i = 1 To locitot
        freq_tab(79, i) = 0
        For j = 0 To allele_no - 1
            freq_tab(79, i) = freq_tab(79, i) + freq_tab(j, i)
        Next j
Next i
'divide the number of each allele by the sum of the total counting of a locus
For i = 1 To locitot
        For j = 0 To allele_no - 1
            freq_tab(j, i) = freq_tab(j, i) / freq_tab(79, i)
        Next j
Next i
End Sub
Sub g_test_and_parameter(individuals, locitot, allele_no, freq_tab, all_alleles)
'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
'get the allele frequency for one locus
'rearrange the allele list to become continuous allele array
'easy for counting observed and expected genotype frequency
'prepare the parameter array and for the parameters
'array to load the allele number for each locus
Dim locus_allele_number(14)
Worksheets("freq tab").Select
```

```
Range("B2").Select
For loci = 1 To locitot
    locus_allele_number(loci - 1) = ActiveCell.Offset(0, loci - 1)
Next loci
For loci = 1 To locitot
    Dim parameters(8)
    Homozygosity = 0
    Heterozygosity = 0
    Effective_no_alleles = 0
    PIC = 0
    PD = 0
    PE = 0
    Pro_identity = 0
    Paternity_index = 0
    locus_allele_no = locus_allele_number(loci - 1)
    'prepare the allele array for a locus
    ReDim locus_allele_tab(25, 1)
    j = 0
    For i = 0 To allele_no - 1
        If freq_tab(i, loci) > 0 Then
            locus_allele_tab(j, 0) = freq_tab(i, 0)
            locus_allele_tab(j, 1) = freq_tab(i, loci)
            j = j + 1
        End If
    Next i
    'make the genotype template for genotype counting
```

```
    ReDim locusgenotypes(300, 2)
    k = 0
    For i = 0 To locus_allele_no - 1
        For j = i To locus_allele_no - 1
            locusgenotypes(k, 0) = locus_allele_tab(i, 0) & "/" & locus_allele_tab(j,
0)
            locusgenotypes(k, 2) = locus_allele_tab(i, 1) * locus_allele_tab(j, 1)
            If i <> j Then
                locusgenotypes(k, 2) = 2 * locusgenotypes(k, 2)
            End If
            k = k + 1
        Next j
    Next i
    genotypes = k
    'transform the 2 cells format in one cell format for a locus
    ReDim locusonecellformat(5999)

    For i = 0 To individuals - 1
        'calculation of Homoygosity
        If all_alleles(i, (loci - 1) * 2) = all_alleles(i, (loci - 1) * 2 + 1) Then
            Homozygosity = Homozygosity + 1
        End If
        locusonecellformat(i) = all_alleles(i, (loci - 1) * 2) & "/" & all_alleles(i, (loci -
1) * 2 + 1)
    Next i
    'calculation of the expected genotype frequencies
```

```
genotype_ct = 0
For i = 0 To individuals - 1
    For j = 0 To k - 1
        If locusonecellformat(i) = locusgenotypes(j, 0) Then
            locusgenotypes(j, 1) = locusgenotypes(j, 1) + 1
            genotype_ct = genotype_ct + 1
        End If
    Next j
Next i
'G test
'G= 2 * sum (fi *ln*(fi/cap fi))
Gtest = 0
obs_divided_exp_sum = 0
temp1 = 0
temp2 = 0
temp3 = 0
tempsum = 0
'Worksheets("temp").Select
'Range("A1").Select
For i = 0 To genotypes - 1
    temp1 = locusgenotypes(i, 1) / (locusgenotypes(i, 2) * genotype_ct)
    temp2 = 0
    If temp1 > 0 Then
        temp2 = Log(temp1)
    End If
    'ActiveCell.Offset(i, loci - 1) = (locusgenotypes(i, 2) * genotype_ct)
```

```
        temp3 = locusgenotypes(i, 1) * temp2
        tempsum = tempsum + temp3
    Next i
    Gtest = 2 * tempsum
    degree_freedom = genotypes - locus_allele_no
    Worksheets("Basic Info").Select
    Range("B8") = Gtest
    Range("B9") = degree_freedom
    parameters(0) = Range("B10")
    ' calculate forensic parameters
    Homozygosity = Homozygosity / genotype_ct
    parameters(1) = Homozygosity
    Heterozygosity = 1 - Homozygosity
    parameters(2) = Heterozygosity
    Effective_no_alleles = 1 / Homozygosity
    parameters(3) = Effective_no_alleles
    'calculation of PIC
    squaresum = 0
    quarticssum = 0
    square_of_squaresum = 0
    For i = 0 To locus_allele_no - 1
        squaresum = squaresum + locus_allele_tab(i, 1) * locus_allele_tab(i, 1)
    Next i
    square_of_squaresum = squaresum * squaresum
    For i = 0 To locus_allele_no - 1
        quarticssum = quarticssum + locus_allele_tab(i, 1) * locus_allele_tab(i, 1) *
```

```
locus_allele_tab(i, 1) * locus_allele_tab(i, 1)
        Next i
        PIC = 1 - squaresum - square_of_squaresum + quarticssum
        parameters(4) = PIC
        PD = 1 - 2 * square_of_squaresum - quarticssum
        parameters(5) = PD

        PE = Heterozygosity * Heterozygosity * (1 - (1 - Heterozygosity) * Heterozygosity
* Heterozygosity)
        parameters(6) = PE
        Pro_identity = squaresum
        parameters(7) = Pro_identity

        Paternity_index = 1 / (2 * Homozygosity)
        parameters(8) = Paternity_index
        For i = 0 To 8
              freq_tab(60 + i, loci) = parameters(i)
        Next i
Next loci
End Sub
Sub freq_tab_output(allele_no, locusnames, freq_tab, locitot)
'xxxxxxxxxxxxxxxxxxxxxxxx
Worksheets("freq tab").Select
Range("A1:P1").Select
Selection.ClearContents
Range("A3:P82").Select
```

```
Selection.ClearContents
Range("B1").Select
For i = 0 To locitot - 1
    ActiveCell.Offset(0, i) = locusnames(i)
Next i
Range("A2") = allele_no
Range("A3").Select
For i = 0 To locitot
    For j = 0 To 59
        If freq_tab(j, i) = 0 Then
            ActiveCell.Offset(j, i) = "-"
        Else
            ActiveCell.Offset(j, i) = freq_tab(j, i)
        End If
    Next j
Next i
End Sub
Sub parameter_output(freq_tab, locitot)
'xxxxxxxxxxxxxxxxxxxxxxxx
Worksheets("freq tab").Select
Range("A63") = "p value"
Range("A64") = "Homozygosity="
Range("A65") = "Heterozygosity="
Range("A66") = "Effective number of alleles="
Range("A67") = "PIC="
Range("A68") = "PD="
```

```
Range("A69") = "PE="
Range("A70") = "Probability of identity="
Range("A71") = "Paternity index="
Range("A3").Select
For i = 1 To locitot
    For j = 60 To 79
        ActiveCell.Offset(j, i) = freq_tab(j, i)
    Next j
Next i
End Sub
Sub correction5_2N(freq_tab, locitot)
'xxxxxxxxxxxxxxxxxxxxxxx
Worksheets("freq tab").Select
Range("B3").Select
For i = 0 To locitot - 1
    For j = 0 To 59
        If ActiveCell.Offset(j, i) < 5 / freq_tab(79, i + 1) Then
        ActiveCell.Offset(j, i) = 5 / freq_tab(79, i + 1)
    End If
    Next j
Next i
End Sub
```

Appendix 7: Five publications and seven conference presentations

A study on the false avuncular inclusion rates in human identification

Chang En Pu M.Sc. [1*] and Adrian Linacre Ph.D.[2]

1. Scientific and Technical Research Center, Ministry Justice Investigation Bureau, Taipei County, Taiwan.

2. Centre for Forensic Science, WestCHEM, University of Strathclyde, Glasgow, UK.

*Corresponding author: Chang En Pu, Forensic DNA section, Scientific and Technical research center, Ministry Justice Investigation Bureau, Taipei County 231, Taiwan, R.O.C. Tel: +886-2-29112241 ext 3703, fax: +886-2-29138599, E-mail: pu_macros@yahoo.com.tw

Abstract

Avuncular testing may be required when the uncle of a child is DNA tested to determine paternity if the father is not available. Also a putative father may in fact be the uncle of a child and share an allele at all the genetic loci tested by chance. This chance of a coincidental match (a false inclusion) will increase greatly in a duo case where the mother is not available for testing.

We report on duo scenarios from both real populations and computer generated populations using the 15 STR markers within the Identifiler® kit. The populations are used to model the chance of false inclusions and false exclusions to differentiate fathers from uncles. STR profiles of members of the population were paired at

random to generate STR profile of children and then one of the children's profile was paired with another STR profile to generate an STR profile of a grandchild. This allowed a child's and putative uncle's DNA profiles to be compared. When an allele was shared at all loci tested between pairs separated by one generation a combined paternity index (CPI) was calculated. In total there were 187,402 comparisons for the four populations processed and CPI values were calculated using 15, 13, 10 and 9 STR loci. Ratio distributions of CPIs were found to be overlapping for child/uncle pairs and child/father pairs. The specificity (1-% of false inclusions) and sensitivity (1-% of false exclusion) under different minimum CPI requirements were estimated empirically. When using 15 STR loci, the specificity at a CPI cutoff value of 100 was only over 96.1% for all the 4 populations and only over 90.9% when 10 of the 15 STR loci used.

By setting the specificity at 95%, the minimum CPI requirement was 606 for the Chinese population, 665 for the Caucasian population, 348 for African American population and 364 for Hispanic population using CODIS 13 STR. Based upon the data obtained it is evident that with more loci the specificity and sensitivity will increase. We report on how varying CPI values should be used dependent upon the number of STR loci used.

Key Words: forensic science, short tandem repeat (STRs), paternity, duo, uncle, avuncular.

Introduction

A high probability can be obtained when performing a standard paternity test using DNA profiles from a child, mother and father. A problem may arise if the biological father is unavailable for testing and the brother of the putative father is used as a proxy, or if the person assumed to be the father is actually the brother of the father. In previous studies a brother of the biological father was found to match at all 16 STR loci tested in 3.3 % of cases [1]. This false inclusion rate increased to 6.3 % if the biological mother was not available for testing [2]; such without-mother situations occur frequently [3]. In a study of 93 child and biological fathers, where the paternal uncle was also available for testing, five out of 125 child/uncle pairs exhibited a shared allele at the loci tested with Combined Paternity Index (CPI) values between 3,652 and 33,545 [4]. A preset CPI minimum is chosen to prevent as few as possible false inclusions and false inclusions. These CPI values currently range from 0 to 10,000 [5]. We report on a study to determine the specificity and sensitivity of the current setting of different CPI cutoff values in avuncular and paternity testing and optimal CPI values are suggested.

Materials and Methods

DNA profiles from 450 random members of the Chinese population in Taiwan were obtained by using the ABI AmpF*l*STR Identifiler® PCR Amplification Kit (Applied Biosystem, Foster City, CA, USA). All the 450 copies of the 15 STR profiles were posted to the worksheets of Microsoft Excel and processed by a built-in Visual Basic program written by the authors of this study. Every individual of the population was paired with every other individual to form pairs of STR profiles ((449x448)/2). Every pair was then processed by the program to have two children (A & B). A third STR Profile was introduced from the same random population and paired with child A to generate a DNA profile of child C. Under this scenario child C and B would be uncle and nephew/niece. The STR profiles of B and child C were compared to check if B has matching alleles at all loci tested, as would be the case if he was the biological father of C. The CPI was calculated for the child/uncle (avuncular) pairs and for real child/father (paternity) pairs by a standard formula [6]. Some of the loci of the 15 STR Identifiler® kit were removed to simulate the STR profiling data obtained from using fewer STR loci, e.g. D2S1338 and D19S433 were omitted to mimic the CODIS 13 system, others were omitted to mimic the SGM plus®, Profiler Plus®, and Profiler® kit systems.

DNA profiles from 15 STR loci from Caucasian (n= 301), African American (n= 256) and Hispanic (n= 140) populations were obtained from the Short Tandem Repeat DNA Internet Database [7]. These profiles were also processed by the same program to generate real duos and avuncular pairs to determine the rate of any false inclusions. There were 44,850 comparisons, 32,385 comparisons and 9,591 comparisons for these three populations respectively. The CPI calculation based on allele frequency table of the Chinese population was from previous studies [8, 9, 10] and the Caucasian, African Americans and Hispanic population allele frequencies were downloaded from the same origin as their DNA profiles from [7]. All the frequencies of alleles were adjusted by using 5/2N rule [11].

The specificity of the paternity test was calculated as 1 – the % of false positives and sensitivity of the CPI was based upon 1 – the % of false negatives [12]. The rate of false positives seen as an inclusion in an uncle being incorrectly designated as the father, was determined by the percentage of avuncular pairs sharing an allele at all STR loci tested. The percentage of false negatives (seen as an exclusion of paternity) was the proportion of child/father pairs that might not be recognized as real duos based upon their CPIs lower than the preset cutoffs. The optimal cutoffs were obtained by the method stated in (16).

Results and discussion

Success rates for uncles to impersonate the father were estimated under different scale of STR loci

The success rates (false inclusion rates) for uncles to be confused with fathers after a total of 187,402 comparisons for 4 populations, and the 1- PER (the probability of excluding a brother of the true father) being the formula of PER developed in [2], are shown in table 1. When using 15 STR loci there was about 4 % inclusion rate within the Chinese population for uncles sharing an allele with the child at the alleles tested, as would be expected if they were the father in a paternity tests. As the number of STR loci tested decreased; such as when using the nine STR loci in the Profiler system the false inclusion rate increased to 19.31%. The other three populations showed the same trend and in all cases the empirical false inclusion rate was marginally higher than the 1- PER value predicted based on the allele frequency tables. When using 1- false inclusion rates to calculate exclusion rates, the tabulated data is in line with previous studies [1, 2, 3]. A CPI value of 1,000 has been suggested previously [1] although avuncular pairs have been reported to exceed this value [4]. Setting a CPI cutoff value will always have the potential to falsely include an uncle as a biological father [13].

Specificity and Sensitivity of different CPI requirements for differentiating uncles from real fathers

In instances where there is a potential false inclusion or false exclusion a CPI value was calculated. The two parameters affected by the false inclusion of uncles as a father using varying numbers of STR loci under different CPI cutoffs are shown in table 2. The large sample size created by using a virtual population allows both the sensitivity and specificity to be determined. As the CPI cutoff increased, there was an increase in specificity and a subsequent decrease in sensitivity. Using the 15 STR loci with all 4 populations and when no cutoff values were set, specificity reached 96 % with 100 % sensitivity. If CPI cutoffs value varied from 100 to 10,000, then the specificity varied from 96.09 % to 97.96 %, but the sensitivity was reduced from 99.78 % to 75.01 %. If setting the specificity at 95 % and the 13 STR loci comprising the CODIS loci are used, the cutoff values required are 606, 604, 348 and 364 for the Chinese, Caucasian, African American and Hispanic populations respectively. As fewer loci are used there was no sharp difference in cutoff values but the sensitivities were greatly reduced. With a requirement for a specificity of 99 % then the cut off values would be between 57,621 and 109,297 and the sensitivities would be around 50 % for these 4 populations. This illustrates how a requirement for high specificity leads to poor sensitivity and reflects the high false inclusion rates found for uncles compared to unrelated members of the population. These data may be of assistance to those laboratories that deal predominantly with paternity tests related to immigration

events. Other forensic laboratories involved with human identification may wish to note the optimal cutoffs in table 2 for keeping the minimum differences between specificity and sensitivity.

In conclusion, a different CPI cutoff may be applied for paternity tests when evaluating the possible interfering from the uncle is necessary.

Acknowledgments

The authors wish to thank Wu F.C., Wu G.C., Cheng M.Y., and Meng L.M. for their kindly help for profiling all the local samples and providing very helpful discussion.

Reference

1. Thomson JA, Ayres KL, Pilotti V, Barret MN, Walker JIH, Debenham PG. Analysis of disputed single-parent/child and sibling relationships using 16 STR loci. Int J Legal Med. 2001; 115:128-134.
2. Fung WK, Chung YK, Wong DM. Power of exclusion revisited: probability of excluding relatives of the true father from paternity. Int J Legal Med. 2002; 116(2):64-7.
3. Wenk RE, Gjertson DW, Chiafari FA and Houtz T. The specific power of parentage exclusion in a child's blood relatives. Transfusion. 2005; 45(3):

440-4.

4. Wurmb-Schwark N, Malyusz V, Simeoni E, Lignitz E, Poetsch M. Possible pitfalls in motherless paternity analysis with related putative fathers. Forensic Sci Int. 2006; 159(2-3): 92-7.

5. ANNUAL REPORT SUMMARY FOR TESTING IN 2004, Prepared by the Relationship Testing Program Unithttp: //www.aabb.org/ Documents/ Accreditation/ Parentage_Testing_Accreditation_Program/ rtannrpt04.pdf (accessed May 1 2008)

6. AABB, Guidance for standards for relationship testing laboratories, MD, USA, American association of blood banks, 2006, pp 139.

7. Short Tandem Repeat DNA Internet Database, http://www.cstl.nist.gov/div831/strbase/ (accessed May 1 2008)

8. Pu CE, Wu FC, Cheng CL, Wu KC, Chao CH, and Li JM. DNA STR profiling of Chinese population in Taiwan determined by using an automated sequencer, *Forensic Sci. Int.* 97:47-51, 1998.

9. Pu CE. et al. DNA profiling of Chinese population in Taiwan on the STR loci D8S1179, D16S539, D18S51 and D21S11 52th annual meeting, American Academy of Forensic Sciences, Reno, NV, USA, 2000.

10. Lee JC, Chang YY, Su CW, Tzeng CH, Pu CE, Han TH, Chang CW, Liu DC, Lin FR, Lin YC, Chen MH. The Screening of STR and YSTR Loci in Taiwanese Han Population, Proceedings of Taiwan Academy of Forensic Sciences, 2006, 27-38.(in Chinese)

11. Butler JM. Kinship and parentage testing. In: Forensic DNA Typing, Elsevier Academic Press, MA, USA, 2005, pp 529-537.

12. Kirkwood BR and Sterne JAC. Measurement error: assessment and implications. In: Medical statistics, MA, USA, Blackwell publishing, pp. 429-446, 2003.

13. Pu CE. Linacre A. CPI distribution and cutoff values for duo kinship testing. Chinese J. Physiology 50(5):232-239, 2007.

14. Gorlin JB, Polesky HF. The use and abuse of the full-sibling and half-sibling indices. Transfusion. 2000; 40(9):1148-9.

15. Fung WK, Wong DM, Chung YK. How well do serial tandem repeat loci perform in excluding paternity in relatives of the biologic father among immigration cases? Transfusion. 2003; 43(7):982-3.

16. Find ‘best’ cutoffs for maximizing sensitivity/specificity. Available online at http://www.natureserve.org/visitLocal/conftraining/edm_workshop/roc_help.pdf. accessed at May 2007.

Table 1: FIR and 1- PER for the uncle impersonating the father.

		Identifiler		CODIS		SGM plus		Profiler plus		Profiler	
	comparisons	FIR	1-PER	FIR	1-PER	FIR	1-PER	FIR	1-PER	FIR	1-PER
Chinese	100,576	3.99%	2.76%	6.98%	5.01%	8.99%	7.17%	11.78%	9.62%	19.31%	15.65%
Caucasians	44,850	3.58%	2.29%	6.61%	4.26%	8.31%	6.00%	11.00%	8.56%	17.53%	13.41%
African Americans	32,385	3.00%	1.62%	5.84%	3.35%	7.29%	4.47%	12.44%	7.89%	16.54%	12.18%
Hispanic	9,591	3.27%	2.27%	5.40%	4.05%	7.95%	6.23%	10.17%	8.76%	15.13%	12.93%

Identifiler, SGM plus, SGMplus, Profiler plus, Profiler: STR kits provided by Applied Biosystems LtD. USA

FIR: False Inclusion Rate.

PER: the probability of excluding a brother of the true father(2).

Table 2: Specificity and sensitivity affected by false inclusion under different CPI cutoffs.

	Identifiler			CODIS			SGM plus			Profiler plus			Profiler		
	cutoff	spe*	sen**	cutoff	spe	sen	cutoff	spe	sen	cutoff	spe	sen	cutoff	spe	sen
Ch	100	96.09%	99.78%	100	93.53%	98.56%	100	92.14%	95.78%	100	90.89%	90.03%	100	90.05%	71.40%
	1,000	96.66%	95.11%	1,000	95.58%	83.04%	1,000	95.71%	68.55%	1,000	96.30%	49.61%	1,000	97.46%	26.89%
	10,000	97.96%	75.01%	10,000	98.18%	47.34%	10,000	98.68%	28.48%	10,000	99.08%	14.85%	10,000	99.61%	5.35%
	0	96.01%	100.00%	606	95.00%	88.43%	638	95.00%	76.08%	545	95.01%	61.87%	372	95.01%	44.93%
	65,486	99.00%	47.38%	31,246	99.00%	30.28%	15,580	99.00%	22.97%	8,759	99.00%	16.03%	3,383	99.00%	12.31%
	701	96.51%	96.61%	286	94.21%	94.29%	158	92.69%	92.74%	92	90.71%	90.77%	40	86.19%	86.22%
Ca	100	96.45%	99.63%	100	93.73%	98.92%	100	92.42%	96.92%	100	91.04%	91.89%	100	90.23%	75.10%
	1,000	96.87%	96.39%	1,000	95.51%	86.06%	1,000	95.71%	70.77%	1,000	96.34%	49.97%	1,000	97.62%	26.56%
	10,000	98.04%	75.72%	10,000	98.26%	48.26%	10,000	98.78%	26.55%	10,000	99.37%	11.82%	10,000	99.76%	4.25%
	0	96.42%	100.00%	604	95.00%	91.26%	665	95.01%	77.37%	568	95.00%	62.32%	364	95.01%	46.99%
	57,621	99.00%	50.00%	26,232	99.00%	30.58%	12,901	99.00%	23.22%	5,918	99.00%	17.96%	2,641	99.00%	13.32%
	824	96.79%	96.89%	374	94.50%	94.60%	186	93.11%	93.16%	107	91.14%	91.20%	46	87.28%	87.33%
AA	100	97.00%	99.90%	100	94.39%	98.84%	100	93.10%	98.39%	100	90.41%	88.88%	100	90.13%	77.51%
	1,000	97.23%	98.25%	1,000	95.83%	86.78%	1,000	95.61%	79.92%	1,000	95.91%	49.21%	1,000	96.76%	32.65%
	10,000	98.00%	86.32%	10,000	98.05%	53.41%	10,000	98.57%	35.42%	10,000	99.03%	14.57%	10,000	99.45%	6.76%
	0	97.00%	100.00%	348	95.00%	94.61%	677	95.02%	84.96%	676	95.00%	57.05%	470	95.00%	46.35%
	73,864	99.00%	58.64%	40,163	99.00%	31.62%	18,414	99.00%	25.64%	9,675	99.00%	15.07%	4,347	99.00%	13.12%
	1,351	97.30%	97.33%	319	94.94%	95.03%	285	93.88%	93.94%	89	90.14%	90.19%	48	87.54%	87.62%
HA	100	96.76%	100.00%	100	94.69%	99.61%	100	92.59%	98.56%	100	91.18%	94.46%	100	89.67%	83.74%
	1,000	96.95%	98.09%	1,000	95.68%	93.63%	1,000	95.11%	78.22%	1,000	95.82%	59.08%	1,000	96.91%	35.98%
	10,000	97.81%	85.35%	10,000	97.78%	61.97%	10,000	98.43%	32.81%	10,000	98.98%	16.31%	10,000	99.52%	6.89%
	0	96.73%	100.00%	364	95.01%	97.68%	923	95.01%	79.27%	707	95.01%	66.05%	502	95.02%	51.69%
	109,297	99.01%	51.91%	42,787	99.01%	35.33%	22,291	99.01%	21.13%	10,055	99.01%	16.10%	3,844	99.01%	15.71%
	1,141	97.04%	97.13%	606	95.28%	95.37%	281	93.45%	93.44%	138	91.60%	91.69%	73	88.68%	88.70%

Ch: Chinese, Ca: Caucasians, AA: African Americans, HA: Hispanic.

spe*: specificity sen**: sensitivity

Systematic evaluation of minimum CPI requirement and its effects on specificity and sensitivity for paternity duo testing

Chang En Pu M.Sc. [1*] and Adrian Linacre Ph.D.[2]

1. Scientific and Technical Research Center, Ministry Justice Investigation Bureau, Taipei County, Taiwan.

2. Centre for Forensic Science, WestCHEM, University of Strathclyde, Glasgow, UK.

*Corresponding author: Chang En Pu, Forensic DNA section, Scientific and Technical research center, Ministry Justice Investigation Bureau, Taipei County 231, Taiwan, R.O.C. Tel: +886-2-29112241 ext 3740, fax: +886-2-29138599, E-mail: pu_macros@yahoo.com.tw

Abstract

If only one putative parent and a child are available for genetic linkage studies, termed a duo situation, the discrimination power of the DNA test is on average reduced by half. This increases the chance of a false inclusion and a false exclusion.

We report on duo scenarios from both real and computer based populations using the 15 STR markers within the Identifiler® kit. In order to understand the capability of the duo scenario for paternity testing, members of the population were paired at random to generate children from whom combined paternity indices (CPI) were calculated. There were 202,050 of these real duos (RDs) created from the Chinese

population, 90,300 RDs from the Caucasian population and 65,280 RDs from the African American population. CPI values were calculated using a varying number of STR loci and the random pairs from the three populations were also checked for the existence of coincidental matched pairs (CMPs). Ratio distributions of CPIs were found to be overlapping for RDs and CMPs, thus the specificity (1-% of false inclusions) and sensitivity (1-% of false exclusion) under different minimum CPI requirements were estimated empirically. When using 15 STR loci, the specificity at a CPI cutoff value of 0 was over 99.98% for 3 the populations and only 97.58% when 8 of the 15 loci used.

By setting the specificity at 99% (CPI=100), 99.9% (CPI=1,000) and 99.99% (CPI=10,000), the CPI cutoff values were determined. With a specificity value of 99.99%, the minimum CPI requirement was 139 for the Chinese population, 94 for the Caucasian population and 0 for African American population. Based upon the data obtained we report on how the number of loci used will require the use of varying CPI cutoff values.

Key Words: forensic science, short tandem repeat (STRs), paternity, false inclusion duo, specificity, sensitivity.

Introduction

In a survey conducted by American Association of Blood Banks (AABB), an apparent increase in the number of duo cases has been found (1). A duo case occurs when a DNA sample can be obtained from only one parent for comparison to one obtained from one offspring. Based upon the data obtained a Combined Paternity Index (CPI) can be determined to report on the probability of a first degree genetic relationship. For many of the laboratories involved in paternity testing a minimal CPI requirement is preset for screening test results as a guideline for first degree genetic linkage.

In the same report by the AABB, the minimum CPI value that was required for determining a duo case varied from 100 (25 out of the 39 laboratories) to 10,000 (1 out of the 39 laboratories). CPI cutoff values varied from “whatever is obtained” (17 out of the 35 laboratories) to 1,000 (1 out of the 35 laboratories) for family reconstruction cases. When a CPI value greater than the cut off figure is obtained there is a high degree of confidence that the samples tested are from first degree relatives. There is still the chance that two unrelated people will share one allele at all the 15 loci resulting in a possible false inclusion (2,3), and the use of fewer loci inducing more false inclusions was also shown in other studies (4, 5, 6, 7, 8). Previously only a limited study on the evaluation specificity of some CPI cutoff

values have been reported (8). We report on a study to determine the specificity and sensitivity of the current setting of different CPI cutoff values using: 99%, 99.9% and 99.99% as the specificity of the test.

Materials and Methods

DNA profiles from 450 random members of the Chinese population in Taiwan were generated using the ABI AmpF*l*STR Identifiler® PCR Amplification Kit (Applied Biosystem, Foster City, CA, USA). All the 450 copies of the 15 STR profiles were added to the Microsoft Excel data sheets and processed by a built-in Visual Basic program written by authors of this study. Every individual of the population was paired with every other individual to form pairs, e.g. for the Chinese population in this study (450 x 449)/2 equaling 101,025 pairs were made. Every pair was processed by the program by a standard formula to have two children to be used for counting as a duo CPI (9). Totally 202,050 CPIs were obtained. Some of the loci of the 15 STR core kit were removed to simulate the STR profiling data obtained from using fewer STR loci, e.g. D2S1338 and D19S433 were omitted to mimic the CODIS 13 system. And for mimicking the DNA degradation situation, loci that with longer PCR products were taken off one by one, e.g. starting from the “longest” loci CSF1PO, and then D2S1338, D18S51 etc., to ultimately obtain only 8 loci. From the 101,025 random

pairs generated using loci in a number of commercially available kits or mimicking DNA degradation, CPIs values were obtained for those pairs with coincidental matches where at least one allele was shared at all loci examined.

DNA profiles from the 15 STR loci from a Caucasian (n= 301) and American African (n= 256) population were obtained from the Short Tandem Repeat DNA Internet Database (10). These profiles were also processed by the same program to generate real duos and to determine the number of matching pairs. There were 90,300 real duos generated for CPI counting and 45,150 random pairs analyzed for Coincidental Matching Pairs (CMPs) for the Caucasian population. For the African American population, 65,280 real duos were made from which 32,640 pairs were used to generate the number of CMPs. The allele frequency table used for the Chinese population was from previous studies (11, 12, 13) and the Caucasian and African Americans allele frequencies were downloaded from the same origin as their DNA profiles (10). All the frequencies of alleles were adjusted by using 5/2N rule (15).

The specificity of the duo test was calculated as 1 – the % of false positives and sensitivity of the CPI was based upon 1 – the % of false negatives. The rate of false positives, seen as an inclusion in a paternity test, was determined by the percentage of pairs sharing STR loci coincidentally, and their CPI values were greater than any preset minimum CPI requirements. The percentage of false negatives (seen as an

exclusion of paternity) was the proportion of real duos that might not be recognized as real duos based upon their CPIs lower than the preset cutoff.

Results and discussion

Distribution of CPI of Real Duos (RD) and Coincidental Matched Pairs (CMP) was found to be overlapping

The ratio distribution of CPI values of 202,025 Chinese Real Duos (RDs), 90,300 Caucasian RDs and 65,280 RDs for the STR loci in the Identifiler®, SGM plus® and Profiler® kits are shown in Figure 1. The box-plots for the distribution of CPIs of CMPs are also illustrated. With an increasing number of STR loci used larger CPIs distributions are obtained as expected. There were no clear boundaries to differentiate the distribution of RDs and CMPs when smaller CPI cutoff values are obtained.

For many relationship testing laboratories, a minimum CPI value requirement or cutoff point is used for determining the existence of a first degree genetic relationship. Due to the RDs with low CPI values and CMPs with high CPI values observed in this study, false exclusions could occur by setting the cutoff higher than required, and false inclusion could occur if set too low. Further evaluation of specificity and sensitivity of different CPI cutoff values under different situations is required.

Number of Real Duos and Coincidental Matched Pairs that Meet the Different CPI Requirements

25 CMPs were observed in the 101,025 random pairs from the 450 members of the Chinese population taken at random using 15 STR loci. Eight CMPs for the Caucasian population (from 45,150 pairs) and 4 CMPs (from 32,540 pairs) for the African American population were obtained (table 1). When using STR kits with fewer loci, more CMPs were found. The nine loci in the Profiler® kit resulted in 1,474 CMPs for the Chinese population, 522 CMPs for the Caucasian population and 315 CMPs for African American population.

Specificity and Sensitivity of different CPI Requirements

In table 2 and 3, the specificity and sensitivity for duos using varying numbers of STR loci are illustrated. As the CPI cutoff increased, there was a subsequent decrease in sensitivity and an increase in specificity. Using 15 STR loci the CPI cutoff value varied from 0 (stated as “whatever is obtained”) to 10,000, the sensitivity varied from 100% to 74.48%, the specificity from 99.98% to 100.00% for the Chinese population. The change was more pronounced with fewer STR loci confirming the findings of a previous study (8).

If a minimum CPI value of 100 was chosen using 15 STR loci, the specificity

versus sensitivity for the Chinese population was 99.98% versus 99.72%, for the Caucasian population it was 99.99% versus 99.76% and for the African American population it was 99.99% versus 99.96%. The percentages are reduced if fewer STR loci are used. If CPI=100 for 7 loci (shown as the column 15-8 in table 3), the specificity versus sensitivity for the Chinese population was 99.82% versus 42.83%, for the Caucasian population it was 99.85% versus 45.36% and for African American population it was 99.79% versus 56.01%. Based upon these data setting a minimum CPI requirement may not be appropriate for all scenarios.

Evaluation of CPI cutoff based on Different Specificity in different situation

By setting the specificity at 99% (CPI=100), 99.9% (CPI=1,000) and 99.99% (CPI=10,000), the correlated CPI cutoff could be calculated for varying number of STR loci (shown in tables 4 and 5). A specificity set a 99.99% using the 15 STR loci was obtained with a CPI of 139 for the Chinese population, 94 for the Caucasian population and “whatever is obtained” for the African American population. When using the CODIS 13 STR loci and the same specificity of 99.99%, the CPI cutoff for the Chinese population was 628, 565 for the Caucasian population and 2,024 for the African American population. With these 13 STR loci the sensitivity would be reduced to 86.29%, 91.06%, and 82.03% for the three populations respectively.

By using the optimal cutoff selection method proposed by Zou (17), the optimal CPI cutoff that maximizes the specificity and sensitivity for all the situations were all around “CPI=1” (data not shown). This is in line with the findings of similar low CPI cutoff values reported by a previous study using only a few hundred real duos (8). For automatic database searching, the efficiency of the test depends on how many candidate pairs that are picked by computer program need further confirmation. Ultimately it is the role of the decision maker to set the cutoff when balancing the specificity and sensitivity as improving the capability of one will result in the decrease in the capability of the other. No optimum CPI minimum value can be set for paternity duo tests that will never result in false inclusions or false exclusions. The laboratories that joined the AABB survey strongly supported that testing without a mother should be processed only when the mother is unavailable or if she is deceased. Otherwise, every effort should be made to include the mother in the test (1).

In conclusion, a different CPI cutoff may be applied for duo cases based upon the number of STR loci used.

Acknowledgments

The authors wish to thank Wu F.C., Wu G.C., Cheng M.Y., and Meng L.M. for their kindly help for profiling all the local samples.

Reference

1. ANNUAL REPORT SUMMARY FOR TESTING IN 2004, Prepared by the Relationship Testing Program Unithttp: //www.aabb.org/ Documents/ Accreditation/ Parentage_Testing_Accreditation_Program/ rtannrpt04.pdf (accessed May 1 2008)

2. Konjhodzic R, Omerovic E, Davoren J, Vanek D, Crews J and Huffine E. Performance Comparison of the Penta D and Penta E Loci With the D2S1338 and D19S433 Loci in the Massive DNA Identification Effort in Former Yugoslavia, In: Proceedings of the 55th Annual Meeting of the American Academy of Forensic Sciences, Chicago, IL, USA, 2003, pp 37.

3. Rizvic A and Huffine E. Coordinating the Identification Efforts of the Missing from Former Yugoslavia, In: Proceedings of the 55th Annual Meeting of the American Academy of Forensic Sciences, Chicago, IL, USA, 2003, pp 64-65.

4. Birus I, Marcikic M, Lauc D, Dzijan S and Lauc G. How high should paternity index be for reliable identification of war victims by DNA typing? *Croat. Med. J.* 44:322-6, 2003.

5. De Ungria MC, Frani AM, Magno MM, Tabbada KA, Calacal GC, Delfin FC and Halos SC. Evaluating DNA tests of motherless cases using a Philippine

genetic database. *Transfusion* 42:954-957, 2002.

6. Gornik I, Marcikic M, Kubat M, Primorac D, and Lauc G. The identification of war victims by reverse paternity is associated with significant risks of false inclusion. *Int. J. Legal. Med.* 116:255-7, 2002.

7. Presciuttini S, Ciampini F, Alu M, Cerri N, Dobosz M, and Domenici R Allele sharing in first-degree and unrelated pairs of individuals in the Ge.F.I. AmpFlSTR® Profiler Plus™ database, *Forensic Sci. Int.* 131: 85-89, 2003.

8. Pu C.E., Linacre A. CPI distribution and cutoff values for duo kinship testing. Chinese J. Physiology 50(5):232-239, 2007.

9. AABB, Guidance for standards for relationship testing laboratories, MD, USA, American association of blood banks, 2006, pp 139.

10. Short Tandom Repeat DNA Internet Database, http://www.cstl.nist.gov/div831/strbase/ (accessed May 1 2008)

11. Pu CE, Wu FC, Cheng CL, Wu KC, Chao CH, and Li JM. DNA STR profiling of Chinese population in Taiwan determined by using an automated sequencer, *Forensic Sci. Int.* 97:47-51, 1998.

12. Pu, CE. et al. DNA profiling of Chinese population in Taiwan on the STR loci D8S1179, D16S539, D18S51 and D21S11 52th annual meeting, American Academy of Forensic Sciences, Reno, NV, USA,

2000.

13. Lee JC, Chang YY, Su CW, Tzeng CH, Pu CE, Han TH, Chang CW, Liu DC, Lin FR, Lin YC, Chen MH. The Screening of STR and YSTR Loci in Taiwanese Han Population, Proceedings of Taiwan Academy of Forensic Sciences, 2006, 27-38.(in Chinese)

14. Pu, CE and Linacre A. CPI distribution and cut-off value for duo paternity building. In: Proceedings of the 58th Annual Meeting of the American Academy of Forensic Sciences, Seattle, WA, USA, 2006, pp 172-173.

15. Butler JM. Kinship and parentage testing. In: Forensic DNA Typing, Elsevier Academic Press, MA, USA, 2005, pp 529-537.

16. Kirkwood BR and Sterne JAC. Measurement error: assessment and implications. In: Medical statistics, MA, USA, Blackwell publishing, pp. 429-446, 2003.

17. Zou KH, Wells III WM, Kikinis R and Warfield SK. Three validation metrics for automated probabilistic image segmentation of brain tumors, Statist. Med. 23:1259-1282, 2004.

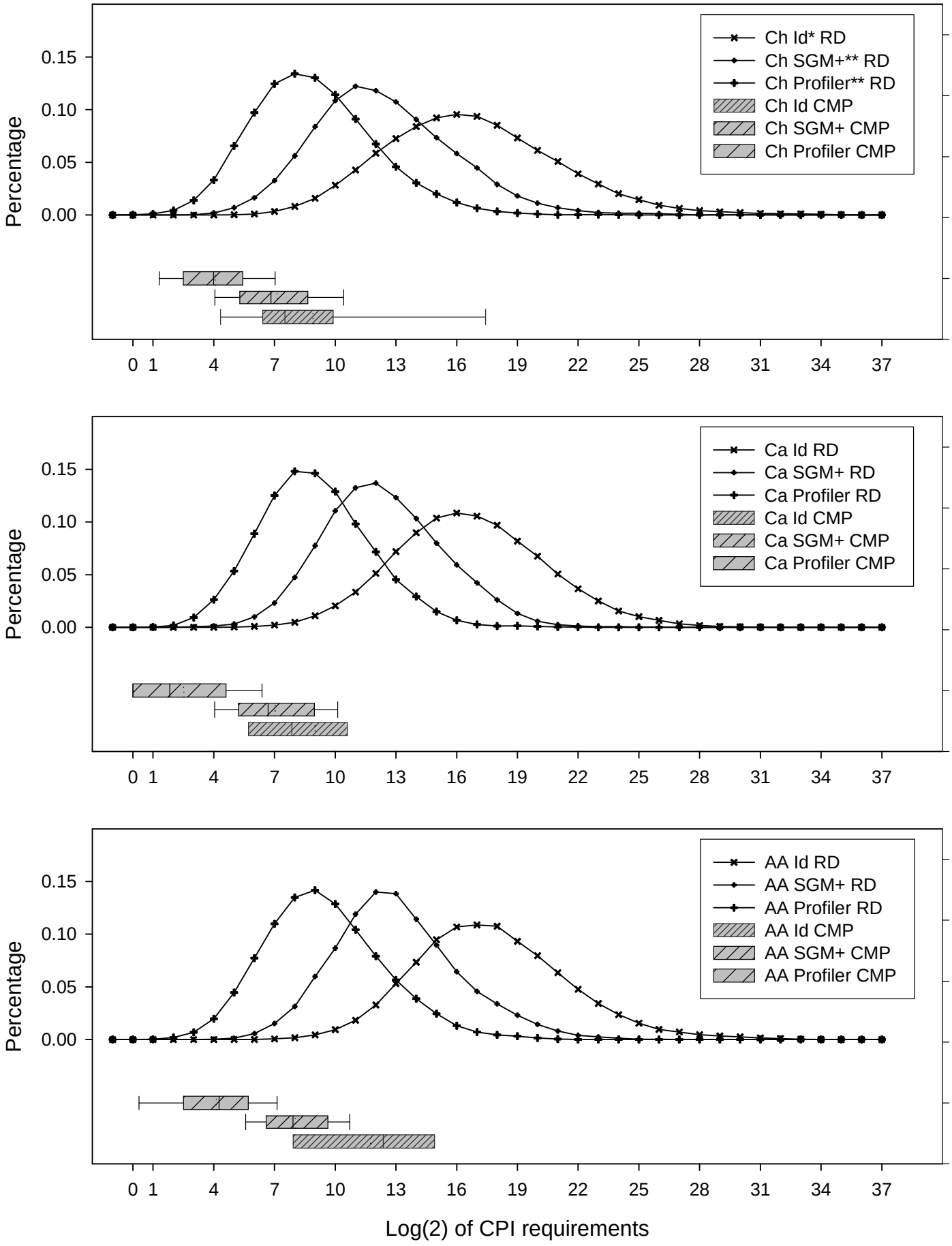


Fig 1. Ratio distribution and boxplot of CPI for 3 systems and 3 populations (Ch: Chinese, Ca: Caucasian, AA: African Americans, *: Identifiler, **: SGM plus and ***: Profiler are PCR Amplification kit made by Applied Biosystem , Foster City, LA, USA)

Table 1. Number of RDs and CMPs that meet the minimum CPI requirements for 3 populations on 5 STR systems

System name	Identifiler		CODIS 13		SGM+		Profiler+*		Profiler	
Minimum CPI requirement	RD	CMP	RD	CMP	RD	CMP	RD	CMP	RD	CMP
Population	Chinese									
What ever is obtained	0	25	0	98	0	162	0	434	0	1,474
10	2	25	105	90	160	153	712	392	5,585	919
100	571	17	4,564	47	8,864	87	21,629	138	59,456	176
101	576	17	4,620	47	8,982	87	21,894	137	59,810	175
150	1,172	14	7,424	37	13,743	75	31,661	103	74,918	133
200	1,743	12	10,104	30	18,248	67	39,780	82	86,242	95
300	3,015	11	15,182	28	26,104	51	53,589	56	102,081	63
500	5,552	9	23,358	17	38,745	36	72,804	35	121,301	31
1,000	11,161	5	38,076	10	60,505	22	100,041	22	144,491	9
1,001	11,172	5	38,101	10	60,552	22	100,074	22	144,521	9
10,000	51,557	3	104,456	2	137,984	2	168,727	4	188,703	1
number of RD	202,050									
number of CMP comparing	101,025									
Population	Caucasian									
What ever is obtained	0	8	0	31	0	67	0	159	0	522
10	0	8	11	30	46	64	118	151	1,657	357
100	218	4	1,135	20	2,425	34	7,724	73	23,007	72
101	220	4	1,148	20	2,467	34	7,809	73	23,202	72
150	389	4	2,007	15	4,087	28	11,640	53	30,178	55
200	584	4	2,793	10	5,714	27	15,586	39	35,813	41
300	1,005	4	4,472	10	8,875	21	21,273	24	43,731	29
500	1,748	4	7,267	8	14,407	16	30,441	14	53,622	15
1,000	3,477	2	12,980	4	24,232	8	43,631	4	65,215	7
1,001	3,485	2	12,991	4	24,244	8	43,647	4	65,237	7
10,000	20,183	1	44,910	0	63,181	1	78,648	0	86,155	0
number of RD	90,300									
number of CMP comparing	45,150									
Population	African Americans									
What ever is obtained	0	4	0	21	0	30	0	110	0	315
10	0	4	6	21	128	30	189	103	975	224
100	27	4	503	19	5,258	21	6,259	45	14,495	52
101	28	4	509	19	5,323	21	6,345	44	14,617	52
150	71	3	869	16	7,671	19	9,204	38	19,172	31
200	105	3	1,256	15	9,965	18	11,647	29	22,867	23
300	228	3	2,066	10	13,491	12	15,798	19	28,154	12
500	437	3	3,530	9	18,864	9	21,790	12	34,840	7
1,000	1,046	3	6,709	7	26,633	5	30,719	6	43,169	4
1,001	1,050	3	6,717	7	26,639	5	30,734	6	43,176	4
10,000	8,987	2	28,036	1	50,208	0	54,751	2	60,144	1
number of RD	65,280									
number of CMP comparing	32,640									

Profiler+*: abbreviation of Profiler Plus AmpliSTR kit from Applied Biosystem, USA.

Table 2. Sensitivity and specificity versus minimum CPI requirements for 3 populations on 5 STR systems

System name	Identifiler		CODIS 13		SGM+		Profiler+		Profiler	
Minimum CPI requirement	spe*	sen**	spe	sen	spe	sen	spe	sen	spe	sen
Population	Chinese									
What ever is obtained	99.98%	100.00%	99.90%	100.00%	99.84%	100.00%	99.57%	100.00%	98.54%	100.00%
10	99.98%	100.00%	99.91%	99.95%	99.85%	99.92%	99.61%	99.65%	99.09%	97.24%
100	99.98%	99.72%	99.95%	97.74%	99.91%	95.61%	99.86%	89.30%	99.83%	70.57%
101	99.98%	99.71%	99.95%	97.71%	99.91%	95.55%	99.86%	89.16%	99.83%	70.40%
150	99.99%	99.42%	99.96%	96.33%	99.93%	93.20%	99.90%	84.33%	99.87%	62.92%
200	99.99%	99.14%	99.97%	95.00%	99.93%	90.97%	99.92%	80.31%	99.91%	57.32%
300	99.99%	98.51%	99.97%	92.49%	99.95%	87.08%	99.94%	73.48%	99.94%	49.48%
500	99.99%	97.25%	99.98%	88.44%	99.96%	80.82%	99.97%	63.97%	99.97%	39.96%
1,000	100.00%	94.48%	99.99%	81.16%	99.98%	70.05%	99.98%	50.49%	99.99%	28.49%
1,001	100.00%	94.47%	99.99%	81.14%	99.98%	70.03%	99.98%	50.47%	99.99%	28.47%
10,000	100.00%	74.48%	100.00%	48.30%	100.00%	31.71%	100.00%	16.49%	100.00%	6.61%
Population	Caucasian									
What ever is obtained	99.98%	100.00%	99.93%	100.00%	99.85%	100.00%	99.65%	100.00%	98.84%	100.00%
10	99.98%	100.00%	99.93%	99.99%	99.86%	99.95%	99.67%	99.87%	99.21%	98.17%
100	99.99%	99.76%	99.96%	98.74%	99.92%	97.31%	99.84%	91.45%	99.84%	74.52%
101	99.99%	99.76%	99.96%	98.73%	99.92%	97.27%	99.84%	91.35%	99.84%	74.31%
150	99.99%	99.57%	99.97%	97.78%	99.94%	95.47%	99.88%	87.11%	99.88%	66.58%
200	99.99%	99.35%	99.98%	96.91%	99.94%	93.67%	99.91%	82.74%	99.91%	60.34%
300	99.99%	98.89%	99.98%	95.05%	99.95%	90.17%	99.95%	76.44%	99.94%	51.57%
500	99.99%	98.06%	99.98%	91.95%	99.96%	84.05%	99.97%	66.29%	99.97%	40.62%
1,000	100.00%	96.15%	99.99%	85.63%	99.98%	73.17%	99.99%	51.68%	99.98%	27.78%
1,001	100.00%	96.14%	99.99%	85.61%	99.98%	73.15%	99.99%	51.66%	99.98%	27.76%
10,000	100.00%	77.65%	100.00%	50.27%	100.00%	30.03%	100.00%	12.90%	100.00%	4.59%
Population	African Americans									
What ever is obtained	99.99%	100.00%	99.94%	100.00%	99.91%	100.00%	99.66%	100.00%	99.03%	100.00%
10	99.99%	100.00%	99.94%	99.99%	99.91%	99.80%	99.68%	99.71%	99.31%	98.51%
100	99.99%	99.96%	99.94%	99.23%	99.94%	91.95%	99.86%	90.41%	99.84%	77.80%
101	99.99%	99.96%	99.94%	99.22%	99.94%	91.85%	99.87%	90.28%	99.84%	77.61%
150	99.99%	99.89%	99.95%	98.67%	99.94%	88.25%	99.88%	85.90%	99.91%	70.63%
200	99.99%	99.84%	99.95%	98.08%	99.94%	84.73%	99.91%	82.16%	99.93%	64.97%
300	99.99%	99.65%	99.97%	96.84%	99.96%	79.33%	99.94%	75.80%	99.96%	56.87%
500	99.99%	99.33%	99.97%	94.59%	99.97%	71.10%	99.96%	66.62%	99.98%	46.63%
1,000	99.99%	98.40%	99.98%	89.72%	99.98%	59.20%	99.98%	52.94%	99.99%	33.87%
1,001	99.99%	98.39%	99.98%	89.71%	99.98%	59.19%	99.98%	52.92%	99.99%	33.86%
10,000	99.99%	86.23%	100.00%	57.05%	100.00%	23.09%	99.99%	16.13%	100.00%	7.87%

spe*: specificity.

sen**: sensitivity.

Table 3. Sensitivity and specificity versus minimum CPI requirements mimicking degradation from long loci for 3 populations on 8 degradation situations

Degradation situation*	15-0		15-1		15-2		15-3		15-4		15-5		15-6		15-7		15-8	
Minimum CPI requirement	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen
Population	Chinese																	
What ever is obtained	99.98%	100.00%	99.97%	100.00%	99.92%	100.00%	99.84%	100.00%	99.62%	100.00%	99.37%	100.00%	98.97%	100.00%	97.91%	100.00%	96.38%	100.00%
10	99.98%	100.00%	99.97%	100.00%	99.93%	99.96%	99.86%	99.87%	99.69%	99.35%	99.52%	98.01%	99.27%	96.46%	98.79%	93.18%	98.31%	87.63%
100	99.98%	99.72%	99.98%	99.46%	99.96%	98.02%	99.93%	95.56%	99.88%	89.23%	99.83%	80.44%	99.81%	72.20%	99.79%	56.66%	99.82%	42.83%
101	99.98%	99.71%	99.98%	99.45%	99.96%	97.98%	99.93%	95.52%	99.88%	89.16%	99.84%	80.31%	99.81%	72.03%	99.80%	56.49%	99.83%	42.56%
150	99.99%	99.42%	99.98%	99.00%	99.96%	96.86%	99.94%	93.34%	99.91%	85.23%	99.88%	74.70%	99.86%	65.27%	99.88%	48.37%	99.91%	34.59%
200	99.99%	99.14%	99.99%	98.51%	99.97%	95.74%	99.95%	91.34%	99.92%	81.78%	99.90%	70.37%	99.90%	60.22%	99.92%	42.72%	99.94%	29.26%
300	99.99%	98.51%	99.99%	97.59%	99.98%	93.70%	99.96%	87.85%	99.95%	76.46%	99.93%	63.71%	99.94%	52.72%	99.95%	35.01%	99.96%	22.63%
500	99.99%	97.25%	99.99%	95.79%	99.98%	90.28%	99.97%	82.61%	99.96%	68.87%	99.97%	55.01%	99.97%	43.76%	99.98%	26.32%	99.99%	15.59%
1,000	100.00%	94.48%	99.99%	92.08%	99.99%	84.05%	99.99%	73.80%	99.98%	57.80%	99.98%	43.29%	99.98%	32.27%	99.99%	16.74%	99.99%	9.14%
1,001	100.00%	94.47%	99.99%	92.07%	99.99%	84.04%	99.99%	73.79%	99.98%	57.78%	99.98%	43.27%	99.98%	32.25%	99.99%	16.73%	99.99%	9.13%
10,000	100.00%	74.48%	100.00%	68.44%	100.00%	53.44%	100.00%	38.35%	100.00%	22.93%	100.00%	13.52%	100.00%	7.97%	100.00%	3.05%	100.00%	1.43%
Population	Caucasian																	
What ever is obtained	99.98%	100.00%	99.98%	100.00%	99.93%	100.00%	99.84%	100.00%	99.65%	100.00%	99.40%	100.00%	99.03%	100.00%	98.14%	100.00%	96.52%	100.00%
10	99.98%	100.00%	99.98%	99.99%	99.93%	99.99%	99.84%	99.91%	99.65%	99.64%	99.40%	99.01%	99.03%	98.32%	98.14%	96.04%	98.43%	91.49%
100	99.99%	99.76%	99.98%	99.67%	99.93%	98.88%	99.84%	96.57%	99.65%	92.24%	99.40%	86.14%	99.03%	78.37%	98.14%	62.89%	99.85%	45.36%
101	99.99%	99.76%	99.98%	99.67%	99.93%	98.86%	99.84%	96.53%	99.65%	92.15%	99.40%	86.01%	99.03%	78.18%	98.14%	62.67%	99.86%	45.09%
150	99.99%	99.57%	99.98%	99.39%	99.93%	98.08%	99.85%	94.69%	99.71%	88.67%	99.56%	80.86%	99.34%	71.44%	98.93%	53.95%	99.90%	36.32%
200	99.99%	99.35%	99.99%	99.10%	99.96%	97.37%	99.92%	93.00%	99.89%	85.69%	99.87%	76.62%	99.84%	66.13%	99.85%	47.60%	99.94%	30.66%
300	99.99%	98.89%	99.99%	98.56%	99.96%	95.87%	99.92%	89.99%	99.89%	80.50%	99.88%	69.74%	99.84%	58.00%	99.85%	38.72%	99.97%	23.39%
500	99.99%	98.06%	99.99%	97.46%	99.97%	93.25%	99.94%	84.78%	99.92%	72.61%	99.90%	60.26%	99.89%	47.74%	99.91%	28.63%	99.99%	16.16%
1,000	100.00%	96.15%	99.99%	94.88%	99.97%	87.79%	99.95%	75.42%	99.94%	60.51%	99.93%	46.59%	99.91%	34.61%	99.94%	17.83%	100.00%	9.02%
1,001	100.00%	96.14%	99.99%	94.87%	99.98%	87.78%	99.96%	75.41%	99.96%	60.49%	99.94%	46.56%	99.95%	34.60%	99.96%	17.80%	100.00%	9.02%
10,000	100.00%	77.65%	99.99%	72.21%	99.99%	54.31%	99.97%	36.02%	99.97%	21.27%	99.97%	12.59%	99.97%	7.10%	99.98%	1.80%	100.00%	0.86%
Population	African Americans																	
What ever is obtained	99.99%	100.00%	99.98%	100.00%	99.95%	100.00%	99.87%	100.00%	99.75%	100.00%	99.53%	100.00%	99.14%	100.00%	98.27%	100.00%	97.58%	100.00%
10	99.99%	100.00%	99.98%	100.00%	99.95%	99.99%	99.88%	99.95%	99.77%	99.73%	99.61%	99.51%	99.36%	99.00%	98.97%	96.77%	98.65%	95.34%
100	99.99%	99.96%	99.98%	99.90%	99.97%	99.32%	99.93%	97.93%	99.87%	94.06%	99.86%	89.88%	99.79%	83.04%	99.77%	67.31%	99.79%	56.01%
101	99.99%	99.96%	99.98%	99.89%	99.97%	99.31%	99.93%	97.90%	99.87%	93.97%	99.86%	89.79%	99.79%	82.96%	99.77%	66.98%	99.79%	55.78%
150	99.99%	99.89%	99.98%	99.73%	99.97%	98.79%	99.94%	96.69%	99.90%	91.07%	99.89%	85.65%	99.84%	76.97%	99.84%	58.77%	99.88%	45.96%
200	99.99%	99.84%	99.98%	99.63%	99.97%	98.31%	99.95%	95.35%	99.92%	88.46%	99.91%	81.91%	99.88%	72.03%	99.89%	52.36%	99.92%	39.17%
300	99.99%	99.65%	99.98%	99.24%	99.98%	97.34%	99.95%	93.11%	99.94%	83.90%	99.94%	76.10%	99.92%	64.58%	99.92%	43.99%	99.96%	30.37%
500	99.99%	99.33%	99.98%	98.58%	99.98%	95.59%	99.97%	89.11%	99.95%	76.90%	99.97%	67.51%	99.96%	54.62%	99.95%	34.32%	99.97%	21.38%
1,000	99.99%	98.40%	99.99%	96.81%	99.99%	91.63%	99.97%	81.58%	99.98%	65.71%	99.98%	54.81%	99.98%	41.16%	99.98%	22.80%	99.99%	12.29%
1,001	99.99%	98.39%	99.99%	96.81%	99.99%	91.61%	99.97%	81.56%	99.98%	65.69%	99.98%	54.78%	99.98%	41.15%	99.98%	22.79%	99.99%	12.27%
10,000	99.99%	86.23%	99.99%	79.14%	100.00%	62.70%	100.00%	43.53%	100.00%	27.01%	100.00%	18.41%	100.00%	10.62%	100.00%	3.37%	100.00%	1.52%

Degradation situation*: degradation of loci following the sequence of CSF1PO, D2S1338, D18S51, FGA, D7S820, D16S539, D21S11, D13S317 in the Identifiler kit.

Table 4. Predicted minimum CPI requirements of specificity of 99.90% and 99.99% for 3 populations and 5 STR systems

System name	Identifiler		CODIS 13		SGM+		Profiler+		Profiler	
Minimum CPI requirement	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen
Population	Chinese									
What ever is obtained	99.98%	100.00%	99.90%	100.00%	99.84%	100.00%	99.57%	100.00%	-	-
7.7	-	-	-	-	-	-	-	-	99.00%	98.12%
62	-	-	-	-	99.90%	97.56%	-	-	-	-
139	99.99%	99.49%	-	-	-	-	-	-	-	-
145	-	-	-	-	-	-	99.90%	84.76%	-	-
182	-	-	-	-	-	-	-	-	99.90%	59.09%
628	-	-	99.99%	86.29%	-	-	-	-	-	-
680	-	-	-	-	-	-	-	-	99.99%	34.61%
1,381	-	-	-	-	99.99%	64.42%	-	-	-	-
1,433	-	-	-	-	-	-	99.99%	43.74%	-	-
Population	Caucasian									
What ever is obtained	99.98%	100.00%	99.93%	100.00%	99.85%	100.00%	99.65%	100.00%	-	-
4.2	-	-	-	-	-	-	-	-	99.00%	99.70%
44	-	-	-	-	99.90%	99.18%	-	-	-	-
94	99.99%	99.79%	-	-	-	-	-	-	-	-
166	-	-	-	-	-	-	99.90%	85.53%	-	-
171	-	-	-	-	-	-	-	-	99.90%	63.86%
565	-	-	99.99%	91.06%	-	-	-	-	-	-
759	-	-	-	-	-	-	99.99%	57.62%	-	-
1,010	-	-	-	-	-	-	-	-	99.99%	27.61%
1,109	-	-	-	-	99.99%	71.35%	-	-	-	-
Population	African Americans									
What ever is obtained	99.99%	100.00%	99.94%	100.00%	99.91%	100.00%	99.66%	100.00%	99.03%	100.00%
140	-	-	-	-	-	-	-	-	99.90%	71.96%
168	-	-	-	-	-	-	-	-	-	-
703	-	-	-	-	-	-	-	-	99.99%	40.01%
1,088	-	-	-	-	99.99%	57.85%	-	-	-	-
2,024	-	-	99.99%	82.03%	-	-	-	-	-	-
2,236	-	-	-	-	-	-	99.99%	37.53%	-	-

Table 5. Predicted minimum CPI requirements of specificity of 99.90% and 99.99% for 3 populations and 8 different degradation situation

Degradation situation		15-0		15-1		15-2		15-3		15-4		15-5		15-6		15-7		15-8	
Minimum CPI requirement		spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen	spe	sen
Population		Chinese																	
What ever is obtained		99.98%	100.00%	99.97%	100.00%	99.92%	100.00%	99.84%	100.00%	99.62%	100.00%	99.37%	100.00%	-	-	-	-	-	-
	1	-	-	-	-	-	-	-	-	-	-	-	-	99.00%	99.90%	-	-	-	-
	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.00%	90.01%	-	-
	20.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.00%	76.08%
	37	-	-	-	-	-	-	99.90%	98.70%	-	-	-	-	-	-	-	-	-	-
	124	-	-	-	-	-	-	-	-	99.90%	87.22%	-	-	-	-	-	-	-	-
	139	99.99%	99.49%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	135	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	36.67%
	186	-	-	99.99%	98.66%	-	-	-	-	-	-	99.90%	71.54%	-	-	-	-	-	-
	170	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	45.98%	-	-
	198	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	60.39%	-	-	-	-
	707	-	-	-	-	99.99%	87.40%	-	-	-	-	-	-	-	-	-	-	-	-
	945	-	-	-	-	-	-	99.99%	74.58%	-	-	-	-	-	-	-	-	-	-
	749	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	20.40%	-	-
	470	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	16.32%
	1,171	-	-	-	-	-	-	-	-	99.99%	55.25%	-	-	-	-	-	-	-	-
	1,080	-	-	-	-	-	-	-	-	-	-	99.99%	42.05%	-	-	-	-	-	-
	1,240	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	29.01%	-	-	-	-
Population		Caucasian																	
What ever is obtained		99.98%	100.00%	99.98%	100.00%	99.93%	100.00%	99.84%	100.00%	99.65%	100.00%	99.40%	100.00%	99.03%	100.00%	-	-	-	-
	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.00%	94.92%	-	-
	19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.00%	82.47%
	41	-	-	-	-	-	-	99.90%	98.87%	-	-	-	-	-	-	-	-	-	-
	68	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	94	99.99%	99.79%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	99	-	-	99.99%	99.68%	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	104	-	-	-	-	-	-	-	-	99.90%	91.91%	-	-	-	-	-	-	-	-
	133	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	56.67%	-	-
	137	-	-	-	-	-	-	-	-	-	-	99.90%	82.15%	-	-	-	-	-	-
	142	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	37.56%
	157	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	70.76%	-	-	-	-
	394	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	359	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	20.63%
	722	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	22.54%	-	-
	793	-	-	-	-	99.99%	89.81%	-	-	-	-	-	-	99.99%	38.80%	-	-	-	-
	915	-	-	-	-	-	-	-	-	-	-	99.99%	48.29%	-	-	-	-	-	-
	1,181	-	-	-	-	-	-	-	-	99.99%	57.41%	-	-	-	-	-	-	-	-
	1,246	-	-	-	-	-	-	99.99%	72.19%	-	-	-	-	-	-	-	-	-	-
Population		African Americans																	
What ever is obtained		99.99%	100.00%	99.98%	100.00%	99.95%	100.00%	99.87%	100.00%	99.75%	100.00%	99.53%	100.00%	99.14%	100.00%	-	-	-	-
	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.00%	96.57%	-	-
	17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.00%	90.42%
	21	-	-	-	-	-	-	99.90%	99.75%	-	-	-	-	-	-	-	-	-	-
	139	-	-	-	-	-	-	-	-	99.90%	91.71%	-	-	-	-	-	-	-	-
	156	-	-	-	-	-	-	-	-	-	-	99.90%	85.15%	-	-	-	-	-	-
	161	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	44.30%
	222	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	70.18%	-	-	-	-
	230	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.90%	49.49%	-	-
	724	-	-	99.99%	97.80%	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	881	-	-	-	-			-	-	-	-	-	-	-	-	-	-	99.99%	13.72%
	982	-	-	-	-	99.99%	91.75%	-	-	-	-	-	-	-	-	-	-	-	-
	1425	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	34.77%	-	-	-	-
	1,438	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.99%	17.66%	-	-
	1,495	-	-	-	-	-	-	-	-	-	-	99.99%	47.47%	-	-	-	-	-	-
	1,708	-	-	-	-	-	-	-	-	99.99%	56.26%	-	-	-	-	-	-	-	-
	1,983	-	-	-	-	-	-	99.99%	71.70%	-	-	-	-	-	-	-	-	-	-

Available online at www.sciencedirect.com
ScienceDirect
Journal of Forensic and Legal Medicine 15 (2008) 373–377
JOURNAL OF FORENSIC AND LEGAL MEDICINE
www.elsevier.com/jflm

Original Communication

Increasing the confidence in half-sibship determination based upon 15 STR loci

Chang En Pu MS (Scientist Specialist and Section Chief) [a,*], Adrian Linacre PhD (Senior Lecturer) [b]

[a] *Scientific and Technical Research Center, Ministry Justice Investigation Bureau, Taipei County 231, Taiwan, ROC*
[b] *Centre for Forensic Science, WestCHEM, University of Strathclyde, Glasgow, UK*

Received 4 September 2007; received in revised form 22 December 2007; accepted 8 February 2008
Available online 11 April 2008

Abstract

A half-sibship relationship is when two siblings share only one parent. It may be necessary to determine if two individuals are half-siblings in cases of immigration, inheritance, genetic counseling or the identification of human remains. In such instances a combined half-sibship index (CHSI) can be calculated. Support for this kinship is also based upon the number of shared-alleles at DNA loci. We report on the combination of the calculation of CHSI with the all-shared-alleles (ASA) to enhance the specificity of any half-sibship test. The 15 STR loci (including CODIS 13) that comprise the Identifiler® loci were applied to three populations using 355,620 simulated pairs of half-siblings and 178,815 unrelated pairs. Based upon the data obtained, the sensitivity and specificity can be evaluated to determine the existence of half-sibship. This report highlights the uncertainty problems inherent in this form of indirect kinship testing and recommends a combination evaluation of CHSI and ASA.

Keywords: Short tandem repeat (STR); Kinship; Half-sibship; Combined half-sibship index (CHSI); Sensitivity; Specificity

1. Introduction

A claim of half-sibling relationship may be made in cases of inheritance, immigration[1] and genetic counseling.[2] A genetic link as a half-sibling may assist in the identification of human remains if no closer genetic relative is available for testing. A half-sibling is defined as two individuals sharing one parent, this may be a mother when it is termed uterine, or one father when it is called agnate or consanguine. Theoretically there is a 0.5 chance that two half-siblings will share 1 allele at any one locus and a 0.5 chance that they will share no alleles.[4] The confidence that two individuals are half-siblings, or not, will increase if more loci are examined, or if there are more potential half-siblings for comparison. When determining the probability whether or not two individuals are half-siblings it is necessary to consider the allele frequencies of any alleles that are shared. From this calculation a combined half-sibship indices (CHSI)[3] is reported. If CHSI is less than 1 then the evidence supports two individuals as not being related as half-siblings, otherwise the data supports the existence of a half-sibling relationship. There remains much uncertainty after this calculation.[1,3,4] Uncertainty remains with the evaluation of the resulting LR figure in terms of its reliability in determining whether two samples have come from half-siblings.

We report on the use 15 STR, that comprise the Identifiler® loci (including the 13 CODIS loci), to study three populations containing 355,620 simulated half-sibling pairs and 178,815 random pairs generated from DNA profiles created within the populations. Using this large number of sample pairs, it is possible to evaluate the sensitivity and specificity for the 15 STR core set in discriminating between half-siblings and random pairs. The discriminating power combining the CHSI and all-shared-alleles (ASA) is reported that would assist a laboratory with an interpretation of test results.

* Corresponding author. Tel.: +886 2 29112241x3740; fax: +886 2 29138599.
E-mail address: pu_macros@yahoo.com.tw (C.E. Pu).

1752-928X/$ - see front matter © 2008 Elsevier Ltd and FFLM. All rights reserved.
doi:10.1016/j.jflm.2008.02.004

2. Materials and methods

450 STR DNA profiles from random Chinese individuals within the Taiwan population were obtained using the ABI AmpF*l*STR Identifiler® PCR Amplification Kit (Applied Biosystem, Foster City, CA, USA).

The profiles were processed by Microsoft Excel Macros controlled by built in Visual Basic a program written by authors of this study. Every individual of the population was paired with every other individual to form random pairs, e.g. for Taiwan population in this study $(450 \times 449)/2$, equaling 101,025 pairs, were made. For generating half-siblings, every individual in the population was paired with two other individuals, e.g., $(449 \times 448)/2$, equaling 100,576 triples (one person with two mates). One offspring was generated from each of the two pairs within a triple, resulting in 201,152 half-sibling pairs. DNA profiles from the 15 STR loci from a Caucasian ($n = 301$) and American African ($n = 256$) population were obtained from Short Tandem Repeat DNA Internet Database.[6] These data were processed in the same way as those for the Taiwanese population to generate half-sibling pairs and random pairs. For the Caucasian population 89,700 half-sibling pairs were generated and 45,150 random pairs were obtained and for African American population 64,770 half-sibling pairs and 32,640 random pairs were generated. The CHSI were calculated according to standard formulae,[3] and allele frequency tables used for Caucasian and African American populations were downloaded from the same origin as the DNA profiles[6] and the frequency table for the Chinese population was from previous studies.[7–9] All the frequencies of alleles were adjusted by using 5/2N rule[10] for this study.

The rate of false negatives was calculated as the percentage of real half-sibship that could be excluded based upon any given cutoff point of CHSI. The rate of false positives equaled the percentage of random pairs of DNA profiles where their CHSI was greater than any chosen CHSI threshold value. The sensitivity of the test is based upon 1 – the % of false negatives and the specificity is based upon 1 – the % of false positives.[11]

ASA were determined by counting all the alleles shared by paired profiles. In the case of both alleles at any one locus being shared a score of 2 is registered. If 1 allele is in common a score of 1 is registered.

The above formulae were also applied on four situations, like low CHSI with low ASA, low CHSI with high ASA, high CHSI with low ASA and high CHSI with high ASA for real half-siblings and random pairs to evaluate the synergy effects combing both CHSI and ASA.

3. Results and discussion

3.1. Ratio distribution of CHSI for three populations by using 15 STR core set

The CHSI ratio distribution of simulated half-sibling pairs and random pairs for three populations is shown in Fig. 1. Bipolar models with a widespread ratio distribution were found for all of the three populations. Simulated siblings pairs with very low CHSI values (1.40E−03 for the Chinese population) and random pairs with very high

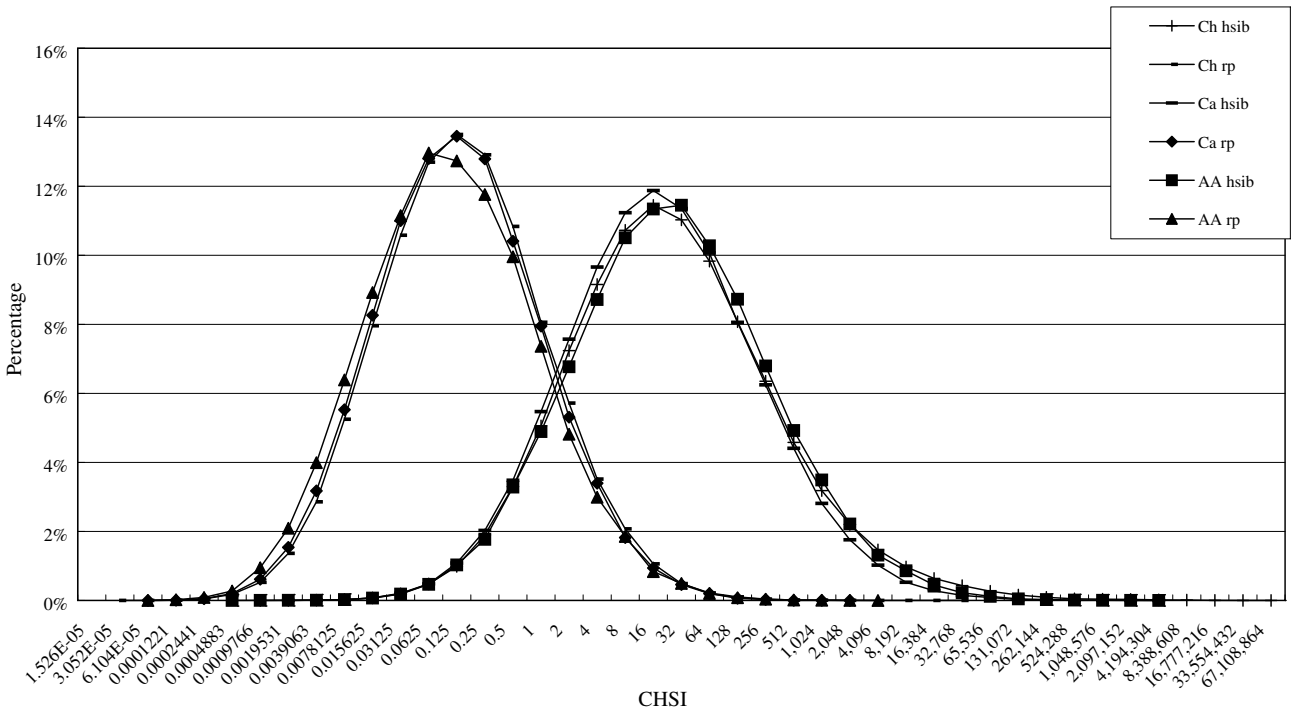


Fig. 1. CHSI ratio distribution for three populations using the 15 STR system (Ch: Chinese, Ca: Caucasians, AA: African Americans, hsib: half-siblings; rp: random pairs).

CHSI values (2.29E+05 for African American population) were observed in this study, together with other extreme CHSI values are reported in Table 1. From the data obtained for the 15 STR loci (Table 2), 12.02%, 12.86% and 11.73% of CHSI values from simulated half-sibling pairs were found to be less than 1 for Chinese, Caucasians and African Americans populations, respectively. Random pairs that produced CHSI values greater than 1 for these populations were 13.24%, 12.22% and 11.33% for these three populations.

3.2. Sensitivity and specificity under different CHSI cut-offs for 15 STR systems

The ability of the DNA test to correctly classify kinship testing results into two categories is assessed by their specificity and sensitivity.[11] In Table 2, the sensitivity and specificity using a range of CHSI values from 0.03125 to 1000 are illustrated. The CHSI requirements as cutoff values followed previous recommendations.[5] As the CHSI cutoff values increased there was a corresponding decrease in sensitivity and increase in specificity. According to Table 2, when adopting CHSI value of 1 as the cutoff value, then in 15 STR core set system for the three populations the sensitivity was 87.98% and the specificity was 86.76% for the Chinese population, for Caucasian population the sensitiv-

Table 1
Highest and lowest CHSIs for three populations using 15 STR systems

		Pairs	Highest	Lowest
Chinese	Simulated half-siblings	201,152	1.90E+07	1.40E−03
	random pairs	101,025	1.04E+04	5.34E−05
Caucasians	Simulated half-siblings	89,700	1.35E+06	2.30E−03
	random pairs	45,150	6.90E+02	7.77E−05
African Americans	Simulated half-siblings	64,770	1.37E+06	8.13E−04
	random pairs	32,640	2.29E+05	7.15E−05

Table 2
Evaluation of sensitivity, specificity and CHSI threshold value for half-sibship determination using 15 STR system

Population	Chinese		Caucasians		African Americans	
CHSI threshold value	[a]SEN (%)	[b]SPE (%)	[a]SEN (%)	[b]SPE (%)	[a]SEN (%)	[b]SPE (%)
0.03125	99.70	28.76	99.69	30.37	99.71	33.88
0.0625	99.22	41.46	99.21	43.18	99.24	46.85
0.125	98.23	54.96	98.12	56.63	98.21	59.59
0.25	96.34	67.87	96.10	69.42	96.44	71.35
0.5	93.04	78.70	92.62	79.83	93.16	81.31
1	87.98	86.76	87.14	87.78	88.27	88.67
2	80.74	92.48	79.58	93.09	81.50	93.48
3	75.63	94.80	74.20	95.30	76.62	95.39
10	57.19	98.49	54.92	98.67	58.83	98.69
33	37.93	99.63	34.97	99.71	39.02	99.68
100	23.14	99.92	19.99	99.96	23.37	99.91
150	18.83	99.95	15.70	99.98	18.74	99.94
200	16.19	99.97	13.05	99.99	15.86	99.96
300	12.95	99.98	9.90	100.00	12.33	99.98
330	12.25	99.98	9.23	100.00	11.59	99.98
7500	9.68	99.99	6.79	100.00	8.89	99.99
1000	6.46	100.00	3.95	100.00	5.37	99.99

[a] Sensitivity: % true half-sibs with CHSI values greater than threshold.
[b] Specificity: % random pairs with CHSI values less than threshold.

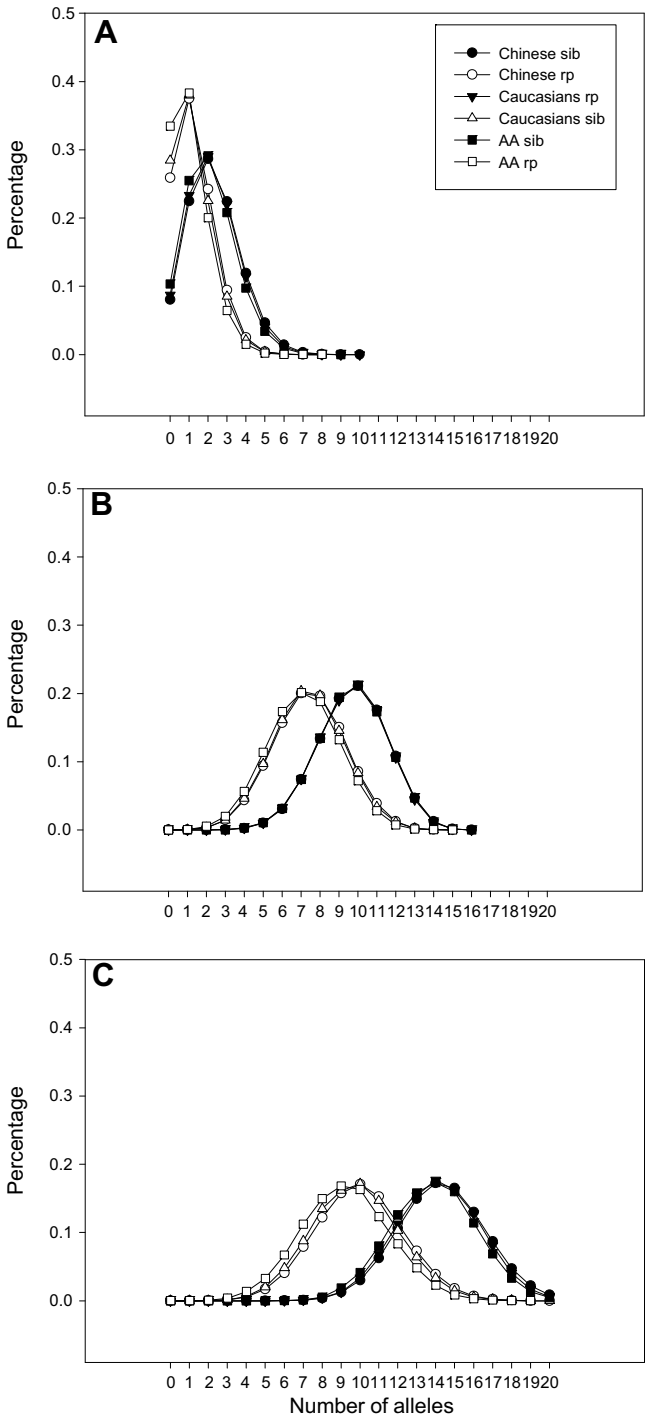


Fig. 2. Ratio distribution of allele sharing instances. A for the number of 2-allele sharing loci, B for the number of loci with one allele shared and C when all shared alleles are counted. Half siblings are denoted by hsib and random pairs by rp.

Table 3
Synergy effects of All-Shared-Alleles cutoffs and CHSI cutoffs for half-sibship determination by using 15 STR

Chinese	0.125		0.25		0.5		1		1.25		1.5		2		3		10	
ASA	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)
7	98.23	54.96	96.34	67.87	93.04	78.71	87.98	86.76	85.87	88.89	84.02	90.39	80.74	92.48	75.63	94.80	57.19	98.49
8	98.22	55.03	96.33	67.89	93.04	78.71	87.98	86.76	85.87	88.89	84.02	90.39	80.74	92.48	75.63	94.80	57.19	98.49
9	98.11	55.57	96.27	68.04	93.00	78.75	87.96	86.78	85.85	88.90	84.00	90.40	80.73	92.49	75.62	94.81	57.19	98.49
10	97.42	58.16	95.82	68.98	92.74	79.07	87.80	86.87	85.72	88.98	83.89	90.46	80.64	92.52	75.56	94.82	57.17	98.49
11	94.96	65.15	93.92	72.24	91.45	80.27	86.98	87.30	85.02	89.28	83.28	90.68	80.13	92.66	75.18	94.89	57.03	98.50
12	89.00	75.86	88.57	78.72	87.15	83.38	83.86	88.58	82.26	90.19	80.79	91.39	78.07	93.14	73.62	95.16	56.39	98.54
13	78.46	86.14	78.37	86.78	77.84	88.49	76.13	91.22	75.15	92.23	74.19	93.03	72.29	94.23	68.93	95.75	54.22	98.64
14	63.57	93.34	63.56	93.41	63.46	93.73	62.91	94.60	62.53	95.00	62.11	95.35	61.15	95.96	59.21	96.84	48.89	98.84
15	46.36	97.24	46.36	97.24	46.35	97.29	46.24	97.43	46.14	97.52	46.01	97.62	45.71	97.81	44.96	98.16	39.51	99.17
16	29.86	99.02	29.86	99.02	29.86	99.02	29.85	99.03	29.84	99.04	29.82	99.06	29.76	99.10	29.58	99.17	27.53	99.52
17	16.86	99.70	16.86	99.70	16.86	99.70	16.86	99.71	16.86	99.71	16.86	99.71	16.85	99.71	16.83	99.72	16.29	99.81
18	8.18	99.93	8.18	99.93	8.18	99.93	8.18	99.93	8.18	99.93	8.18	99.93	8.18	99.93	8.18	99.94	8.10	99.94
Caucasian																		
7	97.47	56.63	95.46	69.42	92.00	79.83	86.57	87.78	84.34	89.69	82.46	91.15	79.05	93.09	73.71	95.30	54.55	98.67
8	97.46	56.70	95.45	69.44	92.00	79.84	86.56	87.78	84.34	89.69	82.46	91.15	79.05	93.09	73.71	95.30	54.55	98.67
9	97.31	57.26	95.37	69.57	91.95	79.85	86.54	87.78	84.32	89.69	82.44	91.15	79.03	93.09	73.70	95.30	54.55	98.67
10	96.57	60.16	94.91	70.55	91.69	80.12	86.40	87.85	84.21	89.73	82.35	91.17	78.96	93.10	73.66	95.30	54.54	98.67
11	93.78	67.91	92.75	74.22	90.21	81.54	85.52	88.33	83.46	90.07	81.71	91.40	78.45	93.25	73.30	95.37	54.44	98.68
12	87.29	78.64	86.88	81.01	85.46	84.94	82.15	89.70	80.52	91.06	79.08	92.14	76.29	93.70	71.67	95.60	53.85	98.71
13	76.17	88.10	76.09	88.61	75.59	89.96	73.98	92.27	73.03	93.09	72.12	93.80	70.22	94.86	66.80	96.25	51.75	98.83
14	60.90	94.41	60.89	94.46	60.81	94.73	60.31	95.44	59.94	95.77	59.53	96.02	58.64	96.54	56.73	97.27	46.35	99.01
15	43.31	97.76	43.31	97.76	43.30	97.79	43.22	97.90	43.15	97.96	43.05	98.01	42.78	98.15	42.09	98.42	36.85	99.29
16	27.20	99.21	27.20	99.21	27.19	99.21	27.18	99.22	27.17	99.23	27.16	99.24	27.11	99.26	26.95	99.34	25.16	99.63
17	14.52	99.75	14.52	99.75	14.52	99.75	14.52	99.76	14.52	99.76	14.52	99.76	14.51	99.76	14.50	99.78	14.11	99.84
18	6.49	99.94	6.49	99.94	6.49	99.94	6.49	99.94	6.49	99.94	6.49	99.94	6.49	99.94	6.49	99.94	6.43	99.95
African Americans																		
7	97.35	59.60	95.53	71.35	92.10	81.31	87.05	88.67	85.05	90.66	83.26	91.75	80.08	93.48	75.12	95.39	57.27	98.69
8	97.32	59.74	95.52	71.38	92.09	81.31	87.05	88.67	85.04	90.66	83.26	91.75	80.08	93.48	75.12	95.39	57.27	98.69
9	97.13	60.71	95.40	71.66	92.02	81.39	87.02	88.68	85.02	90.67	83.24	91.76	80.07	93.49	75.11	95.39	57.27	98.69
10	95.82	64.72	94.43	73.27	91.42	81.90	86.66	88.81	84.73	90.74	82.97	91.80	79.86	93.52	74.97	95.40	57.23	98.69
11	92.09	73.69	91.33	77.86	89.15	83.76	85.14	89.52	83.42	91.16	81.80	92.11	78.92	93.69	74.27	95.49	56.98	98.69
12	84.38	83.73	84.14	85.04	83.05	87.63	80.56	91.20	79.28	92.41	78.03	93.13	75.73	94.36	71.79	95.89	55.96	98.73
13	71.83	91.69	71.79	91.89	71.48	92.54	70.45	93.97	69.80	94.57	69.07	94.97	67.74	95.66	65.13	96.68	52.71	98.86
14	56.13	96.48	56.13	96.49	56.07	96.55	55.78	96.86	55.56	97.06	55.30	97.18	54.72	97.44	53.47	97.88	45.72	99.09
15	38.95	98.72	38.95	98.72	38.95	98.72	38.90	98.74	38.84	98.76	38.79	98.78	38.63	98.84	38.26	98.99	34.76	99.45
16	23.19	99.55	23.19	99.55	23.19	99.55	23.18	99.55	23.18	99.55	23.16	99.56	23.14	99.56	23.07	99.59	22.05	99.70
17	11.95	99.87	11.95	99.87	11.95	99.87	11.95	99.87	11.95	99.87	11.95	99.87	11.95	99.87	11.94	99.87	11.73	99.88
18	5.18	99.97	5.18	99.97	5.18	99.97	5.18	99.97	5.18	99.97	5.18	99.97	5.18	99.97	5.18	99.97	5.16	99.97

ity was 87.14% and the specificity was 87.78% and for African American population the sensitivity was 88.27% and the specificity was 88.67%. Using a CHSI of 1 would falsely exclude more than 10% of real half-sibling, indicating the potential problems with this type of DNA testing.

3.3. Ratio distribution of allele sharing for three populations

The ratio distribution of allele-sharing for simulated half-sibling pairs and random pairs is shown in Fig. 2, three of the allele sharing situation were compared. Graphs C illustrate the ASA situation where there is greater separation between half-sibling pairs and random pairs (with smaller percentage of overlapping area in the bipolar model). These data indicate the potential value of ASA in the determination of half-sibship.

3.4. Evaluation of synergy effects of two criteria for the half-sibship determination

The CHSI vs. ASA ratio distribution was calculated for all the simulated half-siblings and random pairs, and the sensitivity and specificity for each CHSI and ASA combination was evaluated. Values of CHSI varying from 0.125 to 10 were analyzed against the values of 7–18 of ASA using the 15 STR system (Table 3). A combination of these two data sets increased the specificity in an inclusion of the relationship. A CHSI of 0.125 and ASA of 13 resulted in a specificity of 86.14%, whereas if only considering ASA the specificity was only 54.96% (Table 2) for the Taiwanese population. When CHSI = 1 and ASA = 13, the specificity increased to 91.22% instead of 86.76% for the Taiwanese population (see Table 2, when based on CHSI = 1 only), 92.27% instead of 87.78% for Caucasian population and 93.97% instead of 88.67% for African American population. Combining the two criterion increases the specificity for half-sibling determination especially for the cases of "low CHSI with high ASA", otherwise they are maybe denied as half-siblings by using CHSI cutoffs only.

There is alternative of resolving half-sibship cases with greater confidence by using more autosomal STR loci, using mitochondrial DNA analysis and using STR loci on either of the sex determining chromosomes.

4. Conclusion

The use of CHSI and all-shared-alleles together increases the specificity with a resulting increase in the confidence in half-sibship determination for low CHSI cases.

Acknowledgements

The authors wish to thank Wu F.C., Wu G.C., Cheng M.Y., and Meng L.M. for establishing part of the STR profile database.

References

1. Allen RW, Fu J, Reid TM, Baird M. Considerations for the interpretation of STR results in cases of questioned half-sibship. *Transfusion* 2007;**47**:515–9.
2. <http://www.cbsnews.com/stories/2006/03/17/60minutes/main1414965.shtml> [accessed 1.06.07].
3. Wenk RE, Traver M, Chiafari FA. Determination of sibship in any two persons. *Transfusion* 1996;**36**:259–62.
4. Wenk RE, Chiafari FA. Distinguishing full siblings from half-siblings in limited pedigrees. *Transfusion* 2000;**40**:1148–9.
5. Gaytmenn R, Hildebrand DP, Sweet D, Pretty IA. Determination of the sensitivity and specificity of sibship calculations using AmpFISTR Profiler Plus. *Int J Legal Med* 2002;**116**:161–4.
6. Short Tandom Repeat DNA Internet Database. <http://www.cstl.nist.gov/div831/strbase/>.
7. Pu CE, Wu FC, Cheng CL, Wu KC, Chao CH, Li JM. DNA STR profiling of Chinese population in Taiwan determined by using an automated sequencer. *Forensic Sci Int* 1998;**97**:47–51.
8. Pu CE et al. *DNA profiling of Chinese population in Taiwan on the STR loci D8S1179, D16S539, D18S51 and D21S11 52th annual meeting*. Reno (NV, USA): American Academy of Forensic Sciences; 2000.
9. Lee JC, Chang YY, Su CW, Tzeng CH, Pu CE, Han TH, et al. The screening of STR and YSTR loci in Taiwanese han population. In: Proceedings of Taiwan academy of forensic sciences; 2006. p. 27–38 [in Chinese].
10. Butler JM. Kinship and parentage testing. In: Butler JM, editor. *Forensic DNA typing*. MA, USA: Elsevier Academic Press; 2005. p. 477.
11. Kirkwood BR, Sterne JAC. Measurement error: assessment and implications. In: Kirkwood BR, Sterne JAC, editors. *Medical statistics*. MA, USA: Blackwell Publishing; 2003. p. 429–46.

Available online at www.sciencedirect.com
ScienceDirect
Journal of Forensic and Legal Medicine 15 (2008) 329–334
JOURNAL OF FORENSIC AND LEGAL MEDICINE
www.elsevier.com/jflm

Original Communication

Systematic evaluation of sensitivity and specificity of sibship determination by using 15 STR loci

Chang En Pu MS (Scientist Specialist and Section Chief) [a,*], Adrian Linacre PhD (Senior Lecturer) [b]

[a] *Forensic DNA Section, Scientific and Technical Research Center, Ministry Justice Investigation Bureau, No. 74 Chun_hya Road, Hsin-tien, Taipei County 231, Taiwan*
[b] *Centre for Forensic Science, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, UK*

Received 17 June 2007; received in revised form 30 August 2007; accepted 31 December 2007
Available online 11 April 2008

Abstract

Paternity disputes and other forms of kinship testing are routinely resolved using short tandem repeat (STR) DNA loci. Sibship determination is encountered in instances where the DNA profiles of two individuals are compared to determine if they are siblings. If either parent is available for testing then the situation is simplified but if neither parent of the two individuals is available for DNA testing, a combined sibling indices (CSI) for the determination of sibship between two people can be determined. Support for kinship is also based upon the sharing of alleles, particularly when both alleles are shared at the same locus, termed two-allele-sharing-loci (TASL). We report on the combination of CSI and TASL to enhance the determination of sibship. The 15 STR loci that comprise the Identifiler® loci were applied to three populations using pairs of full siblings or unrelated pairs. Based upon the data obtained, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) can be applied to determining whether two DNA profiles come from full or non-sibling pairs. This report highlights the problems inherent in this form of kinship testing and recommends a combination use of CSI and TASL for sibship determination.

Keywords: Short tandem repeat (STR); Kinship; Sibship; Combined sibship index (CSI); Sensitivity; Specificity

1. Introduction

Sibship determination is encountered in instances such as linking human remains to a relative and when neither of the biological parents of the two individuals is available for testing. Theoretically at any one locus there is a 0.5 chance that two siblings will share one allele, a 0.25 chance that they will share neither allele and a 0.25 chance that they will share both alleles. The chance that two unrelated individuals share either one or both alleles at any one locus is dependent upon the frequency of the alleles.[1,2] A probability of sibship can be determined based upon the frequency of the matching alleles in the population and will increase when a high number of high discriminating loci are examined.[3] Uncertainty using DNA testing to resolve sibship increases if the parents were heterozygous rather than homozygous.[1,4] There is no evidence of sibship if there is no two-allele-sharing-locus (TASL) between the two profiles,[5] however confidence is increased if a number of TASL existed.

Based upon the degree of sharing of alleles between two DNA profiles it is possible to determine a combined sibship indices (CSI).[1] When the index is less than 1 the two individuals might not be related as siblings. If the index is over 1 then the data supports the existence of a sibling relationship. Other cut-off point have been recommended such a CSI $\geqslant 3$.[2] These figures are guides as it was found that 1.6% of random pairs of DNA profiles had CSI greater than 1[2] when using 15 STR loci. In a study using 16 STR

* Corresponding author. Tel.: +886 2 29112241x3740; fax: +886 2 291 38599.
E-mail address: pu_macros@yahoo.com.tw (C.E. Pu).

1752-928X/$ - see front matter © 2008 Elsevier Ltd and FFLM. All rights reserved.
doi:10.1016/j.jflm.2007.12.018

loci 0.1% of unrelated pairs of DNA profiles were found to have CSI > 100 and 0.01% for CSI > 1000.[6] Using the 15 STR loci used in the Identifiler® kit, none of the non-sibling pairs were found with CSI ≥ 1, while all sibling pairs have CSI > 10.[7] In a different study using the same loci, 6.06% of sibling pairs exhibited CSI < 1 and 9.1% of random unrelated pairs had CSI > 1.[8] The variation of percentage of random pairs with CSI > 1 found above maybe owing to the different levels of consanguinity that might be present in those populations.

We have extended the studies using the 15 STR comprising the Identifiler® loci to study three populations using 357,630 full sibling pairs and 178,815 non-sibling pairs generated from DNA profiles of random population. Using this high number of sample pairs, it is possible to evaluate the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the 15 STR loci for discriminating between full and non-siblings. The relationship between the CSI values and TASL is reported.

2. Materials and methods

DNA profiles from a Chinese population were obtained using the ABI AmpFISTR Identifiler® PCR Amplification Kit (Applied Biosystem, Foster City, CA, USA). The STR products were analyzed with an ABI Prism 3100 XL Genetic Analyzer.

STR profiles (450) of random members of the Taiwanese Chinese population were processed by Microsoft Excel Macros controlled by Visual Basic program written by authors of this study. Every member of the population was paired with every other member to form random pairs, e.g. for Chinese population in this study (450 × 449)/2, equaling 101,025 pairs, were made. Every pair was set to have two children, resulting in 202,050 sibling pairs being generated. DNA profiles from the 15 STR loci from a Caucasian ($n = 301$) and American African ($n = 256$) population were obtained from short tandem repeat DNA internet database.[9] These data were processed in the same way as those of the Taiwanese population to generate sibling pairs and random pairs. Inevitably for computer based populations no account of substructure is made and mating occurs randomly. The combined sibship indices (CSI) were calculated for each simulated sibling pairs or random pairs by using standard formulae,[1] and allele frequency tables used for calculation of CSI were adjusted by using 5/2N rule.[10]

The rate of false negatives equaled the percentage of real sibship testing cases (in this study the simulated sibling pairs) that would be excluded based upon any given cut-off point of CSI or TASL. The rate of false positives equaled the percentage of random pairs of DNA profiles where their CSI or TASL was greater than any recommended cut-off value. The sensitivity of the test is based upon 1 – the % of false negatives, the specificity of the test is based upon 1 – the % of false positives, the positive predictive value (PPV) = the proportion of subjects correctly identified as siblings and the negative predictive value

Table 1
Maximum and minimum CSI for three populations using 15 STR systems

Population	Profiles	Type	Pairs	Maximum	Minimum
Chinese	450	Simulated siblings	202,050	4.06E+16	1.68E−05
		Random pairs	101,025	4.79E+04	2.33E−09
Caucasians	301	Simulated siblings	90,300	1.23E+14	3.42E−04
		Random pairs	45,150	1.52E+04	6.38E−09
African Americans	256	Simulated siblings	65,280	3.54E+14	9.52E−04
		Random pairs	32,640	6.24E+08	3.43E−09

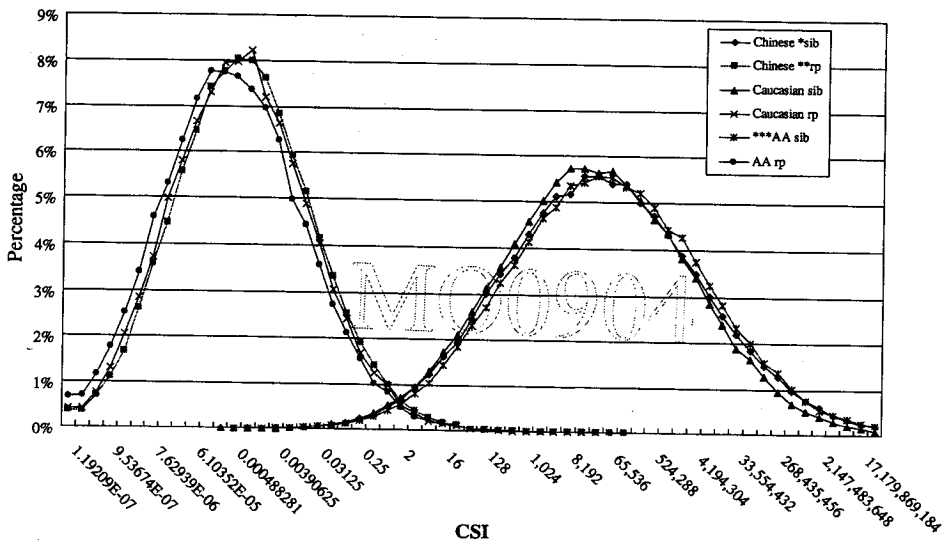


Fig. 1. Ratio distribution of CSI for three populations (*sib: siblings; **rp: random pairs; ***AA: African Americans).

Table 2
Evaluation of sensitivity, specificity, PPV, NPV and optimal CSI cut-off points for sibship determination using 15 STR system

Population	Chinese						Caucasians						African Americans					
CSI	[a]SEN (%)	[b]SPE (%)	[c]Abs-*X*	[d]Sqrt-*Y*	PPV (%)	NPV (%)	SEN (%)	SPE (%)	Abs-*X*	Sqrt-*Y*	PPV (%)	NPV (%)	SEN (%)	SPE (%)	Abs-*X*	Sqrt-*Y*	PPV (%)	NPV (%)
0.03125	99.849	87.905	0.1194	1.3303	89.195	99.828	99.868	88.977	0.1089	1.3376	90.059	99.852	99.859	90.291	0.0957	1.3463	91.139	99.844
0.067	99.747	91.538	0.0821	1.3538	92.180	99.724	99.750	92.330	0.0742	1.3592	92.860	99.730	99.755	93.269	0.0649	1.3657	93.679	99.738
0.125	99.615	93.795	0.0582	1.3682	94.136	99.592	99.590	94.454	0.0514	1.3726	94.725	99.568	99.619	95.150	0.0447	1.3776	95.358	99.601
0.25	99.364	95.713	0.0365	1.3796	95.864	99.339	99.338	96.146	0.0319	1.3825	96.265	99.316	99.403	96.713	0.0269	1.3869	96.799	99.386
0.5	99.012	97.131	0.0188	1.3870	97.184	98.993	98.952	97.420	0.0153	1.3886	97.459	98.936	99.096	97.736	0.0136	1.3918	97.766	99.084
1	98.500	98.125	0.0037	1.3904	98.132	98.494	98.392	98.410	0.0002	1.3916	98.409	98.392	98.660	98.585	0.0008	1.3947	98.586	98.659
2	97.816	98.788	0.0097	1.3902	98.777	97.837	97.659	98.966	0.0131	1.3904	98.952	97.689	98.068	99.093	0.0102	1.3942	99.084	98.088
3	97.298	99.050	0.0175	1.3884	99.033	97.345	97.174	99.207	0.0203	1.3887	99.191	97.230	97.603	99.277	0.0167	1.3922	99.265	97.642
10	95.151	99.613	0.0446	1.3776	99.595	95.358	94.876	99.670	0.0479	1.3761	99.653	95.110	95.760	99.691	0.0393	1.3823	99.678	95.920
10.3	95.074	99.624	0.0455	1.3771	99.606	95.288	94.791	99.672	0.0488	1.3755	99.655	95.033	95.695	99.694	0.0400	1.3819	99.681	95.861
33	91.953	99.852	0.0790	1.3574	99.839	92.542	91.466	99.880	0.0841	1.3543	99.869	92.129	92.785	99.893	0.0711	1.3634	99.885	93.264
100	87.735	99.935	0.1220	1.3298	99.926	89.068	87.009	99.949	0.1294	1.3252	99.941	88.497	88.919	99.942	0.1102	1.3377	99.935	90.019
150	85.825	99.953	0.1413	1.3174	99.946	87.580	85.012	99.969	0.1496	1.3123	99.964	86.962	87.109	99.966	0.1286	1.3259	99.961	88.578
200	84.429	99.958	0.1553	1.3084	99.951	86.522	83.627	99.978	0.1635	1.3034	99.974	85.928	85.846	99.972	0.1413	1.3177	99.968	87.598
300	82.331	99.969	0.1764	1.2951	99.963	84.980	81.349	99.980	0.1863	1.2889	99.976	84.278	83.791	99.975	0.1618	1.3045	99.971	86.049
330	81.818	99.971	0.1815	1.2918	99.965	84.611	80.788	99.984	0.1920	1.2854	99.981	83.882	83.346	99.975	0.1663	1.3016	99.971	85.720
500	79.492	99.976	0.2048	1.2773	99.970	82.979	78.303	99.987	0.2168	1.2700	99.983	82.170	81.150	99.975	0.1883	1.2877	99.970	84.137
1000	75.218	99.984	0.2477	1.2512	99.979	80.137	73.754	99.989	0.2623	1.2425	99.985	79.209	77.004	99.982	0.2298	1.2620	99.976	81.300
Minimum	75.218	87.905	0.0037	–	89.195	80.137	73.754	88.977	0.0002	–	90.059	79.209	77.004	90.291	0.0008	–	91.139	81.300
Maximum	99.849	99.984	–	1.3904	99.979	99.828	99.868	99.989	–	1.3916	99.985	99.852	99.859	99.982	–	1.3947	99.976	99.844

[a] SEN: sensitivity.
[b] SPE: specificity.
[c] Abs-*X*: Abs(SEN − SPE).
[d] Sqrt-*Y*: Sqrt($SEN^2 + SPE^2$).

(NPV) = the proportion of subjects correctly identified as non-siblings.[4]

The optimum CSI cut-off points were obtained by minimizing, using sensitivity − specificity,[11] and maximizing by using the square root of (sensitivity^2 + specificity^2).[12]

3. Results and discussion

3.1 Ratio distribution of CSI for three populations

The CSI ratio distribution of simulated sibling pairs and random pairs are depicted in Fig. 1. Bipolar models with a widespread ratio distribution were found for all of the three populations. From the data obtained, 1.500%, 1.608% and 1.340% of CSI values from simulated sibling pairs were found to be less than 1 for Chinese, Caucasians and African Americans populations, respectively, and 1.875%, 1.590% and 1.415% of random pairs were found to have CSI larger than 1 (Table 2). Simulated siblings pairs with very low CSI values (1.68E − 05 for Chinese population) and random pairs with very high CSI values (6.24E + 08 for African American population) were observed in this study. Other extreme CSI values were also found in Table 1. While it is never possible to be definitive on sibship, a large uncertainty value can exist. It was proposed in[8] that when two DNA profiles produce CSI values between 0.67 and 10 further loci should be analyzed. In this present study the lowest CSI value for a sibling pair was significantly less than 0.067 and CSI for random pairs was much larger than 10.3. Our data indicates that simply applying a minimum CSI requirement may result in false exclusions if set too high or false inclusions if too low.

3.2 Sensitivity, specificity, PPV and NPV under different CSI cut-offs

The ability of the DNA test to correctly classify kinship testing results into two categories (sibling or not sibling) is assessed by specificity and sensitivity.[13] In Table 2 the sensitivity and specificity using a range of CSI values is illustrated. CSI cut-off values at 0.067, 3 and 10.3 are used as they follow previous recommendations.[2,4,8,14] As the CSI cut-off values increased there was a corresponding decrease in sensitivity and increase in specificity. When adopting a

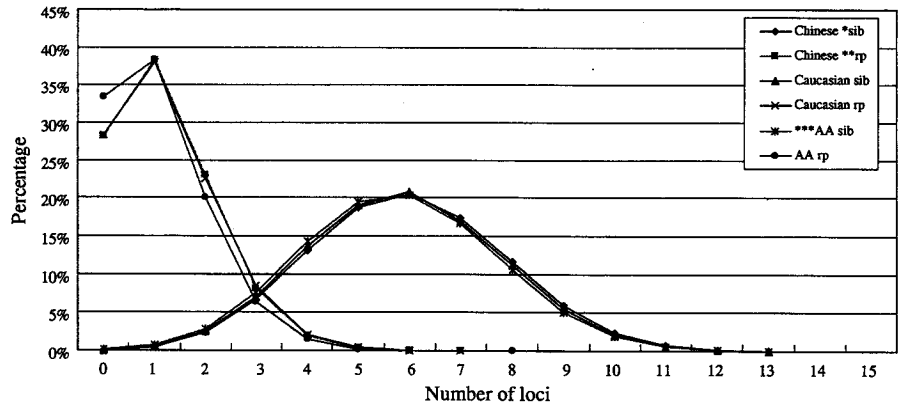


Fig. 2. Ratio distribution of TASL for three populations (*sib: siblings; **rp: random pairs; ***AA: African Americans).

Table 3
Sensitivity, specificity, PPV and NPV versus two allele sharing cut-off points

Population	Chinese				Caucasians				African Americans			
Loci	SEN (%)	SPE (%)	PPV (%)	NPV (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)
1	99.952	28.147	58.178	99.831	99.956	28.430	58.274	99.837	99.917	33.447	60.021	99.753
2	99.445	66.459	74.778	99.171	99.419	66.412	74.747	99.129	99.226	71.752	77.841	98.933
3	97.075	89.484	90.226	96.835	96.904	88.913	89.733	96.632	96.437	91.801	92.165	96.264
4	90.414	97.689	97.507	91.064	89.889	97.415	97.205	90.595	88.932	98.254	98.074	89.876
5	77.376	99.662	99.566	81.499	76.230	99.526	99.382	80.720	74.710	99.752	99.669	79.775
6	58.641	99.957	99.927	70.733	57.224	99.940	99.896	70.026	55.277	99.963	99.934	69.090
7	38.132	99.997	99.992	61.778	36.412	99.996	99.988	61.127	34.910	99.994	99.982	60.571
8	20.828	100.000	100.000	55.812	19.542	100.000	100.000	55.414	18.284	99.997	99.983	55.030
9	9.295	100.000	100.000	52.437	8.475	100.000	100.000	52.212	7.736	100.000	100.000	52.012
10	3.284	100.000	100.000	50.835	2.907	100.000	100.000	50.737	2.627	100.000	100.000	50.666
11	0.879	100.000	100.000	50.221	0.788	100.000	100.000	50.197	0.700	100.000	100.000	50.176
12	0.169	100.000	100.000	50.042	0.155	100.000	100.000	50.038	0.146	100.000	100.000	50.036
13	0.025	100.000	100.000	50.006	0.025	100.000	100.000	50.006	0.023	100.000	100.000	50.006
14	0.003	100.000	100.000	50.001	0.002	100.000	100.000	50.000	0.000	100.000	100.000	50.000
15	0.000	100.000	100.000	50.000	0.000	100.000	100.000	50.000	0.000	100.000	100.000	50.000

Table 4
Synergy effects of TASL cut-offs and CSI cut-offs for sibship determination

TASL	CSI = 0.125		CSI = 0.5		CSI = 1		CSI = 2		CSI = 3		CSI = 10		CSI = 33		CSI = 100	
	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)	SEN (%)	SPE (%)
Chinese																
1	99.590	93.943	98.997	97.173	98.488	98.150	97.807	98.797	97.290	99.054	95.147	99.614	91.952	99.852	87.734	99.935
2	99.180	94.862	98.659	97.443	98.191	98.293	97.559	98.883	97.073	99.113	95.004	99.632	91.872	99.854	87.686	99.936
3	96.949	96.883	96.633	98.238	96.319	98.767	95.862	99.145	95.489	99.307	93.801	99.685	91.028	99.872	87.114	99.941
4	90.389	98.777	90.292	99.205	90.156	99.401	89.940	99.559	89.764	99.632	88.810	99.817	86.944	99.912	83.996	99.957
5	77.372	99.727	77.358	99.789	77.331	99.824	77.281	99.864	77.229	99.885	76.909	99.925	76.153	99.963	74.670	99.978
6	58.641	99.959	58.640	99.964	58.638	99.966	58.630	99.970	58.624	99.975	58.567	99.981	58.403	99.991	58.009	99.994
7	38.132	99.997	38.132	99.997	38.132	99.997	38.131	99.998	38.131	99.998	38.127	99.998	38.113	99.998	38.057	99.999
8	20.828	100.000	20.828	100.000	20.828	100.000	20.828	100.000	20.828	100.000	20.828	100.000	20.827	100.000	20.823	100.000
Caucasian																
1	99.568	94.534	98.937	97.429	98.382	98.412	97.652	98.968	97.167	99.209	94.876	99.670	91.467	99.880	87.010	99.949
2	99.124	95.178	98.585	97.635	98.083	98.481	97.408	99.008	96.947	99.231	94.735	99.670	91.396	99.880	86.966	99.949
3	96.776	96.746	96.444	98.202	96.136	98.802	95.636	99.181	95.267	99.342	93.483	99.697	90.493	99.887	86.390	99.949
4	89.859	98.587	89.736	99.092	89.607	99.333	89.377	99.508	89.200	99.579	88.208	99.772	86.274	99.903	83.208	99.953
5	76.226	99.626	76.208	99.743	76.184	99.783	76.143	99.812	76.105	99.825	75.790	99.900	75.008	99.951	73.553	99.969
6	57.223	99.942	57.220	99.945	57.220	99.949	57.215	99.956	57.207	99.958	57.168	99.971	56.980	99.984	56.564	99.984
7	36.411	99.996	36.411	99.996	36.411	99.996	36.411	99.996	36.411	99.996	36.410	99.996	36.394	99.996	36.347	99.996
8	19.542	100.000	19.542	100.000	19.542	100.000	19.542	100.000	19.542	100.000	19.542	100.000	19.542	100.000	19.538	100.000
African Americans																
1	99.557	95.245	99.047	97.770	98.617	98.603	98.032	99.102	97.575	99.283	95.749	99.691	92.783	99.893	88.919	99.942
2	98.984	95.938	98.572	97.953	98.197	98.695	97.670	99.145	97.255	99.305	95.533	99.697	92.644	99.893	88.831	99.942
3	96.345	97.549	96.132	98.594	95.905	99.047	95.538	99.354	95.250	99.452	93.940	99.740	91.481	99.905	88.047	99.945
4	88.912	99.059	88.848	99.369	88.787	99.550	88.623	99.672	88.494	99.709	87.819	99.850	86.345	99.942	84.000	99.963
5	74.710	99.804	74.701	99.828	74.698	99.856	74.660	99.893	74.635	99.905	74.429	99.942	73.938	99.966	72.895	99.975
6	55.277	99.966	55.276	99.966	55.276	99.966	55.274	99.969	55.271	99.969	55.242	99.979	55.155	99.982	54.914	99.985
7	34.910	99.994	34.910	99.994	34.910	99.994	34.910	99.994	34.910	99.994	34.907	99.994	34.903	99.994	34.874	99.994
8	18.284	99.997	18.284	99.997	18.284	99.997	18.284	99.997	18.284	99.997	18.284	99.997	18.284	99.997	18.284	99.997

simple CSI value of 1, below which indicates non-sibling and above supports a sibling pair, then for the three populations the sensitivity was 98.500% and the specificity was 98.125% for the Chinese population, for Caucasian population the sensitivity was 98.392% and the specificity was 98.410% and for African American population the sensitivity was 98.660% and the specificity was 98.585%. The PPV and NPV were all over 98% for the three populations.

3.3 Ratio distribution, sensitivity and specificity of TASL

The TASL ratio of simulated sibling pairs and random pairs is depicted in Fig. 2. Bipolar models were found for all of the three populations.

The sensitivity and specificity for 1–15 of TASL are illustrated in Table 3. As the number of cut-off loci increased there was a corresponding decrease in sensitivity and increase in specificity. At the TASL cut-off $\geqslant 5$, the specificity was greater than 99% for all of the populations, but the sensitivity was only around 75%. Based upon these data, TASL alone is not an optimal screening method for the sibship determination. A previous instance was reported where no TASL were found for a pair of DNA profiles[5] leading to the exclusion of sibship; based upon our data (Table 3) the NPV was 99.837% (for Caucasians) regardless of CSI value. The use of a TASL value of 1 as the cut-off only in place of CSI calculating was found to be more sensitive and easy for sibship screen of mass comparing, since the sensitivity was 99.95% compared to a sensitivity value of 98.500% for a CSI of 1 for Chinese, and the comparing could done without the help of computer. When pairs were screened then CSI calculation applied afterward and Table 4 could be used to determine the specificity.

3.4 Evaluation of synergy effects of two criteria for the sibship determination

The sensitivity and specificity of both the CSI and TASL values were combined. Values of CSI varying from 0.125 to 100 were analyzed against the values of 1–9 for TASL (Table 4). A combination of these two data sets increased the confidence of in an exclusion or inclusion. A CSI of 0.125 and TASL of five resulted in a specificity of 99.727% for Chinese population, 99.626% for the Caucasian population and 99.804% for African American population. Instances of medium CSI values were reflected as the true specificity, e.g. when CSI = 3 and TASL = 5, resulted in the specificity increasing to 99.885% instead of 99.050% (see Table 2, when based on CSI = 3 only), 99.825% instead of 99.207% for Caucasian population and 99.905% instead of 99.277% for African American population.

The data obtained in this study indicates that values of CSI and TASL to be adopted as the cut-offs may vary from population to population. The values used may also vary if the DNA test is used for identification of human remains or in criminal cases if the burden of proof required also varies. The ideal situation to resolve sibship cases with greater confidence would be to use more autosomal STR loci if possible or to use mitochondrial DNA analysis or STR loci on either of the sex determining chromosomes.

4. Conclusion

The 15 loci STR core set system used in this study could not distinguish all 357,630 full sibling pairs from 178,815 non-sibling pairs, as a small percentage of false positives and false negatives were found. The use of CSI and TASL together increased the confidence in sibship determination. Using TASL = 1 for the first-step-screening for disaster cases was also proposed.

Acknowledgments

The authors wish to thank Wu F.C., Wu G.C., Cheng M.Y., Meng L.M. and Cheng S.M. for their kindly help for genotyping all the local samples.

References

1. Wenk RE, Traver M, Chiafari FA. Determination of sibship in any two persons. *Transfusion* 1996;**36**:259–62.
2. Tzeng CH, Lyou JY, Chen YR, Hu HY, Lin JS, Wang SY, et al. Determination of sibship by PCR-amplified short tandem repeat analysis in Taiwan. *Transfusion* 2000;**40**:840–5.
3. Wenk RE, Chiafari FA. Distinguishing full siblings from half-siblings in limited pedigrees. *Transfusion* 2000;**40**:1148–9.
4. Gaytmenn R, Hildebrand DP, Sweet D, Pretty IA. Determination of the sensitivity and specificity of sibship calculations using AmpFISTR profiler plus. *Int J Legal Med* 2002;**116**:161–4.
5. Henke J, Henke L. Which short tandem repeat polymorphisms are required for identification? Lessons from complicated kinship cases. *Croat Med J* 2005;**46**(4):593–7.
6. Thomson JA, Ayres KL, Pilotti V, Barrett MN, Walker JIH, Debenham PG. Analysis of disputed single-parent/child and sibling relationships using 16 STR loci. *Int J Legal Med* 2000;**115**:128–34.
7. Reid TM, Wolf CA, Kraemer CM, Lee SC, Baird ML, Lee RF. Specificity of sibship determination using the ABI identifiler multiplex system. *J Forensic Sci* 2004;**49**:1–3.
8. Giroti RI, Verma S, Singh K, Malik R, Talwar I. A gray zone approach for evaluation of 15 short tandem repeat loci in sibship analysis: a pilot study in Indian subjects. *J Clin Forensic Med.* doi:10.1016/j.jcfm.2006.08.002.
9. *Short tandom repeat DNA internet database.* http://www.cstl.nist.gov/div831/strbase.
10. Butler JM. Kinship and parentage testing. In: Butler JM, editor. *Forensic DNA typing.* MA, USA: Elsevier Academic Press; 2005. p. 529–37.
11. *Find 'best' cut-offs for maximizing sensitivity/specificity.* Available from: http://www.natureserve.org/visitLocal/conftraining/edmworkshop/rochelp.pdf. 2007.
12. Zou KH, Wells III WM, Kikinis R, Warfield SK. Three validation metrics for automated probabilistic image segmentation of brain tumors. *Statist Med* 2004;**23**:1259–82.
13. Kirkwood BR, Sterne JAC. Measurement error: assessment and implications. In: Kirkwood BR, Sterne JAC, editors. *Medical statistics.* MA, USA: Blackwell Publishing; 2003. p. 429–46.
14. AABB, *2002 parentage testing annual report.* MD, USA: American Association of Blood Banks; 2007. http://www.aabb.org/Documents/Accreditation/Parentage_Testing_Accreditation_Program/ptannrpt02.pdf.

CPI Distribution and Cut-Off Values for Duo Kinship Testing

Chang-En Pu[1] and Adrian Linacre[2]

[1]*Scientific and Technical Research Center*
Ministry of Justice, Investigation Bureau
Taipei, Taiwan, R.O.C.
[2]*Center for Forensic Science*
Department of Pure and Applied Chemistry
University of Strathclyde, Glasgow, UK.

Abstract

DNA-based tests commonly use 13 STR (short tandem repeat) loci in human identification and paternity testing – the Combined DNA Index System or CODIS. Its average degree of accuracy of paternity identification is greater than 0.9999 under the circumstance of a mother, a child and a putative father. However, the possibility of false inclusions increases under circumstances such as [1] only two members of a family group are available – a duo case during determination of paternity or [2] identification of human remains while only one living relative is present.

In Taiwan, the National Unidentified Human Remains Database uses the CODIS 13 STR for the identification of family members. Two or more reference samples in the DNA database have been found to share one allele at all loci tested. Then the Combined Paternity Index (CPI) is used to determine and provide an estimate of kinship in such cases. Combining 499,500 sets of DNA data for the 13 STR CODIS loci, totally 431 (0.086%) cases are false inclusions where all 13 loci shared at least one allele. Simulated partial DNA profiles (not all 13 loci yielded results) were created to mimic the mutation and degradation process. All 431 real duo cases were analyzed to evaluate sensitivity and specificity. This report provided four kinship-matching situations with CPI cutoff values when the number of allele-sharing loci exceeded 11. CPI values greater or lesser than the suggested cutoff point will provide a greater degree of confidence in determining whether two samples are derived from first-degree relatives.

Key Words: forensic science, short tandem repeat (STRs), paternity, false inclusion duo, sensitivity, specificity

Introduction

The Combined DNA Index System (CODIS), originally established by the FBI in 1998, is composed of 13 autosomal short tandem repeat (STR) loci in addition to a sexual (gender) identity test (6). These loci have become the standard examination for human identification and paternity testing in many countries throughout the world. The 13 STR loci used within the CODIS have an average power of paternity exclusion greater than 0.9999 for most populations based upon mother, child and father combinations, a paternity trio case (2, 9). A duo paternity case is defined as the availability of only two sets of genetic relatives, such as one parent and one offspring. Such duo cases occur in paternity testings, *e.g.* either when the mother is unavailable to provide a sample, or in cases of identification of human remains by linking its DNA to a living relative only. In such duo cases, if allele sharing is found in all of the 13 loci, the probability of parentage, known as the Combined Paternity Index (CPI), is calculated. Even if high CPI values are obtained, there is still a chance that two unrelated people will share one allele at all 13 CODIS loci resulting in a false inclusion (1, 3). In Taiwan, the

Corresponding author: Chang-En Pu, Section Chief of the Forensic DNA section, Scientific and Technical Research Center, Ministry Justice Investigation Bureau, Taipei County 231, Taiwan, R.O.C. Tel: +886-2-29112241 ext. 3740, Fax: +886-2-29138599, E-mail: pu_macros@yahoo.com.tw
Received: August 4, 2006; Revised (Final): April 6, 2007; Accepted: April 10, 2007.

CODIS 13 STRs used during determination of kinship (first-degree) match the National Unidentified Human Remains Database. There have been instances where a DNA profile matched an unrelated individual along with a known member of the family on this reference database. False inclusions such as this have been found in previous studies (4, 5, 8, 10). If fewer than 13 CODIS loci are used, then the chances of a false inclusion will increase. Fewer loci are generated if less than the optimal amount of DNA is present, such as in degraded samples. Such a situation occurs commonly in the isolation of DNA from human skeletal remains.

At any single locus the Paternity Index (PI) can be calculated between two samples that share at least one allele. The PI for each locus tested can be multiplied to generate the combined PI (CPI), thus increasing the odds in favor of the person being a parent of a child compared to other unrelated individuals. The probability of paternity can be determined from the CPI by a standard method[1]. For many laboratories, a minimal CPI value is used; increased values indicate a high degree of confidence that the samples tested are genetically linked as first degree relatives and not that of a false inclusion.

In a recent study conducted by 34 laboratories for the American Association of Blood Banks (AABB), the minimum CPI value that was required for determining a Mother not Tested (MNT) case, which is the same as a duo case, varied from "Whatever is obtained" (2 of the 34) to 10,000 (1 of the 34). In 23 laboratories, the minimal CPI value was 100[2]. The choice of an appropriate minimal CPI value will affect the sensitivity (1-rate of false negatives) and specificity (1-rate of false positives) of the DNA test. There is an inverse relationship between these two measures, when by setting cutoffs their capabilities can be adjusted, but increasing the capability to improve one will have the effect of decreasing the capability of the other (7) We report on a study to determine a CPI value that can be used in duo paternity cases that will minimize the occurrence of a false inclusion yet not exclude real first degree relatives.

Query 1: Explain what is 1-rate

Materials and Methods

From January 2003 to December 2005, 431 real paternity cases provided by the Science and Technical Research Center for the Investigation Bureau of the Ministry of Justice, Taiwan were analyzed. The Combined Paternity Index (CPI) using the CODIS 13 STR from real paternity duo cases was determined by a standard formula (please see footnote 2 in page 3).

To simulate degraded DNA, either one or two loci from the 13 loci were erased randomly to create 12 loci and 11 loci matches, respectively. To simulate mutational events duo pairs matching, all 13 loci were used and a non-match was created at one locus. This resulted in a match at 12 loci and 1 mutation at the 13th loci (or 12mut/13). CPI values were modified by incorporating the rate of mutation into the probability of paternity calculation (please see footnote 1 in page 3).

One thousand members of the Chinese population in Taiwan were randomly selected and processed by Microsoft Excel Macros using the Visual Basic program. Every member of the population was paired with every other member, *e.g.* (1,000 × 999)/2 equaling 499,500 pairs. CPI values were obtained for pairs where at least one allele was shared at all 13 loci. CPI values were obtained from mimicked degraded samples with 12 allele sharing out of 12 loci and 11 alleles shared from 11 loci. CPI values were also obtained from mimicked 12mut/13 cases. When processing pairing by computer, no substructure of the population was considered.

The rate of false negatives which equaled the percentage of real paternity cases would be excluded for any given cut off point. The rate of false positives equaled the percentage of co-incidental matches above any given cut off point. The sensitivity of the test is based upon 1 - the % of false negatives. The specificity of the test is based upon 1 - the % of false positives.

The optimum CPI cutoff point was obtained by choosing the maximum of square root of sensitivity^2 + specificity^2 (11).

Results

The Smallest and Largest CPI Values Found

Using the CODIS 13 STR loci for paternity test, real paternity duo cases with very low CPI values (20.81) and coincidental kinship-matched pair with very high CPI values (47,042.98) were observed in this study (Table 1).

Number of Duos and Coincidental Pairs that Meet the Different CPI Requirements

A total of 431 coincidental matches were observed with at least one matching allele at all 13 loci for the 499,500 computer-generated pairs of the random population. When only 12 loci were used, such as in the degraded samples, there were 4,286 coincidental matches out of the 499,500 pairs observed, while

[1] AABB, 2002 Parentage testing annual report. MD, USA, American association of blood banks,°@2002. [Available at http://www.aabb.org /Documents/ Accreditation/ Parentage_Testing_Accreditation_Program/ptannrpt02.pdf. (Accessed March 2007).

[2] AABB, Guidance for standards for relationship testing laboratories, MD, USA, American association of blood banks, 2006, pp 139.

Table 1. The smallest and largest CPI values resulted from real duo paternity cases and coincidental matched duos under CODIS 13 STRs systems

Real Paternity Duo						
Locus	Child	Father	PI	Child	Father	PI
STR D3S1358	16,18	16,16	1.63	15,17	15,15	1.40
STR vWA	16,18	18,19	1.36	19,19	14,19	5.29
STR FGA	22,23	19,22	1.39	24,24	19,24	2.78
STR TH01	9,9	6,9	1.09	7,9	7,9	1.43
STR TPOX	8,11	8,9	0.43	8,8	8,8	1.73
STR CSF1PO	10,13	10,12	0.95	12,12	10,12	1.50
STR D5S818	10,12	11,12	1.13	10,11	10,10	2.62
STR D13S317	8,12	8,13	0.89	10,12	11,12	1.65
STR D7S820	11,12	12,12	2.18	9,13	8,13	7.25
STR D8S1179	11,15	11,12	2.43	13,14	14,14	2.58
STR D21S11	29,33.2	29,31	0.93	29,33	30,33	50.00
STR D18S51	13,16	13,14	1.60	15,19	15,19	8.05
STR D16S539	9,11	9,11	1.89	10,11	10,10	3.95
		Least CPI= 20.81			Largest CPI= 9,756,809.81	
Coincidental Match (False Positive might be established)						
Locus	Ind. a	Ind. b	PI	Ind. c	Ind. d	PI
STR D3S1358	15,17	15,16	0.70	16,16	16,16	3.27
STR vWA	14,18	18,18	2.71	16,18	16,16	3.16
STR FGA	22,23	20,22	1.39	20,23	23,23	2.29
STR TH01	7,9	9,9	1.09	7,8	8,9	3.94
STR TPOX	8,12	8,9	0.43	8,11	8,8	0.86
STR CSF1PO	11,12	10,12	0.75	9,10	10,12	0.95
STR D5S818	7,11	10,11	0.75	13,13	10,13	3.70
STR D13S317	11,11	11,12	2.03	11,11	10,11	2.03
STR D7S820	11,12	11,13	0.70	10,10	10,12	3.14
STR D8S1179	10,13	13,14	1.19	12,13	13,13	2.39
STR D21S11	30,31.2	30,32	0.95	31,33.2	31,32	2.23
STR D18S51	15,16	14,15	1.30	15,15	13,15	2.59
STR D16S539	11,12	9,11	1.02	9,11	9,11	1.89
		Least CPI= 1.49			Largest CPI= 47,042.98	

PI: Paternity Index was calculated as formulae suggested by AABB (please see footnote 2 in page 3)
CPI: Combined Paternity Index, the product of multiplication of all paternity indices.
Ind. a, b, c, d: Individuals from one thousand random sample population.

6,555 pairs was found when 11 alleles were shared from 11 loci (Table 2).

Sensitivity and Specificity of CPI Frequency Ratio Distribution

In table 3, the sensitivity and specificity of the DNA test were illustrated. As the CPI cutoff values increased, there was a subsequent decrease in sensitivity and increase in specificity. In 13/13 scenario, as the increment of CPI cutoff varied from 10 to 10,000, the sensitivity varied from 100% to 47.332%; the specificity from 99.921% to 99.999%. The effect was sharper in 12mut/13, 12/12 and 11/11 scenario. If 100 was chosen as the minimum CPI requirement, the sensitivity varied from 19.258% for 12mut/13, 91.647% for 11/11, 96.752% for 12/12 to 98.144% for 13/13.

Evaluation of CPI Cutoff for Different Kinship Matching Situations

By increasing the CPI cutoff values in increments of 1 starting from 0 to 16,384, the optimum CPI value was obtained empirically by choosing the maximum of the square root of (sensitivity^2 + specificity^2), when there is one allele shared at all 13 loci, and under the circumstances of degraded DNA and/or a single mutation (Table 4). The optimized cutoffs for 12mut/

Table 2. Number of real duos and coincidental matched pairs that meet the minimum CPI requirement

Minimum CPI requirement	Real duo 13/13	Coin. M. 13/13	Real duo 12mut/13	Coin. M 12mut/13	Real duo12/12	Coin. M.12/12	Real duo 11/11	Coin. M. 11/11
Whatever is obtained	0	431	0	4,286	0	4,286	0	6,555
10	0	393	208	152	0	3,784	0	5,714
100	8	244	348	11	14	1,979	36	2,879
101	8	243	348	11	14	1,971	36	2,863
150	11	201	366	8	30	1,582	46	2,259
200	17	176	375	4	37	1,340	55	1,876
300	29	142	389		50	1,043	71	1,415
500	43	109	402	1 75	716	102	951	
1,000	80	54	415	0	113	413	149	534
1,001	80	54	415	0	114	413	150	534
10,000	227	7	429	0	272	28	319	39

Minimum CPI requirement: The minimum CPI requirement used in different laboratories in USA (please see footnote 1 in page 3)
Real duo 13/13: Real paternity duos with all 13 allele-sharing loci.
Coin. M. 13/13: Coincidental matched pairs with all 13 allele-sharing-loci out of 13.
Real duo 12mut/13: Real paternity duos with one locus in CODIS 13 was randomly excluded to mimic one-locus-mismatch situation and CPI mutation corrected.
Coin. M. 12mut/13: Coincidental matched pairs with only 12 allele-sharing-loci out of 13, and the CPI of the non-matched loci were mutation corrected.
Real duo 12/12: Real paternity duos with one locus in CODIS 13 was randomly erased to mimic one-locus-degraded situation, CPIs were calculated without mutation correction.
Coin. M.12/12: Coincidental matched pairs with only 12 allele-sharing-loci out of 13, CPIs were calculated without mutation correction.
Real duo 11/11: Real paternity duos with two loci in CODIS 13 was randomly erased to mimic two-locus-degraded situation, CPIs were calculated without correction.
Coin. M.11/11: Coincidental matched pairs with only 11 allele-sharing-loci out of 13.

Table 3. Sensitivity and specificity of CPI frequency ratio distribution of two sets of data; one from a computer based study and one from known sets of paternity duos

Minimum CPI requirement	13/13			12mut/13			12/12			11/11		
	sensitivity	specificity	sqrt (x)	sensitivity	specificity	sqrt (x)	sensitivity	specificity	sqrt (x)	sensitivity	specificity	sqrt (x)
Whatever is obtained	100.000%	99.914%	1.41361	100.000%	99.142%	1.40816	100.000%	99.142%	1.40816	100.000%	98.688%	1.40497
10	100.000%	99.921%	1.41366	51.740%	99.970%	1.12566	100.000%	99.242%	1.40886	100.000%	98.856%	1.40615
100	98.144%	99.951%	1.40080	19.258%	99.998%	1.01836	96.752%	99.604%	1.38859	91.647%	99.424%	1.35219
101	98.144%	99.951%	1.40080	19.258%	99.998%	1.01836	96.752%	99.605%	1.38860	91.647%	99.427%	1.35222
150	97.448%	99.960%	1.39600	15.081%	99.998%	1.01129	93.039%	99.683%	1.36356	89.327%	99.548%	1.33750
200	96.056%	99.965%	1.38635	12.993%	99.999%	1.00840	91.415%	99.732%	1.35289	87.239%	99.624%	1.32422
300	93.271%	99.972%	1.36726	9.745%	99.999%	1.00473	88.399%	99.791%	1.33314	83.527%	99.717%	1.30078
500	90.023%	99.978%	1.34535	6.729%	100.000%	1.00226	82.599%	99.857%	1.29592	76.334%	99.810%	1.25654
1,000	81.439%	99.989%	1.28958	3.712%	100.000%	1.00069	73.782%	99.917%	1.24206	65.429%	99.893%	1.19413
1,001	81.439%	99.989%	1.28958	3.712%	100.000%	1.00069	73.550%	99.917%	1.24069	65.197%	99.893%	1.19286
10,000	47.332%	99.999%	1.10635	0.464%	100.000%	1.00001	36.891%	99.994%	1.06582	25.986%	99.992%	1.03313

13/13: There were 13 allele-sharing loci between a pair.
12mut/13: There were 12 out of 13 allele-sharing loci between a pair, the CPI for the mismatched loci were mutation-corrected.
12/12: There were 12 out of 13 allele-sharing loci between a pair.
11/11: There were 11 out of 13 allele-sharing loci between a pair.
Calculation of sensitivity and specificity was based on the ratio of pair numbers in table 2 and their CPI value.
sqrt(x) : sqrt (sensitivity 2+ specificity 2).

Table 4. Evaluation of CPI cutoff for different matching situation.

	12mut/13			13/13			12/12			11/11		
CPI	sensitivity	specificity	sqrt (x)	sensitivity	specificity	sqrt (x)	sensitivity	specificity	sqrt (x)	sensitivity	specificity	sqrt x)
0	100.000%	99.142%	1.408159	100.000%	99.914%	1.413604	100.000%	99.142%	1.408159	100.000%	98.688%	1.404965
0.25	94.664%	99.562%	1.373816	100.000%	99.914%	1.413604	100.000%	99.142%	1.408161	100.000%	98.688%	1.404968
0.5	89.559%	99.686%	1.340084	100.000%	99.914%	1.413604	100.000%	99.143%	1.408166	100.000%	98.689%	1.40497
1	83.991%	99.798%	1.304383	100.000%	99.914%	1.413604	100.000%	99.146%	1.408189	100.000%	98.694%	1.405008
2	76.102%	99.873%	1.255637	100.000%	99.915%	1.413609	100.000%	99.154%	1.408247	100.000%	98.710%	1.405121
5	64.269%	99.944%	1.188246	100.000%	99.917%	1.413628	100.000%	99.189%	1.408493	100.000%	98.768%	1.405530
10	51.740%	99.970%	1.125653	100.000%	99.921%	1.413657	100.000%	99.242%	1.408867	100.000%	98.856%	1.406148
11	50.116%	99.974%	1.118321	100.000%	99.922%	1.413663	100.000%	99.251%	1.408930	100.000%	98.871%	1.406252
12	49.652%	99.976%	1.116265	100.000%	99.923%	1.413669	100.000%	99.258%	1.408978	100.000%	98.885%	1.406351
13	48.260%	99.978%	1.110166	100.000%	99.923%	1.413670	100.000%	99.268%	1.409045	100.000%	98.899%	1.406453
14	47.564%	99.981%	1.107180	100.000%	99.924%	1.413674	100.000%	99.277%	1.409112	99.768%	98.912%	1.404889
15	46.404%	99.982%	1.102257	100.000%	99.926%	1.413688	100.000%	99.285%	1.409170	99.768%	98.924%	1.404977
16	45.708%	99.983%	1.099354	100.000%	99.926%	1.413688	100.000%	99.293%	1.409227	99.768%	98.939%	1.405082
17	44.548%	99.985%	1.094596	100.000%	99.927%	1.413696	100.000%	99.302%	1.409286	99.768%	98.952%	1.405175
18	44.084%	99.985%	1.092723	100.000%	99.927%	1.413698	100.000%	99.308%	1.409327	99.768%	98.965%	1.405263
19	43.387%	99.987%	1.089945	100.000%	99.927%	1.413700	99.768%	99.314%	1.407728	99.768%	98.976%	1.405345
20	41.299%	99.987%	1.081809	100.000%	99.928%	1.413703	99.768%	99.321%	1.407772	99.768%	98.986%	1.405416
25	35.731%	99.989%	1.061818	99.768%	99.930%	1.412079	99.768%	99.352%	1.407996	99.072%	99.036%	1.400833
30	33.411%	99.991%	1.054256	99.768%	99.932%	1.412091	99.768%	99.386%	1.408237	98.840%	99.086%	1.399546
40	29.466%	99.994%	1.042452	99.768%	99.936%	1.412118	99.304%	99.437%	1.405311	96.984%	99.163%	1.387054
50	25.986%	99.996%	1.033177	99.768%	99.939%	1.412140	98.840%	99.481%	1.402348	95.824%	99.231%	1.379457
100	19.258%	99.998%	1.018352	98.144%	99.951%	1.400801	96.752%	99.604%	1.388590	91.647%	99.424%	1.352194
128	16.473%	99.998%	1.013458	97.680%	99.957%	1.397593	94.432%	99.655%	1.372895	90.487%	99.502%	1.344938
256	12.065%	99.999%	1.007246	94.896%	99.969%	1.378369	90.023%	99.770%	1.343809	84.687%	99.683%	1.307996
512	6.729%	100.000%	1.002259	89.791%	99.979%	1.343807	82.367%	99.858%	1.294448	75.870%	99.814%	1.253755
1,024	3.712%	100.000%	1.000689	81.206%	99.989%	1.288114	73.086%	99.919%	1.237954	64.501%	99.895%	1.189090
2,048	2.552%	100.000%	1.000326	74.942%	99.994%	1.249607	62.645%	99.956%	1.179640	52.204%	99.944%	1.127565
4,096	1.856%	100.000%	1.000172	64.269%	99.997%	1.188690	51.276%	99.980%	1.123624	40.603%	99.975%	1.079058
8,192	0.464%	100.000%	1.000011	52.204%	99.998%	1.128049	39.675%	99.990%	1.075736	28.538%	99.987%	1.039803
16,384	0.000%	100.000%	1.000000	39.443%	100.000%	1.074974	29.234%	99.997%	1.041828	19.026%	99.996%	1.017896

sqrt(x) : sqrt(sensitivity 2+ specificity 2). The sensitivity and specificity percentage for the CPIs between the CPI categories were omitted.

13 matches, 13/13 matches, degraded DNA with 12/12 matches, further degraded DNA with only 11/11 matches were "whatever is obtained", 20, 18, and 13, respectively.

Discussion

For many relationship (identity) testing laboratories, a minimum CPI value requirement or cutoff point is used, above which there is a high degree of confidence that the samples tested are genetically linked as first degree relatives and that a false inclusion will be prevented. Because real paternity duo cases with very low CPI values and coincidental kinship-matched pair with very high CPI values were observed in this study, false exclusions could happen by setting too high of cutoffs, and false inclusion could happen if setting too low of cutoffs. Simply applying a fixed requirement of minimum CPI for different matching situation may not be a good choice.

The ability of the DNA test to correctly classify events into two categories (inclusions and exclusions) is assessed by sensitivity and specificity (7). The sensitivity and specificity of the CODIS 13 STRs system for paternity duo testing was determined for real cases and computer generated pairs from populations under different scenarios. As the CPI cutoff values increased, there was a resultant decrease in the chance of a false inclusion (sensitivity) and increase in the false exclusion (specificity). The effect was more significant when fewer loci were available, *e.g.* in the degraded sample situation. For many laboratories, a cut off point of a minimum CPI value of 100 was used (23 in 34 laboratories in the AABB survey (please see footnote 2 in page 3) and recognized as generally accepted minimum standard for an inclusion of paternity (3) in the 13/13 scenario, with a sensitivity of 98.144% and specificity of 99.951%. As the specificity was not 100%, 0.05% of the cases were false inclusions. At a CPI of 100 the sensitivity was only 98.14%, resulting in 1.86% of real duo paternity being classified as exclusions. For example, a high CPI such as 10,000 would result in 52.668% (1-47.332%) of real duo paternity cases being reported as an exclusion. The optimized CPI value, being a balance between sensitivity and specificity, should be suggested.

From AABB data (please see footnote 2 in page 3) most of the DNA laboratories (19 of 29 laboratories) used "whatever is obtained" as CPI requirement for family reconstruction cases. Applying this criterion to these data, the sensitivity was 100% for 13/13 matches, 12/12 matches and 11/11 matches, and the specificity was 99.914%, 99.142% and 98.688% for each scenario respectively. If the cutoff of CPI = 0 was increased to CPI = 20 for 13/13 scenario, then the specificity could be increased to 99.928% and sensitivity maintained 100%, for the other two match situations, similar effects were found.

Clear differences in the sensitivity of the test when using a CPI cutoff value of 100 were observed for 13 allele shared (98.144%) compared to a single mutation (12mut/13) (19.258%). A single mutation, which is a routine observation for paternity testing laboratories (3), would result in approximately 80% of paternity case being reported as exclusions, if the cutoff were not adjusted.

By using the optimal cutoff selection method proposed by zou *et al.* (11), the optimum CPI values were obtained empirically for different scenarios. They may serve as the minimum CPI requirements.

In conclusion, a different CPI value should be applied for duo cases when there is full allele sharing at all 13 loci compared to a case where there is a possible mutation in one of the loci. For a paternity duo test and a family reconstruction test, the cutoff could be set at a CPI value of 20 if only one allele were shared at all 13 loci. If the sample were degraded and only 12 loci were available then the cuff off CPI value should be 18, and a CPI value of 13 for 11 loci. If a mutation had been suspected in a duo case, more STR tests should be added to confirm the mutation, and then DNA test results and non-DNA finding should be combined to determine the paternity. Applying CPI = 0 as the cutoff directly is not suitable although in table 4 CPI = 0 was listed as the candidate. For no optimum CPI minimum value can be set for paternity duo tests that will never result in false inclusions or false exclusions, although the application of these suggested CPI cutoffs will maximize the paternity screening. The ideal situation to resolve duo cases with greater confidence would be to add more autosomal STR loci to increase the CPI value (the specificity will be higher), and if possible to use mitochondrial DNA analysis or STR loci on sex Chromosome for further confirmation.

Acknowledgments

The authors wish to thank Wu F.C., Wu G.C., Cheng M.Y., Meng L.M. and Cheng S. M. for their kind assistance for genotyping all the local samples, and give special thanks to Meng L.M. for her advice on the data.

References

1. Birus, I., Marcikic, M., Lauc, D., Dzijan, S. and Lauc, G. How high should paternity index be for reliable identification of war victims by DNA typing? *Croat. Med. J.* 44: 322-326, 2003.
2. Budowle, B., Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A. and Keys, K.M. Population data on the thirteen CODIS core short tandem repeat loci in African American, U.S. Caucasians, Hispanics,

Bahamians, Jamaicans, and Trinidadians. *J. Forensic Sci*. 44: 1277-1286, 1999.
3. Butler, J.M. Kinship and parentage testing. In: Forensic DNA Typing, Elsevier Academic Press, MA, USA, 2005, pp. 529-537.
4. De Ungria, M.C., Frani, A.M., Magno, M.M., Tabbada, K.A., Calacal, G.C., Delfin, F.C. and Halos, S.C. Evaluating DNA tests of motherless cases using a Philippine genetic database. *Transfusion* 42: 954-957, 2002.
5. Gornik, I., Marcikic, M., Kubat, M., Primorac, D. and Lauc, G. The identification of war victims by reverse paternity is associated with significant risks of false inclusion. *Int. J. Legal. Med*. 116: 255-257, 2002.
6. Killeen, A.A. Identity testing. In: Principles of molecular pathology, NJ, USA, Humana Press, 2004, pp. 297-308.
7. Kirkwood, B.R. and Sterne, J.A.C. Measurement error: assessment and implications. In: Medical statistics, MA, USA, Blackwell Publishing, pp. 429-446, 2003.
8. Presciuttini, S., Ciampini, F., Alu, M., Cerri, N., Dobosz, M. and Domenici, R. Allele sharing in first-degree and unrelated pairs of individuals in the Ge. F.I. AmpFlSTR(r) Profiler Plus(tm) database, *Forensic Sci. Int*. 131: 85-89, 2003.
9. Pu, C.E., Wu, F.C., Cheng, C.L., Wu, K.C., Chao, C.H. and Li, J.M. DNA STR profiling of Chinese population in Taiwan determined by using an automated sequencer, *Forensic Sci. Int*. 97: 47-51, 1998.
10. Pu, C.E. and Linacre, A. CPI distribution and cut-off value for duo paternity building. In: Proceedings of the 58th Annual Meeting of the American Academy of Forensic Sciences, Seattle, WA, USA, 2006, pp. 172-173.
11. Zou, K.H., Wells, III W.M. Kikinis R. and Warfield, S.K. Three validation metrics for automated probabilistic image segmentation of brain tumors, *Statist. Med*. 23: 1259-1282, 2004.

D53 A Computer Program for Calculating Forensic Population Study Parameters of STR Loci

Chang En Pu, MS, Ministry Justice Investigation Bureau, No.74 Chun-Hua Road, Hsin-Tien, Taipei, 231, Taiwan; and Adrian Linacre, PhD, Forensic Science Unit, Department of Pure and Applied Chemistry, University of Strathclyde, 16 Richmond Street, Glasgow, Scotland G1 1XQ, United Kingdom*

The goal of this presentation is to provide a program to help calculate a wide range of forensic population parameters and a FPP (false positive parentage) rate. This presentation will impact the forensic community and/or humanity by making the population study more efficient and provide an easy way to evaluate the false positive parentage rate to avoid false identification, especially for DNA database operation.

The computer program (STRstatistics 2005.1) is presented, which is capable of calculating a wide range of commonly used forensic population study parameters. These include: p value of G-tests for HWE proportion; the number of types of a particular allele; the occurrence frequency of alleles, expected and observed heterozygosity (H); polymorphism information content (PIC); power of discrimination (PD); probability of a match (PM); power of exclusion (PE) for trio and duo paternity tests; typical paternity index (PIt) and typical power of exclusion (PEt). The evaluation of data by these means is frequently a requirement in forensic practice, particularly when examining a new population.

At present, there are limitations to the computer programs that are available for forensic population studies, such as locus by locus handling (rather than batch handling), limited sample volume, and data format transformation. Many other genetic processing related computer programs were not designed exclusively for forensic evaluation of population study and therefore only provide analysis for a few forensic population parameters, therefore requiring additional calculation tools to be used. The STRstatistics 2005.1 program runs on the basis of the initial STR data such as that directly imported from Applied Biosystems Genotyper® software as well as an Excel format or by manual addition. Microsoft® Excel® Macros and built-in functions controlled by Visual Basic language written by the authors was used to handle the Hardy-Weingberg test and other forensic calculations. The application requires only that the users post or import their 15 STR genotypes from a population onto a Microsoft® Excel® worksheet, then press the hot key to activate the Macros. The allele frequency and forensic parameter table will be generated ready for publishing or for use as a population database. The program is capable of handling data of 1,000 individuals and 15 loci simultaneously, from which the informative forensic parameters will be tabulated automatically. The "STRstatistics 2005.1" Microsoft® Excel® template contains several worksheets. The "ori STR" worksheet provides brief instructions for using the template, and describes some limitations of the template. The genotype data for 15 STR loci, which comprises 30 columns with 2 columns for each locus, may be pasted onto the "ori STR" worksheet. Up to 40 alleles for each locus are acceptable. Genotypes containing text alleles (e.g., nc or 9.x) or with more than 2 alleles will be treated as in text (nonnumeric) mode and ignored in the auto-run program analyses. The final results table is produced as "publish tab" but can be modified manually by users to meet the required formats for publication. The program is freely available to any forensic scientist interested. Please e-mail requests to the corresponding author.

Forensic Science, STRs, Population Study

CPI distribution and cut-off value for duo paternity building

Chang En Pu, MS[1]*, Meng-Yi Chen, MS[1], Lim-Ming Meng, MS[1], Fang-Chin Wu, BS[1], Adrian Linacre, PhD[2]

1. Scientific and Technical Research Center, Ministry Justice Investigation Bureau, 74, Chung-Hua Road, Hsin-Tien City, Taipei County, Taiwan, R.O.C.

2. Forensic Science Unit, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, UK.

*Corresponding author: Fax: +886 2 29138599, e-mail: pu_macros@yahoo.com.tw

Abstract

The STR loci comprising CODIS has an average power of paternity exclusion larger than 0.9999 based upon mother, child and father combinations (a trio case). This figure is true for many populations. In cases requiring the identification of human remains if only one living relative (either of the parents or offspring) is available, this represents a duo case for parentage building. In duo cases when allele sharing is found in all the 13 loci, the probability of parentage could be determined. However it is hard to avoid a false parentage evaluation if the pair happened to share an allele at all 13 loci. In Taiwan, the National Unidentified Bodies CODIS 13 STR Database has approximately 1250 bodies and 350 families for comparison originally. Using STR typing and blood-relative comparing instances a body first-degree matched to more than one individual was found not rarely, however the CPI (Cumulated Paternity Index) was extremely low. It is necessary to evaluate the false parentage rate and set a cut-off value of CPI and vise versa to analyze the distribution of CPIs from real paternity cases, hope to help evaluate the paternity and lead to identification.

According to the published frequencies of STR alleles, the cumulated power of exclusion (PE) for duo for Chinese in Taiwan is 98.13%. The data showed that about 1.87 % random individuals could not be excluded from being a first-degree blood relative to the population. For proving this, CODIS 13 population data of 1,000 Chinese in Taiwan was collected and paired resulting in 499,500 pairs. Microsoft Excel Macros controlled by a Visual Basic program written by authors was used to handle the allele sharing comparison and CPI calculation. There were 462 (0.0925%) pairs found with all 13 allele sharing loci. False parentage relation was noted when the CPI for pairs ranged from 2.56 to 6,835,432.78, and the median CPI was 484.69 meaning that if the CPI of 484.69 were used as the cut-off, 50% of the false pairs would not be recognized as first-blood-relative, and if the CPI cut-off increased to 1,000, 62.9% false pairs could be eliminated, however the false exclusion rate for real duos was 5.7%(cut-off = 484.69) and then increased to 10.8%(cut-off = 1,000) respectively. The dilemma could be resolved by profiling more STR systems when duos were found with low CPI or adding anthropology and other information to make the confirmation. This is especially the case for mass and open comparing operation of STR database for the unidentified bodies.

Keywords: forensic science, short tandem repeat (STRs), CODIS 13, false parentage, unidentified body

A work-sheet based computer program for forensic population study of STR loci

Chang En Pu, MS[1]* and Adrian Linacre, PhD[2]

1. Scientific and Technical Research Center, Ministry Justice Investigation Bureau, 74, Chung-Hua Road, Hsin-Tien City, Taipei County, Taiwan, R.O.C.
2. Forensic Science Unit, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, UK.

*Corresponding author: Fax: +886 2 29138599, e-mail: pu_macros@yahoo.com.tw

We present the computer program STRstatistics 2005.1, which is capable of calculating forensic population study parameters including: p value of G-tests for HWE proportion; the number of types of a particular allele; the occurrence frequency of alleles, expected and observed heterozygosity (H); polymorphism information content (PIC) ; power of discrimination (PD) ; probability of a match (PM) ; power of exclusion (PE) for trio and duo paternity tests ; typical paternity index (PIt) and typical power of exclusion (PEt). The evaluation data by these means is frequently a requirement in forensic practice, particularly when examining a new population. At present there are limitations to the computer programs that are available for forensic population studies, such as locus by locus handling (not batch handling), limited sample volume, data format transformation, and many other genetic processing related computer programs were not designed exclusively for forensic evaluation of population study and therefore only provide analysis for a few forensic population parameters, requiring additional calculation tools to be used. The STRstatistics 2005.1 program runs on the basis of the initial STR data such as that directly imported from Applied BioSystem's Genotyper software as well as an Excel format or by manual addition. Microsoft Excel Macros and built-in functions controlled by Visual Basic language written by the authors was used to handle the Hardy-Weingberg test and other forensic calculations. The application requires only that the users post or import their 15 STR genotypes from a population onto a Microsoft Excel worksheet, then press the hot key to activate the Macros. The allele frequency and forensic parameter table will be generated ready for publishing or for use as a population database. The program is capable of handling data of 1,000 individuals and 15 loci simultaneously, from which the informative forensic parameters will be tabulated automatically. The "STRstatistics 2005.1" Microsoft Excel template contains several worksheets. The "ori STR" worksheet provides brief instructions for using the template, and describes some limitations of the template. The genotype data for 15 STR loci, which comprises 30 columns with 2 columns for each locus, may be pasted onto the "ori STR" worksheet. Up to 40 alleles for each locus are acceptable. Genotypes containing text alleles (e.g., nc or 9.x) or with more than 2 alleles will be treated as in text (nonnumeric) mode and ignored in the auto-run program analyses. However, the final results frequency table in "publish tab" can be modified manually by users to meet the required formats for publication. For a test of the program, 15 loci STR genotypes of a Chinese population in Taiwan was pasted into the program, allele frequencies and other parameters were generated in a Excel worksheet table in minutes. The data generated by the program was rechecked by the authors manually, also assisted by Excel, to confirm the efficiency and accuracy of the program. Besides these known parameters, another one named FPR (false parentage rate) and all the random pairs that happen to have first-degree-blood-relative

match in the sampling population will also be tabulated for reference. The easy handling and once-for–all parameter's program could be a useful tool for forensic scientists for evaluating STR systems or renewing population statistics. The program is freely available to any forensic scientist interested. Please e-mail requests to the corresponding author.

Keywords: Forensic Science, STRs, forensic parameters, computer programs.

Study on the DNA STR profiling for artificially degraded bones

Chang-En Pu, MS, Ling-Min Meng, MS, Chun-Ming Chao, Shiu-Ming Cheng, BS, MS, Fang-Chin Wu, BS

Scientific and Technical Research Center,

Ministry Justice Investigation Bureau, 74, Chung-Hua Road, Hsin-Tien City, Taipei County, Taiwan, R.O.C.

*Corresponding author: Fax: +886 2 29138599, e-mail: pu_macros@yahoo.com.tw

Abstracts

The DNA in human body is biological material and very vulnerable to the environmental factors when taken outside the body or when the body were dead. The insulting factors including temperature, humidity, sun light etc., adding on many affecting factors to the specimens, resulting in the difficulties for DNA purification and STR typing.

The unidentified bodies when found were usually in a very bad conditions, especially the bodies that were purposely burned, dropped into water, disposed at open field, most of the time only skeletons could remain. From the experiences of forensic DNA of MJIB DNA lab, there were at least 10% of the human remains that could not be DNA typed owing to the badly degraded samples. DNA STR typing methods should be enhanced to increase success rate of STR typing.

We used 35 clavicle bone specimens divided into five groups (7 for each group), groups were treated such as buried in soil, immersed in sea water, immersed in fresh water, exposed to the atmosphere with shelter and exposed to the atmosphere without shelter respectively. The samples were processed for a time course of 5 months, for each month 35 samples were taken out for sawing 1 cm long away for DNA extraction, then the STR multiplex kits were used for CODIS 13 typing, for each group if every bone were successfully STR typed, there should be 98 STR typing for one group for one month, and totally 490(7 x 14 x 5)STR typing for a group for 5 months.

The data showed that the success rate decreased as the degrading time lasted longer, and the general success rate for STR typing for the above groups were 9/490, 52/490, 34/490, 113/490 and 69/490, the groups without mixing with waters directly had the higher rate, and the buried in soil group had the lowest rate maybe owing to the much more saprobes in soil or maybe owing to the inhibitors from soil for PCR reaction.

The primer sequences for the loci CSF1PO and D7S820 were modified to have the PCR products shortened, and they were applied onto the 35 five months' treated samples, the success rate for CSF1PO was increased from 3/35 to 14/35, and D7S820 from 3/35 to 11/35, and for real case handling when no STR types were obtained mtDNA was still applied to get enough information for maternity matching. The

redesigning of the primers' sequences and the adding of mtDNA sequencing could enhance the unidentified body DNA recognizing.

Key Words: Forensic Science, Forensic DNA, aged and degraded bone specimens, unidentified body

Study on automatic forensic DNA STR blood-relative comparing for un-identified body database processing

Chang En Pu, MS[1]* and Adrian Linacre, PhD[2]

1. Scientific and Technical Research Center, Ministry Justice Investigation Bureau, 74, Chung-Hua Road, Hsin-Tien City, Taipei County, Taiwan, R.O.C.
2. Forensic Science Unit, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, UK.

*Corresponding author: Fax: +886 2 29138599, e-mail: pu_macros@yahoo.com.tw

Abstracts

In recent years, many disasters happened in Taiwan, on Feb. 16, 1998, China Airline CI006 crashed, about 200 people died, more than half of the bodies were broken into pieces, thousands of people died in 921 earth quake in 1999, on Oct. 31, 2000, Singapore Airline SQ006 crashed before taking off, 82 persons died, for each disaster many much more family members appeared for building genetic relationship to the dead to lead to the body identification, it was a lot of forensic DNA work in a very short time.

During that period of time, the DNA comparison and matching were done by manual, in another CI 676 case, a committee of more than 10 experts of Medical doctors, forensic scientists, professors worked together for about 15 days to match the DNA data between dead body and families, the time and labor could have been saved.

The CODIS 13 systems are the most popular forensic DNA system in the forensic world, stable and specific, and also very suitable for body identification. If the STR typing of one of a family was found to share allele for all 13 loci with a dead body, after calculating the CPI, the identification of the dead body could be built. The comparison can be done manually, but for sibship comparison, because for each loci, the allele sharing could either exist or not, it is very time-consuming to do it manually, besides this, the STR databases of unidentified bodies and families from all over the country should be compared every week, it is necessary to have a software that help the comparing works..

Software EXCEL provided by Microsoft are widely used in forensic laboratory for STR database handling, but the software can not be used for determining paternity or sibship directly, authors established a software system to determine genetic relationship directly and automatically from the allele table imported from the auto-sequencer, when used on the current national databases of 900 unidentified

bodies and 245 families, less than 2 seconds was needed to finish all the blood-relative comparing, the efficiency was increased apparently.

Key Words: STR polymorphism, genetic comparison, unknown body identification.

B52 Evaluating the False Parentage Rate and CPI Cut-off of CODIS 13 STR for Seven Populations

Chang En Pu, MS, Meng-Yi Chen, MS, Lim-Ming Meng, MS, and Fang-Chin Wu, BS, Ministry Justice Investigation Bureau, PO BOX 3562, No 74, Chun-Hua Road, Hsin-Tien, Taipei 231, Taiwan; Adrian Linacre, PhD, Forensic Science Unit, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, UK, 204 George Street, Glasgow, Scotland G1 1XW, United Kingdom*

After attending this presentation, attendees will be able to recognize that false parentage determination after DNA test is possible.

This presentation will impact the forensic community and/or humanity by demonstrating the cutoff value of CPI for parentage could be established to avoid false parentage, especially important for unidentified body recognition on immigration blood relative testing.

We report on the use of STR typing for the genetic linkage of unidentified human remains and the problems associated with false paternity results. STR loci are chosen and used on based upon their power of discrimination and ability to multiplex with other STR loci. Most of the STR loci used in forensic science both for criminal and civil investigations. The CODIS 13 core STR systems has an average power of paternity exclusion (PE, to exclude a random man) larger than 0.9999(in trio cases) in many populations. In the identification of unidentified bodies only one relative (either of parents or either of son/daughter) is available for testing. This results in duo cases for parentage building. When allele sharing is found in all the 13 loci, the probability of parentage could be determined preliminarily. However it is hard to avoid a false confirmation of the alleged father or false identification of the unidentified bodies. In Taiwan, the national unidentified bodies CODIS 13 STR database has approximately 680 bodies and 200 families. Sometimes a body matches to more than one individuals (from different families) and cases with extremely low Cumulated Paternity Index (CPI) were found. It is necessary to evaluate the false parentage rate and set a cut-off value of CPI for avoiding false determination of parentage.

The CODIS 13 population data of 177 African American, 194 US Caucasian, 202 Southwestern Hispanic, 153 Bahamian, 157 Jamaican, 76 Trinidadian and 1,000 Chinese in Taiwan was collected from published websites of FBI USA or by authors. The cumulated power of exclusion (PE) for duo for African America = 98.31%, US Caucasian = 98.23%, Southwestern Hispanic = 97.97%, Bahamian = 98.48%, Jamaican = 98.32%, Trinidadian = 98.61% and Chinese in Taiwan = 98.13%. The data showed that about 1.4 % to 2% of random men could not be excluded from being an alleged father for the studied populations. A matching test model was designed to evaluate the practically possible false parentage rate.

All the collected individuals were paired resulting in 15,576, 18,721, 20,301, 11,628, 12,246, 2,850, and 499,500 pairs for each population respectively. Each pair was checked for allele sharing locus by locus and the CPI for those pairs with 13 loci sharing was also calculated. Microsoft Excel Macros controlled by a program written by the authors were used to handle the comparison and calculation. CPI calculations were based on the distribution frequencies for the respective populations. These were 8(0.0514%), 10(0.0534%), 9(0.0443%), 3(0.0258%), 7(0.0572%), 3(0.1053%) and 250(0.0501%) pairs found with 13 allele sharing loci for the seven populations. False parentage was noted when the CPI for pairs ranged from 1.76 to 1,950,430. These were 4.48% CPI lower than 100, and 78.27 % lower than 1,000. If the suitable CPI cut-off value were used such as CPI=1,000, there would be less of false parentage.

Though we could reduce the false positive rate by increasing the cut-off value, the false exclusion rate (a real father excluded as a random man) would be increased accordingly. In this study many duo paternity

cases in Taiwan were found with CPI lower than 1,000 when only CODIS 13 STRs were typed. If CPI=1,000 was suggested as the cut-off value these low CPI cases could be classified as false parentage cases. More real paternity cases have to be observed to set an optimum range of value for CPI cut-off, or more STR systems are required when matches were found with low CPI. This is especially the case for mass and open matching operation of STR database for the unidentified bodies. If less than 13 STR typing were obtained from the unknown skeleton owing to degradation of DNA, the CPI would be much lower, and the risk of false identification would be much higher. In such cases mtDNA sequencing and other investigating techniques should be added to enhance the discrimination.

Short Tandem Repeat, CODIS 13, Parentage Test

G59 Experiences of Human Bodies Identified by DNA Typing in Singapore Airlines SQ006 Crash in Taipei

Chang-En Pu, MS, Ministry Justice Investigation Bureau, PO Box 3562, Taipei, Taiwan; Chi-Hsiang Chao, Division Chief, Ministry Justice Investigation Bureau, Taipei, Taiwan; Fang-Chin Wu, Wen-Hsiung Ho, Kuo-Chun Wu, Ling-Min Meng, Researcher, Ministry Justice Investigation Bureau, Taipei, Taiwan; and Kai-Ping Shaw, Director of Forensic Pathology, Institute of Forensic Medicine, Ministry of Justice, Taipei, Taiwan*

The goal of this research project is to establish the standard operation procedure to manage the mass disaster during the experience of Singapore Airline SQ-006 Air crashed Accident by using CODIS 13 STRs profilers to identify badly wounded bodies.

On October 31, 2000, approximately 2317 Taiwan time (1517 UTC), a Singapore Airlines Flight SQ-006, with Singapore registration 9V-SPK, Boeing 747-400 airplane entered the incorrect runway at Chiang-Kai-Shek (CKS) Airport, Taiwan. Heavy rain and strong wind from typhoon "Xiang Sane" prevailed at the time of the accident. The airplane was destroyed by its collision with the runway construction equipment and by post impact fire. There were a total of 179 people on board with 159 passengers, 3 flight crewmembers and 17 cabin attendants. Official report, 83 people died (including 4 cabin crews), and 44 people injured. After the airplane SQ-006 of Singapore Airline got an accident at CKS international airport before taking off, 79 persons from 19 different countries died instantly. Although some of the bodies were identifiable, the prosecutor's office of Tau-Yuan district court (40 km away located southern Taipei metropolitan) decide to request the DNA typing method to confirm each body identity in addition to odontological and anthropologic identification.

Immediately after this accident, the Criminal Investigation Bureau, Ministry of Interior and medical examiners' of IFM were asked to type the sample from deceased's bodies and STR typing were applied, and MJIB and IFM were assigned to collect the family members' STR for further paternity matching. For the purpose to increase the efficiency of DNA typing, the forensic scientists of molecular biologist were organized into blood collector group, DNA extraction group, STR typing group and STR matching work stations as a DNA typing and identifying mass production chain. The CODIS 13 STRs were applied to fulfill the DNA data interchanging.

The first group of families arrived on the next day 8:30am about 9 hours after the crash accident and their blood samples were drawn for DNA typing to establish genetic relationship to the deceased, with the access of the 24 hour running-basis working team, 20 hours after the crash, the first group of 12 bodies were identified. After five days' mass production operation in the DNA laboratory, 80 bodies matching their relatives from 19 different countries were identified including one once rescued decedent died about 24 hours after incidence. Another two decedents died in the hospital without identification problem. Due to lack of lineal or collateral consanguinity information of one decedent to claim the body until the odontological and anthropological characteristics were confirmed. One of the 80 decedents was identified by STR direct matching (STR typing sent from Singapore) with previous blood sample left in the blood bank of Singapore. An unique case of a decedent was identified after the CODIS 13 data of decedent's family member who were unable to be Taiwan in time. 15 of the decedents were identified by sibship profiles due to collateral consanguinity. 65 of the decedents were recognized by paternity test with linear consanguinity. Forensic DNA matching is an accurate and efficient tool for badly crashed bodies identifying in addition to the odontological and anthropological method. To establish the professional guideline as the standard operation procedure of DNA typing and body identification is a crucial role to cope the mass disaster effectively during the urgent operation period. The Singapore

Airlines SQ-006 Air crashed accident by using STR CODIS 13 profilers to identify badly wounded bodies during five days' experience is worth for strategic schema.

000250

Forensic Science, CODIS 13, Air Crash