Sequence variation of the Amelogenin gene on the Y-chromosome

BY

IRMA FERREIRA, B.Sc., B.Sc.(HONS), M.Sc.

Thesis submitted for the degree Philosophiae Doctor (Ph.D.)
in Biochemistry at the North-West University (Potchefstroom Campus)

PROMOTER: Professor Antonel Olckers
Centre for Genome Research, North-West University (Potchefstroom Campus)

CO-PROMOTER: Doctor Gordon Wayne Towers
Centre of Excellence for Nutrition, North-West University (Potchefstroom Campus)

November 2010

Verandering in volgorde van die Amelogeniengeen op die Y-chromosoom

DEUR

IRMA FERREIRA, B.Sc., B.Sc.(HONS), M.Sc.

Proefskrif voorgelê vir die graad Philosophiae Doctor (Ph.D.)
in Biochemie aan die Noordwes-Universiteit (Potchefstroom Kampus)

PROMOTOR: Professor Antonel Olckers
Sentrum vir Genomiese Navorsing, Noordwes-Universiteit (Potchefstroom Kampus)

MEDEPROMOTOR: Doktor Gordon Wayne Towers
Sentrum van Uitnemendheid vir Voeding, Noordwes-Universiteit (Potchefstroom Kampus)

November 2010

Loof die Here, want Hy is goed,
want sy goedentierenheid is tot in ewighed
Ps 107:1

This thesis is dedicated to my husband, Hennie
and sons, Louis and Damien

ABSTRACT

The accurate determination of gender of biological samples has valuable applications in medical and forensic investigations. Gender determination based on length variations in the X-Y homologous amelogenin gene, is part of most commercial multiplex DNA profiling kits. The first report of a failure of the amelogenin sex test was in 1998 when two normal males were typed as female. Subsequently, several amelogenin Y (AMELY) negative males have been reported. This study represents the first report of this phenomenon in the black South African population.

This study determined the size of the Y-chromosome deletion that resulted in the failure of the amelogenin sex test in two black South African AMELY-negative males by typing specific DNA markers surrounding the amelogenin locus. Through deletion size and Y-chromosome microsatellite haplotypes, the relationship between the samples was investigated. The samples were sequenced at the amelogenin gene and typed for thirteen sites on the short arm of the Y-chromosome. In order to determine the Y-chromosome haplotypes, eleven Y-chromosome microsatellite markers were typed.

These samples had the same size deletion of approximately 3 Mb. The Y-chromosome haplotypes indicated that these were probably independent events. The frequency of AMELY-negative males is rare in this population sample of 8,344 individuals, with a frequency of 0.065% in the black South African sample population. Notwithstanding, tests performed for detecting the presence of male DNA based on the presence of the amelogenin gene alone should be reconsidered, as this study confirms that these deletions do occur in the African population. The impact of the results generated in this study on the medical and forensic practise of DNA testing is significant.

OPSOMMING

Die akkurate bepaling van die geslag van biologiese monsters het waardevolle toepassings in mediese en forensiese ondersoeke. Geslagsbepaling gebaseer op die lengteveranderinge in die X-Y homoloë amelogeniengeen, vorm deel van die meeste kommersiële multipleks DNS profileringssisteme. Die aanvanklike berig van ‘n defektiewe amelogenien geslagstoets was in 1998 toe twee normale mans as vroulik getipeer is. Daaropvolgend is ‘n aantal amelogenien Y (AMELY) negatiewe mans gerapporteer. Hierdie studie is die eerste beskrywing van hierdie verskynsel in die swart Suid-Afrikaanse populasie.

In hierdie studie is die grootte van die delesie wat die defektiewe amelogenien geslagtoets veroorsaak het in twee swart Suid-Afrikaanse mans bepaal, deur die tipering van DNS merkers wat die amelogeniengeen omsluit. Die delesiegrootte en Y-chromosoom mikrosatteliet haplotipes is gebruik om die verhouding tussen die monsters te ondersoek. DNS volgordebepaling van die amelogenien peilerbindingsareas is gedoen en dertien lokusse op die kort arm van die Y-chromosoom is getipeer op hierdie monsters. Om die Y-chromosoom haplotipes te bepaal, is elf Y-chromosoom mikrosattelietmerkers getipeer.

Die delesiegrootte was dieselfde vir beide monsters en ongeveer 3 Mb in grootte. Die Y-chromosoom haplotipes is aanduidend van twee onafhanklike gebeurtenisse. Die frekwensie van AMELY-negatiewe mans was laag in hierdie populasiemonster van 8,344 individue, met ‘n algemene frekwensie van 0.065% in die swart Suid-Afrikaanse populasiemonster. Toetse wat die teenwoordigheid van manlike DNA bepaal wat gebaseer is op die teenwoordigheid van slegs die amelogeniengeen, behoort nietemin hersien te word aangesien hierdie studie bevestig dat hierdie delesies in populasies van Afrika voorkom. Die trefkrag van hierdie studie se resultate is aansienlik in die mediese en forensiese velde van DNS toetsing.

TABLE OF CONTENTS

APPENDIX A

LIST OF ABBREVIATIONS AND SYMBOLS

Symbols are listed in alphabetical order

℃	degrees Celsius
%	percent
µ	micro: 10^{-6}
®	registered trademark
TM	trademark

Abbreviations are listed in alphabetical order

A or a	adenine (in DNA sequence)
AI	amelogenesis imperfecta
Ala	alanine
Alu repeat	short interspersed repetitive elements in mammalian genomes, containing *Alu 1* recognition sequence
AMEL	amelogenin
AMELX	amelogenin gene on the X-chromosome
AMELY	amelogenin gene on the Y-chromosome
AmpliTaq DNA polymerase	AmpliTaq®[1] DNA polymerase FS: variant of Taq DNA polymerase
Asp	aspartate
AZF	azoospermia factor
AZFa	azoospermia factor region a
AZFb	azoospermia factor region b
AZFc	azoospermia factor region c
bp	base pair
BPY1	basic protein Y1
BPY2	basic protein Y2
BSA	bovine serum albumin
C or c	cytosine (in DNA sequence)
cDNA	complementary cDNA
CDY	chromodomain Y-linked
CCAAT	promoter element consisting of the following sequence: 5' GGCCAATCT 3'
cm	centimetre: 10^{-2} metre
CSF1PO	repeat polymorphism at c-fms proto-oncogene
DAZ	deleted in azoospermia
DBY	DEAD box Y-linked
DEAD	aspartate-glutamate-alanine-aspartate
DFFRY	ubiquitin-specific protease (Drosophila fat-facets related Y)
DNA	deoxyribonucleic acid
DYZ1	Y-chromosome specific repeat DNA family
DYZ2	Y-chromosome specific repeat DNA family
ddNTP	2',3'-dideoxynucleotide triphosphate
dH_2O	distilled water
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanidine triphosphate
dNTP	deoxynucleotide triphosphate
dTTP	deoxythymidine triphosphate
E1F1AY	translation initiation factor 1A Y
et al.	*et alii*: and others

[1] AmpliTaq® DNA polymerase, is a registered trademark of Roche Molecular Systems Inc., Alameda, CA, USA.

EDTA	ethylene diamine tetra-acetic acid: $C_{10}H_{16}N_2O_8$
EtBr	ethidium bromide: 2,7-diamino-10-ethyl-9-phenyl-phenanthridinium
EtOH	ethanol: CH_3CH_2OH
5'-FAM	5-carboxyfluorescein
FGA	repeat polymorphism at fibrinogen alpha chain gene
FL	fluorescein
FTA	Flinders Technology Associates
g	gram
G or g	guanine (in DNA sequence)
GBY	gonadoblastoma locus on the Y-chromosome
GenBank	GenBank®[1]: United States repository of DNA sequence information
Glu	glutamate
Hg	haplogroup
Hi-Di	formamide solution proprietary to Applied Biosystems®[2]
ILS600	internal lane standard 600
IR	inverted repeats
IR3	inverted repeats 3
JOE	6-carboxy-4',5'-dichloro-2',7'-dimethoxy-fluorescein
kb	kilo-base
KCl	potassium chloride
LINE	long interspersed nuclear element
µg	microgram
$\mu g.\mu l^{-1}$	microgram per microlitre
µL	microlitre
µM	micromolar
M	molar
Mb	megabase
$MgCl_2$	magnesium chloride
min	minute
ml	millilitre
mM	millimolar
MSY1	male-specific region of the Y chromosome
NAHR	nonallelic homologous recombination
NaN_3	sodium azide
NED	proprietary to Applied Biosystems®
ng	nanogram
ng. μl^{-1}	nanogram per microlitre
no.	number
NRY	non-recombining region of the Y-chromosome
OH	hydroxyl
p	short arm of chromosome
PAR1	pseudoautosomal region 1
PAR2	pseudoautosomal region 2
PBS	phosphate buffered saline
PCDHY	protocadherin-adhesion gene on the Y-chromosome
PCR	polymerase chain reaction
PentaD	repeat polymorphism on chromosome 21q
PentaE	repeat polymorphism on chromosome 15q
pH	a measure of acidity: numerically equal to the negative logarithm of H^+ concentration expressed in molarity
PRKY	protein kinase Y-linked
PRY	putative tyrosine phosphatase protein-related Y-linked
q	long arm of chromosome
RBM	RNA-binding motif
RBMY	RNA-binding motif Y-linked

[1] GenBank® is a registered trademark of the National Institute of Health and Human Services for the Genetic Sequence Data Bank, Bethesda, MD, USA.
[2] Applied Biosystems® is a registered trademark of Life Technologies, Carlsbad, CA, USA.

RNA	ribonucleic acid
RPS4Y	protein of small ribosomal subunit Y-linked
s	seconds
SA	South African
Ser	serine
SAPS	South African Police Service
SEY1/SYBL1	synaptobrevin-like 1
SMCY	selected mouse cDNA Y-linked
SRY	sex determining region on the Y-chromosome
STR	short tandem repeat
STS	sequence tagged site
T or t	thymine (in DNA sequence)
TATA	promoter element consisting of the following sequence: 5'-TATA-3'
TB4Y	thymosin beta-4 Y-chromosomal isoform (actin sequestration)
TBE	tris®[1] borate-EDTA buffer
TBL1Y	transducin (beta)-like 1 protein Y-linked
TH01	repeat polymorphism at tyrosine hydroxylase gene
Thr	threonine
TMR	carboxy-tetramethylrhodamine
TPOX	repeat polymorphism at thyroid peroxidase
TPR	tetratricopeptide repeat
Tris-HCl	2-amino-2-(hydroxymethyl)-1,3-propanediol hydrochloride: $C_4H_{11}NO_3.H_2O$
Triton X-100	octylphenolpoly(ethylene-glycoether)$_n$:$C_{34}H_{62}O_{11}$ for n=10
TSPY	testis-specific protein Y-linked
TTY1	testis transcript Y1
TTY2	testis transcript Y2
Tyr	tyrosine
U	units
UK	United Kingdom
UTY	ubiquitous TPR motif Y-linked
USA	United States of America
USP9Y	deubiquinating enzyme Y-linked
UV	ultraviolet
V	volts
VCY	variable charged protein Y-linked
vWA	repeat polymorphism at von Willebrand factor gene
w/v	weight per volume
x g	gravitational acceleration
XKRY	XK-related Y-linked (membrane transport)
YHRD	Y-STR haplotype reference database
Yp	short arm of the Y-chromosome
Yq	long arm of the Y-chromosome
Y-STR	short tandem repeat on the Y-chromosome
Y-STS	sequence tagged site on the Y-chromosome
ZFY	zinc finger transcription factor

[1] Tris® is the registered trademark of the United States Biochemical Corporation, Cleveland, OH, USA.

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation to the following individuals and institutions:

To the **individuals** that participated in this study, for their invaluable contribution.

My thanks and appreciation goes to my supervisor, **Professor Antonel Olckers**, for her insight, advice, time, mentorship, encouragement and friendship without which this project would not have been possible. To my co-supervisor, **Dr Wayne Towers**, for his invaluable input and time.

To **Dr George Gericke** for his encouragement and support and **AMPATH** laboratories for approving this study.

My appreciation to the **Centre for Genome Reseach**, **DNAbiotec (Pty) Ltd** and the **North-West University (Potchefstroom Campus)** for the infrastructure and friendly assistance in academic and financial matters.

My gratitude to **colleagues, friends** and **family** for their support and friendship throughout the duration of the study.

I would like to thank my **husband and sons** for their love, understanding and support.

CHAPTER ONE

Introduction

Commercial multiplex deoxyribonucleic acid (DNA) typing kits incorporate a gender-specific marker for the purpose of human gender identification. Many of the kits in common practise use the amelogenin locus for this purpose due to its simplicity and ease to integrate into a routine procedure (La Fountain *et al*., 1998; Masibay *et al*., 2000; Holt *et al*., 2002; Krenke *et al*., 2002; Wallin *et al*., 2002).

Accurate gender determination is important in many disciplines. Short tandem repeat (STR) analysis is generally used in forensics and is increasingly being used in pathology (Shadrach *et al*., 2004), including testing for prenatal diagnosis of X-linked diseases, monitoring patients that have had bone marrow transplants to ensure the engraftment has been successful (Murphy *et al*., 2007) and even for resolving sample mix-ups (Van Deerlin and Leonard, 2000). In forensic DNA analysis, gender determination of the offender is often the initial information obtained and misleading information can lead to serious consequences in criminal investigations (Von Wurmb-Schwark *et al*., 2006).

Amelogenin is a single copy gene with homologues on both the X and Y-chromosomes (Salido *et al*., 1992) that differ in size and sequence (Nakahori *et al*., 1991). A 6 bp deletion on the X-chromosome generates two fragments of different size when two specific primer sets are used for amplification and this is used to discriminate between the amelogenin gene on the X-chromosome (AMELX) and amelogenin gene on the Y-chromosome (AMELY) alleles (Sullivan *et al*., 1993). Many deletions and rearrangements have been described within the Y-chromosome (Cadenas *et al*., 2007), but since the recent reports of AMELY allele dropouts in phenotypically unaffected males (discussed in Chapter Two), the reliability of the amelogenin gene test for gender identification has been questioned. When a male sample exhibits a dropout of the AMELY allele it will be falsely genotyped as female. Brinkman (2002) states that the implications of wrong gender assignment are so serious that commercial companies should include an additional Y-marker in the multiplex kits.

Current reports on AMELY-negative males indicate that the dropout is due to a deletion (see Chapter Two). Results from typing loci surrounding the AMELY locus suggest that there might be multiple different forms of the deletion. It appears as if AMELY deletions are more often observed in certain populations, in particular those from the Indian subcontinent. Limited information is available to determine if the different forms of the deletions present in other populations, are related by descent.

The broad aim of this study is to determine the deletion size in four black South African individuals, consisting of two father and son pairs, and to compare it with those previously reported. A further aim is to determine how rare these deletions are in the black South African population and their relationship to each other.

The study presented in this thesis is the first report and investigation of AMELY deletions in a black population. In Chapter Three, the methods used for investigation of the AMELY-negative males are described. These protocols were used to achieve the objectives listed in Chapter Two. The results obtained from this study are presented in Chapter Four and conclusions reached from these results are described in Chapter Five.

CHAPTER TWO

Sequence variation of the amelogenin gene on the Y-chromosome

---

In humans, the Y-chromosome establishes maleness. Sequence differences between the X and Y-homologues of the amelogenin gene have been used to differentiate males from females in numerous types of gender determination analyses. Various commercial polymerase chain reaction (PCR) multiplex kits have incorporated the amelogenin locus for gender identification. Several studies have however reported mutations in the Y-homologue, which can cause the mistyping of males as females. Accurate gender determination is especially important in prenatal diagnosis of X-linked diseases and in forensic investigations, thus incorrect gender assignations can have far-reaching consequences.

2.1 EVOLUTION OF THE Y-CHROMOSOME

Studies suggest that although the X and Y-chromosomes currently differ in size and quantity of genes in mammals, they were initially a homologous pair of autosomes (Graves, 1995). It is thought that certain autosomes obtained a sex-determining role and because of the lack of recombination between the X and Y-chromosomes, each evolved differently (Charlesworth, 1996). The Y-chromosome is small and represents only 2-3% of the haploid genome (Graves, 1995). In contrast to the Y-chromosome, the X-chromosome is large and contains several thousand genes.

Since most of the Y-chromosome does not participate in meiotic recombination, preventing the accumulation of harmful mutations by natural selection became inefficient (Charlesworth, 1996), thus selection will only act to inhibit the degeneration of genes with male-specific functions. As most of the genes specific to the Y-chromosome are transcribed in the testes, this indicates a function in spermatogenesis (Stuppia *et al.*, 1998). Genes present on both the X and Y-chromosomes are probably the result of genes whose Y-linked copies have not degenerated. Several X-linked genes have inactive homologues on the Y-chromosome (Lahn and Page, 1999). Dosage compensation is

necessary to ensure similar levels of gene products from X-linked genes in males and females (Charlesworth, 1996).

Lahn and Page (1999) determined in sex chromosomes that the region surrounding the sex determining region Y (SRY) gene stopped recombining first, after which non-recombining regions evolved along most of the chromosome's length. As the Y-chromosome does not recombine during meiosis, classical linkage mapping is impossible and the high density of repeated sequences makes physical mapping and sequencing difficult (Bachtrog and Charlesworth, 2001). Physical mapping of the Y-chromosome has largely depended on naturally occurring deletions. The creation of a deletion map and the ordering of the DNA loci, are useful in locating genes and for studying the structural diversity of the Y-chromosome within and among human populations (Quintana-Murci and Fellous, 2001). Vergnaud *et al*. (1986) presented the first molecular map of the Y-chromosome. They subdivided the Y-chromosome into seven intervals corresponding to the naturally occurring deletions (see Figure 2.1). Vollrath *et al*. (1992) subsequently constructed a more precise deletion map based on the detection of about 200 sequence-tagged sites (STSs). These STSs along the Y-chromosome have been used to define the smallest deleted regions associated with specific phenotypes, thus identifying genes on the Y-chromosome and the origins of Y-chromosome disorders (Jobling *et al*., 1996; Foresta *et al*., 2001; Tilford *et al*., 2001). An essentially complete physical map of the Y-chromosome was generated by Foote *et al*. in 1992.

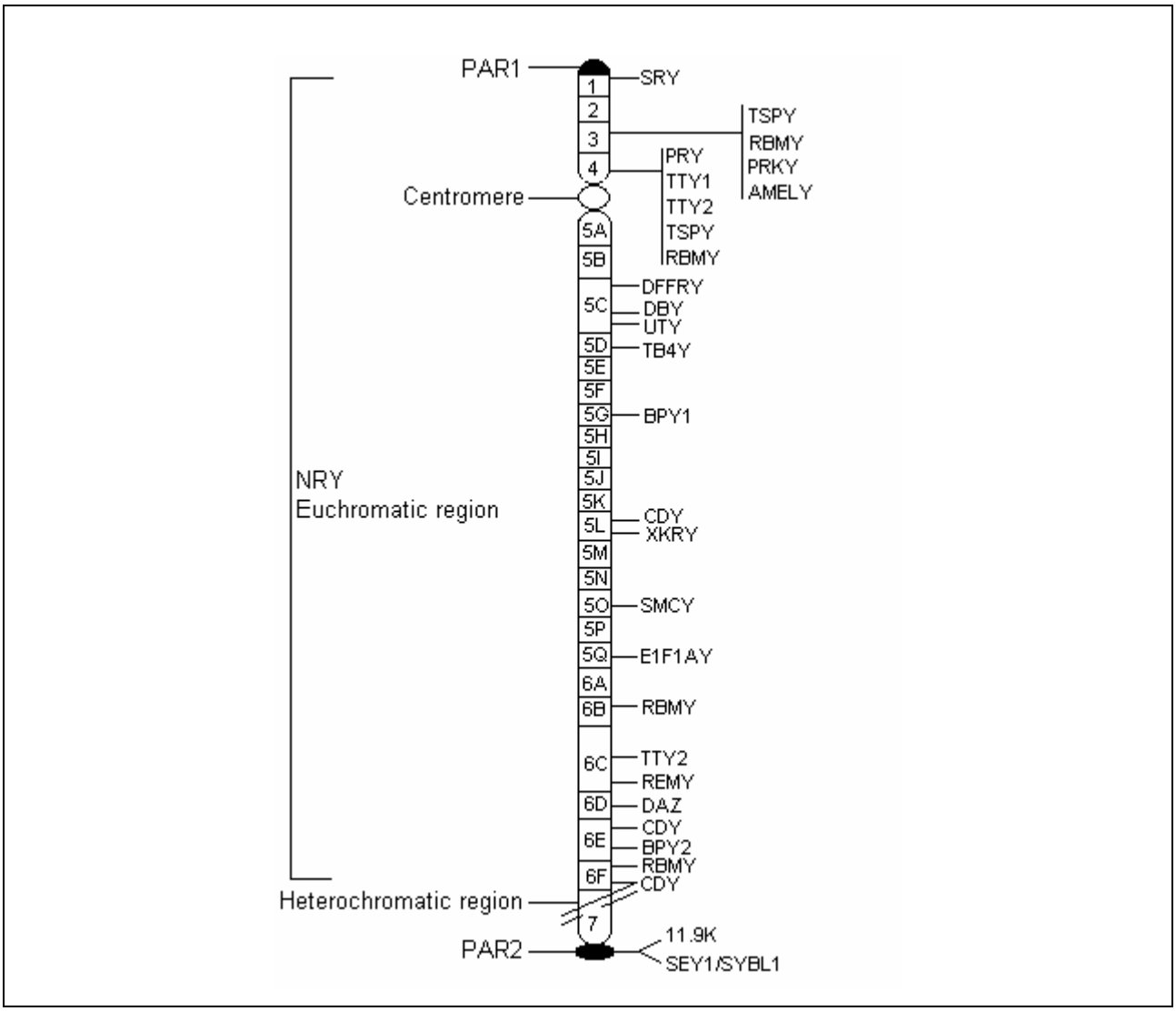
2.2 STRUCTURE OF THE Y-CHROMOSOME

Cytogenetic observations based on chromosome-banding studies identified different regions i.e. the pseudoautosomal regions (PAR1 and PAR2) as well as the euchromatic and heterochromatic regions. PAR1 and PAR2 represent only about 5% of the Y-chromosome (Rozen *et al*., 2003). The greater part of the Y-chromosome is the non-recombining region (NRY). This includes the euchromatic and the heterochromatic regions. The euchromatic region has highly repetitive sequences and contains genes responsible for important biological functions. These genes are discussed in Section 2.3. A schematic representation of the Y-chromosome is given in Figure 2.1.

The pseudo-autosomal regions consist of PAR1 which is located at the terminal region of the short arm (Yp) and PAR 2, which is located at the terminal region of the long arm (Yq).

PAR1 and PAR2 consist of 2,600 and 320 kilo-bases (kb) respectively. During meiosis in a male individual, the pseudoautosomal regions of the Y-chromosome pair and exchange genetic material with the pseudoautosomal regions of the X-chromosome (Burgoyne, 1982). These regions are also referred to as the X-Y homologous regions. Genes located in the pseudoautosomal regions are therefore inherited in the same manner as autosomal genes (Lahn and Page, 1999). As these genes do not present with strict sex linkage the regions are called pseudoautosomal regions (PAR). These regions contain genes that are active on the Y-chromosome and are not subject to inactivation in females (Polani, 1982).

Figure 2.1: Schematic representation of the Y-chromosome



Adapted from Quintana-Murci and Fellous (2001) and Foresta *et al*. (2001). PAR1 = Pseudoautosomal region 1; NRY = Non-recombining region Y-linked; PAR2 = Pseudoautosomal region 2; SRY = sex determining region Y-linked; TSPY = Testis-specific protein Y-linked; PRKY = Protein kinase Y-linked; AMELY = Amelogenin Y-linked; PRY = Putative tyrosine phosphatase protein-related Y-linked; TTY1 = Testis transcript Y1; TTY2 = Testis transcript Y2; RBMY = RNA-binding motif Y-linked; DFFRY = Ubiquitin-specific protease (Drosophila fat-facets related Y); DBY = Dead box Y-linked; UTY = Ubiquitous TPR motif Y-linked; TB4Y = Thymosin beta-4 Y-chromosomal isoform; BPY1 = Basic protein Y1; CDY = Chromodomain Y-linked; XKRY = XK-related Y-linked; SMCY = Selected mouse cDNA Y-linked; E1F1AY = Translation initiation factor 1A Y-linked; DAZ = Deleted in azoospermia; BPY2 = Basic protein Y2; SEY1/SYBL1 = Synaptobrevin-like 1.

The euchromatic region is distal to PAR1 and includes the paracentromeric region of the short arm, the centromere and the paracentromeric region of the long arm. The heterochromatic region consists of the distal region of the long arm, corresponding to Yq12. This region is assumed to be genetically inert and is polymorphic in length in different male populations. The heterochromatic region is composed of repetitive satellite DNA and predominantly consists of the 3.5 kb (DYZ1) and the 2.5 kb (DYZ2) repeat families which together account for 50-70% of the Y-chromosome (Bachtroch and Charlesworth, 2001). The Y-chromosome contains many short interspersed repetitive elements in mammalian genomes, containing *Alu 1* recognition sequence (Alu repeats) and long interspersed nuclear element (LINE) repetitive elements, which are present throughout the human genome (International Human Genome Sequencing Consortium, 2001). Several phenotypes have been associated with the non-recombining region of the Y-chromosome. That the Y-chromosome was involved in male sex determination came initially from the observations that XY or XYY individuals develop testes whereas XX or X0 individuals develop ovaries (Jacobs and Strong, 1959).

2.3 GENES ON THE Y-CHROMOSOME

Turner syndrome is characterised by a female 45X karyotype. The characteristics of this syndrome are growth failure, infertility, anatomical abnormalities and selective cognitive deficiency. Turner syndrome is caused by a haplo-insufficiency of genes on the X-chromosome that are common to both the X and Y-chromosome (Jacobs and Strong, 1959). These genes must escape X-inactivation to account for the difference between 45X and 46XX. In 46XY, these genes have a male counterpart on the Y-chromosome and sufficient product is thus produced. There appears to be different loci on the X and Y-chromosome associated with Turner syndrome characteristics (Rao *et al*., 1997; Ellison *et al*., 1996; Barbaux *et al*., 1995). The following sections highlight the major genes present on the Y-chromosome.

2.3.1 Sex determination

In 1990, the gene necessary for testes development, named SRY, was identified on the Y-chromosome (Sinclair *et al*., 1990). This gene is located on the short arm of the Y-chromosome. It consists of one exon that encodes a protein of 204 amino acids. Subsequently, more genes that map to the Y-chromosome have been identified (Ma *et al*.,

1993). The genetic map of the non-recombining region of the Y-chromosome is represented in Figure 2.2. There are two main classes of genes (Bachtrog and Charlesworth, 2001). Genes in the first class are the housekeeping genes, which have homologues on the X-chromosome. Genes in the second class, however, have a testis-specific function and thus no X-linked homologues.

Figure 2.2: Genetic map of the non-recombining region of the Y-chromosome

Function	Copy no.	Genes	Genes	Copy no.	Function
		PAR1			
Transcription factor - sex determination	1	SRY	RPS4Y	1	Protein of small ribosomal subunit
			ZFY	1	Zinc finger transcription factor
Testis transcript 1	m	TTY1	PCDHY	1	Protocadherin – adhesion
Cyclin B binding protein	m	TSPY	PRKY	1	Ser/Thr protein kinase
			AMELY	1	Tooth enamel formation
Protein tyr phosphatase	m	PRY			
Testis transcript 1	m	TTY1			
Testis transcript 2	m	TTY2			
Cyclin B binding protein	m	TSPY			
		Centromere			
			USP9Y	1	Deubiquinating enzyme
			DBY	1	DEAD-box RNA helicase
			UTY	1	TPR-motif
			TB4Y	1	Actin sequestration
			VCY	2	Variable charged protein
Chromodomain protein	m	CDY			
Membrane transport Protein	m	XKRY	SMCY	1	Transcription factor
			EIF1AY	1	Translation initiation factor
			RBMY	30	RNA-binding protein
Protein tyr phosphatase	m	PRY			
Testis transcript 2	m	TTY2			
RNA-binding protein	4	DAZ	RBMY	30	RNA-binding protein
Basic protein	m	BPY2			
Protein tyr phosphatase	m	PRY			
Chromodomain protein	m	CDY			
		Heterochromatin			
		PAR2			
Y-chromosome genes not present on the X-chromosome			Y-chromosome genes with homologues on the X-chromosome		

Adapted from Bachtrog and Charlesworth (2001). m = multiple copies; DEAD = aspartate-glutamate-alanine-aspartate; TPR = tetratricopeptide repeat; Tyr = tyrosine; Ser = serine; Thr = threonine; RNA = ribonucleic acid.

2.3.2 Spermatogenesis

Tiepolo and Zuffardi in 1976 described the function of the Y-chromosome in spermatogenesis after observing cytogenetically visible deletions in six azoospermic

individuals. They postulated the existence of a locus termed Azoospermia factor (AZF), which was required for successful spermatogenesis. AZF was localised to Yq11.23 and was further subdivided into 3 regions namely Azoospermia factor region a or AZFa, Azoospermia factor region b or AZFb and Azoospermia factor region c or AZFc (Vogt *et al*., 1996; Repping *et al*., 2003; Carvalho *et al*., 2004). A number of genes have been identified in the AZFa region. Three of these genes seem to be housekeeping genes and the *Drosophila* developmental gene fats facets (DFFRY) gene is hypothesised to be important for gametogenesis (Brown *et al*., 1998). Five genes have been described in the AZFb region. One of them, the RNA-binding motif (RBM) gene has been proposed as a candidate gene for infertility (Ma *et al*., 1993). The AZFc region contains the Deleted in Azoospermia (DAZ) gene cluster, among others. This gene is expressed in the testis and has also been theorised to be relevant in gametogenesis (Reijo *et al*., 1996).

2.3.3 Oncogenesis

Loss and rearrangement of the Y-chromosome have been linked to several types of cancer, such as lung cancer (Centre *et al*., 1993), oesophageal carcinoma (Hunter *et al*., 1993), bladder cancer (Sauter *et al*., 1995), male sex cord stromal tumours (De Graaff *et al*., 1999) and prostate cancer (Vijayakumar *et al*., 2006). No proto-oncogenes, tumour suppressor genes or mismatch repair genes have however been localised to the Y-chromosome. The only cancer locus linked to the Y-chromosome is the gonadoblastoma locus on the Y-chromosome (GBY) of which the most likely candidate gene appears to be testis specific protein Y (TSPY) or Cyclin B binding protein (Tsuchiya *et al*., 1995).

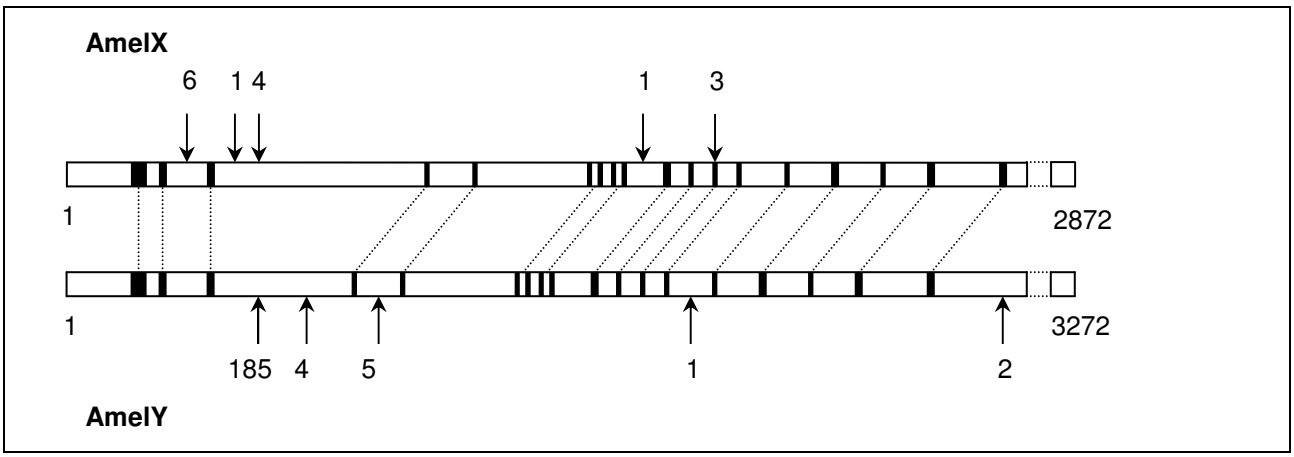
2.3.4 Amelogenin

The amelogenin gene produces a protein important in the development of dental enamel. It appears to regulate the formation of enamel crystallites by providing the hydrophobic environment necessary for the calcium hydroxyapatite crystals (Salido *et al*., 1992). Lau *et al*. (1989) mapped the mouse amelogenin gene to the X-chromosome. Homologous sequences to the mouse amelogenin cDNA have been determined in humans on the short arm of the X-chromosome, i.e. the p22.1-p22.3 region, as well as in the pericentric region of the Y-chromosome (Salido *et al*., 1992). The human amelogenin X gene is 2,872 base pairs (bp) long and is located on Xp22, while the amelogenin Y gene is 3,272 bp long and is located on the Yp11.2 region of the Y-chromosome (Bailey *et al*., 1992; Nakahori *et al*., 1991).

X-linked amelogenesis imperfecta (AI), is a genetic disorder affecting the formation of enamel, and has been mapped to Xp22 (Lagerström *et al*., 1990). This suggests that defects in the amelogenin gene may be responsible for X-linked AI. A deletion in the amelogenin gene has been determined in a family with X-linked AI (Lagerström *et al*., 1991). Results indicate that the amelogenin locus on the Y-chromosome encodes a functional protein, but the expression level of this gene is 10% that of the amelogenin locus on the X-chromosome. The different expression levels of the amelogenin loci is probably due to alterations in the promoter regions, which share ca. 80% sequence similarity. Both promoters contain the identical TATA and CCAAT boxes (Salido *et al*., 1992). Other elements, which may influence the transcription of the amelogenin genes, have not been identified. Regions of homology and regions of deletions can be determined when the amelogenin sequences are aligned. GenBank®[1] accession no. M55418 for the amelogenin X gene and GenBank® accession no. M55419 for the amelogenin Y gene, are used for alignment of the sequences (Haas-Rochholz and Weiler, 1997). As presented in Figure 2.3, a total of 19 regions of absolute homology can be displayed varying in size from 22 to 80 bp.

The amelogenin gene has been studied in murine (Lau *et al*., 1989), bovine (Buel *et al*., 1995), and porcine (Salido *et al*., 1992) models as well as in humans (Nakahori *et al*., 1991). Large parts of the protein sequence are highly conserved among these species (Salido *et al*., 1992).

Figure 2.3: Schematic representation of the regions of homology and deletions between the amelogenin X and amelogenin Y genes



Adapted from Haas-Rochholz and Weiler (1997). The black boxes represent regions of homology; the regions of deletions are represented as arrows. The numbers indicate the size of the deletions in bp.

[1] GenBank® is a registered trademark of the National Institute of Health and Human Services for the Genetic Sequence Data Bank, Bethesda, MD, USA.

2.4 AMELOGENIN- BASED GENDER DETERMINATION

A number of commercial DNA profiling PCR kits use the amelogenin (AMEL) gene for gender determination. These kits co-amplify the AMEL locus in combination with autosomal STR loci (Eng *et al*., 1994; La Fountain *et al*., 1998; Pouchkarev *et al*., 1998; Zehner *et al*., 1998; Masibay *et al*., 2000; Hayashi *et al*., 2000; Holt *et al*., 2002; Krenke *et al*., 2002; Wallin *et al*., 2002; Vauhkonen *et al*., 2004). The AMEL gene sex determination test, tests for both versions of this gene. There is a deletion of 6 bp between the AMEL genes on the X and Y-chromosomes (Mannucci *et al*., 1994), as presented in Figure 2.3. This test is widely used because it is sensitive, easy to interpret and can be co-amplified in a single multiplex reaction (Chang *et al*., 2003). As the amplification of both chromosomes occurs in a single reaction, the amplification of the X-chromosome also acts as an internal positive control (Shadrach *et al*., 2004).

The most frequently used PCR-based sex test is the one described by Sullivan *et al*. (1993). By using the same primers, PCR products of 106 bp and 112 bp are generated from the X and Y-chromosomes respectively. These primers bind to the first intron of the gene. Because the AMEL primer sets are adaptable to multiplexing, they were included in DNA profiling systems such as AmpFℓSTR®[1] (Applied Biosystems) and PowerPlex®[2] (Promega).

The first report of phenotypically unaffected males presenting as females with the AMEL sex test was by Santos *et al*. (1998). They observed two Sri Lankan males, from a group of 24, wherein only the AMELX sequence was amplified. They suggested that these males were AMELY-negative due to a deletion in the Y-chromosome that included the AMELY sequence. Since this report, Roffey *et al*. (2000) and Henke *et al*. (2001) have observed the same phenomenon. Roffey *et al*. (2000) determined an AMELY-negative male in Australia and Henke *et al*. (2001) reported on a Moroccan father-son pair that was AMELY-negative. These authors proposed that the AMELY sequence did not amplify as a result of a point mutation occurring in the primer annealing region. When the Austrian National DNA database was checked as a routine procedure, six individuals were determined to be AMELY-negative among 29,432 phenotypic males (Steinlechner *et al*., 2002). This discrepancy led Steinlechner *et al*. (2002) to design an alternative amelogenin primer set. It was designed with sequence data obtained from the GenBank® sequence

[1] AmpFℓSTR® is a registered trademark of Life Technologies, Carlsbad, CA, USA.
[2] PowerPlex® is a registered trademark of the Promega Corporation, Madison, WI, USA.

database (accession numbers M55418 and M55419). These primers result in amplified products of 219 bp and 225 bp from the X and Y-chromosome, respectively. The locations of the primer sequences designed by Sullivan *et al.* (1993) and Steinlechner *et al.* (2002) are indicated in Chapter 3 Section 3.5.

Amplification with the new primers gave the same result, thus indicating that a deletion in the Y-chromosome encoded gene was the cause of the failure to amplify AMELY rather than a polymorphism at one of the priming sites. Previously, Roffey *et al*. (2000) reported the unsuccessful amplification of Y-chromosome sequences from a phenotypically unaffected male. As PCR amplification was successful with alternate amelogenin primers, he attributed this initial failure to a mutation in the primer binding site.

In a study of 270 males from India, Thangaraj *et al*. (2002) discovered five males that tested negative for AMELY. The SRY gene was present in all these samples thus confirming maleness. Southern blot hybridisation allowed for the determination of a deletion in the p-arm of the Y-chromosome, thus opposing the hypothesis that a point mutation was present in the primer binding site. One of the males were from a paternity case that presented with a 50% paternal contribution, demonstrating that he was fertile despite the deletion on the short arm of the Y-chromosome. In a study of 113 Malay, 113 Chinese and 112 Indian males from the Malaysian population, four Indian males and one Malay male presented with a deletion of AMELY (Chang *et al*., 2003). Amplification with a primer set that spanned the full region of the common amelogenin primer-annealing sites, demonstrated that the absence of AMELY was not due to a point mutation at the primer-binding sites, but was caused by a deletion in the Y-chromosome. As a result of this study, a larger male population of 334 Malays, 331 Chinese and 315 Indians were subjected to STR analysis with the incorporated amelogenin sex test (Chang *et al*., 2007). In this study, 18 AMELY-negative males were observed. Twelve males were from the database and six were from current casework analyses. Self-reported ethnicity revealed that fourteen individuals were from the Indian group, four from the Malay group and none in the Chinese group. These frequencies presented with concordance to the previous data published by these authors (Chang *et al*., 2003).

The high frequency of AMELY-negative males reported in the Indian sub-continent prompted Kashyap *et al*. (2006) to screen the amelogenin locus of 4,257 males. These individuals belonged to 104 different populations in India. Ten samples were determined to be AMELY-negative. The AMELY-negative males belonged to both caste groups and tribal

populations. Testing of these samples with primers from Sullivan *et al*. (1993) and Steinlechner *et al*. (2002) indicated the absence of the Y-chromosome specific product, suggesting a deletion of the amelogenin region was responsible for the failure of the sex test (Kashyap *et al*., 2006). In a study of 77 males from Nepal, Cadenas *et al*. (2007) reported an absence of the AMELY allele in five individuals. Amplification with a different primer pair yielded the same result and these results were thus characterised as deletions and not point mutations in the primer-annealing region.

From the results reported by Chang *et al*. (2003, 2007), it appeared as if AMELY-negative males were absent in the Chinese population. Kao *et al*. (2007) however reported two phenotypically normal male samples in Taiwan that presented with only AMELY amplification and three individuals that presented with AMELX only amplification. They also reported that three female samples revealed XY amplification. These results were obtained from approximately 80,000 males and 20,000 females.

In the Israeli Defence Force, blood samples of soldiers are stored on Flinders Technology Associates (FTA) filter paper cards for possible future reference. During a quality control check, a sample taken from a male soldier was determined to be AMELY-negative and therefore he was genotyped as being female (Michael and Brauner, 2004). Karyotyping confirmed that the soldier contained a normal Y-chromosome. Two different primer sets failed to produce an AMELY-related PCR product, suggesting that this sample contained a deletion (Michael and Brauner, 2004). This was the only AMELY-negative male determined in 96 samples.

Mitchell *et al*. in 2006 reported on five Australian males that were AMELY-negative. These males also tested negative with the alternate primer pair developed by Sullivan *et al*. (1993). Two of the individuals were of Indian origin, one of Italian origin and the remaining two of Sri Lankan origin. No information was available on their fertility status. Mitchell *et al*. (2006) approached the major forensic laboratories throughout Australia for information on AMELY-negative males in order to determine the frequency. Frequency estimates of AMELY-negative males in the Australian population were 0.02% in an estimated 109,000 males tested (Mitchell *et al*., 2006). The results of the frequency analysis of AMELY-negative males from the different publications are summarised in Table 2.1.

From Table 2.1 it can be determined that there is a higher frequency of AMELY-negative males that are of Indian or Sri Lankan origin. A much lower frequency was observed in the

Caucasian population groups. Dropout of the AMELY allele has been revealed to be due to a deletion of the amelogenin region on the Y-chromosome. Shadrach *et al*. (2004) observed that of 327 males tested, one failed to present with the AMELX allele. A product was however observed with alternate primers on the X-chromosome. Subsequent sequence analysis revealed a C to G mutation at the nucleotide complementary to the 3' end of the reverse amelogenin primer. As this mutation was in an intron and not at the splice acceptor/donor sites, no functional change is expected.

Table 2.1: Frequency distribution of AMELY-negative males in different populations

Population	No. males tested	No. negative	% negative	Reference
Austrian	29,432	6	0.018	Steinlechner *et al.,* 2002
Israeli	96	1	1.040	Michael *et al.,* 2004
Italian	13,493	2	0.015	Lattanzi *et al.,* 2005
Australian	109,000	22	0.020	Mitchell *et al.,* 2006
England	2,000	2	0.100	Chang *et al.,* 2007
Spain	1,000	1	0.100	Chang *et al*., 2007
Sri Lanka	24	2	8.330	Santos *et al.,* 1998
India (general)	270	5	1.850	Thangaraj *et al.,* 2002
Malaysian Malays	113	1	0.880	Chang *et al.,* 2003
Malaysian Malays	334	2	0.600	Chang *et al.,* 2007
Singapore Malays	182	1	0.600	Yong *et al*., 2007
Malaysian Indians	112	4	3.570	Chang *et al.,* 2003
Malaysian Indians	315	11	3.490	Chang *et al.,* 2007
Singapore Indians	175	3	1.760	Yong *et al*., 2007
Indian	4,257	10	0.230	Kashyap *et al.,* 2006
South Indian	100	1	1.000	Chang *et al.,* 2007
Nepal	77	5	6.490	Cadenas *et al.,* 2007
Chinese	113	0	0.000	Chang *et al*., 2003
Chinese	331	0	0.000	Chang *et al.,* 2007
Chinese	210	0	0.000	Yong *et al*., 2007
Taiwan	80,000	3	0.004	Kao *et al.,* 2007

2.5 MICROSATELLITE HAPLOTYPES OF AMELY-NEGATIVE MALES

In the absence of recombination, the combination of allelic states at loci over the length of the Y-chromosome is represented by haplotypes, which are generally inherited from generation to generation (Jobling *et al*., 2003). The Y-chromosome passes from father to son and largely escapes meiotic recombination due to the fact that meiotic recombination occurs between homologous chromosomes. The X and Y-chromosomes only have similar

sequences over a small portion of their length, the pseudoautosomal regions and meiotic recombination only occurs at these regions (Strachan *et al*., 2004). A Y-chromosome haplotype is therefore usually inherited unchanged from generation to generation and can be used to determine paternal lineages. The amplification of Y-chromosomal STR markers is performed to confirm the male phenotype in AMELY-negative males, to determine if Y-STR haplotypes were shared between individuals harbouring a Y-chromosome deletion and to possibly determine the extent of the deletion.

Amplification of eight Y-STR markers for the five AMELY-negative samples described by Steinlechner *et al*. (2002) was successful. These results are given in Table 2.2. From the results it appears that the suspected deletion was limited to the amelogenin region on the Y-chromosome. Five different haplotypes were determined and according to Steinlechner *et al*. (2002) this could mean that the deletion is likely to have a common ancestor. Thangaraj *et al*. (2002) examined Y-STR loci on six AMELY-negative samples. Four different haplotypes were revealed (Table 2.2), with two individuals sharing a haplotype. The two individuals that shared a haplotype belonged to the same religious group, indicating possible common ancestry. Y-STR haplotype analysis suggested to Thangaraj *et al*. (2002) that there were at least four different paternal lineages carrying this deletion. Lattanzi *et al*. (2005) also observed two different Y-STR haplotypes in the two Italian AMELY-negative males (Table 2.2).

According to Mitchell *et al*. (2006), none of the five AMELY-negative samples within their investigation shared an identical Y-STR haplotype (Table 2.2). Certain haplotypes were relatively similar and differed only at two to three loci indicating a common ancestor. One of the haplotypes was very different from the others and Mitchell *et al.* (2006) believed that this reflected a distinct paternal lineage. They also speculated that individuals with similar haplotypes probably shared the same deletion. Analysis with four Y-STR markers yielded eight different haplotypes in the ten AMELY-negative samples published by Kashyap *et al*. (2006). These haplotypes are presented in Table 2.2. In two cases, two of the samples shared the same haplotypes. In the study by Cadenas *et al*. (2007), five AMELY-negative males were determined. Y-STR analysis was conducted on these samples. Four related haplotypes were discovered (Table 2.2) to share seven to ten alleles. As the haplotypes were similar, the authors suggested that the deletions arose in the same paternal lineage. From these results it seems likely that a number of the AMELY deletions were inherited by descent, but others occurred in different paternal lineages.

Table 2.2: Y-chromosome haplotypes in AMELY-negative males.

Origin	19	385	388	389I	389II	390	391	392	393	434	435	437	438	439	448	456	458	635	DYAA7.2	Y-GATA	Reference
Moroccan	14	---	---	10	26	23	10	11	12	---	---	---	---	---	---	---	---	---	---	---	Henke *et al*., 2001
Austrian	16	13,14	---	12	28	22	10	12	13	---	---	---	---	---	---	---	---	---	---	---	Steinlechner *et al*., 2002
Austrian	14	13,15	---	13	29	22	10	11	12	---	---	---	---	---	---	---	---	---	---	---	
Austrian	13	15,18	---	13	30	23	10	11	12	---	---	---	---	---	---	---	---	---	---	---	
Austrian	14	14,14	---	12	27	22	10	11	13	---	---	---	---	---	---	---	---	---	---	---	
Austrian	13	11,14	---	13	29	24	10	13	13	---	---	---	---	---	---	---	---	---	---	---	
Austrian	No	No	---	No	No	No	No	No	No	---	---	---	---	---	---	---	---	---	---	---	
Indian	14	---	---	13	30	22	10	---	14	---	---	---	---	---	---	---	---	---	---	---	Thangaraj *et al*., 2002
Indian	14	---	---	13	30	22	10	---	14	---	---	---	---	---	---	---	---	---	---	---	
Indian	16	---	---	14	30	24	11	---	12	---	---	---	---	---	---	---	---	---	---	---	
Indian	17	---	---	13	30	25	10	---	12	---	---	---	---	---	---	---	---	---	---	---	
Indian	14	---	---	13	30	21	11	---	14	---	---	---	---	---	---	---	---	---	---	---	
Malay	---	---	---	---	---	25	---	---	---	---	---	---	10	13	---	---	---	---	---	---	Chang *et al*., 2003
Indian	---	---	---	---	---	25	---	---	---	---	---	---	9	11	---	---	---	---	---	---	
Indian	---	---	---	---	---	24	---	---	---	---	---	---	9	13	---	---	---	---	---	---	
Indian	---	---	---	---	---	25	---	---	---	---	---	---	9	12	---	---	---	---	---	---	
Indian	---	---	---	---	---	23	---	---	---	---	---	---	9	13	---	---	---	---	---	---	
Indian	---	---	---	---	---	24	---	---	---	---	---	---	9	11	---	---	---	---	---	---	
Italian	16	---	---	14	30	22	11	11	14	---	---	---	10	11	---	---	---	---	---	---	Lattanzi *et al*., 2005
Italian	13	---	---	13	29	24	11	13	13	---	---	---	12	12	---	---	---	---	---	---	
Italian	13	---	---	13	---	24	11	13	13	9	11	9	12	12	---	---	---	---	---	---	Mitchell *et al*., 2006
Asian	15	---	---	12	---	24	11	11	13	9	11	9	9	12	---	---	---	---	---	---	
Asian	15	---	---	12	---	23	10	11	12	9	11	9	9	12	---	---	---	---	---	---	
Indian	15	---	---	12	---	24	11	11	12	9	11	9	9	11	---	---	---	---	---	---	
Indian	16	---	---	12	---	24	10	11	12	---	---	---	9	---	---	---	---	---	---	---	
Indian	15	---	---	12	29	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	Kashyap *et al*., 2006
Indian	15	---	---	12	29	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Indian	15	---	---	13	30	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Indian	15	---	---	13	30	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Indian	15	---	---	13	29	24	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Indian	14	---	---	13	29	24	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Indian	15	---	---	11	30	23	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Indian	15	---	---	13	26	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Indian	14	---	---	13	30	23	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Indian	15	---	---	11	28	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Indian	15	13,17	---	12	29	25	10	11	12	---	---	15	9	11	18	14	No	22	---	12	Chang *et al*., 2007
Indian	16	13,17	---	12	28	24	10	11	12	---	---	15	9	13	18	13	No	24	---	11	
Indian	15	14,17	---	12	28	25	11	11	12	---	---	14	9	12	18	13	No	20	---	11	
Indian	17	14,17	---	12	28	23	10	11	12	---	---	15	9	13	19	14	No	22	---	No	

Origin	19	385	388	389I	389II	390	391	392	393	434	435	437	438	439	448	456	458	635	DYAA7.2	Y-GATA	Reference
Malay	13	11,18	---	12	26	25	11	12	13	---	---	14	10	13	21	15	No	19	---	12	Chang *et al.*, 2007 (continued from above)
Indian	15	13,16	---	12	27	24	10	11	12	---	---	15	9	11	19	13	No	23	---	No	
Indian	15	13,16	---	12	28	23	10	11	12	---	---	15	9	12	19	13	No	21	---	11	
Indian	15	14,17	---	11	28	24	10	11	12	---	---	15	9	12	19	13	No	21	---	11	
Indian	14	15,17	---	13	31	24	10	11	14	---	---	14	10	11	19	16	No	18	---	12	
Indian	15	14,17	---	12	28	23	11	11	12	---	---	15	9	12	18	13	No	21	---	11	
Indian	16	13,17	---	12	28	25	10	11	12	---	---	15	9	11	18	13	No	21	---	11	
Indian	17	13,18	---	12	29	25	10	12	12	---	---	15	9	12	18	13	No	21	---	11	
Malay	15	13,18	---	12	28	24	10	11	12	---	---	15	9	12	17	13	No	20	---	11	
Malay	15	13,17	---	12	28	24	10	11	12	---	---	14	9	13	18	13	No	21	---	10	
Indian	16	13,17	---	12	28	24	10	11	12	---	---	15	9	12	21	13	No	21	---	11	
Indian	14	15,18	---	13	31	26	11	11	15	---	---	14	10	11	19	15	No	18	---	12	
Indian	16	13,16	---	12	28	23	10	11	12	---	---	15	9	13	18	13	No	22	---	11	
Malay	16	13,17	---	12	28	24	10	11	12	---	---	15	9	12	19	13	No	20	---	11	
Nepal	15	---	No	12	29	23	10	11	12	---	---	---	---	12	---	---	---	---	7	---	Cadenas *et al.*, 2007
Nepal	15	---	15	12	29	23	10	11	12	---	---	---	---	13	---	---	---	---	7	---	
Nepal	13	---	15	12	29	23	9	11	12	---	---	---	---	12	---	---	---	---	7	---	
Nepal	15	---	15	12	29	24	10	11	12	---	---	---	---	12	---	---	---	---	7	---	
Nepal	15	---	15	12	29	23	10	11	12	---	---	---	---	13	---	---	---	---	7	---	
Maldives	14	15,17	12	13	31	24	10	11	14	12	11	14	10	11	19	---	---	---	---	---	Jobling *et al.*, 2007
Indian	14	14,18	12	13	31	24	10	11	14	12	11	14	10	11	19	---	---	---	---	---	
English	15	15,15	13	13	31	23	10	12	13	11	11	14	10	11	20	---	---	---	---	---	
Afghan	15	13,16	15	12	28	25	10	11	12	11	11	15	9	13	19	---	---	---	---	---	
Australia	14	13,13	15	12	29	24	11	11	12	11	11	15	7	11	18	---	---	---	---	---	
Australia	15	13,16	15	12	28	24	9	11	12	11	11	15	9	12	19	---	---	---	---	---	
Maldives	15	14,18	16	13	30	24	10	11	12	11	11	15	9	12	19	---	---	---	---	---	
Australia	14	11,14	12	13	29	24	10	13	13	11	11	15	12	12	19	---	---	---	---	---	
Australia	14	11,14	12	13	29	24	10	13	13	11	11	15	12	12	19	---	---	---	---	---	
Australia	14	11,14	12	13	29	24	10	13	13	11	11	15	12	12	19	---	---	---	---	---	
Australia	14	11,14	12	13	29	24	10	13	13	11	11	15	12	12	19	---	---	---	---	---	
English	14	11,14	12	13	29	23	11	13	13	11	11	15	12	12	19	---	---	---	---	---	
Bedouin	14	13,17	17	13	29	22	11	11	12	11	11	14	10	11	20	---	---	---	---	---	

No = no amplification; --- = not performed.

Chang *et al.* (2003) determined the Y-STR haplotypes harboured by the individuals in their study with deletions of the amelogenin locus (Table 2.2). They observed that the four AMELY-negative Indian males had a 9-repeat allele at the DYS438 locus, similar to the two Sri Lankan individuals described by Santos *et al.* (1998). It was hypothesised that this allele and the AMELY deletion represent an old and stable haplotype in the Indian

population (Chang *et al*., 2003). They also noted that four European individuals did not have a similar haplotype, indicating possible independent origins (Chang *et al*., 2003). In 2007, Chang *et al*. published data on an additional twelve AMELY-negative males. They included the initial six individuals which were reported in 2003, and also included more Y-STR loci. An absence of the DYS458 locus was observed in all samples. Although the haplotypes determined were similar, none were identical (Table 2.2), indicating no correlation between the deletion and a distinct haplotype.

Thirteen AMELY-negative individuals were determined by Jobling *et al*. (2007). Y-STR haplotypes were determined for the deletion chromosomes, as a clustering of haplotypes within a Y-chromosome haplogroup would support a common ancestry (Jobling *et al*., 2007). Deletion chromosomes of class I (as discussed in Section 2.6) belonging to haplogroup R1b3 shared a single haplotype, while haplogroup H(xH2) class I chromosomes also had closely related haplotypes, implying a single deletion event in each of these cases (Jobling *et al*., 2007). Haplogroup I had two deletion chromosomes with different haplotypes, suggesting independent origins. The class II deletion chromosomes (see Section 2.6) were both within haplogroup R1b3, suggesting a common ancestry. Y-STR haplotypes of the two individuals in class II differed at only four Y-STR loci (Jobling *et al*., 2007). Data published to date, indicates there may be a possible founder event for some of the deletions, however most AMELY deletions occur due to independent events happening sporadically elsewhere (Chang *et al*., 2007).

2.6 CHARACTERISATION OF DELETIONS

The extensive use of the amelogenin PCR assay for gender determination, has presented many examples of males with deletions of the AMELY locus. To date, 65 of these deletions have been reported (Jobling *et al*., 2007). The following section highlights several characteristics of these deletions, particularly the size, possible origin, likely mechanism and genes lost in the deletions.

2.6.1 Size

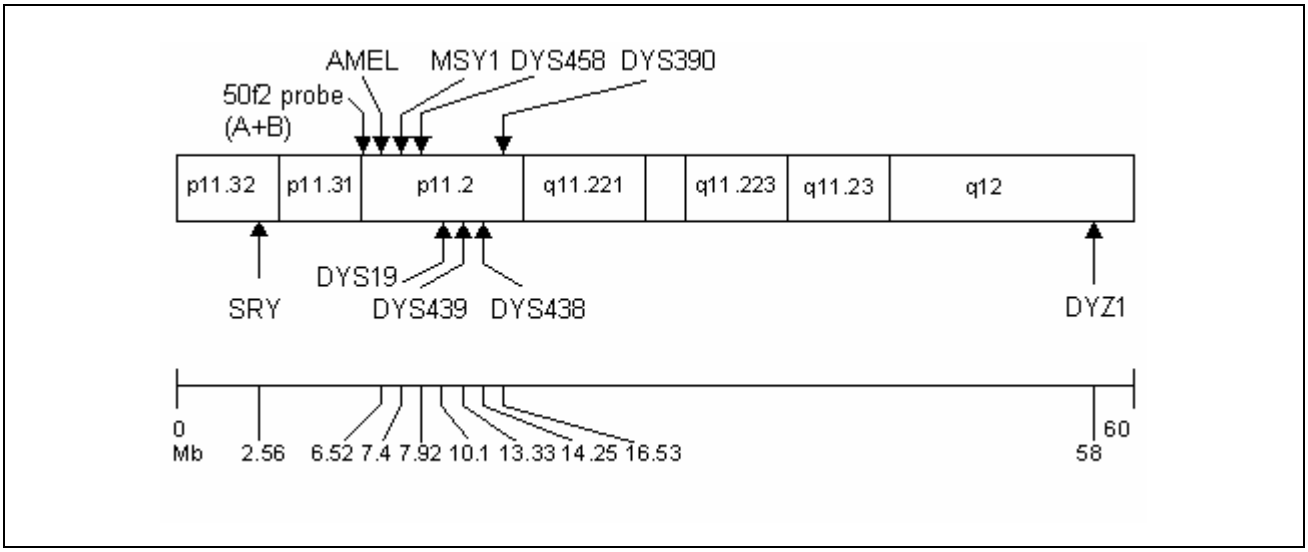
Results obtained by Roffey *et al*. (2000) revealed a phenotypically normal male genotyped as female. Other primer sets that amplify the amelogenin gene were used to determine if the allele dropout was due to a point mutation. The individual was also typed as female

upon analysis with these primer sets. Another sex test based on the DYZ1 repeat sequences, however, demonstrated that the individual carried a Y-chromosome. This led the authors to suggest that a single point mutation could be the cause of the failure of annealing as the different primers have an overlapping region (Roffey *et al*., 2000).

Amplifiable Y-STR loci were determined in five of the six AMELY-negative samples reported by Steinlechner *et al*. (2002). The assumed deletion present in these samples therefore seemed to be limited to the amelogenin-related region. For one sample, none of the Y-STRs could be amplified (however, the SRY gene did amplify), thus indicating that the deletion polymorphism may span from Yp11.2 on the short arm of the Y-chromosome up to Yq11.21 on the long arm. Alternatively, a translocation of the SRY gene to the short arm of the X-chromosome may have occurred (Steinlechner *et al*., 2002). Further investigation was not possible due to a limited amount of sample. To elucidate the cause of the AMELY-negative status in this individual it would be necessary to perform karyotyping to determine the presence of a Y-chromosome.

The SRY gene was present in all five AMELY-negative males described by Thangaraj *et al*. (2002). Using Southern blot hybridisation (with the 50f2 probe), Thangaraj *et al*. (2002) demonstrated a deletion of the 50f2 A and B loci (illustrated in Figure 2.4) in all the samples. This included about 1 Mb in the p-arm of the Y-chromosome as presented in Figure 2.4.

Figure 2.4: Schematic diagram demonstrating the location of AMEL with respect to other amplified loci on the Y-chromosome



Adapted from Chang *et al*. (2003). The location of male specific markers on the Y-chromosome is from the left: sex-determining region Y (SRY) gene, the Y-chromosome specific probe 50f2 (Jobling, 1994), amelogenin (AMEL) gene, minisatellite MSY1 locus, five Y-STR loci (DYS458, DYS19, DYS439, DYS438, DYS390) and the DYZ1 locus. The approximate distances between loci are also indicated.

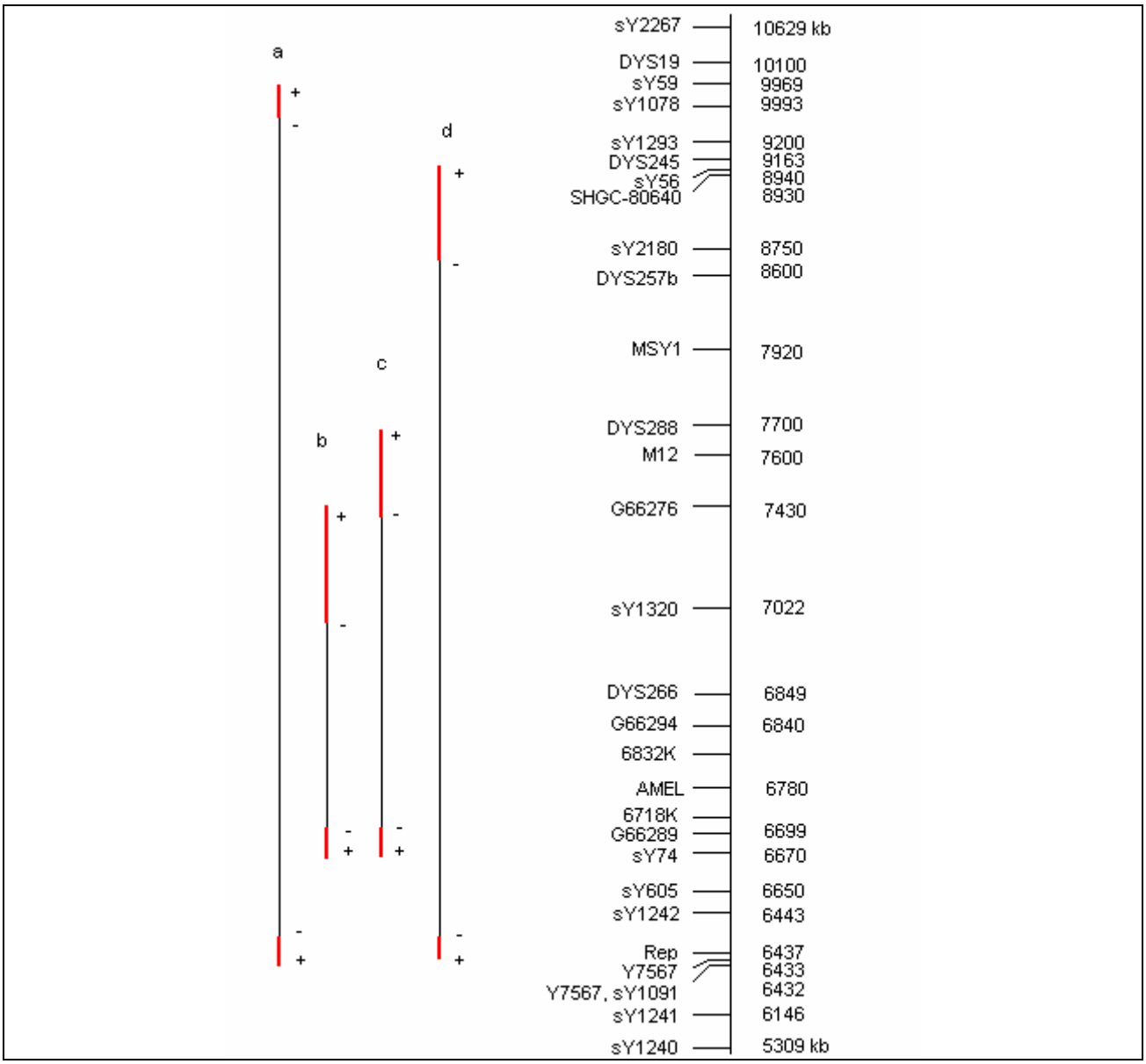
In addition, one infertile male also presented with a deletion of the AZFb and AZFc regions (Yq11.23) on the q-arm, which are implicated in male infertility. Markers that did amplify in these individuals included the DYS19, DYS389, DYS389II, DYS390, DYS391, and DYS393 loci. All five of the AMELY-negative males in the study by Chang *et al*. (2003) also failed to amplify the MSY1 locus (Figure 2.4). This confirmed the observation by Thangaraj *et al*. (2002) that a large deletion occurred on the p-arm of the Y-chromosome. Again, markers DYS438, DYS439, and DYS390 were able to be amplified. The ten AMELY-negative samples reported by Kashyap *et al*. (2006) also demonstrated amplification at the SRY locus and at the following Y-STR markers i.e. DYS19, DYS389I and II, and DYS390. This confirmed that the deletion was limited to the amelogenin region. Chang *et al*. (2007) analysed the Y-haplotypes harboured by the 18 AMELY-negative males and determined a large deletion in Yp11.2 that included the amelogenin, MSY1 and DYS458 loci.

Lattanzi *et al*. (2005) examined the deletion in two Italian AMELY-negative samples and determined the distal breakpoint to be within the 11.5 kb between sY1091 and sY1242. At the proximal end, the major TSPY array was partially deleted and sY59 was present, placing the breakpoint between sY1079 and sY59. Mitchell *et al*. (2006) determined the size of the deletion in five AMELY-negative males by using markers surrounding the AMELY locus (Figure 2.5). Results indicated that the sequence at the G66289 locus was present in all five samples, however none could be amplified at 6,718 kb marker locus. This placed the breakpoint of the deletion for all five samples, 59-78 kb upstream of the AMELY locus (Mitchell *et al*., 2006). Locus G66276 was amplified in two samples, but was absent in three samples. These three samples demonstrated amplification at locus DYS288. The results indicated that for three of the samples, the deletion was 712-1,001 kb in size and for two of the AMELY-negative samples, the deletion was 304 to 731 kb in size (Mitchell *et al*., 2006). These deletions are indicated in Figure 2.5. A sibling of one of the AMELY-negative males described by Mitchell *et al*. (2006) was also tested and presented with the same deletion. This observation and the report of this individual having three AMELY-negative male relatives demonstrate that the deletion in this instance is inherited (Mitchell *et al*., 2006).

In order to determine the breakpoints of the deleted regions in five AMELY-negative males, Cadenas *et al*. (2007) used a number of Y-specific markers surrounding the amelogenin locus. The samples gave no results for the following markers, i.e. sY1242, sY605, sY74, M12, MSY1, DYS257b and sY2180. Successful amplification was achieved

for the Rep locus which is 5 kb from sY1242. At the opposite end, the SHGC-80640 locus was amplified in all samples. The positions of the breakpoints are illustrated in Figure 2.5. The distal breakpoint was at the 5 kb mark, between sY1242 and Rep and the proximal breakpoint was at the 180 kb point, between sY2180 and SHGC-80640. The deletion breakpoints were the same in all five samples and suggested a deletion of 2.3 Mb (Cadenas *et al*., 2007).

Figure 2.5: Position of AMELY deletions with regard to amplified loci. The breakpoints of the deletions are indicated



Adapted from Mitchell *et al*. (2006) and Cadenas *et al*. (2007). a = Lattanzi *et al*. (2005), b and c = Mitchell *et al*. (2006), d = Cadenas *et al*. (2007). + indicates positive STS amplification and – indicates no amplification. The solid black line indicates the area of no amplification and therefore the size of the deletion. The solid red line indicates the area of uncertainty within the breakpoint intervals. The distance between the loci is indicated on the right.

Comparisons among the AMELY-negative males indicated that many deletions have occurred at the AMELY locus. The different studies employed different markers making it

difficult to identify a common deletion between them. The five AMELY-negative males identified by Cadenas *et al*. (2007) had identical deletions, thus strengthening the suggestion of a common ancestor. The distal breakpoint of the deletions described by Lattanzi *et al*. (2005) was the same as that described by Cadenas *et al*. (2007) suggesting that the deletions are not independent mutations. The Y-chromosome haplotypes of the two individuals described by Lattanzi *et al*. (2005) differ in eight of the nine Y-STR loci determined. The Y-chromosome haplotypes described by Cadenas *et al*. (2007) differ only in four of the ten Y-STR loci. In these two groups, eight common Y-STRs were analysed and very few were similar. The identical deletion breakpoints could also be the result of the presence of a breakage hot spot (Cadenas *et al*., 2007). The latter seems to be the more likely explanation, as Vijayakumar *et al.* (2006) also identified recurrent deletions at Yp11.2 in prostate tumours, supporting the existence of a hot spot for rearrangements in this region.

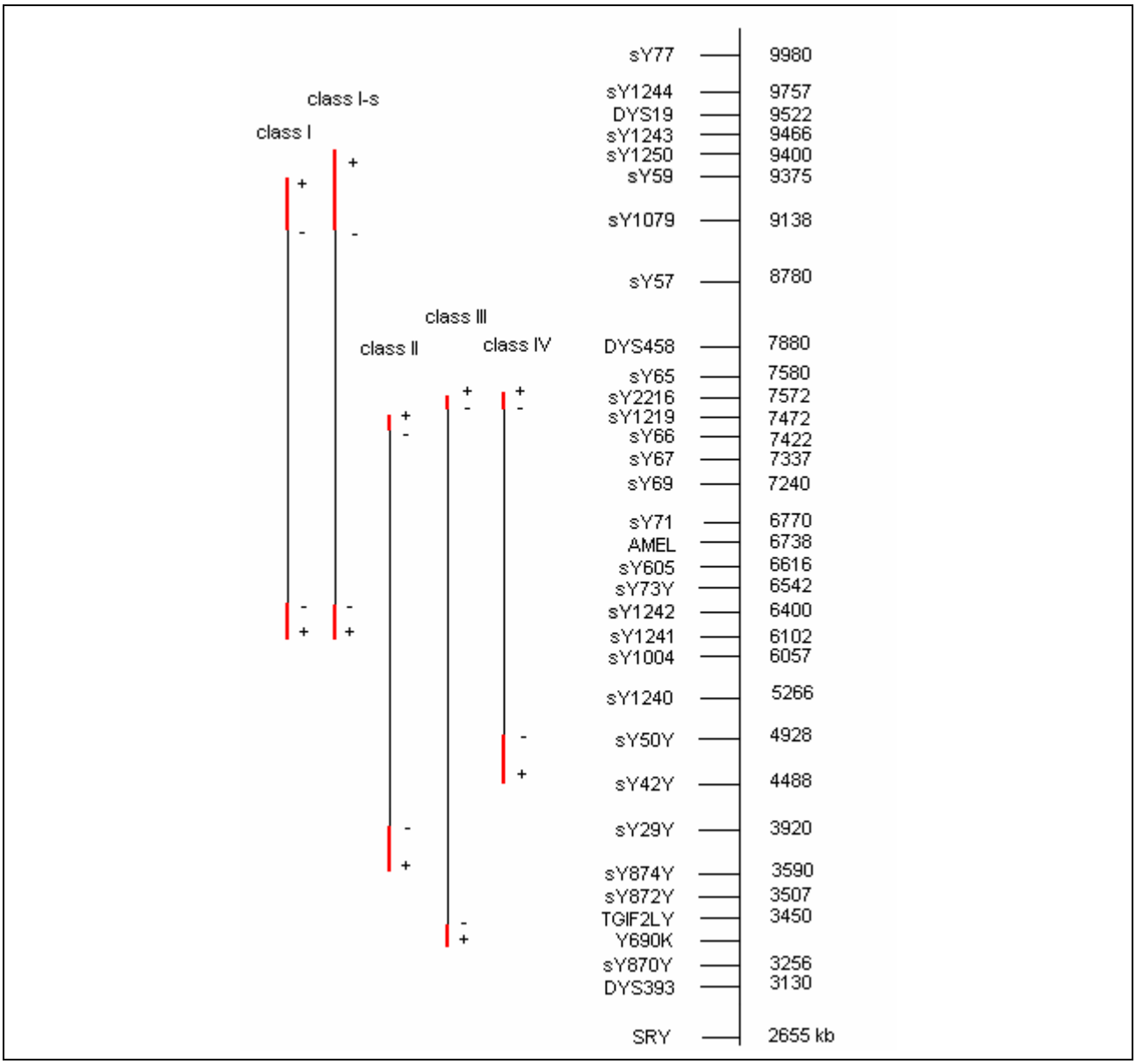
Yong *et al*. (2007) characterised the Yp11.2 deletion from 21 AMELY-negative males from Malaysia and Singapore by sequence analysis. Similar to Lattanzi *et al*. (2005) sY1091 and Y7567R demonstrated positive amplification (Figure 2.5). The authors argue that these may be false positives as this region forms part of the inverted repeat (IR) 3 which has a homologous paralogue (Yong *et al*., 2007). They therefore believe that the most distal true positive STS is sY1241, upstream of the distal IR3. Proximally, STS sY59 and sY1078 were positive and sY1293 was negative for amplification (Figure 2.5). sY59 and sY1078 are sited at the 3' end of the major TSPY array and sY1293 is 5' to the major TSPY array. This placed the breakpoint boundaries in the minor and major TSPY arrays (Yong *et al*., 2007). The distal breakpoint was placed in the TB1 repeat of the minor TSPY array. The proximal breakpoint could not be definitively positioned because of the repetitiveness of the major TSPY array, but the authors believe that the proximal breakpoint is located at the distal part of the major TSPY array. A deletion involving the proximal end of the TSPY array should have produced a reduction in TSPY copy number and they detected no statistical difference between AMELY-negative males and AMELY-positive males in TSPY copy number (Yong *et al*., 2007).

In an attempt to compare the different deletion results, Jobling *et al*. (2007) screened a collection of 45 AMELY-negative males from 12 populations. Two of these samples were first described by Santos *et al*. (1998) while 30 of the other samples had been previously reported (Henke *et al*., 2001; Chang *et al*., 2003; Lattanzi *et al*., 2005; Mitchell *et al*., 2006;

Parkin *et al*., 2007 and Chang *et al*., 2007). The remaining 13 samples were from Jobling *et al*. (2007).

Jobling *et al*. (2007) used 33 Y-chromosome markers to map the deletions among 45 Y-chromosomes and differentiated five classes (Class I, I-sY59, II, III and IV). These classes are demonstrated in Figure 2.6. Class I was the most general (84%) and had a proximal breakpoint in the TSPY major array, between STS sY1079 and sY59. The distal breakpoint was in a 298 kb interval between sY1242 and sY1241. Further refinement of

Figure 2.6: Schematic representation of deletion mapping data



Adapted from Jobling *et al*. (2007). + indicates positive STS amplification and – indicates no amplification. The solid black line indicates the area of no amplification and therefore the size of the deletion. The solid red line indicates the area of uncertainty within the breakpoint intervals. The distance between the loci is indicated on the right and was obtained from NCBI map viewer (build 37.1, reference assembly).

this breakpoint was considered impractical due to the existence of the highly homologous copy of the IR3 locus (Skaletsky *et al*., 2003), which is proximal to the deleted segment. The second deletion class (I-sY59) was identified in three individuals and was related to the first. The distal breakpoint was the same, but proximally the sY59 STS was absent. STS sY59 lies just inside the proximal end of the TSPY array and Jobling *et al.* (2007) decided to classify these deletions also as Class I. Class I deletions therefore accounted for 91% of the deleted chromosomes. Class II deletions occurred in two individuals and Class III and IV deletions were present in one individual respectively. The proximal breakpoint of the Class II deletion was between sY66 and sY1219 with the distal breakpoint between sY29Y and sY874Y. Class III and IV had the same distal breakpoint between sY1219 and sY1216. The proximal breakpoints however differed. The Class III breakpoint was between TGIF2LY and Y690K and the Class IV breakpoint between sY50Y and sY42Y. These breakpoints are all illustrated in Figure 2.6.

2.6.2 Origin

In order to assess whether the four deletion classes reflect four separate deletion events, the Y-chromosomal haplogroups were determined. If the deletion chromosomes within a class belong to different Y-haplogroups, they probably would have arisen independently (Jobling *et al*., 2007). Jobling *et al*. (2007) determined six different Y-haplogroups amongst the Class I deletions, demonstrating that the event is recurrent. Of the 41 samples with Class I deletion chromosomes, 30 of them belonged to Y-haplogroup J2e1*. This suggested a founder event is responsible for the Class I lineage within Y-haplogroup J2e1. Class I chromosomes within the R1b3 Y-haplogroup had the same associated Y-haplotype, again suggesting a founder lineage whereas Class I deletion chromosomes within the H(xH2) Y-haplogroup had similar haplotypes, suggesting a single deletion event. The two individuals with Class I deletion chromosomes within Y-haplogroup I had different deletion-associated Y-haplotypes suggesting independence of origin. The two Class II deletion chromosomes belonged to Y-haplogroup R1b3. The Y-haplotypes of Class III deletions were very similar and together with the shared Y-haplogroup suggested a common founder event. The individual harbouring the Class III deletion chromosome also belonged to Y-haplogroup R1b3 and the Class IV deletion belonged to Y-haplogroup J(xJ2). According to Jobling *et al*. (2007), the four deletion classes within this investigation were due to ten separate deletion events.

The findings by Jobling *et al*. (2007) are in agreement with those reported by Chang *et al*. (2007) and Cadenas *et al*. (2007). Chang *et al*. (2007) suggested that the AMELY deletion did not arise due to natural selection either, as they determined no association with a distinct class of Y-haplotypes. The 18 AMELY-negative males they reported, belonged to 3 Y-haplogroups whereas a further twelve belonged to Y-haplogroup J2e. This indicated a founder event for the deletions within Y-haplogroup J2 while the others probably represent independent events. All five of the AMELY-negative individuals identified by Cadenas *et al*. (2007) belonged to the same lineage i.e. J2b2. Although the individuals displayed Y-haplotypes that were not identical, they were very similar, suggesting a common ancestor. Chang *et al*. (2007) deduced that although the AMELY-negative males seemed to have multiple origins among global populations, at least the Indian AMELY-negative males may represent a founder event with regards to the deletion fragment.

2.6.3 Mechanism

The Y-chromosome has numerous repeat sequences distributed across its length. Blanco *et al*. (2000) described flanking repeat sequences within the Y-chromosome which could be involved in creating deletions or insertions by means of intrachromosomal recombination.

Repetitive palindromic regions bracket areas of homologous recombination and are responsible for most of the AZFb and AZFc deletions (Repping *et al*., 2002) whereas nonallelic homologous recombination (NAHR) between flanking direct repeats are the likely mechanism of AZFa deletions (Blanco *et al*., 2000) and a few AZFb and AZFc deletions (Repping *et al*., 2002). NAHR has been suggested as a mechanism for genetic rearrangement, leading to duplications or deletions in the genome (Stankiewics and Lupski, 2002). Copy number variations usually occur in regions that contain or are flanked by large homologous repeats or segmental duplications. NAHR is the process whereby segmental duplications on the same chromosome can facilitate copy number changes of the duplicated regions along with intervening sequences (Inoue and Lupski, 2002). Yq deletions and duplications appear to be the result of repetitive sequence blocks lying in the same orientation that display homologous recombination (Kuroda-Kawaguchi *et al*., 2001; Fernandes *et al*., 2004; Thangaraj *et al*., 2003).

Although the Y-chromosome exhibits cytogenetically visible structural variation, only the variable copy number at the TSPY tandem array (Tyler-Smith *et al*., 1988) and a 3.6 Mb

inversion polymorphism (Affara *et al*., 1986) have been described on Yp. In the AMELY-negative males described by Santos *et al*. in 1998, the authors proposed that the deletions were due to NAHR between the 20.4 kb repeat units of the TSPY A and TSPY B loci, similar to the repeat-mediated deletions of the AZFa-c regions on Yq. Thangaraj *et al*. (2003) hypothesised that it was likely that the AMELY deletions were due to homologous recombination, but also argued that the deletions could be the result of transposable elements. Jobling *et al*. (2007) determined that the proximal breakpoint of AMELY Class I deletions were within the TSPY major array. The distal breakpoint corresponded to the distal IR3 element that contained a copy of the 20.4 kb TSPY repeat in the same orientation as the major array (Jobling *et al*., 2007). Therefore, Jobling *et al*. (2007) argued that the most likely mechanism was NAHR between the TSPY repeat copies with the deleted region being between 3.0-3.8 Mb depending on the position of the breakpoint in the major array. If NAHR was the likely mechanism, the copy number of TSPY repeats in deletion chromosomes would be reduced. Jobling *et al*. (2007) analysed the TSPY copy number in nine Class I deleted chromosomes and determined that the mean copy number in non-deleted chromosomes was 33 copies whereas in deleted chromosomes it was 21 copies. This observation strengthens TSPY-mediated recombination as a mechanism for the deletions. Jobling *et al*. (2007) searched for direct repeats at the breakpoint intervals of Classes II, III and IV, but none were determined and this led them to posit that non-homologous end joining may be the mechanism responsible for these classes rather than homologous recombination.

Examples of repeat-mediated deletions of the Y-chromosome that have been described also demonstrate duplications of that region (Bosch and Jobling, 2003). Murphy *et al*. (2007) recently described two brothers of Italian descent that demonstrated a duplication of a region on Yp. This duplicated region contained the AMELY, transducin (beta)-like 1 protein Y-linked (TBL1Y) and serine-threonine protein kinase Y-linked (PRKY) genes. This corresponds to the area deleted in AMELY-negative males and therefore it appears that this region may be a hotspot for genetic events.

2.6.4 Genes

The PRKY, TBL1Y and AMELY genes belong to the X-degenerate class of Y-chromosome genes. This degeneration has not occurred during recent human evolution, thus it is likely that the genes commonly lost in AMELY deletions are functionally significant (Jobling *et al*., 2007). The AMELY-negative males identified to date all have an unaffected

appearance with no obvious indications of an altered phenotype. This does not however exclude the fact that the effects might be subtler. Heritability of the deletion in certain cases (Henke *et al*., 2001; Thangaraj *et al*., 2002; Lattanzi *et al*., 2005) indicates that it is not associated with infertility.

As previously stated in Section 2.3.4, the AMELY gene encodes a protein expressed during tooth development (Salido *et al*., 1992), which in turn has an X-homologue. The expression of AMELY is 10% that of AMELX. This could explain the normal teeth present upon examination of two individuals with AMELY deletions (Lattanzi *et al*., 2005). In individuals harbouring Class I deletions (Jobling *et al*., 2007), both the ubiquitously expressed PRKY and TBL1Y genes are also deleted. These genes are expressed in the prostate and foetal brain, but their function is currently unknown. Both of the genes have X-homologues. The TSPY gene which has been posited to be involved in NAHR, which is responsible for Class I deletions (Jobling *et al*., 2007), has also been implicated as a putative oncogene involved in gonadoblastoma (Lau *et al*., 2000) and prostate cancer (Vijayakumar *et al*., 2006).

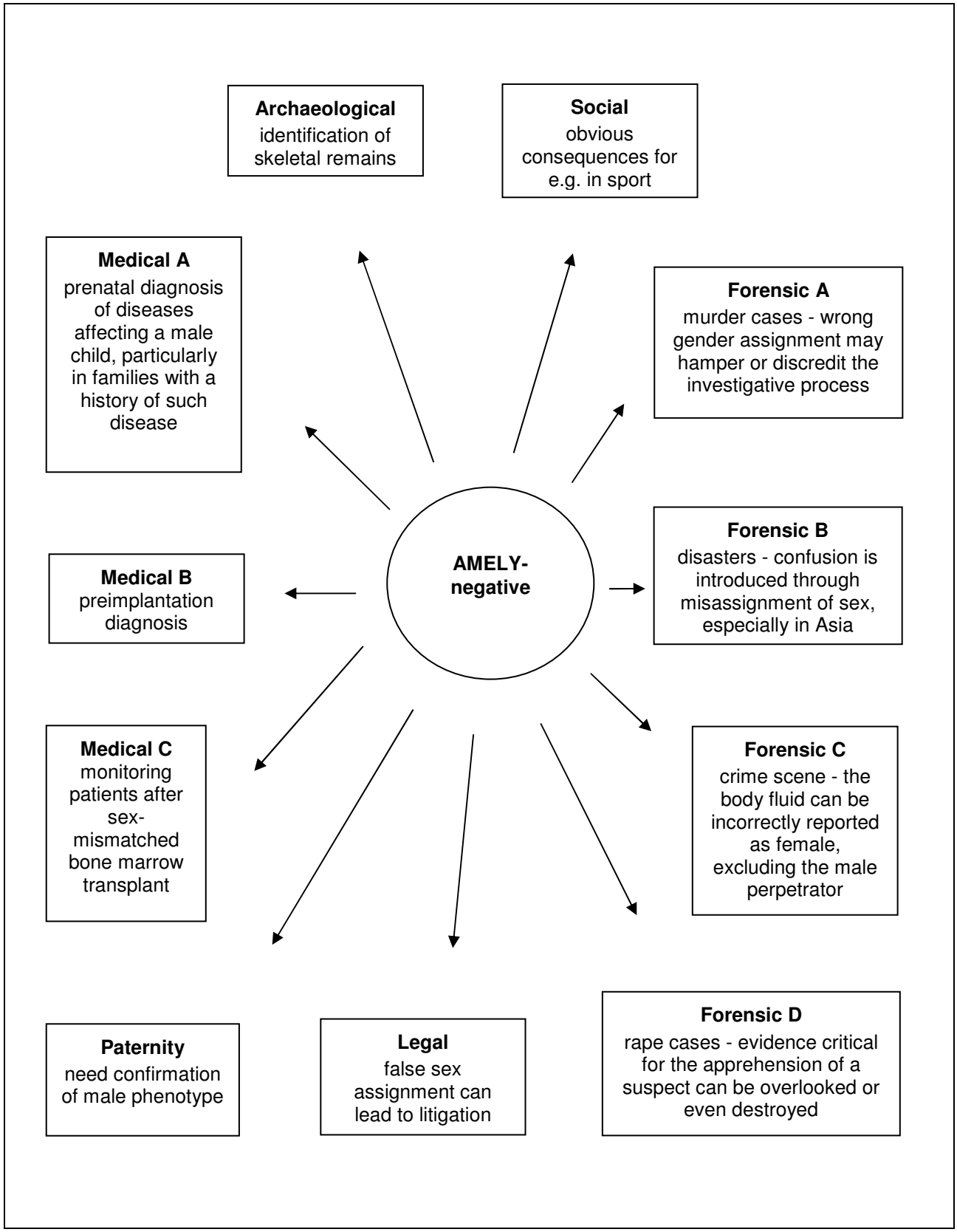
2.7 IMPLICATIONS OF THIS STUDY FOR AMELY DELETIONS

Gender identification is essential in medical diagnosis of sex-linked disease and is crucial in the prenatal diagnosis of X-linked diseases such as Duchenne muscular dystrophy and haemophilia (Thangaraj *et al*., 2002). This is especially relevant if an affected male child is born instead of the unaffected or carrier female child the parents were expecting. Gender identification is also important in the diagnosis of sex chromosome aneuploidies (Young *et al*., 2001) and preimplantation diagnosis. Furthermore, many laboratories use a gender identification test after a sex-mismatched bone marrow transplant, to ascertain the success of the transplant. Monitoring of the patient may be hampered due to the failing of the gender identification test. The implications of wrong gender assignment have an impact on many areas and these are summarised in Figure 2.7.

In archaeology, sex typing of skeletal remains from cemeteries and graves are important for the reconstruction of historical events. For the identification of bodies from a scene of a mass disaster, the wrong gender assignment can lead to confusion and mismatching of remains. STR-analysis was employed to assist in the identification of bodies after the mass disaster near Waco in 1993 (Clayton *et al*., 1995). The amelogenin sex test was

employed to help confirm a proposed identification and to provide gender information where none was available from pathology results (Clayton *et al*., 1995). Wrong gender assignment may therefore hamper the investigation instead of aiding it.

Figure 2.7: Implications of incorrect gender assignment



The implications are even greater if the mass disaster following the South Asian tsunami in 2004 (Jobling *et al*., 2007) and the recent Haiti disaster in 2010 is considered. For criminal evidence, the determination of the gender of the offender is usually the initial information obtained to be used for further investigation. Body fluid from a crime scene may therefore erroneously be reported as being female whereas the perpetrator may be male. The presence of small amounts of male DNA in a rape case may be undetected and evidence as a result might be overlooked or destroyed. This evidence may have been important in the apprehension of a suspect. In rape cases, there is also the possibility of contamination of DNA from both victim and culprit and gender identification is therefore crucial.

Due to the simplicity of the amelogenin test in gender identification, it has gained wide acceptance and was integrated into routinely automated genetic profiling procedures. The unreliability of the amelogenin test however, raises concern over its continued use.

Since the first report of two Sri Lankan Indian males that were typed as females (Santos *et al*., 1998) by means of amplification of the amelogenin locus, other laboratories have also observed this phenomenon. So far, amelogenin dropout males have been noted in Australian Caucasians (Roffey *et al*., 2000), Moroccans (Henke *et al*., 2001), Austrians (Steinlechner *et al*., 2002), Indians (Thangaraj *et al*., 2002), Spaniards, Slovenians and individuals from the U.K. (Chang *et al*., 2003). Although the frequency of amelogenin dropout is low in the population groups other than those from India, the implications of wrong gender assignment are severe and have far-reaching consequences. Gender determination in human identification requires high reliability, and a gender identification test based on the amelogenin locus alone is insufficient.

A gender-determining locus is an essential part of the commercial multiplex kits used in human identification, with amelogenin the most often employed. In a PCR reaction, the same primer pair allows for the discrimination between male and female as this single copy gene differs in size on the X and Y-chromosomes, as discussed in Section 2.3. Reports of phenotypically unaffected males that are negative for the Y-chromosome amelogenin gene as a result of a deletion surrounding the amelogenin locus, are described in Section 2.4. There appears to be multiple forms of the deletion (Section 2.6), but most reports have been in the South Asian population.

In many disciplines the correct gender determination is crucial (Section 2.7) and it is therefore important to determine how unique this deletion is and whether it is only present

in certain populations. This question should be addressed in the South African population to determine if additional Y-chromosome markers should be included in gender-determining tests. As discussed in Section 2.1, most of the Y-chromosome does not recombine at meiosis and is generally inherited unchanged. The non-recombining part of the Y-chromosome preserves the mutational events that have occurred along male lineages. Y-STR haplotypes, discussed in Section 2.5, can be used to determine if any of the AMELY-negative males belong to the same paternal lineage and if an association exists between a specific haplotype and deletion size. The study of this mutation in the South African population may shed light on the history of human populations, as well as clarify the wrong assignment of gender with its wide-ranging implications as depicted in Figure 2.7.

2.8 AIM OF THE STUDY

Subsequent to the screening of 6,130 black South African samples and 2,214 samples from the Caucasian South African population, thus a total of 8,344 samples, four AMELY-negative black South African males were detected. These samples were screened as part of a paternity testing service laboratory programme, of which the author screened at least 40 percent.

The aims of this study were threefold:

a. To determine the size of the deletion fragments within the Y-chromosome of AMELY-negative black South African males by using STS markers surrounding the amelogenin locus.
b. To examine the relationship between the South African AMELY-negative males, by comparison of their deletion sizes and the Y-chromosome haplotypes associated with the deletion fragments.
c. Finally this study undertook a comparison of the deletions present in the South African population to those reported in other population groups.

CHAPTER THREE

Materials and Methods

3.1 PATIENT POPULATION

Samples were received from 6,130 black South African males and 2,214 Caucasian South African males with written informed consent. The individuals were of self-declared origin. All samples submitted to the laboratory for STR analysis were included. This project was approved by the Ethics Committee of the North-West University (Potchefstroom Campus) with Ethics approval number NWU-00038-07-S8.

3.2 DNA ISOLATION

The samples consisted either of peripheral blood collected in ethylene diamine tetra-acetic acid (EDTA) tubes or buccal swabs. DNA was isolated with the Biosprint®[1] 15 DNA Blood kit (QIAGEN, 2003) for the blood samples, and the QIAamp®[1] DNA mini kit (QIAGEN, 2003) for the buccal swabs. DNA was immediately isolated and subsequently stored at -70°C.

DNA was isolated from human whole blood by means of the KingFisher®[2] mL instrument (AEC-Amersham) and the BioSprint® 15 DNA Blood kit. The Kingfisher® BS15 DNA Blood 200 protocol was used. This protocol was performed at room temperature. Briefly, 20 µL of protease was pipetted into the bottom of the well after which 200 µL blood, 200 µL Buffer AL, 200 µL isopropanol and 30 µL of vortexed MagAttract Suspension G were added. This was followed by an incubation step of 7 min to allow for the lysis of the cells and the binding of the DNA to the MagAttract magnetic particles. The bound DNA was washed twice using different wash buffers, AW1 and AW2, followed by an air-dry step, and elution in 500 µL Buffer AE. The supplier, due to the proprietary nature of the buffer components, labelled the different buffers in a generic manner.

[1] Biosprint® and QIAamp® are registered trademarks of QIAGEN Pty. Ltd., Australia.
[2] KingFisher® is a registered trademark of ThermoLabsystems, Finland.

For isolation of DNA from buccal swabs, a spin protocol was used. The QIAamp® DNA purification procedure that was used, consists of three steps and was carried out using spin columns in a standard microcentrifuge at room temperature. The buccal swab was air-dried and placed in a 2 ml microcentrifuge tube. The swab was separated from the stick and 400 µL phosphate buffered saline (PBS) was added. Twenty µL Proteinase K and 400 µL Buffer AL were added and the sample was vortexed for 15 s and incubated at 56°C for 10 min. Then, 400 µL ethanol (96–100%) was added to the sample and vortexed for 15 s. Seven hundred µL of the mixture was applied to the spin column followed by centrifugation for 1 min at 6,000 x g. The spin column was washed twice with 500 µL Buffer AW1 and centrifugation at 6,000 x g for 1 min as well as once with 500 µL Buffer AW2 and centrifugation at 16,060 x g for 3 min to dry the membrane completely. After addition of 150 µL Buffer AE to the spin column, the reaction was incubated for 1 min at room temperature followed by centrifugation at 6,000 x g for 1 min.

DNA quantification was not done. According to the manufacturer, 4.5 to 9 µg DNA can be isolated from 200 µL whole blood. Throughout, 200 µL of blood was used and the DNA was eluted in 500 µL Buffer AE (QIAGEN, 2003). Of this volume, 0.3 µL DNA was used in the PCR reaction with the AmpFℓSTR SGM Plus®[1] kit and 1.25 µL DNA was used in the PCR reaction with the PowerPlex 16®[2] kit. This was established to be sufficient DNA to yield a PCR product for analysis. For the buccal swabs, according to the manufacturer, a single swab yields 0.8 to 2.0 µg DNA. The DNA was eluted in 150 µL Buffer AE (QIAGEN, 2003). Again, 0.3 µL DNA for the AmpFℓSTR SGM Plus® kit and 1.25 µL DNA for the PowerPlex 16® kit was enough to yield sufficient PCR product.

3.3 SHORT TANDEM REPEAT ANALYSIS

STR length polymorphism analysis is generally the standard methodology for the determination of human identity. The most sensitive protocol for STR analysis involves the use of fluorescently labelled primers in a PCR-based method followed by the automated, fluorescent detection of the resultant amplicons (Collins *et al.*, 2004). In this investigation the following methods were used i.e. AmpFℓSTR SGM Plus® and PowerPlex 16® kits to determine the genetic identities of 6,130 black South African individuals and 2,214

[1] AmpFℓSTR SGM® Plus is a registered trademark of Life Technologies, Carlsbad, CA, USA.
[2] PowerPlex 16 System® is a registered trademark of the Promega Corporation, Madison, WI, USA.

Caucasian South African individuals. They are discussed in greater detail in the following sections.

3.3.1 AmpFℓSTR SGM Plus® PCR

The AmpFℓSTR SGM Plus® kit (Applied Biosystems, 1999; Budowle *et al*., 1996; Kimpton *et al*., 1996; Sparkes *et al*., 1996; Cotton *et al*., 2000) is used for human identification using fluorescently labelled STR loci. This kit amplifies ten STR loci: D3S1358, vWA, D16S539, D2S1338, D8S1179, D21S11, D18S51, D19S433, TH01, and FGA in a single multiplex reaction. For gender identification, the kit employs the amelogenin gene whereby amplification with the same primer pair results in different product sizes for the X and Y-chromosomes. The fluorescent dyes 5'-carboxyfluorescein (5'-FAM), 6-carboxy-4',5'-dichloro-2',7'-dimethoxy-fluorescein (JOE), or NED®[1] were used to label one of the primers in the primer pair, and respectively the dyes are detected as blue, green and yellow via the ABI PRISM®[2] 310 or ABI PRISM® 3130 Genetic Analyser (Applied Biosystems, 1999).

DNA samples were amplified via PCR using the AmpFℓSTR SGM Plus® kit according to the manufacturer's recommendations. A 0.2 ml thin wall PCR tube was used for the PCR reaction. A master mixture was prepared by adding the following volumes of reagents: 3.95 µL sterile PCR grade distilled H_2O (dH_2O), 5.25 µL PCR Reaction Mix (buffer salt, magnesium chloride ($MgCl_2$), deoxynucleoside triphosphates (dATP, dCTP, dGTP, dTTP), bovine serum albumin (BSA), 0.05% sodium azide (NaN_3)), 2.75 µL PCR Primer Set to amplify 10 STR and the amelogenin loci and 1.25 U AmpliTaq Gold®[3] DNA Polymerase for a final volume of 12.2 µL. To this volume, 0.3 µL of the DNA sample was added. To the positive control, 0.3 µL AmpFℓSTR SGM Plus® Control DNA 007 (0.10 $ng.\mu L^{-1}$ human male genomic DNA in 0.05% NaN_3 and buffer) was added and 0.3 µL sterile PCR-grade dH_2O to the negative control. The primer sequences are propriety information and therefore the exact primer sequences cannot be reported. However, these primers are standard to the AmpFℓSTR SGM Plus® kit.

PCR amplification was performed on the GeneAmp PCR System 9700® (Applied Biosystems, 1997). The following program was used: An initial incubation step of

[1] NED® is a registered trademark of Life Technologies, Carlsbad, CA, USA. Its chemical structure is not publicly available.
[2] ABI PRISM® is a registered trademark of Life Technologies, Carlsbad, CA, USA.
[3] AmpliTaq® Gold is a registered trademark of Roche Molecular Systems Inc., Pleasanton, CA, USA.

11 minutes at 95ºC, 28 cycles of alternating denaturing (60 seconds at 94ºC), annealing (60 seconds at 59ºC) and extension steps (60 seconds at 72ºC) followed by a final extension step of 45 minutes at 60ºC. After PCR, the amplified PCR product was stored at 2-6ºC and protected from light.

The PCR products were prepared for the ABI PRISM® 310 or ABI PRISM® 3130 Genetic Analyser by adding 1 µL of the PCR product or allelic ladder to 25 µL Hi-Di®[1] Formamide and 0.3 µL Genescan-500[ROX] Size Standard® into 0.2 ml thin wall PCR strip tubes. The tubes were placed into the GeneAmp PCR 9700® instrument, denatured for 3 min at 95ºC and chilled to 4ºC.

3.3.2 PowerPlex 16® System

This system allows for the co-amplification of sixteen short tandem repeat loci: D3S1358, vWA, D16S539, D8S1179, D21S11, D18S51, PentaE, PentaD, CSF1PO, TH01, TPOX, D5S818, D7S820, D13S317 and FGA. For gender identification, the amelogenin gene is also amplified. One primer of each primer pair is labelled with either the fluorescein (FL), carboxy-tetramethylrhodamine (TMR), or JOE dye, which is detected as blue, yellow, and green, respectively, on the ABI PRISM® 310 or ABI PRISM® 3130 Genetic Analyser.

DNA samples were subjected to PCR using the manufacturer's recommendations. A master mix was prepared by adding the following volumes of reagents: 8.35 µL sterile PCR grade dH_2O, 1.25 µL Gold ST∗R®[2] 10X Buffer (500 mM potassium chloride (KCl), 100 mM 2-amino-2-(hydroxymethyl)-1,3-propanediol hydrochloride (Tris-HCl), pH 8.3, 15 mM $MgCl_2$, 1% Triton X-100, 2 mM of each dNTP, 1.6 $mg.ml^{-1}$ BSA), 1.25 µL 10X Primer Set to amplify 16 STR and the amelogenin loci and 2 U AmpliTaq Gold® DNA Polymerase to a final volume of 11.25 µL. To each tube, 1.25 µL of each DNA sample was added. To the positive control, 1.25 µL Control DNA 9947A (0.5 $ng.\mu L^{-1}$ standard female DNA) was added and 1.25 µL sterile PCR grade dH_2O was added to the negative control. Primer sequences represented in Table 3.1 were used in the amplification of a segment of the amelogenin gene.

[1] Hi-Di® is a registered trademark of Life Technologies, Carlsbad, CA, USA.

[2] Gold ST∗R® is a registered trademark of the Promega Corporation, Madison, WI, USA.

Table 3.1: Sequences of the primers used in the detection of a segment of the amelogenin gene

Locus	Orientation	Primer sequence	5' end
Amelogenin	Forward	5'-ccc tgg gct ctg taa aga a-3'	TMR
	Reverse	5'-atc aga gct taa act ggg aag ctg-3'	OH

PCR amplification was performed on the GeneAmp PCR System 9700®. The following program was used: An incubation step of 11 minutes at 95ºC, followed by 10 cycles of 30 seconds at 94ºC (ramp 100%), 30 seconds at 60ºC (ramp 29%), 45 seconds at 70ºC (ramp 23%), 20 cycles of 30 seconds at 90ºC (ramp 100%), 30 seconds at 60ºC (ramp 29%), 45 seconds at 70ºC (ramp 23%) and a final extension step of 30 minutes at 60ºC. After PCR, the amplified PCR product was stored at 2-6ºC and protected from light.

The PCR products were prepared for the ABI PRISM® 310 Genetic Analyser by adding 1 µL of the PCR product or allelic ladder to 24 µL Hi-Di® Formamide and 1 µL Internal Lane Standard 600 (ILS 600)® into 0.2 ml thin wall PCR strip tubes. The tubes were placed into the GeneAmp PCR 9700 instrument, denatured for 3 min at 95ºC and chilled to 4ºC.

Reaction products were detected with the ABI PRISM® 310 or ABI PRISM® 3130 Genetic Analyser. GeneScan®[1] analysis was performed on the raw data and PCR products were sized according to the international nomenclature using the Genotyper®[1] Software package (Applied Biosystems, 1997; Evett *et al*., 1998) or the GeneMapper®[1] Software package (Applied Biosystems, 2003). Allele designations for the PowerPlex 16® System were determined using the PowerTyper 16®[2] Macro (Promega, 2008).

3.4 Y-CHROMOSOME HAPLOTYPING

STR markers on the Y-chromosome are different from the autosomal markers as the Y-STR markers are present on areas of the non-recombining region of the Y-chromosome (Mathias *et al*., 1994; Kayser *et al*., 1997; Jobling *et al*., 1997; Jobling *et al*., 1998; Ayub *et al*., 2000). Due to the fact that the Y-STR markers are on the same chromosome, a haplotype profile of the Y-chromosome is produced.

[1] GeneScan®, Genotyper® and GeneMapper® are registered trademarks of Life Technologies, Carlsbad, CA, USA.
[2] PowerTyper 16® and PowerPlex Y® are registered trademarks of the Promega Corporation, Madison, WI, USA.

Y-STRs were typed with the PowerPlex Y®2 System in order to confirm the gender of specific samples, to determine if a Y-chromosome deletion was present and to determine if the individuals shared a common Y-STR haplotype profile. This system allows for the amplification of 12 STR loci: DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438 and DYS439 in a single reaction. This is achieved by means of FL, TMR, or JOE dye, to label one primer of a primer pair. The dyes are detected as blue, yellow, and green, correspondingly, on the ABI PRISM® 310 or ABI PRISM® 3130 Genetic Analyser instrument.

DNA samples were subjected to PCR amplification using the manufacturer's recommendations. A master mixture was prepared by adding the following volumes of reagents: 8.47 µL sterile PCR grade dH_2O, 1.25 µL Gold ST∗R® 10X Buffer (500 mM KCl, 100 mM Tris-HCL, pH 8.3, 15 mM $MgCl_2$, 1% Triton X-100, 2 mM of each dNTP, 1.6 $mg.ml^{-1}$ BSA), 1.25 µL 10X Primer Pair (locus-specific FL, JOE, and TMR-labelled and unlabeled primers) and 1.4 U AmpliTaq Gold® DNA Polymerase for a final volume of 11.25 µL. To each tube, 1.25 µL of each DNA sample was added. To the positive control, 1.25 µL Control DNA 9948 (0.5 $ng.\mu L^{-1}$ standard male DNA) was added and 1.25 µl sterile PCR grade dH_2O was added to the negative control.

PCR amplification was performed on the GeneAmp PCR System 9700® thermal cycler. The following program was used: An incubation step of 11 minutes at 95ºC, followed by 10 cycles of 30 seconds at 94ºC (ramp 100%), 30 seconds at 60ºC (ramp 29%), 45 seconds at 70ºC (ramp 23%), 20 cycles of 30 seconds at 90ºC (ramp 100%), 30 seconds at 58ºC (ramp 29%), 45 seconds at 70ºC (ramp 23%) and a final extension step of 30 minutes at 60ºC. After PCR, the amplified PCR product was stored at 2-6ºC and protected from light.

The PCR products were prepared for electrophoresis on the ABI genetic analyser by adding 1 µL of the PCR product or allelic ladder to 24 µL Hi-Di® Formamide and 1 µL Internal Lane Standard 600 (ILS 600)® into 0.2 ml thin wall PCR strip tubes or a 96 well plate. The tubes or plate were placed into the GeneAmp PCR 9700® Instrument, denatured for 3 min at 95ºC and chilled to 4ºC.

Reaction products were detected with the ABI PRISM® 310 or ABI PRISM® 3130 Genetic Analyser. GeneScan® analysis was performed on the raw data and PCR products were sized according to the international nomenclature using the Genotyper® Software package (Applied Biosystems, 1997) or the GeneMapper® Software package (Applied Biosystems,

2003). Allele designations for the PowerPlex Y System® were determined using the PowerTyper Y®[1] Macro (Promega, 2008).

3.5 PCR PRIMER BINDING SITES

Samples that presented with abnormal amelogenin peak profiles with the AmpFℓSTR SGM Plus® kit and PowerPlex 16 System® were re-amplified with alternative amelogenin primer sets to determine whether a mutation had occurred within the annealing region of the primers used by the commercial genotyping kits. These alternative primer sequences were obtained from primers described by Steinlechner *et al*. (2002) as listed in Table 3.2.

Table 3.2: Primers used in the amplification of annealing regions of primers used by commercial genotyping kits

Primer name	Orientation	Sequence
Steinlechner	Forward	5′–acc acc tca tcc tgg gca cc–3′
	Reverse	5′–aac aca ggc ttg agg cc–3′

The reaction mixture for the amplification of this specific fragment consisted of the following: 1X PCR Gold® buffer (150 mM Tris-HCl, pH 8.3, 500 mM KCl), 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 2.5 µM primers and 1 U of AmpliTaq Gold® DNA Polymerase. The PCR reaction had a final volume of 20 µL. Thermal cycling was performed in a Perkin Elmer 9700® thermal cycler using the following parameters: 94ºC for 5 min, 32 cycles of 94ºC for 1 min, 62ºC for 1 min and 72ºC for 1 min and a final extension step of 72ºC for 10 min. Successful amplification of the sample was determined with 2% agarose gel electrophoresis. The annealing sites of the primers are indicated in Table 3.3.

[1] PowerTyper Y® is a registered trademark of the Promega Corporation, Madison, WI, USA.

Table 3.3: Partial sequence of the human amelogenin gene on the Y-chromosome from nucleotide 181 to 540

Nucl no.	DNA sequence: M55419
181	ccctataaaa gctaccacct catcctgggc accctggtta tatcaacttc agctatgagg
241	taatttttct ctttactaat tttgatcact gtttgcatta gcagtcccct gggctctgta
301	aagaatagtg ggtggattct tcatcccaaa taaagtggtt tctcaagtgg tcccaatttt
361	acagttccta ccatcagctt cccagtttaa gctctgatgg ttggcctcaa gcctgtgttg
421	ctccagcacc ctcctgcctg accattcgga ttgactcttt cctcctaaat atggctgtaa
481	gtttattcat tcatgaacca ctgctcagga aggttccatg aaagggcaaa aagtcaactc

Nucl no. = nucleotide number. The forward primer sequence of the PowerPlex16® System is indicated by the double underlined text (www) whereas the position of the reverse primer sequence of the PowerPlex16® System is indicated by the single underlined text (www). The forward primer sequence published by Steinlechner *et al.* (2002) is indicated by the red text (www) whereas the position of the reverse primer sequence published by Steinlechner *et al.* (2002) is indicated by the blue text (www). This sequence has been retrieved from GenBank® accession number M55419 and the nucleotides are numbered accordingly.

3.6 CHAIN TERMINATION SEQUENCING

PCR reactions were performed as described under Section 3.5. The PCR product was purified using the QIAquick®[1] gel extraction kit according to the manufacturer's instructions. Five volumes of PB buffer was added to 1 volume of PCR product after which the mixture was applied to a QIAquick® column and centrifuged at 6,000 x g for 1 min. The spin column was washed with 750 µL PE Buffer by centrifugation at 6,000 x g for 1 min. This was followed by centrifugation at 16,060 x g for 3 min to dry the membrane completely. To the column, 50 µL elution buffer was added, after which the sample was incubated at room temperature for 1 min and centrifuged at 6,000 x g for 1 min. The efficiency of the purification was visually determined by agarose gel electrophoresis.

PCR products were sequenced using ABI PRISM® BigDye®[2] Terminator Cycle Sequencing Ready Reaction. Both the forward and reverse strands were sequenced using the forward and reverse primers published by Steinlechner *et al.* (2002), respectively. The sequencing reaction consisted of the following: 2 µL of 5X sequencing buffer, 4 µL ready reaction premix (Tris-HCl, pH 9.0 and $MgCl_2$), 3.2 µM primer and 3-10 ng of the purified PCR product for a total volume of 20 µL. Thermal cycling was performed in a Perkin Elmer 9700® thermal cycler using the following parameters: 96ºC for 1 min, 25 cycles of 96ºC for 10 sec, 50ºC for 5 sec and 60ºC for 4 min and a final hold step of 4ºC indefinitely.

Extension products were purified by sodium acetate/ethanol precipitation. Briefly, 62.5 µL of 95% ethanol, 3 µL of 3M sodium acetate and 14.5 µL deionised water was added to the

[1] QIAquick® is a registered trademark of QIAGEN Pty. Ltd., Australia.

[2] BigDye® is a registered trademark of Life Technologies, Carlsbad, CA, USA.

cycle sequencing product. The sample was centrifuged at 6,000 x g for 20 min at room temperature. The precipitate was washed with 70% ethanol, air-dried for 15 min and resuspended in 6 µL Hi-Di® Formamide. Samples were electrophoresed on a SpectruMedix®[1] (SCE2410) Genetic Analysis System and basecalling was achieved with the Base Spectrum programme of the aforementioned system.

Sequence data was obtained from the Genbank® sequence database (accession numbers M55418 and M55419). Sequence investigation and alignment were done with the BioEdit® sequence alignment editor (Hall, 2001).

3.7 DELETION MAPPING

Y-STS markers were used to evaluate deletions on the Y-chromosome. Each STS provides a straightforward means of determining the presence of a specific site on the Y-chromosome. Sizing of the amelogenin Y-chromosome deletion was achieved by testing for the presence of sequences both upstream and downstream of the AMELY locus. Y-specific markers around the AMELY locus were amplified by PCR and analysed by gel electrophoresis. Primer sequences and PCR conditions were obtained at http://www.ncbi.nlm.nih.gov/nucleotide/ followed by the name of the specific STS locus.

3.7.1 Y-STS sY1240

The primer set presented in Table 3.4 was used to generate the Y-specific PCR product of Y-STS1240. The primer sequences and amplification conditions for locus G75486 were obtained at http://www.ncbi.nlm.nih.gov/nucleotide/G75486.

Table 3.4: Primers used in the amplification of Y-STS sY1240

Locus	STS primer name	Sequence
G75486	sY1240 Forward:	5′-ggg tcc tag ata ggc tcc aag-3′
	sY1240 Reverse:	5′-ttc atg ttg gca gtg att gg-3′

The above primer set yielded a PCR fragment of 385 bp. The DNA sequence of this PCR fragment is indicated in Table 3.5.

[1] SpectruMedix® is a registered trademark of the SpectruMedix LLC, State College, PA, USA.

Table 3.5: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY1240 STS locus

Nucl no.	DNA sequence (PCR product of 385 bp):
1	gggtcctaga taggctccaa ggaaaccaaa ttacttgaag aagcctgaga ttattcttaa
61	agaaatatag atttccttgt atgtttacaa tgaatgagtc ttatttttac atgcatggaa
121	gtatcacagt cactggacac tgtataatga aaacttacaa gttctcattt tgaacaagct
181	tccttttttcc ttagttgaca aaattatcca ggtttctttt agcatgttca caatttcaga
241	ttttatacaa ttacaaactg aactttcctt ataaactgat tgttatagct aatttttcct
301	aataaaattc aaattaagac agcacctgtg aattctattt agagttctat tgatactcaa
361	tgtagccaat cactgccaac atgaa

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

The reaction mixture for amplification of the specific fragment consisted of the following: 1X PCR Gold® buffer (150 mM Tris-HCl, pH 8.3, 500 mM KCl), 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 2.5 µM primers and 1 U of AmpliTaq Gold® DNA Polymerase. The PCR reaction had a final volume of 20 µL. Thermal cycling was performed in a Perkin Elmer 9700® thermal cycler using the following parameters: 94ºC for 3 min, 35 cycles of 94ºC for 1 min, 61ºC for 1 min and 72ºC for 1 min and a final extension step of 72ºC for 5 min. Successful amplification of the sample was determined with 2% agarose gel electrophoresis.

3.7.2 Y-STS sY1241

Primers used in the amplification of Y-STS sY1241 were according to http://www.ncbi.nlm.nih.gov/nucleotide/G75487 and are presented in Table 3.6. The PCR reaction was performed as described in Section 3.7.1.

Table 3.6: Primers used in the amplification of Y-STS sY1241

Locus	STS primer name	Sequence
G75487	sY1241 Forward:	5′-agg cta ctg tga atc acg gcc-3′
	sY1241 Reverse:	5′-gtg cat gtg ttc cct ttg tg-3′

The primer set presented in Table 3.6 was used to generate the Y-specific PCR product of 508 bp as indicated in Table 3.7. The annealing regions of the primers are also indicated.

Table 3.7: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY1241 STS locus

Nucl no.	DNA sequence (PCR product of 508 bp):
1	aggctactgt gaatcacgcc actgcactcc agcctgggca acagagtaag accttgtttc
61	tatctctgtt tctctctgtc tctctcacac acacagacac acacaaataa agggaaagat
121	attcattaac aatctccagt tgtaattgct ttataatgat atcttaatct tgaaatgaat
181	tacttagatt tacataaaat aggcacatct agctgtggta aggatacact gttcatttag
241	tatggagcca gcacagaaat gggagaaaga cagttgtaga gaaactgtca agggaattca
301	agtccccaaa accccagcta agcagcgtga agaagtagaa tagctggggg aatggcgcag
361	cttgctgttt tttggtgctg gttatgacca aaccttagaa aaagcatcat actgtgatga
421	aatcagccac atagaccagt ggaacagaag agagagccca catgtattca actacttttt
481	gacgagggca caaagggaac acatgcac

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

3.7.3 Y-STS sY1242

The primer pair used to generate this Y-specific PCR product is indicated in Table 3.8. The PCR reaction mixture and the thermal cycling conditions are described in Section 3.7.1. These conditions were obtained at http://www.ncbi.nlm.nih.gov/nucleotide/G75488.

Table 3.8: Primers used in the amplification of Y-STS sY1242

Locus	STS primer name	Sequence
G75488	sY1242 Forward:	5′-cgt cgg tat ttt acg aca acg-3′
	sY1242 Reverse:	5′-gca ttt gtt ttt cat gtg cg-3′

The primer pair presented above was used to generate a Y-specific PCR product of 440 bp. This sequence is indicated in Table 3.9.

Table 3.9: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY1242 STS locus

Nucl no.	DNA sequence (PCR product of 440 bp):
1	cgtcggtatt ttacgacacg gaagaatata acatatacat cggtatttta cgacaccgac
61	gattataaca gatacgtcgg tattttacga caccgacgaa tataacatat acgtcggtat
121	tttacggcac ctacgaatat aacatataca ttggtatttt acgacacccc acccccaaa
181	aaaaggcgtc acatttacat aaacataatt atcttaaaag ccagtataat tttaatttt
241	ttgtagtcat caccttcaga ctttatttg gagaagtgat tacggaaatc tgaaatatca
301	aggcctgatg agaacactta aattaaccac actccagaag tccaaatctg aaaagcaaag
361	atgtttctga tataatagtc taaattctgc atttcctctc tattgggcag tatgatattg
421	cgcacatgaa aaacaaatgc

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

3.7.4 Y-STS sY605

Table 3.10 indicates the primers used to generate the Y-specific PCR product for Y-STS sY605. The amplification reaction mixture and conditions are available at site http://www.ncbi.nlm.nih.gov/nucleotide/G65838 and are described in Section 3.7.1.

Table 3.10: Primers used in the amplification of Y-STS sY605

Locus	STS primer name	Sequence
G65838	sY605 Forward:	5′-acc tcc gaa gac tga acc ag-3′
	sY605 Reverse:	5′-ccc ttg agt cca cag agt cc-3′

The Y-specific PCR product of 279 bp indicated in Table 3.11 was obtained after amplification. The annealing regions of the primers are indicated in the table.

Table 3.11: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY605 STS locus

Nucl no.	DNA sequence (PCR product of 279 bp):
1	tacacctccg aagactgaac caggaagaag ctgaatccct gaatacacta ataacaagtt
61	ctgaaattga ggcagtaata aataatctac caaccaaaaa agctcaagac cagatggatt
121	tcccgatcca tcttctattc tattctattc tattcattat tctaattgaa gcaacacaat
181	aattataaac caaacacagg atgttctcac ttagaagtgg gagctatgct atgaggatgc
241	aaatgaataa gaatgatgta atggactctg tggactcaag ggggaaggtg ggaggggtg
301	agagataaaa gggtacacat tgggttcagt gtacactgct tgggtgatgg gtgcaccaaa
361	atcacagaaa tcaccactaa aaacttatcc atgtaaccaa acactacctg tttccc

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

3.7.5 Y-STS sY71

The primer set described in http://www.ncbi.nlm.nih.gov/nucleotide/G12016 and presented in Table 3.12 was used to generate this Y-specific PCR product of 122 bp. A description of the PCR reaction is presented in Section 3.7.1.

Table 3.12: Primers used in the amplification of Y-STS sY71

Locus	STS primer name	Sequence
G12016	sY71 Forward:	5′-cca tct ggc tca atg gtt ag-3′
	sY71 Reverse:	5′-ctg aag gtg gcc att tct ta-3′

The amplified fragment of 122 bp obtained for locus G12016 is specified in Table 3.13. Primer binding regions are indicated for this sequence.

Table 3.13: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY71 STS locus

Nucl no.	DNA sequence (PCR product of 122 bp):
1	gtaaacacac cntgtggagt aattaacatg gttaagagag tattctacga atgtttaacc
61	tcagataccg aggtctaaag taaataccat tacagacaat tttcctggtc gacctccttg
121	cccgcctgcc cacagagggc catctggctc aatggttagt taatggaggt agggtaaaca
181	gacttaactg tggaaacctc tattgaccct agtatttacc ctatgaccta atgctctaag
241	gtaagaaatg gccaccttca gcctgttcaa ttattacaag ctatataacc tttnagcctt
301	ccaaaaaggt cgtgactatt ctctataaac cttccctagt atttccccnt aatatttntg
361	gccaccatcc tnngnngttt cccnaataat tccccntaaa atttnttgnn nccatcctng
421	nggggnntcc cncacctngg ggtcanggna ttttccngnc ntnggntncc cgntggnngg
481	gtaacagggg ggncacnntt tggcccn

Nucl no. = Nucleotide number. The primer sequences are indicated by the underlined text.

3.7.6 Y-STS sY69

The primer set presented in Table 3.14 was utilised to generate this Y-specific PCR product of 234 bp indicated in Table 3.15. The PCR reaction described in http://www.ncbi.nlm.nih.gov/nucleotide/G66516 and Section 3.7.1 was used for amplification.

Table 3.14: Primers used in the amplification of Y-STS sY69

Locus	STS primer name	Sequence
G66516	sY69 Forward:	5′-gga aca gca tct tgc tct gt-3′
	sY69 Reverse:	5′-act atg gga gac caa ggc tc-3′

Table 3.15 indicates the sequence of the PCR product after amplification. Annealing regions of the primer set used is specified in the table.

Table 3.15: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY69 STS locus

Nucl no.	DNA sequence (PCR product of 234 bp):
1	ggaacagcat cttgctctgt tgcccaggct ggagtgcagt ggcacgatct tgctcacaga
61	agcctccacc tcctggcttc aagcaattct catgccttag cctcttgagt agctgggatt
121	acagatgccc agctaatttt tgtattttta ctagagatgg ggtttcacca tgttggccag
181	gctggtctca atctcctaac ctcaagtgat ccatgagcct tggtctccca tagt

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

3.7.7 Y-STS sY1219

A Y-specific PCR product of 298 bp was obtained after amplification with the primer set indicated in Table 3.16. The PCR reaction was performed as described in Section 3.7.1. The PCR conditions were described at http://www.ncbi.nlm.nih.gov/nucleotide/G75484.

Table 3.16: Primers used in the amplification of Y-STS sY1219

Locus	STS primer name	Sequence
G75484	sY1219 Forward	5′-cca gac gtt cta ccc ttt cg-3′
	sY1219 Reverse	5′-ctc cct tgg ttc atg cca tt-3′

The annealing regions of the primer set given in Table 3.16 are indicated in Table 3.17. The sequence of the resulting PCR product is also demonstrated.

Table 3.17: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY1219 STS locus

Nucl no.	DNA sequence (PCR product of 298 bp):
1	ccagacgttc tacccttttcg agattagtta atatgtttac acacagagtt ttctttatag
61	gattataatt tacaatgttt tcacaatttt cttaaacagt ccactttatt ttatttaact
121	ttaagacaac ttttttattc ttaagcaaaa tacatagtta tgccttataa tttttaacta
181	aaaccacttt ttaccatttt tatacacttt tatgcaaatc catgtttagc agtagtgggc
241	gcctgtattc ccagctactc cagaggctga ggcaggagaa tggcatgaac caagggag

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

3.7.8 Y-STS sY2216

The primer set presented in Table 3.18 was used to generate the Y-specific PCR product of 75 bp indicated in Table 3.19. The PCR reaction mixture and amplification conditions were performed as described in Section 3.7.1. Successful amplification of the sample however, was determined with 3% agarose gel electrophoresis.

Table 3.18: Primers used in the amplification of Y-STS sY2216

Locus	STS primer name	Sequence
G66284	sY2216 Forward	5′-tct ggt gag agt ggc atc tg-3′
	sY2216 Reverse	5′-ttt agc aaa ccc tgt ggg ag-3′

The amplification product and primer binding regions of Y-STS sY2216 are indicated in Table 3.19. The amplification conditions can be obtained at http://www.ncbi.nlm.nih.gov/nucleotide/G66284.

Table 3.19: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY2216 STS locus

Nucl no.	DNA sequence (PCR product of 75 bp):
1	gatcacagtt agcttcaaag gctgaaaaag gacacatcat tggggtagtt tgccttcacc
61	catgtcattc aatcacacat gaccacagac atctggtgag agtggcatct gtactgtggt
121	tcttagaatg gagtaagcct cagaatctcc cacagggttt gctaaaatgc agactgccag
181	gcctcaccct cagagggttc agcanatc

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

3.7.9 Y-STS sY65

The primer set presented in Table 3.20 was used to generate the Y-specific PCR product of locus G66515. The PCR reaction was essentially performed as described in Section 3.7.1., with the exception that the annealing temperature was 60ºC.

Table 3.20: Primers utilised in the amplification of Y-STS sY65 locus

Locus	STS primer name	Sequence
G66515	sY65 Forward	5'-act aaa aca cca tta gaa aca aag g-3'
	sY65 Reverse	5'-ctg agc aac ata gtg acc cc-3'

The PCR product obtained with the primers used in Table 3.20 is 309 bp long. The sequence of the PCR product is indicated in Table 3.21. The PCR primer sequences and amplification conditions were obtained at http://www.ncbi.nlm.nih.gov/nucleotide/G66515.

Table 3.21: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY65 STS locus

Nucl no.	DNA sequence (PCR product of 309 bp):
1	actaaaacac cattagaaac aaaggactta aactaggaat taattatttc tctttctctt
61	tccatggcca acaaacattg aaaaaaaatt gccatctttt tttttatttg tttgttagag
121	atggggatct cactctgttt cttagattgt agtgccatgg cacaataatg gctcactgca
181	gcctcaaact cctgggctca agtgatcacc cccatacaga ctcccgagta gctgggaaca
241	caggcacatg ccaccacccc tagctaattt tttattattt gtagagatgg gggtcactat
301	gttgctcag

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

3.7.10 Y-STS sY57

The primer set presented in Table 3.22 was used to generate the Y-specific PCR product indicated in Table 3.23. The PCR reaction was performed according to the conditions described in http://www.ncbi.nlm.nih.gov/nucleotide/G38358 and Section 3.7.9.

Table 3.22: Primers used in the amplification of the Y-STS sY57 locus

Locus	STS primer name	Sequence
G38358	sY57 Forward:	5′-gaa ctt gtc ggg agg caa t-3′
	sY57 Reverse:	5′-tga tac act tcc tcc ttt agt gg-3′

Table 3.23 demonstrates the sequence obtained after amplification. The size of the PCR product is 288 bp.

Table 3.23: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY57 STS locus

Nucl no.	DNA sequence (PCR product of 288 bp):
1	atgcaagctt caggaacttg tcgggaggca atggtgacat ncattgtnnc cttagccnnn
61	gctcacaatc aaccatggtg cactgcgact agctcatgca cattcatcag gcagattcag
121	gcacctggct gtcagagctg tcagccttcc tcagtagagg aaaatnctac agtcggcact
181	ggcctggtat caggaaaata gatgcctgca aaancccact gtgggaccct aaaagtcttg
241	acctcaggtc ccctttgtcc tgtctctgtt gtcaggatcc actaaaggag gaagtgtatc
301	aagactctna ggtggtctgt ggaaactnct cttctgtctc catttttnaa aggctgtgtg
361	caataattgg gtcctgtgga gattggagta tagtc

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

3.7.11 Y-STS sY1079

The primer set presented in Table 3.24 was used to generate the Y-specific PCR product of 246 bp indicated in Table 3.25. The conditions of the PCR reaction are described in Section 3.7.1. and at site http://www.ncbi.nlm.nih.gov/nucleotide/G66115.

Table 3.24: Primers used in the amplification of the Y-STS sY1079 locus

Locus	STS primer name	Sequence
G66115	sY1079 Forward	5′-gga aaa ttt tga tat gcc caa-3′
	sY1079 Reverse	5′-ctg cag aaa tca cat gga cag-3′

The annealing regions of the primers described in Table 3.24 are indicated in Table 3.25. Table 3.25 indicates the sequence of the resulting 246 bp PCR product.

Table 3.25: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY1079 STS locus

Nucl no.	DNA sequence (PCR product of 246 bp):
1	ggaaaatttt gatatgccca aagttcttaa ctttcttttc tttagaagtt tcatatttca
61	gtctaggtat gagatggaat tgactgtgat cattgttttt atttcactgt gactactgag
121	tttctgatac agtgatacta atgtatagac ttaaaagctt gcctcctcct cctcctcctt
181	cttcttcttc tcctcctcct catcctcctc ctccccgctt tggacctgtc catgtgattt
241	ctgcag

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

3.7.12 Y-STS sY1250

Table 3.26 presents the primer pair used to generate the Y-specific PCR product of Y-STS sY1250. A description of the PCR reaction and cycling conditions were given in Section 3.7.1.

Table 3.26: Primers utilised in the amplification of Y-STS sY1250

Locus	STS primer name	Sequence
G75495	sY1250 Forward:	5′-ttt ttc taa cct tgc ctg cg-3′
	sY1250 Reverse:	5′-tgc aga gaa gca gcc tac aa-3′

The sequence of the PCR product of 493 bp is indicated in Table 3.27. The binding regions of the primers are also indicated in this table. The amplification conditions were obtained at http://www.ncbi.nlm.nih.gov/nucleotide/G75495.

Table 3.27: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY1250 STS locus

Nucl no.	DNA sequence (PCR product of 493 bp):
1	tttttctaac cttgcctgcg gtttgcacca tttattacat ttttccaaca accaaaggtt
61	ggctttgtat gttttactaa ttttctctac atcattatcc cctcacttta gtttttcaga
121	attgattctg ttgtttctgt tctaattctt tctttaaata tctagtacat taattttcaa
181	gttgtagaaa catttgtcta tgaactccta ttgcaatatc acttttcctg ctacccacaa
241	atttaatctg taatatttgc agtaccatta atttctatta tgaatttata ttatgataat
301	ccatgagttg ctgagaaata gctgttataa ttttgttgtt caattcccat ttaattttat
361	tttaacttgt gctaactcag ttgaaaattc tttactaatt ttttaaatct catatcaaga
421	cttttattca catcaattgt tctataaatg ctccctcctt gaagaacact ctcttgtagg
481	ctgcttctct gca

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

3.7.13 Y-STS sY1243

The Y-specific amplification product of Y-STS sY1243 was obtained by using the primer set presented in Table 3.28. After amplification, a fragment of 493 bp was obtained.

Table 3.28: Primers used in the amplification of the Y-STS sY1243 locus

Locus	STS primer name	Sequence
G75489	sY1243 Forward:	5′-atc tgc aca ctt ggg tag gc-3′
	sY1243 Reverse:	5′-gag gaa atg cag aat ttg gg-3′

The DNA sequence of the PCR product obtained after amplification, is indicated in Table 3.29. The primer annealing regions are also indicated in Table 3.29.

Table 3.29: Partial sequence of the human Y-chromosome demonstrating the annealing regions of the primer set used to amplify the sY1243 STS locus

Nucl no.	DNA sequence (PCR product of 493 bp):
1	atctgcacac ttgggtaggc aaggcaggta gattaccaga tgtcagaagt tcaggaccag
61	cctggtcaac atagtgaaac cccatctcta ctaaatattc aaaaattagc caggtatggt
121	ggcaagtgcc tggaatccca gctactcggg aggctgaggc aggagaatta gttgaaccca
181	tgaggtggag gttgcagtaa gccaatatca gaccactgca atccagcctg ggccacaaga
241	gcaaaacttt ttctccaccc cacccccca aaaaaggcgt cacatttaca taaacataat
301	tatcttaaaa gccagtataa ttttaatttt attgtagtca tcaccttcag acattgttta
361	ttttggagaa gtgattatgg aaatctgaaa tatcaaggcc tgatgagaat acttaaatta
421	accacactcc agaagtccaa atctgaaaag caaagatgtt tctgatataa tagcccaaat
481	tctgcatttc ctc

Nucl no. = nucleotide number. The primer sequences are indicated by the underlined text.

The PCR reaction was performed according to the method described in Section 3.7.1. The primer sequences and amplification reaction conditions are described at http://www.ncbi.nlm.nih.gov/nucleotide/G75489.

3.8 AGAROSE GEL ELECTROPHORESIS

Detection of PCR products was carried out using a 2% (w/v) mini agarose gel. The gel was made up to a final volume of 100 ml and contained 2 g agarose (Bioline, Celtic), 100 ml 0.5X TBE (10X TBE buffer: 89.15 mM Tris, pH 8.1, 88.95 mM boric acid, 2.498 mM Na_2EDTA) and 0.5 $\mu g.ml^{-1}$ ethidium bromide (Promega). For Y-STS sY2216, successful amplification of the sample was determined by using a 3% agarose gel. In this case, the

only difference was that 3 g agarose was used instead of 2 g agarose. Ten µL of the PCR product was added to 1 µL bromophenol blue loading solution (Promega). Electrophoresis was carried out for 60 min at 10 volts (V) per centimetre ($V.cm^{-1}$) in 0.5X TBE buffer. DNA was visualised with an ultraviolet (UV) transilluminator and the images were captured with a Sony Cyber-shot® digital still camera.

CHAPTER FOUR

Results and Discussion

This study was undertaken to determine the size of the deletion in four AMELY-negative males from the black South African population observed during an investigation of 6,130 black South African individuals and 2,214 Caucasian individuals. This was achieved by the DNA typing of certain STS markers surrounding the amelogenin locus. Y-STR haplotypes were also determined, in order to determine gender and examine the possible relationship between these samples.

Amelogenin negative (AMELY-negative) males have been described in different population groups, as discussed in Section 2.4. Misidentification of the male genotype by the amelogenin test has been reported to be particularly prevalent among individuals of Indian origin (Chang *et al*., 2007). Little data exists on the incidence of AMELY-negative individuals in the global population and that which there is, is only based on small sample sizes as indicated in Table 2.1. The incidence of AMELY-negative males reported in European derived groups is low. The frequency of AMELY-negative males in Australia (Mitchell *et al*., 2006) is similar to those reported for the Austrian population (Steinlechner *et al*., 2002). AMELY-negative males also appear to be rare in the UK, Spain and Slovenia (Chang *et al*., 2003). Reports to date indicate that of the different population groups, South Asians have a higher frequency of AMELY-negative males and the deletion may therefore be population-specific (Chang *et al*., 2003).

Although the incidence of AMELY-negative individuals is low in most populations tested, the serious consequences of an incorrect gender determination in forensic or identity investigations indicates that caution should be taken when the AMELY gene is used for gender determination. Brinkman (2002) argues that AMELY-negative males appear to be rare because of the lack of information on the occurrence of this phenomenon. Due to the paucity of data with regard to the prevalence of AMELY-negative males in black African populations the following investigation was undertaken in a cohort of black South African males. The aim of this investigation is to determine if a deletion at the AMELY locus is present in South Africans and to compare the size of the deletion and the Y-STR

haplotypes associated with the deletion, with previously published data. This could help determine how unique these deletions are and if the deletions are a result of independent events or have arisen from a common source.

4.1 DNA ISOLATION

DNA quantification was not performed seeing as DNA was isolated and amplified as routine procedures in the laboratory according to verified and standardised operating procedures. Due to the fact that the samples collected were not compromised, sample starting volumes and DNA volumes for the PCR reaction were determined by the yield of sufficient PCR product. According to the manufacturer of the DNA isolation kit, 4.5 to 9.0 µg DNA can be isolated from 200 µL whole blood. Throughout, 200 µL blood was used and the DNA was eluted in 500 µL Buffer AE (QIAGEN, 2003). Of this volume, 0.3 µL or 1.25 µL DNA sample was used for the PCR reaction depending on the kit used, as described in Section 3.2. It was determined that sufficient DNA was present if the reaction yielded a PCR product that could be analysed. For the buccal swabs, 1 swab yields 0.8 to 2.0 µg DNA according to the manufacturer. The DNA was eluted in 150 µL Buffer AE. Again, the volume of DNA sample used, as described in Section 3.2, was enough to yield sufficient PCR product.

4.2 STR ANALYSIS

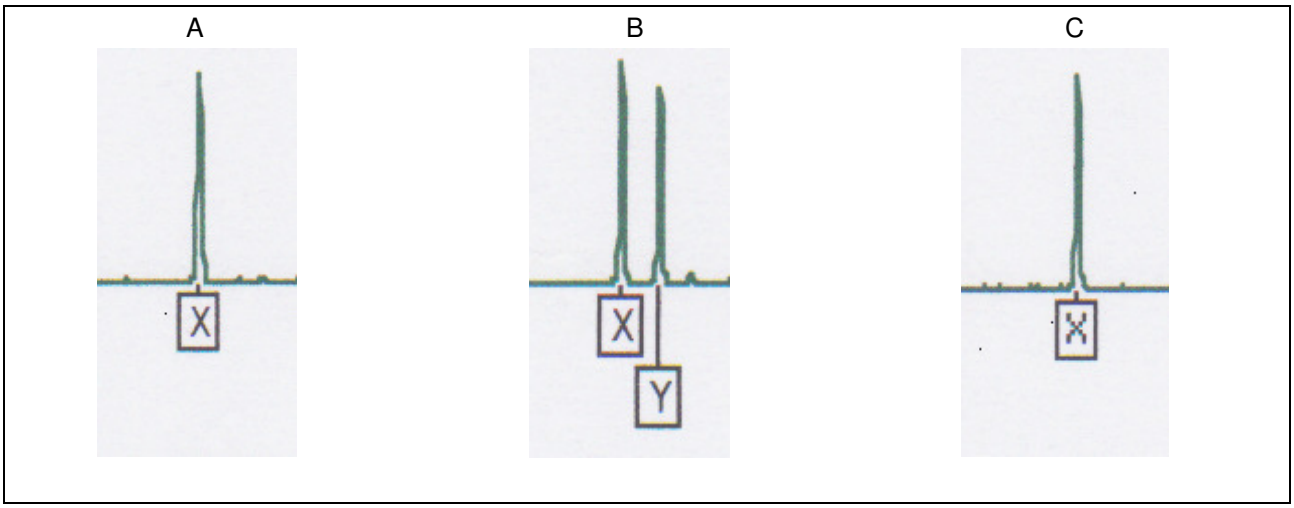
Amelogenin based sex-tests are part of the various PCR multiplex reactions used in DNA identification kits from different commercial suppliers. In this study, analysis of 6,130 male samples from the black South African population and 2,214 male samples from the Caucasian South African population, revealed a deletion of AMELY in four male samples from the black South African population. As these were two father and son pairs, the deletions present in the sons were assumed to be identical by descent to those of the fathers and thus were counted as only two deletion events.

4.2.1 AmpFℓSTR SGM Plus® PCR

Capillary electrophoresis of amplified DNA generated with the AmpFℓSTR SGM Plus® kit, yielded peaks of 103 bp and 109 bp for males with no deletion, after electrophoresing the

resultant products on an ABI PRISM® 310 Genetic Analyser (La Fountain *et al*., 1998). The first peak is from the X-chromosome and the second peak from the Y-chromosome. Figure 4.1 indicates DNA from a male containing, as expected, two peaks as well as that from a female containing one peak. In contrast, DNA from the AMELY-negative males presents with amplification of a 103 bp fragment (X-chromosome) only. The height and area of this peak is similar to that of the female sample, leading to incorrect gender assignment.

Figure 4.1: Representative electropherogram of amelogenin PCR



DNA from a female (A), an AMELY-positive male (B) and an AMELY-negative male (C) were subjected to PCR using the AmpFℓSTR SGM Plus® PCR system. XY designations were determined using the Kazam®[1] Macro. Genescan® software was used to size the X and Y PCR products. Horizontal axes represent PCR product sizes in bp. Vertical axes denote fluorescent intensity.

4.2.2 PowerPlex® 16 System

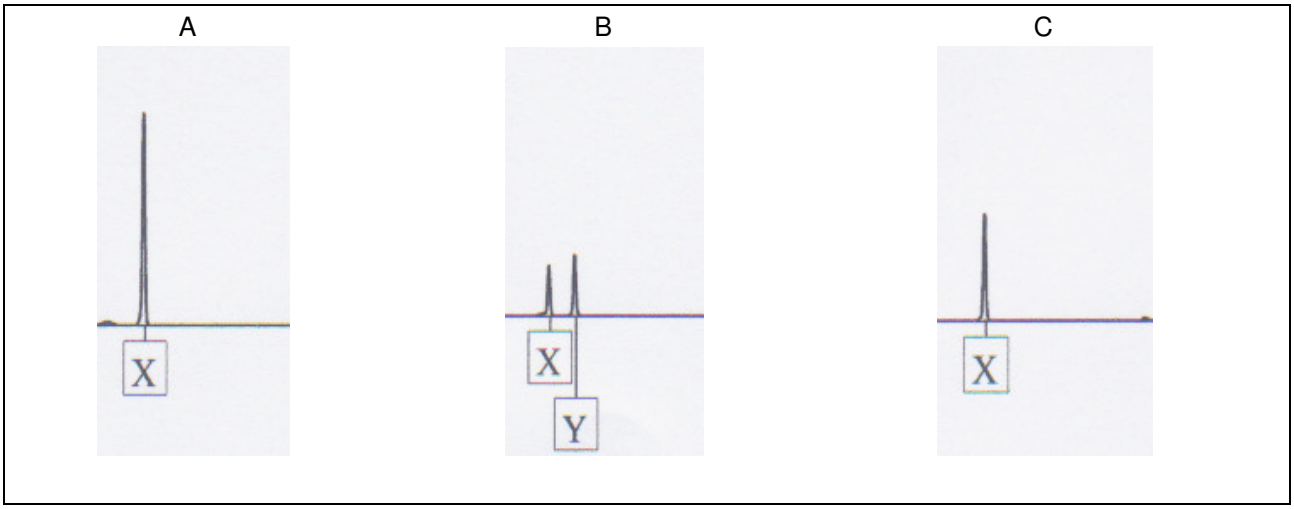
Capillary electrophoresis of DNA generated with the PowerPlex® 16 System kit, yielded peaks of 212 bp and 218 bp for the amelogenin locus. The amplicon generated from the X-chromosome is 212 bp while the Y-chromosome amplicon is 218 bp. Figure 4.2 illustrates samples from an AMELY-positive male containing two peaks, a female containing one peak and an AMELY-negative male demonstrating amplification only at 218 bp (X-chromosome). The characteristics of the peak obtained for the AMELY-negative male is comparable to that of a female sample and that could result in an incorrect gender designation.

These results were obtained from two phenotypically unaffected black males and their sons. In 1998, Santos *et al*. reported two phenotypically unaffected males as

[1] Kazam® is a registered trademark of Life Technologies, Carlsbad, CA, USA.

AMELY-negative. They thought it probable that a deletion of the amelogenin locus on the Y-chromosome was responsible for the wrong gender assignment.

Figure 4.2: Representative electropherogram of amelogenin PCR



DNA from a female (A), an AMELY-positive male (B) and an AMELY-negative male (C) were subjected to PCR using the PowerPlex 16® system. XY designations were determined using the PowerTyper 16® Macro. Genescan® and Genotyper® software were used to size the X and Y PCR products. Horizontal axes represent PCR product sizes in bp. Vertical axes denote fluorescent intensity.

Subsequently, a number of similar deletions have been reported (see Chapter Two). Santos *et al*. (1998) observed an 8% frequency of sex test failures due to a deletion in their Sri Lankan population set. Thangaraj *et al*. (2002) observed a deletion frequency of 1.85% in an Indian population set. An observed frequency of amelogenin sex test failures in a Caucasian Austrian sample was 0.018% (Steinlechner *et al*., 2006). Other authors have obtained similar frequencies for European derived populations. Chang *et al*. (2007) also observed a high incidence of AMELY-negative males in an Indian and Malay population, but none in a Chinese population. The frequency of AMELY-negative males observed in the different populations are summarised in Table 4.1.

The incidence of AMELY-negative males in the black South African population is low (0.065%) and similar to the incidence reported for the European derived population groups. This data supports previous observations that the AMELY-negative alleles appear to be population-specific, with the highest incidence among South Asians. Chang *et al.* (2007) argues that a lineage associated with AMELY-negative males may have spread from the Indian subcontinent whereas in other populations similar mutations have occurred sporadically.

Roffey *et al*. (2000) previously reported a wrong gender assignment for a phenotypically unaffected male with the amelogenin test. This failure was attributed to a mutation in the

primer binding region of the amelogenin primers employed in the kit, because the correct gender was determined by using alternate amelogenin primers. Therefore to ensure that the results generated in this study were not due to variations in the primer binding region, the primer binding sites were investigated by means of an alternative primer pair.

Table 4.1: Frequency distribution of AMELY-negative males in different populations

Population	Cohort size	No. negatives	% negatives	Reference
South African black	6,130	2	0.065	Present study
South African Caucasian	2,214	0	0.000	Present study
Australia	109,000	22	0.020	Mitchell *et al.,* 2006
Europe				
Austrian	29,432	6	0.018	Steinlechner *et al.,* 2002
Italian	13,000	1	0.008	Lattanzi *et al.,* 2005
England	2,000	2	0.100	Chang *et al.,* 2007
Spain	1,000	1	0.100	Chang *et al*., 2007
Asia				
Israeli	96	1	1.040	Michael *et al.,* 2004
Malaysian Malays	113	1	0.880	Chang *et al.,* 2003
Malaysian Malays	334	2	0.600	Chang *et al.,* 2007
Malaysian Indians	112	4	3.570	Chang *et al.,* 2003
Malaysian Indians	315	11	3.490	Chang *et al.,* 2007
India (general)	270	5	1.850	Thangaraj *et al.,* 2002
South Indian	100	1	1.000	Chang *et al.,* 2007
Sri Lankan	24	2	8.330	Santos *et al.,* 1998
Nepal	77	5	6.490	Cadenas *et al.,* 2007
Indian	4,257	10	0.230	Kashyap *et al.,* 2006
Chinese	113	0	0.000	Chang *et al*., 2003
Chinese	331	0	0.000	Chang *et al.,* 2007
Taiwan	80,000	3	0.004	Kao *et al.,* 2007

4.3 PCR PRIMER BINDING SITES

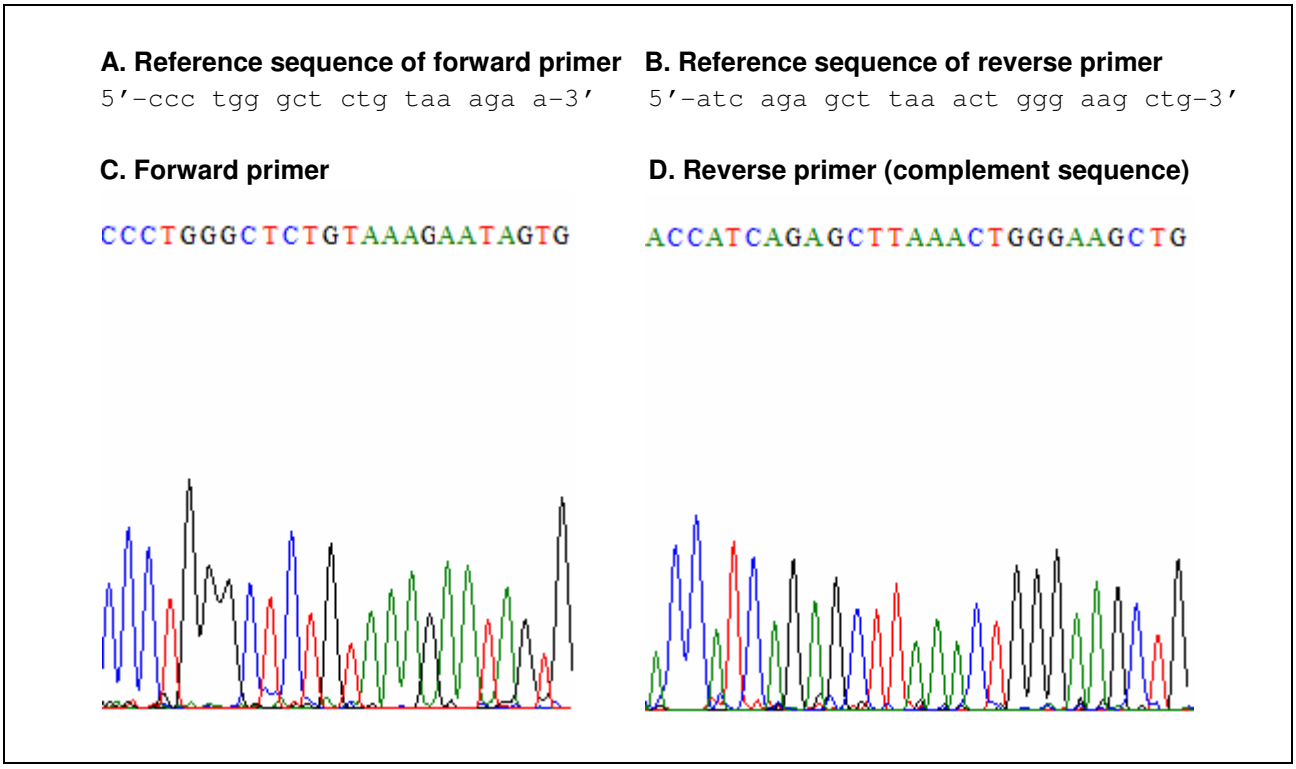
A possible explanation for the failure to amplify amelogenin sequences in the four samples within this investigation could be that a mutation occurred within the annealing region of the primers employed by the AmpFℓSTR SGM Plus® PCR and PowerPlex 16® System kits. A strategy to solve this problem was to amplify the amelogenin locus using the alternative primers described by Steinlechner *et al*. (2002). DNA from the AMELY-negative males was subjected to PCR using these primers, according to the protocol described in Section 3.5.

PCR products were analysed with agarose gel electrophoresis and the resultant fragments were gel-purified and sequenced.

4.4 CHAIN TERMINATION SEQUENCING

Sequencing was achieved as described in Section 3.6 and the resultant electropherograms were stored electronically. Sequencing results were inspected and manually aligned using the freeware BioEdit® sequence alignment editor program. Sequence alignment was accomplished by comparison to the Genbank® sequence of the human amelogenin gene (accession number M55419). Figures 4.3 and 4.4 indicate the sequencing results from one of the AMELY-negative males. Sequence analysis of the AMELY-negative males did not reveal a mutation in the primer binding regions of the primers used by the AmpFℓSTR SGM Plus® PCR and PowerPlex 16® System kits.

Figure 4.3: Representative electropherogram of an AMELY-negative male from this investigation.

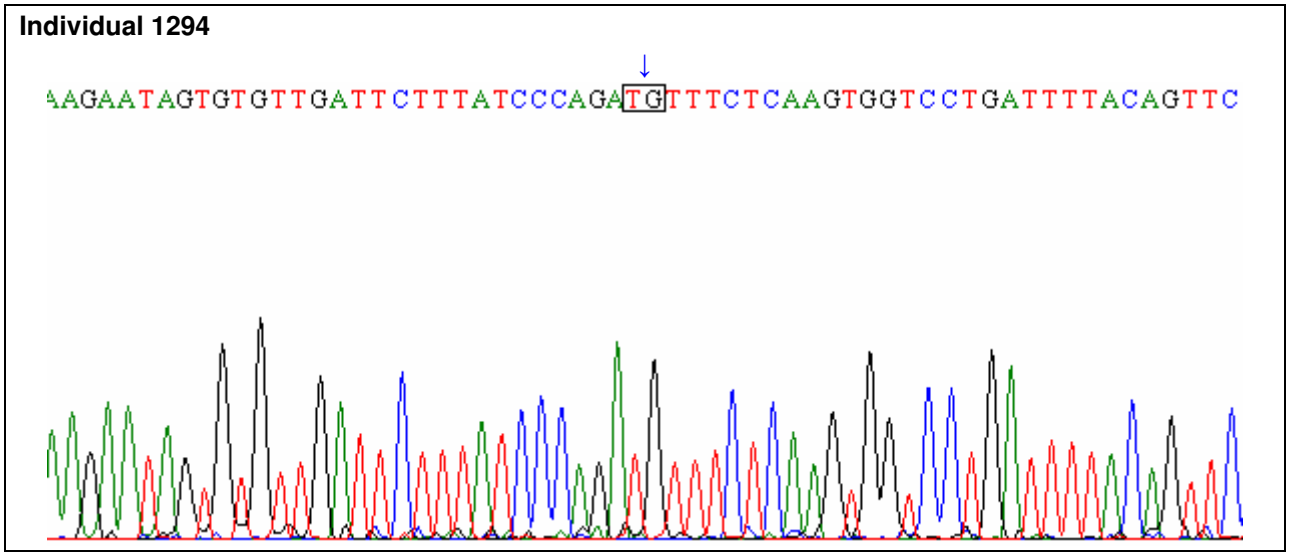


Sequence analysis of the forward strand (C) and reverse strand (D) are given to indicate the primer binding regions employed by the commercial kits. The complement sequence for the reverse primer is indicated. This was achieved with the BioEdit® sequence alignment editor program. Reference sequence = Genbank® accession number M55419.

The absence of a Y-specific product after amplification with the primers used by the commercial kits and the alternative primers external to those, as indicated in Table 3.3, suggests that a length mutation involving both sets of primers occurred rather than a

mutation at one of the priming sites of the amelogenin primer sets. The commercial kits have primers that are bracketed by the alternative external primers. If there was a mutation within the priming sites of one of the commercial kits, the external primers would have generated an amplification product. However, neither of these sets of primers generated products, hence the conclusion that the deletion encompasses all four of these priming sites, of both the commercial and alternative primer sets. Exclusive amplification of the X-chromosome is illustrated in Figure 4.4.

Figure 4.4: Representative electropherogram of an AMELY-negative male from this investigation



Partial sequence analysis is given to illustrate exclusive amplification of the X-chromosome. The position of the 6 bp deletion on the X-homolog is indicated (↓).

Moreover, earlier studies on AMELY-negative males indicated that the failure to amplify a Y-specific product using the described primers was due to a deletion on the Y-chromosome involving the amelogenin gene (Thangaraj *et al*., 2002; Chang *et al*., 2003; Lattanzi *et al*., 2005; Cadenas *et al*., 2007; Mitchell *et al*., 2006; Kashyap *et al*., 2006; Jobling *et al*., 2007). The conclusion made above, that the failure to amplify amelogenin sequences in the four samples within this investigation is due to a deletion on the Y-chromosome, therefore seems appropriate.

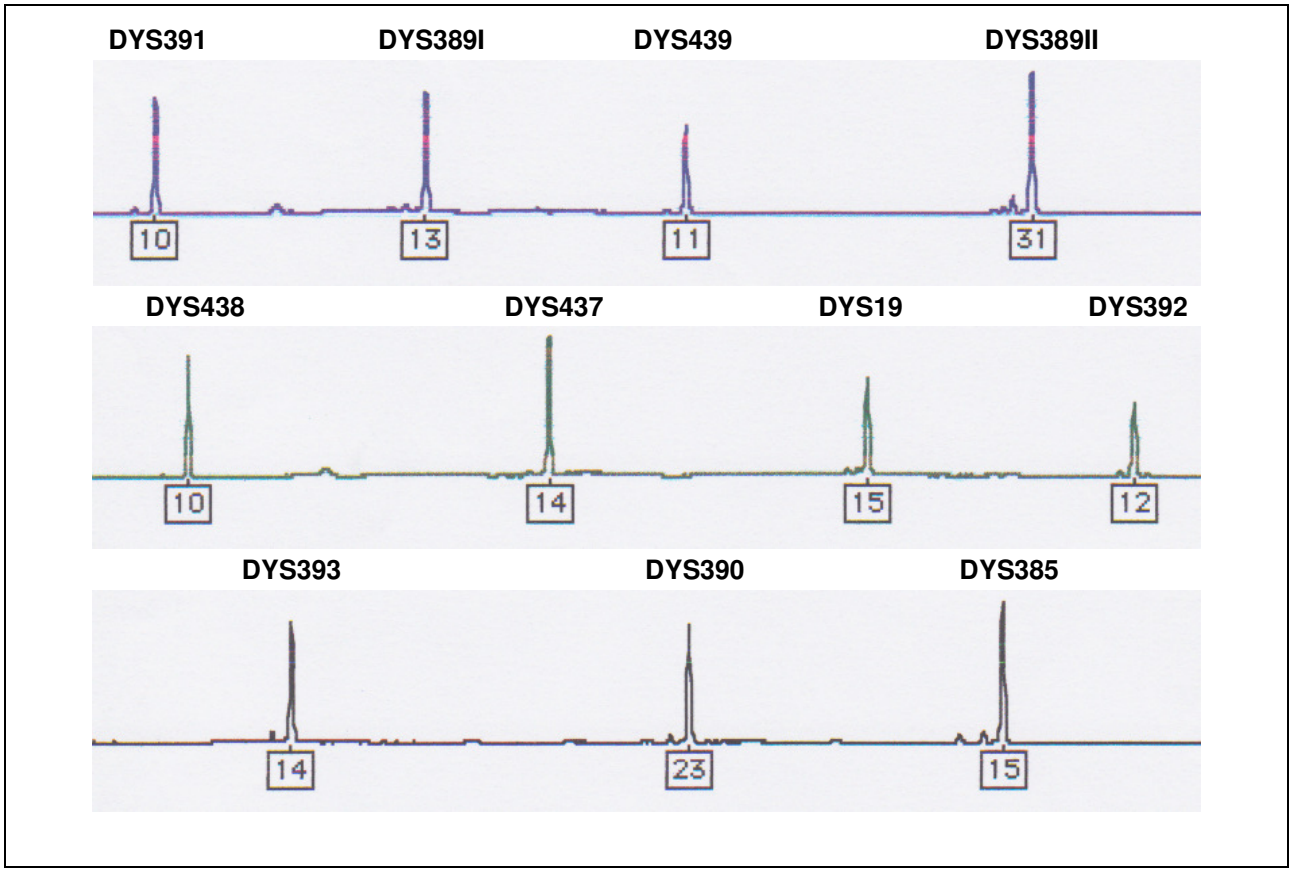
4.5 Y-CHROMOSOME HAPLOTYPING

In order to confirm the male phenotype for all four individuals, Y-STR loci were typed. These Y-STRs were typed with the PowerPlex Y® System, described in Section 3.4. Amplification of the eleven Y-chromosomal STR markers i.e. DYS391, DYS389I,

DYS389II, DYS439, DYS438, DYS437, DYS19, DYS392, DYS393, DYS390 and DYS385, was successful in all AMELY-negative males and the sex could therefore be determined correctly by typing the Y-STR loci. The Y-STR loci are dispersed over both arms of the Y-chromosome and these results therefore also suggested that the assumed deletion is restricted to the amelogenin-related region of the Y-chromosome.

The Y-STR haplotypes for the fathers and their respective sons were identical and only the haplotypes of the fathers are therefore presented. The Y-chromosome DNA haplotype profile of individual 1007 is presented in Figure 4.5.

Figure 4.5: PowerPlex Y® electropherogram of individual 1007



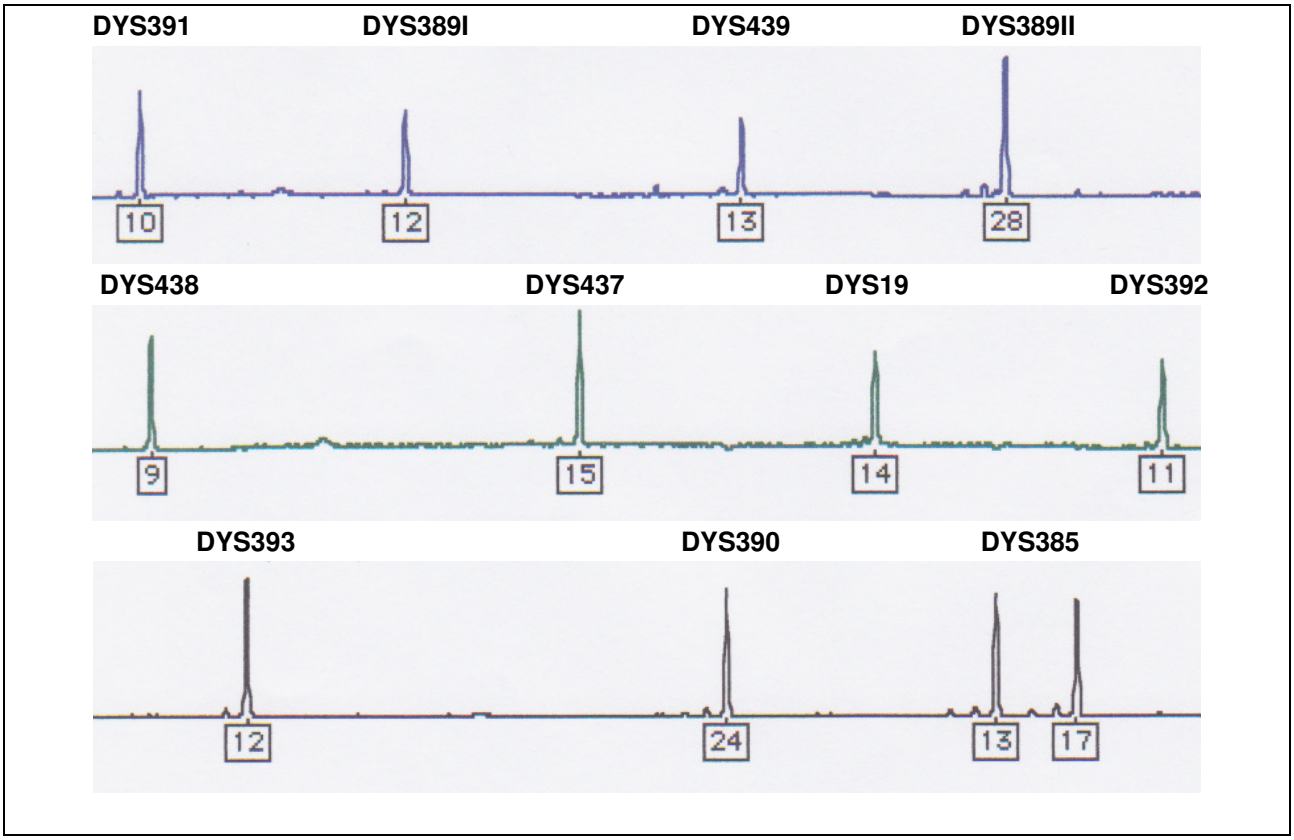
DNA was subjected to PCR using the PowerPlex Y® system. Designations were determined using the PowerTyper Y® Macro. Genescan® and Genotyper® software were used to size the PCR products. Horizontal axes represent PCR product sizes in bp. Vertical axes denote fluorescent intensity.

Figure 4.6 illustrates the Y-STR haplotype profile of individual 1294. There are two peaks at locus DYS385 in individual 1294, but not in individual 1007. Some Y-STR loci occur more than once due to duplicated palindromic regions in the Y-chromosome. A locus-specific set of primers can therefore produce more than one PCR product that may appear to be two loci for a Y-chromosome haplotype. DYS385 is present as two duplicated regions located 40,000 bp apart on Yq. Two alleles can thus be generated when DYS385 is amplified with a single set of primers, appearing as two peaks in the electropherogram.

Both alleles generated can also be the same size, resulting in a single peak in the electropherogram (Butler, 2005).

It is apparent from Figure 4.5 and 4.6 that the Y-STR haplotypes of the two black South African AMELY-negative males are very different. Of the eleven loci, they have only one in common. This difference in Y-STR haplotypes can either be interpreted as the deletion being a relatively old single deletion event which has undergone high levels of genetic differentiation or that the deletion has occurred several times. Related Y-STR haplotypes would suggest that the Y-chromosome deletions probably occurred in the same paternal lineage (Cadenas *et al.*, 2007) therefore the observation of two separate Y-STR haplotypes for the two South Africans implies that the deletions have an independent origin.

Figure 4.6: PowerPlex Y® electropherogram of individual 1294



DNA was subjected to PCR using the PowerPlex Y® system. Designations were determined using the PowerTyper Y® Macro. Genescan® and Genotyper® software were used to size the PCR products. Horizontal axes represent PCR product sizes in bp. Vertical axes denote fluorescent intensity.

Previous studies have determined Y-STR profiles for AMELY-negative males (Chang *et al.*, 2002; Henke, et al., 2001; Steinlechner *et al.*, 2002; Thangaraj *et al.*, 2002; Chang *et al.*, 2003; Lattanzi *et al.*, 2005; Mitchell *et al.*, 2006; Kashyap *et al.*, 2006; Chang *et al.*, 2007; Cadenas *et al.*, 2007; Jobling *et al.*, 2007). These studies consist of a maximum of

71 AMELY-negative males. Although there are many microsatellites in common between the samples, only one microsatellite locus, DYS390, was common in all the samples. The data is summarised in Table 4.2.

The haplotypes of individuals from populations that are geographically associated were grouped in Table 4.2 in order to reveal whether a pattern existed. For the European derived individuals, no clear pattern emerged. Not only were there very different Y-STR haplotypes observed for the European individuals than those for Asians, but they were also different from each other. This would suggest that the Europeans investigated belong to different paternal lineages. Four of the Australian individuals described by Jobling *et al*. (2007) have identical haplotypes and the two other Australian individuals share a number of alleles with these four. Unfortunately, no information was available on the Australian population described by Jobling *et al*. (2007) and it is therefore not known if they were of European descent. The Australian individuals seem to belong to the same paternal lineage or closely related paternal lineage and the mutation in this population described are therefore likely inherited by descent. Mitchell *et al*. (2006) also reported on the Australian population, but indicated in their study that one individual was of Italian descent and four individuals of either Indian or Asian descent. It is indicated as such in Table 4.2. Similarly, it was determined that the Y-STR haplotypes of the five Austrian individuals (Steinlechner *et al*., 2002) and the Indian samples (Thangaraj *et al*., 2002; Kashyap *et al*., 2006) were different and it was suggested that there were different paternal lineages carrying this deletion and that the mutation probably arose independently. This argument seems valid since the diverse haplotypes indicates different origins for the deletion, which arose independently. If this were not the case, the deletion would have segregated on the same haplotype in the different populations. The fact that the amelogenin deletion occurs in apparently unrelated individuals from different populations could suggest a hot spot for mutations in that region of the Y-chromosome. Although it has not been described in the literature as an area prone to rearrangements, this can explain the origin of a similar deletion in different populations.

Only one Y-STR haplotype was determined to be shared between two Indian individuals from different regions (Thangaraj *et al*., 2002). In the Indian and Malay populations, Chang *et al*. (2007) observed no distinct classes of related Y-STR haplotypes in eighteen samples that corresponded with the deletion on the Y-chromosome. In a previous report, Cadenas *et al*. (2007) observed that the Y-STR loci reported for AMELY-negative males in the Indian population studied were similar in a number of shared alleles.

Table 4.2: Y-chromosome STR markers in AMELY-negative males

Origin	19	385	388	389I	389II	390	391	392	393	434	435	437	438	439	448	456	458	635	DYAA7.2	Y-GATA	Ref.
SA # 1007	15	15,15	---	13	31	23	10	12	14	---	---	14	10	11	---	---	---	---	---	---	11.
SA # 1294	14	13,17	---	12	28	24	10	11	12	---	---	15	9	13	---	---	---	---	---	---	11.
English	15	15,15	13	13	31	23	10	12	13	11	11	14	10	11	20	---	---	---	---	---	10.
English	14	11,14	12	13	29	23	11	13	13	11	11	15	12	12	19	---	---	---	---	---	10.
Italian	13	---	---	13	29	24	11	13	13	---	---	---	12	12	---	---	---	---	---	---	5.
Italian	13	---	---	13	---	24	11	13	13	9	11	9	12	12	---	---	---	---	---	---	6.
Austrian	13	11,14	---	13	29	24	10	13	13	---	---	---	---	---	---	---	---	---	---	---	2.
Austrian	14	13,15	---	13	29	22	10	11	12	---	---	---	---	---	---	---	---	---	---	---	2.
Austrian	13	15,18	---	13	30	23	10	11	12	---	---	---	---	---	---	---	---	---	---	---	2.
Italian	16	---	---	14	30	22	11	11	14	---	---	---	10	11	---	---	---	---	---	---	5.
Austrian	16	13,14	---	12	28	22	10	12	13	---	---	---	---	---	---	---	---	---	---	---	2.
Austrian	14	14,14	---	12	27	22	10	11	13	---	---	---	---	---	---	---	---	---	---	---	2.
Austrian	No	No	---	No	No	No	No	No	No	---	---	---	---	---	---	---	---	---	---	---	2.
Australia	14	11,14	12	13	29	24	10	13	13	11	11	15	12	12	19	---	---	---	---	---	10.
Australia	14	11,14	12	13	29	24	10	13	13	11	11	15	12	12	19	---	---	---	---	---	10.
Australia	14	11,14	12	13	29	24	10	13	13	11	11	15	12	12	19	---	---	---	---	---	10.
Australia	14	11,14	12	13	29	24	10	13	13	11	11	15	12	12	19	---	---	---	---	---	10.
Australia	14	13,13	15	12	29	24	11	11	12	11	11	15	7	11	18	---	---	---	---	---	10.
Australia	15	13,16	15	12	28	24	9	11	12	11	11	15	9	12	19	---	---	---	---	---	10.
Bedouin	14	13,17	17	13	29	22	11	11	12	11	11	14	10	11	20	---	---	---	---	---	10.
Afghan	15	13,16	15	12	28	25	10	11	12	11	11	15	9	13	19	---	---	---	---	---	10.
Indian	16	13,17	---	12	28	24	10	11	12	---	---	15	9	13	18	13	No	24	---	11	8.
Indian	16	13,17	---	12	28	24	10	11	12	---	---	15	9	12	21	13	No	21	---	11	8.
Indian	17	14,17	---	12	28	23	10	11	12	---	---	15	9	13	19	14	No	22	---	No	8.
Indian	15	13,16	---	12	28	23	10	11	12	---	---	15	9	12	19	13	No	21	---	11	8.
Indian	15	14,17	---	12	28	23	11	11	12	---	---	15	9	12	18	13	No	21	---	11	8.
Indian	16	13,16	---	12	28	23	10	11	12	---	---	15	9	13	18	13	No	22	---	11	8.
Indian	15	14,17	---	12	28	25	11	11	12	---	---	14	9	12	18	13	No	20	---	11	8.
Indian	16	13,17	---	12	28	25	10	11	12	---	---	15	9	11	18	13	No	21	---	11	8.
Indian	15	13,17	---	12	29	25	10	11	12	---	---	15	9	11	18	14	No	22	---	12	8.
Indian	17	13,18	---	12	29	25	10	12	12	---	---	15	9	12	18	13	No	21	---	11	8.
Indian	15	---	---	12	29	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	7.
Indian	15	---	---	12	29	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	7.
Indian	15	13,16	---	12	27	24	10	11	12	---	---	15	9	11	19	13	No	23	---	No	8.
Indian	16	---	---	12	---	24	10	11	12	---	---	---	9	---	---	---	---	---	---	---	6.
Indian	15	---	---	12	---	24	11	11	12	9	11	9	9	11	---	---	---	---	---	---	6.
Indian	15	14,17	---	11	28	24	10	11	12	---	---	15	9	12	19	13	No	21	---	11	8.
Indian	15	---	---	11	28	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	7.
Indian	14	14,18	12	13	31	24	10	11	14	12	11	14	10	11	19	---	---	---	---	---	10.

Origin	19	385	388	389I	389II	390	391	392	393	434	435	437	438	439	448	456	458	635	DYAA7.2	Y-GATA	Ref.
Indian	14	15,17	---	13	31	24	10	11	14	---	---	14	10	11	19	16	No	18	---	12	8.
Indian	14	15,18	---	13	31	26	11	11	15	---	---	14	10	11	19	15	No	18	---	12	8.
Indian	15	---	---	13	30	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	7.
Indian	15	---	---	13	30	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	7.
Indian	14	---	---	13	30	23	---	---	---	---	---	---	---	---	---	---	---	---	---	---	7.
Indian	14	---	---	13	30	22	10	---	14	---	---	---	---	---	---	---	---	---	---	---	3.
Indian	14	---	---	13	30	22	10	---	14	---	---	---	---	---	---	---	---	---	---	---	3.
Indian	14	---	---	13	30	21	11	---	14	---	---	---	---	---	---	---	---	---	---	---	3.
Indian	17	---	---	13	30	25	10	---	12	---	---	---	---	---	---	---	---	---	---	---	3.
Indian	15	---	---	13	29	24	---	---	---	---	---	---	---	---	---	---	---	---	---	---	7.
Indian	14	---	---	13	29	24	---	---	---	---	---	---	---	---	---	---	---	---	---	---	7.
Indian	15	---	---	13	26	25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	7.
Indian	15	---	---	11	30	23	---	---	---	---	---	---	---	---	---	---	---	---	---	---	7.
Indian	---	---	---	---	---	25	---	---	---	---	---	---	9	12	---	---	---	---	---	---	4.
Indian	---	---	---	---	---	25	---	---	---	---	---	---	9	11	---	---	---	---	---	---	4.
Indian	---	---	---	---	---	24	---	---	---	---	---	---	9	11	---	---	---	---	---	---	4.
Indian	---	---	---	---	---	24	---	---	---	---	---	---	9	13	---	---	---	---	---	---	4.
Indian	---	---	---	---	---	23	---	---	---	---	---	---	9	13	---	---	---	---	---	---	4.
Indian	16	---	---	14	30	24	11	---	12	---	---	---	---	---	---	---	---	---	---	---	3.
Nepal	15	---	No	12	29	23	10	11	12	---	---	---	---	12	---	---	---	---	7	---	9.
Nepal	15	---	15	12	29	23	10	11	12	---	---	---	---	13	---	---	---	---	7	---	9.
Nepal	13	---	15	12	29	23	9	11	12	---	---	---	---	12	---	---	---	---	7	---	9.
Nepal	15	---	15	12	29	24	10	11	12	---	---	---	---	12	---	---	---	---	7	---	9.
Nepal	15	---	15	12	29	23	10	11	12	---	---	---	---	13	---	---	---	---	7	---	9.
Asian	15	---	---	12	---	24	11	11	13	9	11	9	9	12	---	---	---	---	---	---	6.
Asian	15	---	---	12	---	23	10	11	12	9	11	9	9	12	---	---	---	---	---	---	6.
Maldives	15	14,18	16	13	30	24	10	11	12	11	11	15	9	12	19	---	---	---	---	---	10.
Maldives	14	15,17	12	13	31	24	10	11	14	12	11	14	10	11	19	---	---	---	---	---	10.
Malay	13	11,18	---	12	26	25	11	12	13	---	---	14	10	13	21	15	No	19	---	12	8.
Malay	15	13,18	---	12	28	24	10	11	12	---	---	15	9	12	17	13	No	20	---	11	8.
Malay	15	13,17	---	12	28	24	10	11	12	---	---	14	9	13	18	13	No	21	---	10	8.
Malay	16	13,17	---	12	28	24	10	11	12	---	---	15	9	12	19	13	No	20	---	11	8.
Malay	---	---	---	---	---	25	---	---	---	---	---	---	10	13	---	---	---	---	---	---	4.
Moroccan	14	---	---	10	26	23	10	11	12	---	---	---	---	---	---	---	---	---	---	---	1.

Ref. = reference; SA = South African; No = no amplification; --- = not performed. Yellow highlight indicates most common allele; Blue highlight indicates second most common allele; Orange highlight indicates third most common allele; Purple highlight indicates fourth most common allele. References: 1. Henke *et al.* (2001); 2. Steinlechner *et al.* (2002); 3. Thangaraj *et al.* (2002); 4. Chang *et al.* (2003); 5. Lattanzi *et al.* (2005); 6. Mitchell *et al.* (2006); 7. Kashyap *et al.* (2006); 8. Chang *et al.* (2007); 9. Cadenas *et al.* (2007); 10. Jobling *et al.* (2007); 11. Present study.

Cadenas *et al*. (2007) noted that the five AMELY-negative males in their study shared between seven and ten alleles. The similar Y-STR haplotypes that these individuals share suggested that the deletions could have occurred in the same paternal lineage within this population. This aspect sets the Indian population apart from the other populations studied, as this is the only population where a paternal origin for this deletion seems likely because it segregated on identical or closely related haplotypes. In the other populations the deletion segregated on a variety of haplotypes, indicating diverse origins for the same deletion. The fact that the deletion segregates on an identical or closely related haplotype, might indicate a founder deletion in the Indian population. Similarly, the diverse haplotypes that the deletion segregates with in the other populations would indicate independent origins. This heritability might also account for the higher incidence of the deletion in the Indian population compared to the other populations. As indicated in Table 4.2 it is evident that individuals form the Australian population share a number of alleles with the Indian population. As the Australian population is not described, no conclusions can be made.

As mentioned before, the Y-STR haplotypes of the two black South African individuals are very different with only one allele in common. A Y-microsatellite database, Y-STR haplotype reference database or YHRD (www.yhrd.org), was searched in an attempt to identify other possible geographical location(s) of the Y-STR haplotypes observed for the black South African AMELY-negative males. This database is a collaborative effort established to generate reliable Y-STR frequency estimates in the assessment of matches in forensic and genealogical casework and the assessment of male population distributions as reflected by Y-STR haplotypes. It is important to note that this database, as most other similar ones, is heavily biased towards populations of Europe and Asia. Nonetheless, a search of the YHRD revealed that the Y-STR haplotype of individual 1294 was observed twice in an East Asian population. In fact, when this haplotype was compared to the haplotypes of AMELY-negative males published by Chang *et al*. (2007), it was noted that this haplotype was almost identical to a haplotype of an individual belonging to the Indian population (Y- haplogroup J2e1*). These haplotypes are indicated in green in Table 4.2. The only difference in the two haplotypes was at locus DYS19. At locus DYS19, the South African sample has 14 repeats and the Indian sample has 16 repeats. The Y-chromosome is inherited from father to son mostly unchanged and Y-STR markers are thus inherited unaltered from father to son. Male blood relatives therefore all carry the same Y-STR profile. This association is so significant that Y-STR markers are in fact used in kinship analysis to follow paternal lineages (Jobling *et al*., 1997). As most of the Y-chromosome does not undergo recombination, mutations thus generally create new

lineages (Kayser *et al*., 1997). Related Y-STR haplotypes therefore in general suggest that these individuals may be related by descent. If this assumption holds this result may indicate that the two males have an identical deletion. However, it seems unlikely that the African sample investigated here, and the Indian sample described in the literature, are related by descent. Alternatively, the deletions may be identical due to chance. In addition, the deletions may have occurred due to a hot spot mechanism. Previously, Repping *et al*. (2002) demonstrated a hot spot in the long arm of the Y-chromosome for rearrangements that resulted in deletions. It is possible that a similar hot spot exists in this region of Yp. If the amount of loci that are identical is taken into consideration, it is possible for a hot spot to exist in the region of the deletion as the Y-STR loci are dispersed over both arms of the Y-chromosome and are not restricted to the amelogenin-related region of the Y-chromosome.

The Y-STR haplotype of individual 1007 was not detected in the YHRD, but was also compared to the haplotypes of AMELY-negative males published by Jobling *et al*. (2007). Again, an almost identical haplotype was determined. Individual 1007 only differed at locus DYS393 from an English individual (haplogroup I). These haplotypes are indicated in blue in Table 4.2. The English individual has 13 repeats and the South African has 14 repeats at locus DYS393. As before, this related Y-STR haplotype could mean that these two males share an identical deletion by descent. The alternative version that similar deletions could be coincidental should again be considered. No detail is available on the ancestry of both these individuals and it is not possible to determine whether they are related by inheritance and which population group the haplotype segregates with. It appears that a rearrangement hot spot exists in this region of the Y-chromosome and that this deletion has spread in some populations, the South Asian population for example, through descent.

The Y-STR haplotypes of the two black South Africans are so different, that they appear to belong to different paternal lineages. If identical deletions are due to inheritance by descent, the two South Africans would not be expected to carry the same deletion. As these individuals seem to be unrelated, the most likely explanation for the occurrence of the deletions would be that they occurred due to chance. Currently, it is believed that the high frequency of AMELY-negative males in the South Asian population is due to inheritance by descent (Chang *et al*., 2007). Although this may be true for the South Asian population, the existence of a hot spot in Yp could mean that the frequency of AMELY deletions in other populations, including the black South African populations, may be higher than the initial results suggest. Due to the similarity in Y-STR haplotypes of the one

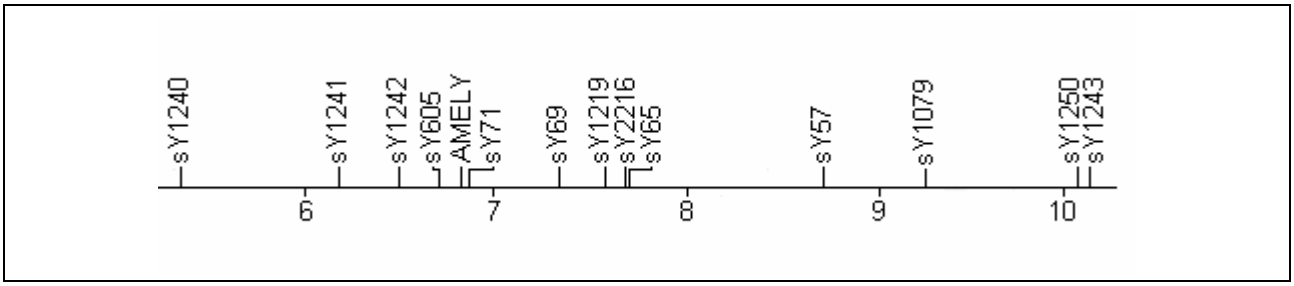
black South African with the Indian individual, and the other black South African with the English individual, it became important to determine the size of the deletion for each of these pairs. If an identical deletion size could be demonstrated within these pairs, it may support the idea of a common ancestry of the deletion, although it seems unlikely given the ethnic origins of the Africans versus the Indian and English individuals. Having said that, no information is available on the ancestry and possible ethnic origin of these individuals. It is thus possible that they do have a distant common ancestor, possibly out of Africa. The ethnic origin of this ancestor is not known. Conversely, if an identical deletion size can be demonstrated in different populations with no apparent common origin, it may support the suggestion of a hot spot for rearrangements in that area of the Y-chromosome.

4.6 DELETION MAPPING

The principal aim was to determine the size of the deletion using a number of sites on the Y-chromosome surrounding the AMELY locus. It was hypothesised that comparisons among the South African samples and those previously reported, might reveal differences in deletion size.

Various STS markers surrounding the AMELY locus were used. These markers are described in Section 3.7. Each STS provided a straightforward means of determining the presence or absence of a specific locus along the length of the Y-chromosome. A schematic representation of the different STS markers used in this study is presented in Figure 4.7, indicating the order of the STS markers as well as their relative positions on this region of the Y-chromosome.

Figure 4.7: Schematic representation of STS markers used for determination of deletion size



Adapted from Jobling *et al.* (2007).

A PCR reaction was considered successful if the following criteria were met:

i. The expected amplification product of the desired size was observed for the positive control, and
ii. Amplification was absent in the negative control, thereby indicating that no contamination by exogenous DNA of the PCR reaction had occurred .

Positive and negative controls were included in each PCR reaction set.

Even though the PCR reaction was successful for all STS markers, various artefacts were observed. The amplification efficiency for all samples for a specific STS, e.g. sY1240 (Figure 4.8), was not always the same. The variation in amplification efficiency was likely due to the fact that the DNA concentration of the samples were not determined, as discussed in Section 3.2. Small differences in DNA concentration can result in different quantities of amplification product. This variation in amplification did not however affect the overall results.

The amplification efficiency also varied between the different STS markers. The efficiency of a PCR reaction not only depends on the quantity and quality of the DNA, but also on a number of different parameters, namely the primer sequences, the buffer conditions and the cycling conditions (Saiki *et al*., 1988). In the study presented here the objective was to determine whether amplification occurred in the AMELY-negative samples, thus demonstrating the presence of the specific locus in the sample and the lower amplification efficiency did therefore not interfere with the result.

Background smears were observed for many of the amplification reactions, e.g. Figure 4.8. These smears generally occur during the amplification process due to the residual non-amplified DNA of various sizes that are present in the reaction (Sambrook and Russel, 2001). In certain of the agarose gels, e.g. Figure 4.8, 4.9, 4.19 and 4.20, shadow fragments were observed as indicated by the white asterisk in Figure 4.8. These are not true fragments, but represent enzyme slippage products. Loading a lesser amount of the PCR product in the well can obliterate these shadow fragments. Alternatively, if a longer extension time is added during the final PCR step, the fragments would be full length. Fragment smears that appear above the fragment of interest is due to degradation, which will also be resolved if less sample is loaded per well, since the agarose detection system is not sufficiently sensitive to detect low concentrations of background fragments that are present. Primer-dimers were also observed for almost all the PCR reactions. Primer-dimers are formed if the conditions of the PCR reaction are not optimal (Saiki *et al*.,

1988). The formation of primer-dimers can be minimised by careful optimisation of the PCR reaction. In this study, however, the formation of primer-dimers did not affect the success of amplification and further optimisation was deemed unnecessary.

Agarose gel electrophoresis was used as a method to identify the size of the PCR product. This was achieved by comparing the mobility of the PCR product to that of a DNA molecular weight marker. Although confirmation of successful amplification was not affected, artefacts were observed in the agarose gels. Distortion of amplified fragments was observed for certain gels, e.g. Figure 4.11 and 4.13. This may have occurred due to residual agarose present in the wells. It is therefore important to remove residual agarose from the wells prior to loading of the samples. For Figure 4.18, the lanes of the amplified fragments appear slanted, as indicated by the white dashed line. This could be the result of the comb not being placed level in the casting tray. In order to prevent slanting, the alignment of the comb and the level of the platform should be verified prior to each electrophoresis run.

An ethidium bromide migration front was detected in some agarose gels, as in Figure 4.19. During electrophoresis, ethidium bromide migrates towards the cathode (Luedtke *et al.*, 2005) because of its positive charge. This can be minimised by either staining the agarose gel after electrophoresis or decreasing the ethidium bromide concentration in the agarose gels. If the gel is stained afterwards, a large volume of buffer containing ethidium bromide has to be handled. As ethidium bromide is a hazardous substance, smaller volumes are safer to handle. The receding ethidium bromide migration front did not, however, affect the results.

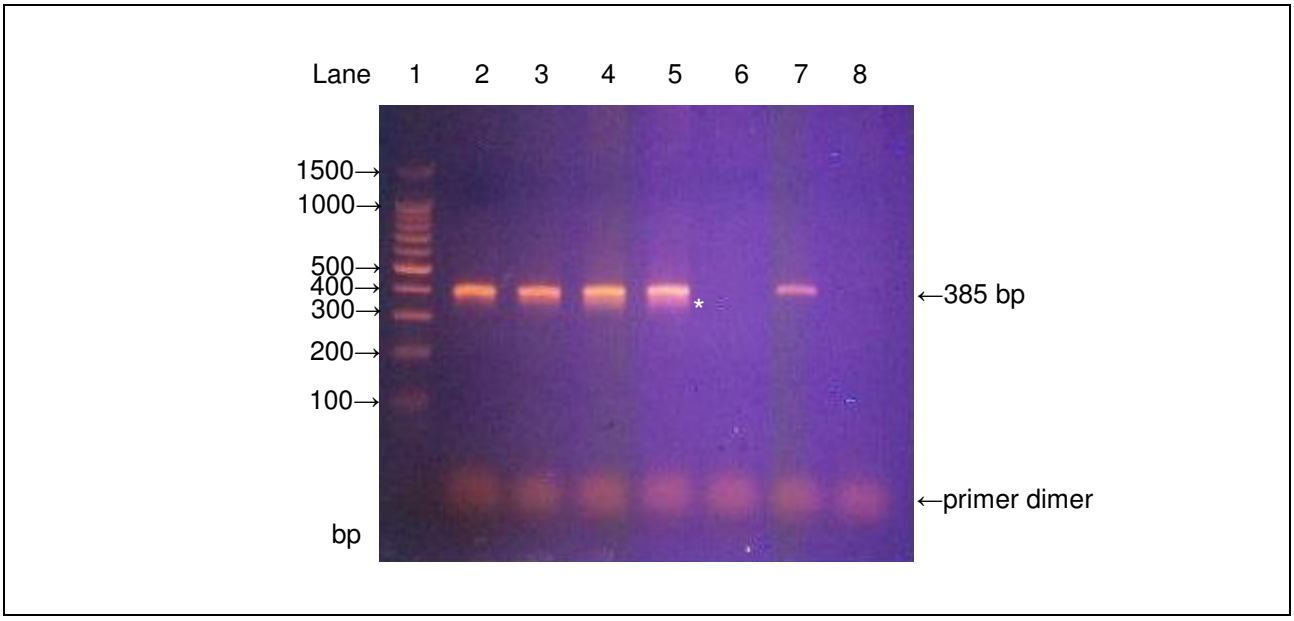
Distortion of the molecular weight marker was observed in some agarose gels, e.g. Figure 4.10, 4.11 and others. This may have occurred due to residual agarose being present in the wells or could be the result of the comb not being aligned or placed level in the casting tray. In order to prevent these artefacts, the correct casting of the agarose gels should be verified prior to each electrophoresis run.

4.6.1 Y-STS sY1240

For the sY1240 locus, the PCR product expected is 385 bp. The expected PCR product was detected in the AMELY-negative males and the positive male control. No PCR

product was detected in the female control. Figure 4.8 illustrates the amplification products of locus sY1240. The asterisk is discussed in Section 4.6, page 64.

Figure 4.8: Agarose gel electrophoresis of the sY1240 locus

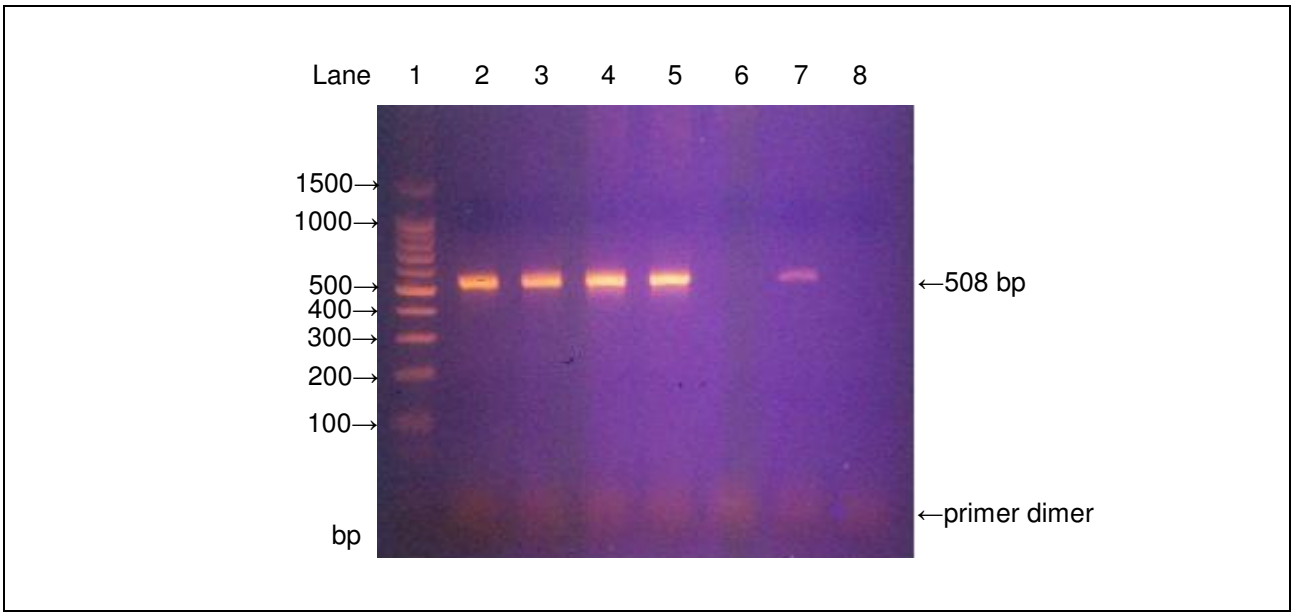


2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control. * indicates an artefact, discussed in Section 4.6.

4.6.2 Y-STS sY1241

The PCR product expected after amplification with primers for the sY1241 locus is 508 bp. AMELY-negative males and the positive male control demonstrated the expected product. A PCR product was absent in the female control. Figure 4.9 illustrates the amplification products.

Figure 4.9: Agarose gel electrophoresis of the sY1241 locus

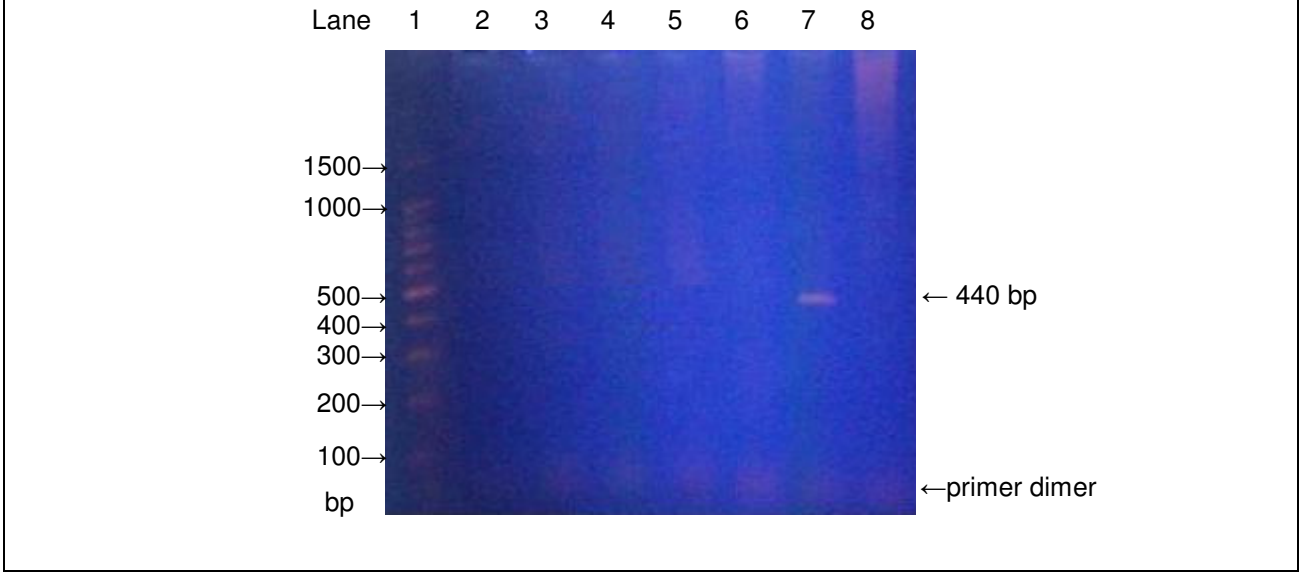


2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control.

4.6.3 Y-STS sY1242

Amplification of the sY1242 locus should yield a PCR product of 440 bp. A PCR product of 440 bp was detected in the positive male control. The AMELY-negative males and the female control failed to demonstrate a PCR product. Figure 4.10 illustrates the resultant PCR products.

Figure 4.10: Agarose gel electrophoresis of the sY1242 locus

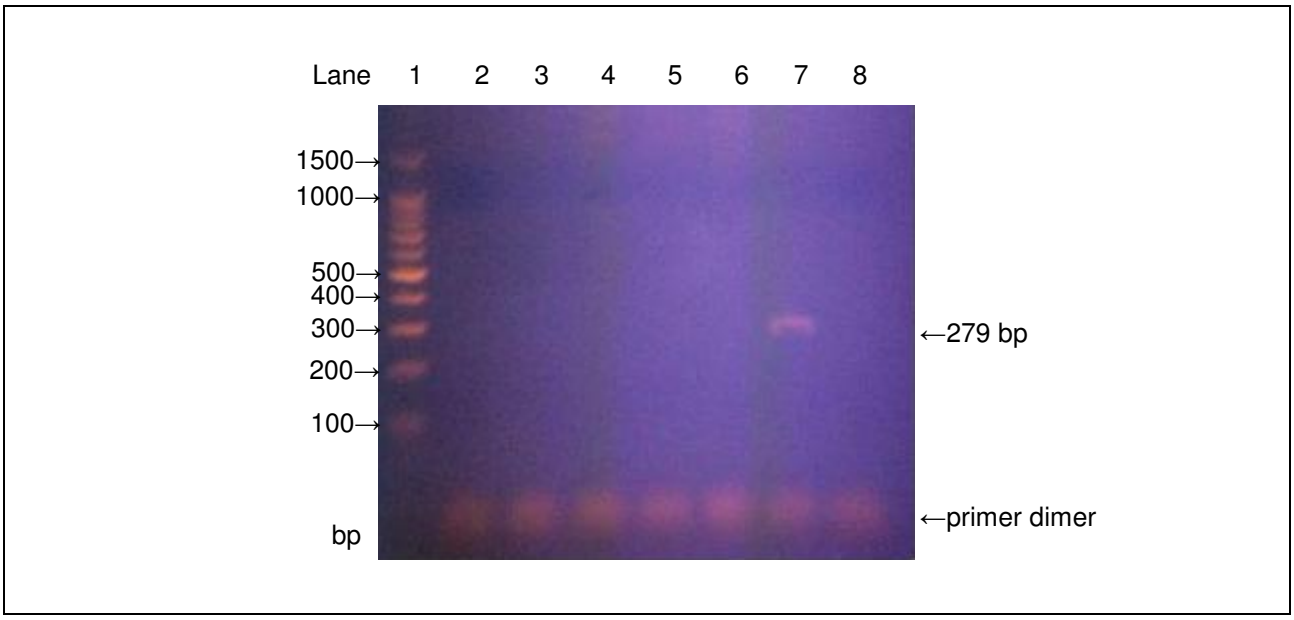


2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control.

4.6.4 Y-STS sY605

A PCR product of 279 bp is expected after amplification with the relevant primers. The PCR product was detected in the positive male control, but not in the AMELY-negative males and the female control. The amplification products are demonstrated in Figure 4.11.

Figure 4.11: Agarose gel electrophoresis of the sY605 locus

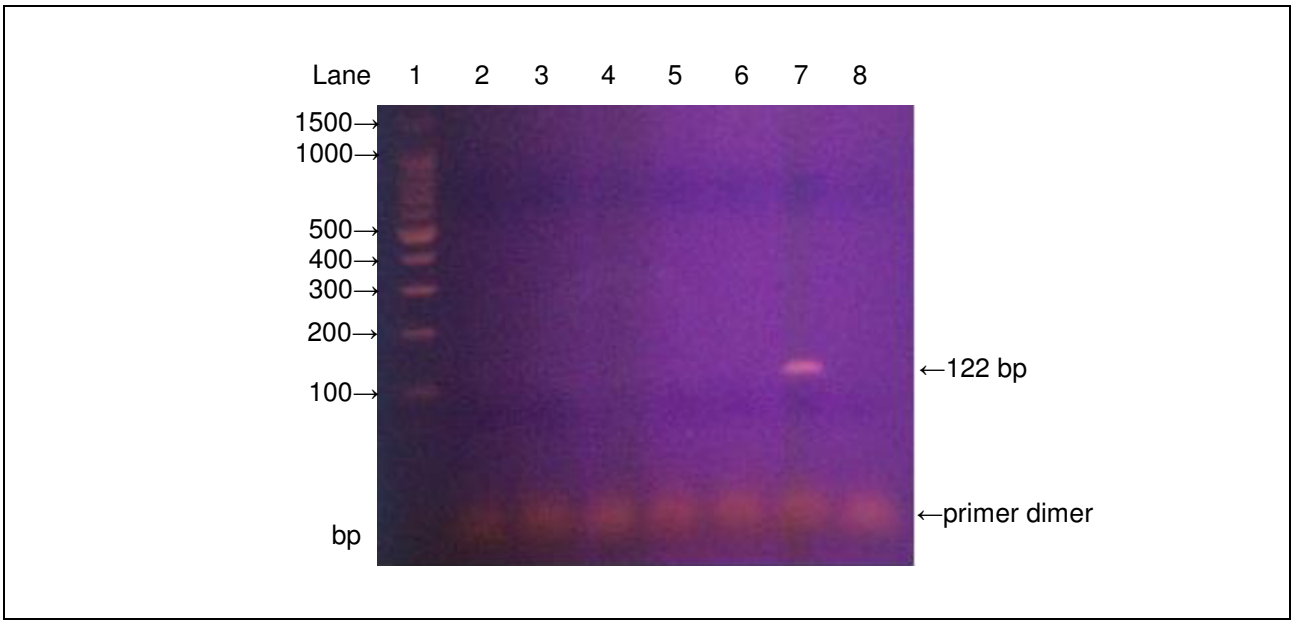


2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control.

4.6.5 Y-STS sY71

Figure 4.12 illustrates the expected PCR product of 122 bp after amplification with primers at the sY71 locus. The PCR product was only detected in the positive male control and no PCR product was detected in the AMELY-negative males or the female control.

Figure 4.12: Agarose gel electrophoresis of the sY71 locus

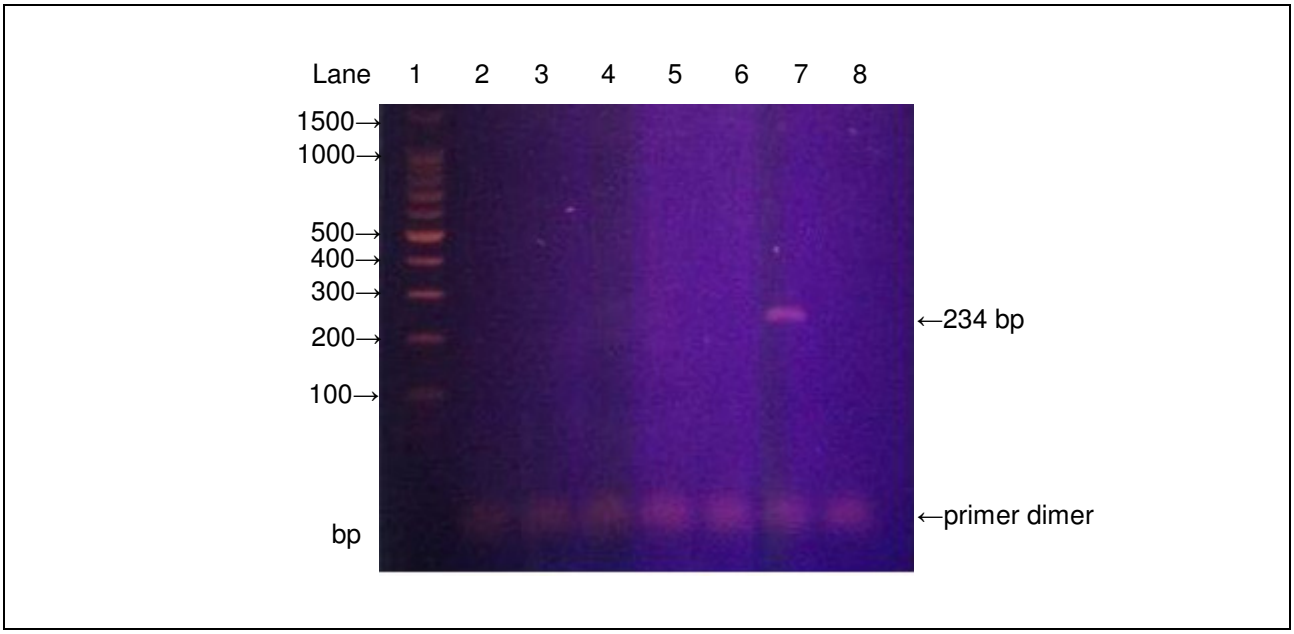


2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control.

4.6.6 Y-STS sY69

The amplification product for the sY69 locus was determined by agarose gel electrophoresis. The expected PCR product of 234 bp was detected in the positive male control. A PCR product was not observed in the AMELY-negative males and the female control.

Figure 4.13: Agarose gel electrophoresis of the sY69 locus

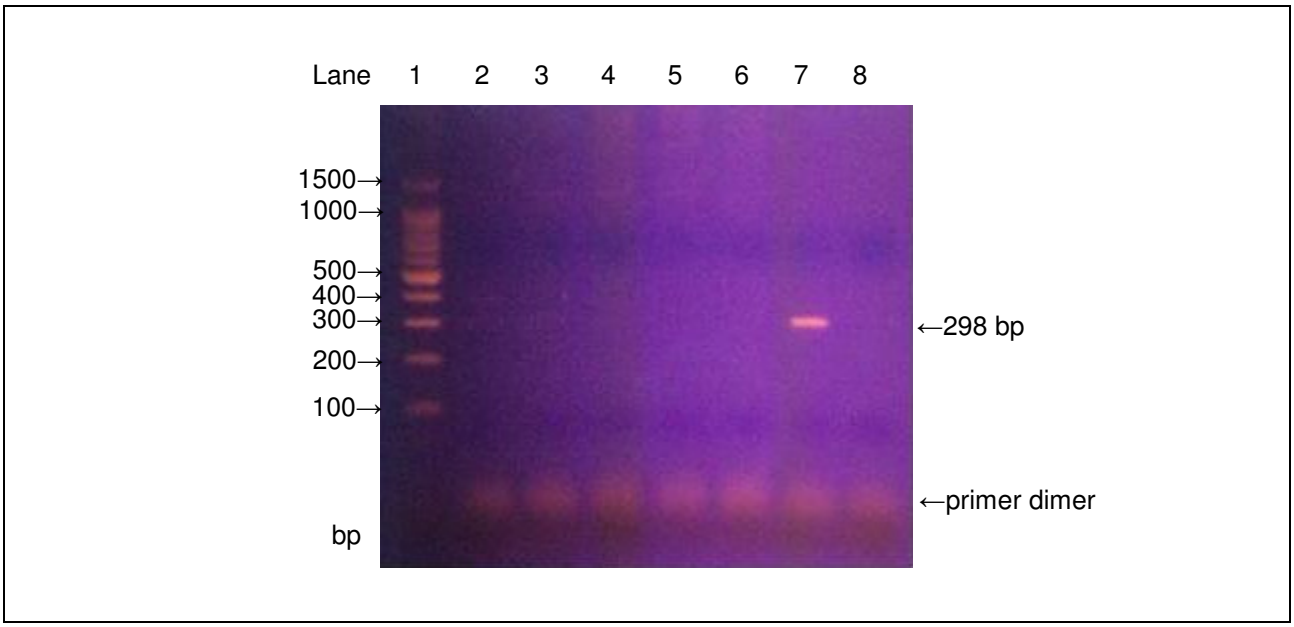


2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control.

4.6.7 Y-STS sY1219

The expected PCR product of 298 bp for locus sY1219 is illustrated by the positive male control in Figure 4.14. No PCR product was detected in the AMELY-negative males and the female control.

Figure 4.14: Agarose gel electrophoresis of the sY1219 locus

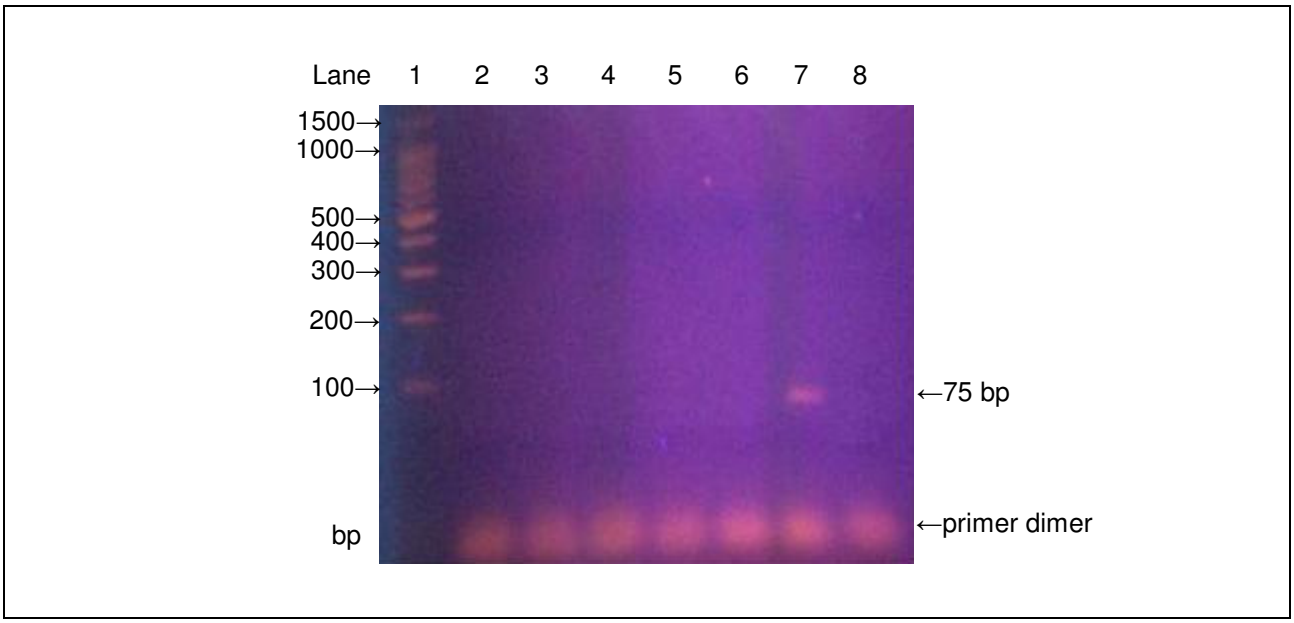


2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control.

4.6.8 Y-STS sY2216

Locus sY2216 PCR amplification yielded the expected product of 75 bp. This PCR product was detected in the positive male control. The AMELY-negative males and the female control did not demonstrate any amplification product.

Figure 4.15: Agarose gel electrophoresis of the sY2216 locus

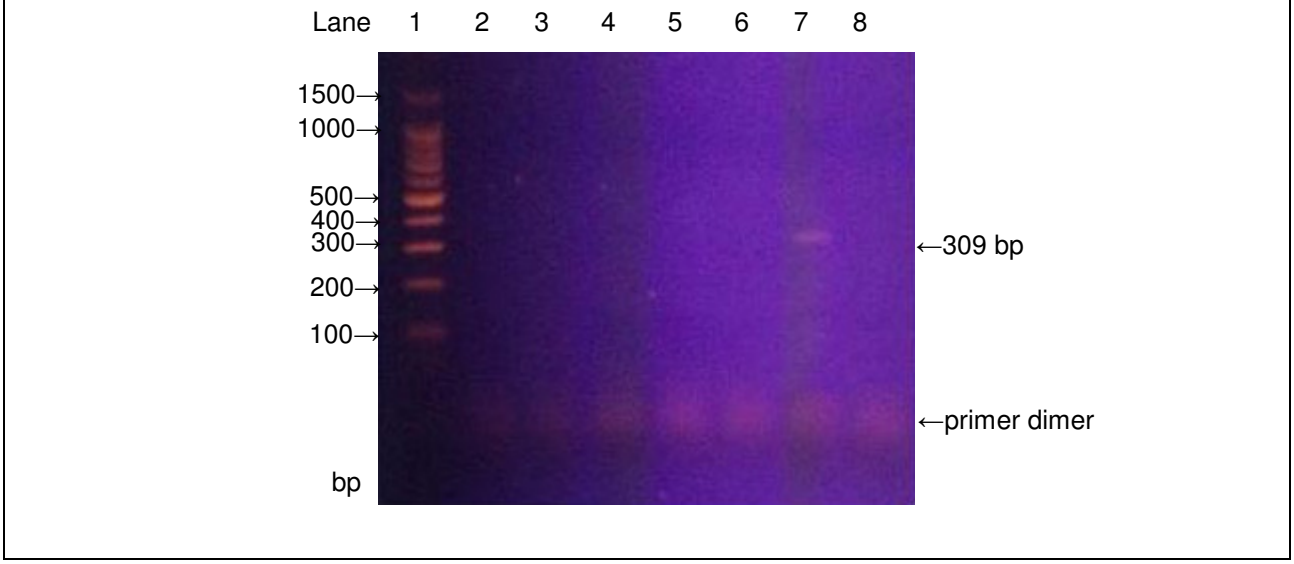


3% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control.

4.6.9 Y-STS sY65

The positive male control presented with the expected PCR product of 309 bp. The fact that the expected PCR product was not detected in the AMELY-negative males and the female control, reveal that locus sY65 is absent.

Figure 4.16: Agarose gel electrophoresis of the sY65 locus

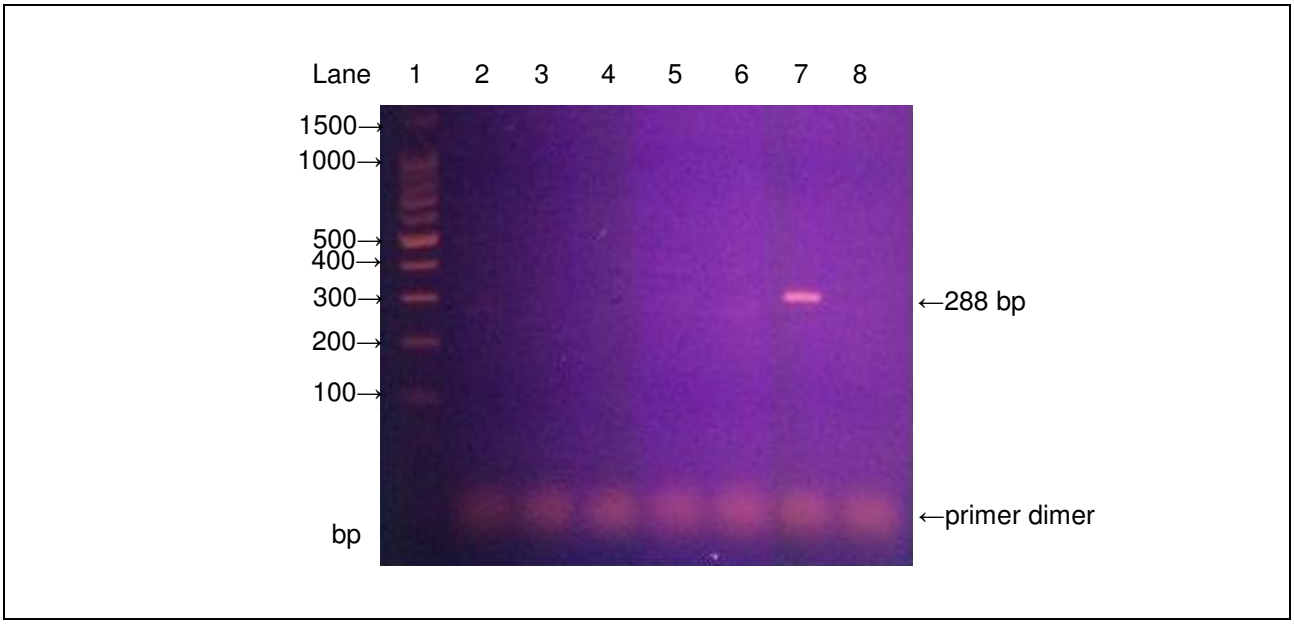


2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control.

4.6.10 Y-STS sY57

The presence of locus sY57 in an individual is demonstrated by the presence of an expected PCR product of 288 bp. Figure 4.17 illustrates the presence of sY57 in the positive male control and the absence of sY57 in the AMELY-negative males and the female control.

Figure 4.17: Agarose gel electrophoresis of the sY57 locus

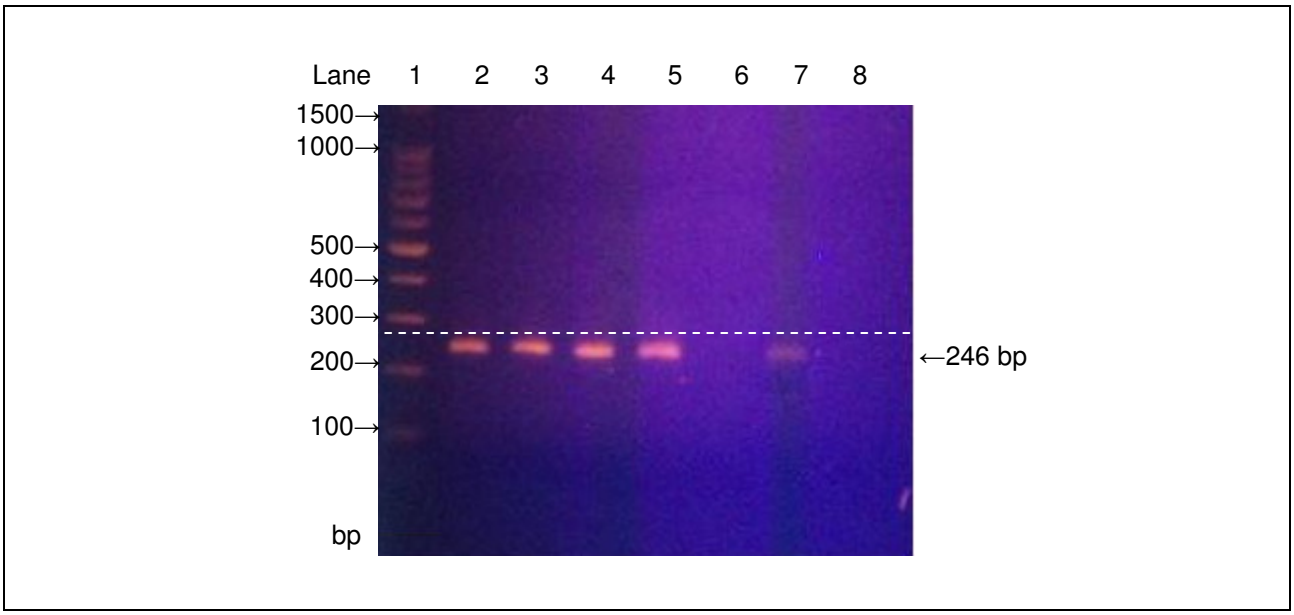


2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control.

4.6.11 Y-STS sY1079

The amplification product of locus sY1079 is illustrated in Figure 4.18 by a fragment of 246 bp. This fragment was detected in the positive male control and the AMELY-negative males but absent in the control female. The dashed white line is discussed in Section 4.6, page 65.

Figure 4.18: Agarose gel electrophoresis of the sY1079 locus

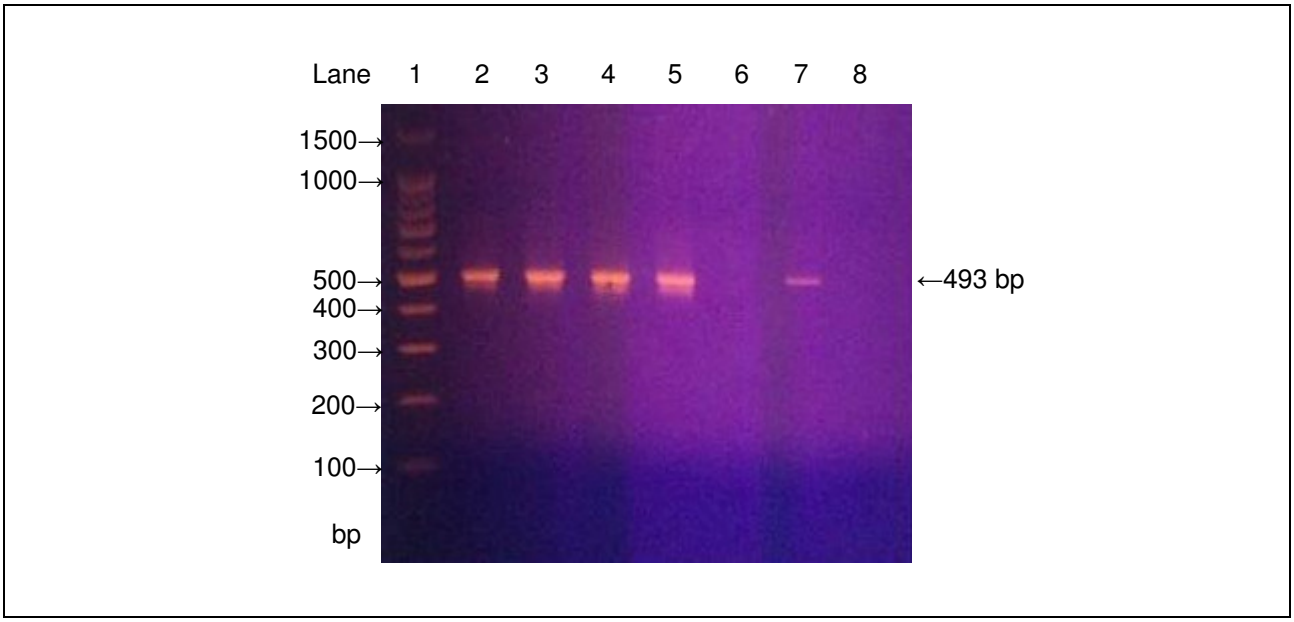


2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control. The dashed white line indicates an artefact, discussed in Section 4.6.

4.6.12 Y-STS sY1250

For locus sY1250, a PCR product of 493 bp is expected if present. No PCR product was detected in the female control, but the expected fragment was detected in the positive male control and the AMELY-negative males.

Figure 4.19: Agarose gel electrophoresis of the sY1250 locus

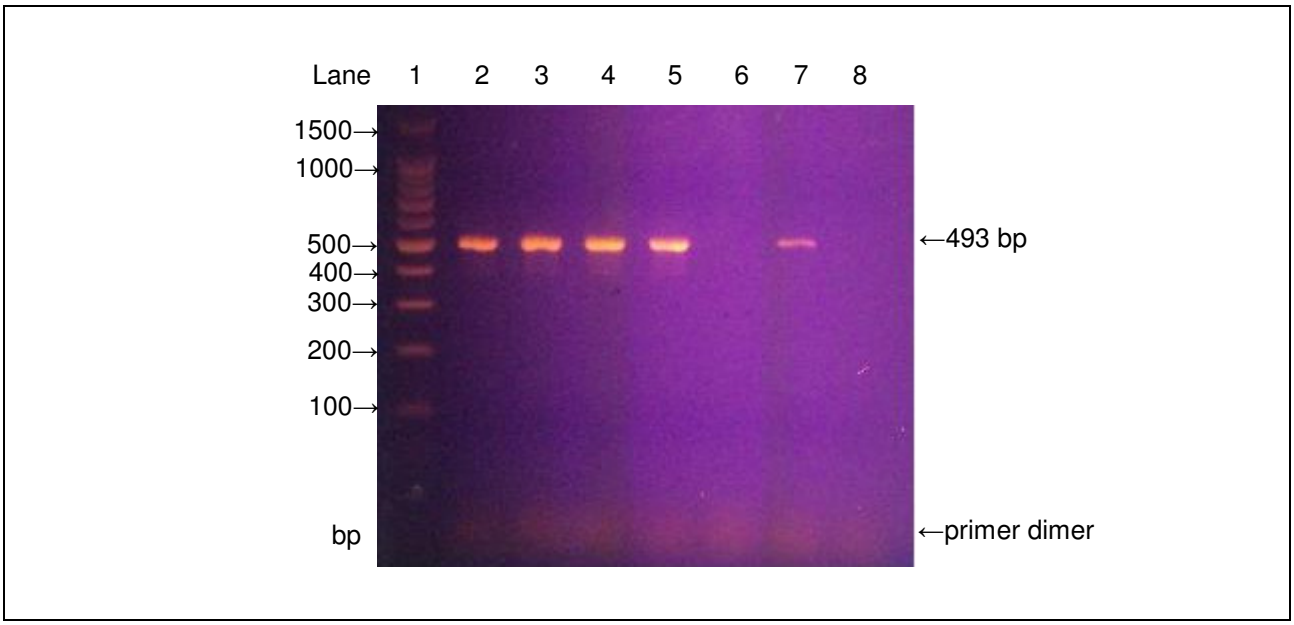


2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control.

4.6.13 Y-STS sY1243

Figure 4.20 illustrates the presence of locus sY1243 in the AMELY-negative males and the positive male control. A fragment of 493 bp is expected if the locus is present. The locus is absent in the female control, as no PCR product was detected.

Figure 4.20: Agarose gel electrophoresis of the sY1243 locus



2% agarose gel electrophoresed at 100 V for 60 minutes in 0.5X TBE buffer. Lane 1 corresponds to the 100 bp molecular weight marker, lanes 2-5 to the AMELY-negative males, lane 6 to the female control, lane 7 to the positive male control and lane 8 the blank control.

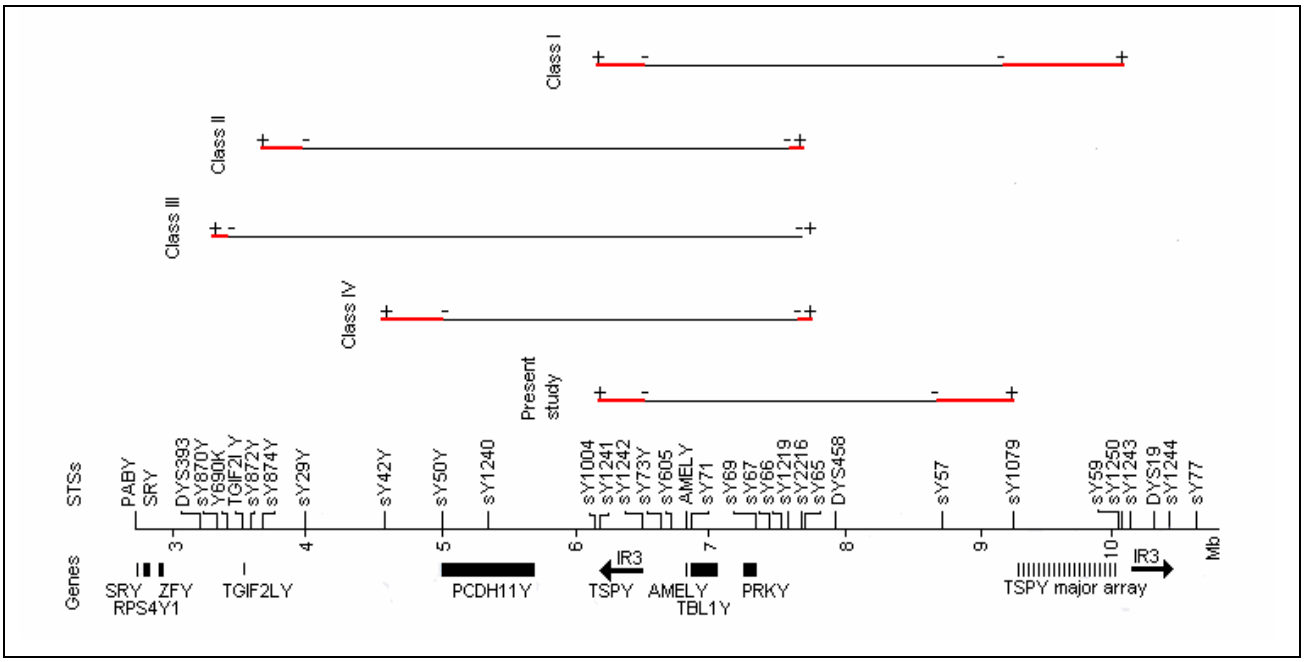
4.7 SUMMARY OF DELETION MAPPING

Results from scoring the 13 sites surrounding the AMELY locus on chromosome Yp, indicate that the same deletion size for all four black South African AMELY-negative males is present. Results obtained with the upstream markers indicated that all four samples contained the sequence at the sY1241 locus but not at the sY1242 locus. This indicates that the breakpoint for both AMELY deletions occurs between 6.1 Mb and 6.4 Mb upstream of the AMELY locus. Downstream of AMELY, the marker sY1079 gave the appropriate size in all four samples. Testing of locus sY57 (DYS257) revealed that it was absent in all samples. The latter finding restricts the proximal breakpoint to between 8.6 Mb and 9.9 Mb downstream from the AMELY locus.

Several studies have demonstrated that the failure of the amelogenin sexing test is due to a large deletion surrounding the amelogenin gene copy on the Y-chromosome (Santos

et al., 1998; Roffey *et al*., 2000; Chang *et al*., 2003; Thangaraj *et al*., 2002; Lattanzi *et al*., 2005; Mitchell *et al*., 2006; Chang *et al*., 2007; Cadenas *et al*., 2007; Jobling *et al*., 2007). Results obtained by typing the Y-specific loci surrounding the AMELY locus in the AMELY-negative males indicate that there are different deletion sizes. The deletion sizes have been detailed in Chapter 2. The different sets of markers used in the various studies however, make direct comparisons difficult. In brief, the deletion reported by Santos *et al*. (1998) included markers DYS7A and B (50f2A and B), 92R7 and LLY22g. The deletion described by Thangaraj *et al*. (2002) also included markers DYS7A and B. Lattanzi *et al*. (2005) characterised the deletion between markers Y7567 and sY1242. Chang *et al*. included MSY1 (2003) and DYS458 (2007) in the deletion they determined. Cadenas *et al*. (2007) reported a deletion from the sY1242 locus to the sY2180 locus. These markers are all indicated in Figure 2.6. In order to compare the deletion sizes, Jobling *et al*. (2007) analysed 45 AMELY-negative males of which 32 individuals were previously reported by Santos *et al*. (1998) in 2 individuals, Henke *et al*. (2001) in 1 individual, Bosch *et al*. (2003) in 1 individual, Lattanzi *et al*. (2005) in 2 individuals, Parkin *et al*. (2007) in 9 individuals, Mitchell *et al*. (2006) in 3 individuals and Chang *et al*. (2007) in 14 individuals. Five apparent deletion classes were distinguished. Figure 4.21 presents the deletion mapping data obtained for the black South African AMELY-negative males and a summary of the five deletion classes reported by Jobling *et al*. (2007).

Figure 4.21: Deletions defined by Y-chromosome markers on AMELY-negative males



Adapted from Jobling *et al*. (2007). + indicates positive STS amplification and – indicates no STS amplification. The solid black line indicates the area of no amplification and therefore the size of the deletion. The solid red line indicates the area of uncertainty within the breakpoint intervals.

The Indian and English individuals with Y-STR haplotypes closely related to that of the two black South Africans (as discussed in Section 4.5), both belong to the Class I deletion type described by Jobling *et al*. (2007). However, from the results obtained, the black South Africans do not have the same deletion that is associated with the Indian and English individuals. The deletion present in the black South Africans though, seems to be related to the Class I deletion (Jobling *et al*., 2007) as it has the same distal breakpoint. This breakpoint is in the distal IR3 element (Jobling *et al*., 2007). As discussed in Chapter 2, the distal IR3 element has a copy of the TSPY repeat in the same direction as the major TSPY array. Proximally the breakpoint in the black South Africans differs to that of Class I, as sY1079 is present in the South African samples, but absent in the Class I samples. STS sY1079 lies just distal to the TSPY major array. This could mean that the deletion was caused by a recombination between the TSPY copy within the minor array and the most distal copy of TSPY within the major array (Jobling *et al*., 2007). It is therefore likely that the Class I deletion and the deletions in the black South Africans are probably not related by descent, but that they occurred separately due to the increased susceptibility to recombination brought about the TSPY array and TSPY repeat IR3. Intra-chromosomal deletions can be caused by homologous recombination between repetitive sequence blocks of the same direction (Freeman *et al*., 2006; Mitchell *et al*., 2006; Jobling *et al*., 2008). AZFc deletions on chromosome Yq are most likely the result of such a mechanism (Kuroda-Kawaguchi *et al*., 2001). It is therefore likely that in a similar mechanism the deletions on Yp are the result of homologous recombination events.

The sons of the AMELY-negative males were also tested for their AMELY status, and were determined to exhibit the same deletion. It is important to note that although the two AMELY-negative black South African males cannot be distinguished from each other based on the deletion mapping results, the deletion is not necessarily identical. In fact, the Y-STR haplotype data indicates that individual 1007 and 1294 are very different. The fact that the same deletion was observed in two apparently unrelated black individuals from the South Africa population corroborate the hypothesis of an increased susceptibility to recombination in that area of the Y-chromosome.

This study has demonstrated that the failure of the amelogenin sexing test in the black South African males is due to a large deletion surrounding the amelogenin gene copy on the Y-chromosome. Results obtained by typing the Y-specific loci surrounding the AMELY locus indicate that the two black South Africans share identical deletion breakpoints, different to those previously described in Section 2.6. This is the first study to report and

illustrate the deletion on the Y-chromosome of AMELY-negative males in a black population.

CHAPTER FIVE

Conclusions

In this study, the absence of a Y-chromosome amelogenin-specific product in four black South African males was demonstrated to be due to a deletion involving the amelogenin locus. This is the first report of an amelogenin deletion in a black South African population. After an extensive and comprehensive literature survey, encompassing at least the past 2-3 decades of literature available, no publication has been determined to report an AMELY deletion in an African population. An exhaustive search including google.com, pubmed.com, sciencedirect.com, google patents and patentlens.net was undertaken. A search of the Y-chromosome website yhrd.com yielded a single African American to the list of AMELY deletions, which was added to this database in the period September to November 2009. No publication describing this individual was however reported.

5.1 AMELOGENIN Y-NEGATIVE FREQUENCIES

The data presented in Chapter 4, reveals that the frequency of AMELY-negative males is low (0.065) in the black South African population, which is similar to European derived groups (see Table 4.1). Incidentally, no AMELY-negative males were observed in a group of 2,214 Caucasian South Africans (data not presented) that was investigated in a similar manner to the black South African individuals included in this investigation.

Previous studies have reported high frequencies of AMELY-negative males in South Asians (Thangaraj *et al*., 2002; Chang *et al*., 2003; Cadenas *et al*., 2007; Kashyap *et al*., 2006; Chang *et al*., 2007; Jobling *et al*., 2007). The black South African population is different from previous studies in South Asians, which presented with high frequencies of AMELY-negative males. The data presented by Thangaraj *et al*. (2002), Chang *et al*. (2003), Cadenas *et al*. (2007), Kashyap *et al*. (2006) and Jobling *et al*. (2007) is in agreement with Chang *et al*. (2007), in that one lineage associated with the AMELY-negative males seems to have spread from the Indian subcontinent. However, Chang *et al*. (2007) also postulated that similar mutations have appeared sporadically elsewhere. The results of this study are in agreement with the latter postulate by Chang

et al. (2007) as the mutations observed in this study indicate independent origins rather than inheritance by descent. These results are surprising if the Out-of-Africa hypothesis is considered. With the frequency of sporadic mutations being the same for different populations, it is expected that the oldest populations will have the greatest variation. Therefore, if sporadic, one would expect the frequency of the deletion to be higher in Africa. The fact that this was not observed could be due to the fact that the black South African population in this study is biased towards males that are able to father children. Thus the deletion may have a higher frequency in the general population, but these deletion forms may be slightly deleterious in terms of fertility and therefore males harbouring them would not be part of this study population. Another reason for the higher frequency of AMELY deletions observed in South Asians could be that the deletion had occurred in separate lineages and the frequency was subsequently increased by an expansion of allele frequency through genetic drift or demographic factors (Yong *et al*., 2007), thereby resulting in the high frequency of the deletion in the South Asian population.

5.2 POSSIBLE ORIGINS OF AMELOGENIN-Y NEGATIVE

The haploid Y-chromosome does not recombine at meiosis over most of its length and therefore contains haplotypes in these regions that are changed only by mutation (Blanco *et al*., 2000). Jobling *et al*. in 1998 reported that XX males with a PRKY/PRKX translocation had a specific Y-chromosome haplotype that was predisposed to this translocation. This observation raised the question during the current study of whether the AMELY deletion might be associated with a certain Y-chromosome haplotype. From the literature it appears that several authors (Thangaraj *et al*., 2002; Chang *et al*., 2003; Lattanzi *et al*., 2005; Mitchell *et al*., 2006; Kashyap *et al*., 2006; Cadenas *et al*., 2007; Jobling *et al*., 2007) also questioned whether the AMELY deletion was associated with the same haplotype. However, earlier studies as outlined in Section 2.5, and data presented in this study in Section 4.5, do not support the hypothesis of an association between a specific haplotype and any of the AMELY deletions. It should be considered however, that only two deletion events are described in this study and this hypothesis might well hold up if more samples from an African population are studied.

It has been established that deletions in the Yq-chromosome are associated with azoospermia (Yen, 2001). It seems however that deletions surrounding the amelogenin

locus have no apparent effect on male fertility (Henke *et al*., 2001; Thangaraj et al., 2002; Lattanzi et al., 2005; Kumagai *et al*., 2008). The present study supports this observation. The two AMELY-negative males in the present study have each fathered a son naturally (Section 4.2). Given that for the AMELY-negative samples included in this discussion (one sample from Henke *et al*. (2001), one sample from Thangaraj *et al*. (2002) and both samples from the current study) were from paternity testing groups, this may have contributed to a bias in the chosen sample set itself. The sample from Lattanzi *et al*. (2005) was from a prenatal test and the one sample from Kumagai *et al*. (2008) was from a sibling-ship test. The biased selection of these samples could imply reduced fertility and possibly incomplete infertility. However, the one AMELY-negative sample described by Lattanzi *et al*. (2005) was from a group of 293 infertile males and one of the AMELY-negative samples described by Thangaraj *et al*. (2002) was from a group of 60 infertile males. Both males from the infertile groups also presented with deletions of Yq (Thangaraj *et al*., 2002; Lattanzi *et al*., 2005) and infertility was attributed to these Yq deletions. Therefore the AMELY deletion may be an indicator of infertility but it is not definitive. As discussed in Section 5.1, an association of AMELY deletions and infertility may account for the low incidence of AMELY-negative males in the black South African sample investigated.

Another reason for the lack of evidence indicating reduced fertility in AMELY-negative males is that the main source of AMELY-negative samples is forensic databases. Few sample sets to date are from paternity testing laboratories or fertility clinics. Fertility and other phenotypic features of AMELY-negative males are not recorded in the forensic datasets and therefore the clinical relevance of the deletion is not clear. As such, reduced fertility in these individuals cannot be completely ruled out.

An area to investigate the possible effects of AMELY deletions is indicated by the fact that in both the present study and previous reports, the TBL1Y and PRKY genes are deleted in AMELY-negative males. The functions of these two genes are unknown (see Section 2.6.4) and the AMELY-negative males appear phenotypically unaffected. Most of the AMELY-negative individuals were screened via STR analysis only, and thus detailed clinical examination of the AMELY-negative males may be worthwhile to determine if the deletion of these genes indeed manifests in the phenotype. Although the deletion of these genes appears to have no phenotypic effect, which may be due to a compensatory effect from the X-chromosome homologues, samples are mostly from forensic databases which have no comprehensive phenotype data, thus warranting a detailed clinical examination of

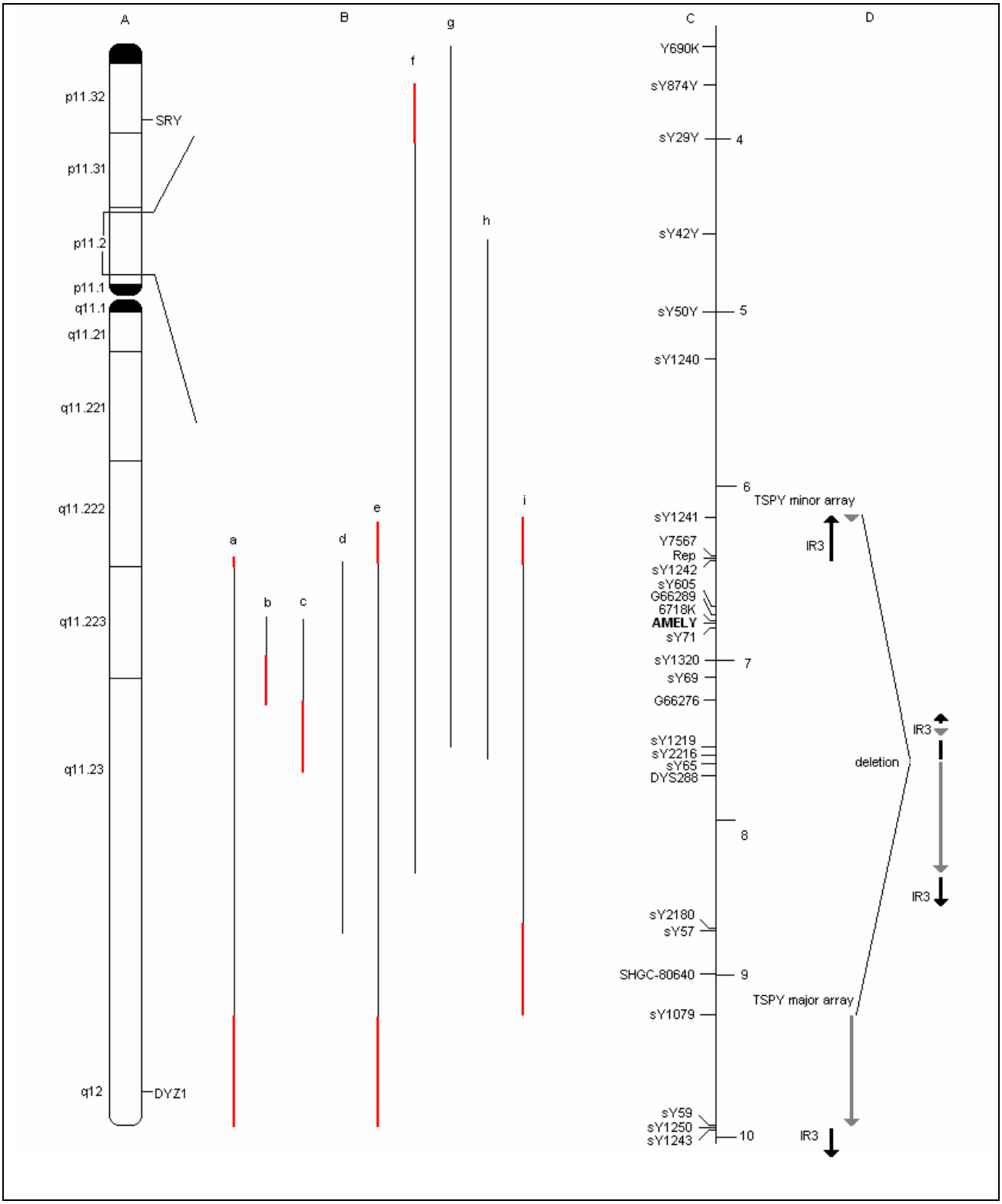
the phenotype of AMELY-negative males. It is however, possible that these results support earlier evidence that deletion events, even if they result in the loss of certain genes, do not necessarily result in the presentation of genomic disorders (Buckland *et al*., 2003). However, investigation into the phenotypic effects of the AMELY deletions is required to better understand their relative absence in certain populations e.g. Africans and their increased frequency in others e.g. South Asians.

5.3 CHARACTERISATION OF THE DELETIONS

Studies targeting the different regions of the amelogenin gene (Lattanzi *et al*., 2005; Mitchell *et al*., 2006; Cadenas *et al*., 2007; Jobling *et al*., 2007) have demonstrated that the failure of the amelogenin test is due to a large deletion that surrounds the amelogenin gene on the Y-chromosome. The deletions mapped in the different studies are schematically represented in Figure 5.1.

Nonallelic homologous recombination seems the most likely mechanism for causing the deletion in the black South African AMELY-negative males determined in this study, similar to the class I deletions illustrated by Jobling *et al*. (2007). As discussed in Section 4.7 and illustrated in Figure 5.1, the deletions presented in this study have the same distal breakpoint as the Class I deletions. The proximal breakpoint however differs to that of Class I, lying just distal to the TSPY major array. This could mean that a recombination event could have occurred between the TSPY minor array and the most distal copy of TSPY within the major array. Figure 5.1 is a schematic illustration of TSPY mediated deletion in the black South African AMELY-negative males.

Figure 5.1: Schematic representation of the Y-chromosome illustrating the size of the AMELY deletion in the black South African population in comparison to the worldwide population



A = Ideogram of the Y-chromosome. The boxed region in Yp11.2 is amplified for further detail. B = Breakpoints of the different deletions are indicated. Black solid lines indicate regions observed to be absent and red solid lines indicate areas of uncertainty within the breakpoint intervals, a = Lattanzi *et al.* (2005); b,c = Mitchell *et al.* (2006); d = Cadenas *et al.* (2007); e-h = Jobling *et al.* (2007); i = present study. C = The relative positions of STS markers are indicated. D = The relative position of inverted repeat IR3, the TSPY major array and the TSPY minor array are illustrated with the probable deletion mechanism.

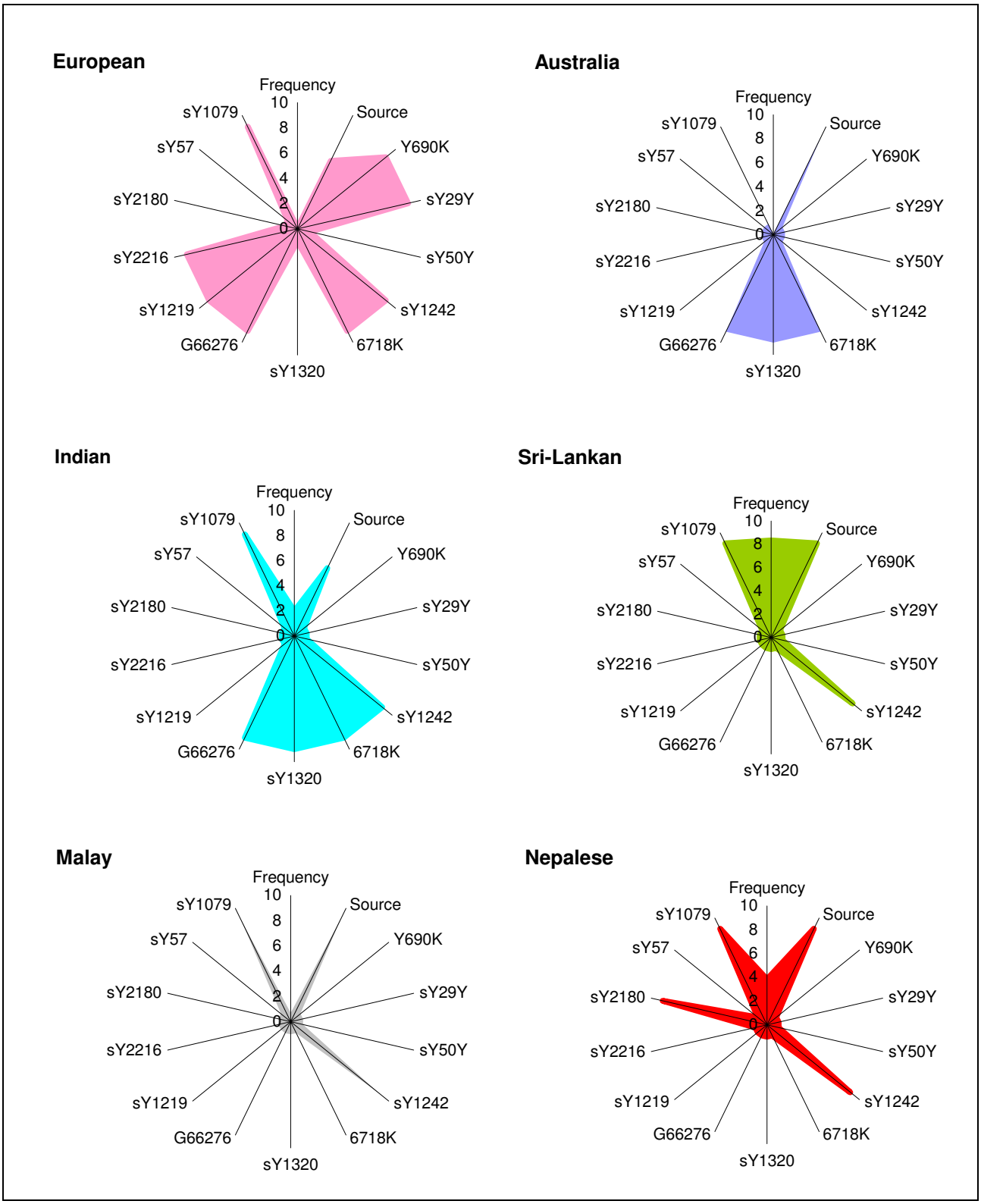
Other markers such as SRY (Santos *et al.*, 1998; Steinlechner *et al.*, 2002; Thangaraj *et al.*, 2002; Lattanzi *et al.*, 2005) and DYZ1 (Roffey *et al.*, 2000) have been suggested for

gender identification, as they are intact in AMELY-negative males. Whether either of these markers will be effective in determining gender is uncertain. Initially the amelogenin gene was hypothesised to be ideal for sexing samples, but it was determined to be inadequate after a large number of individuals have been screened (Santos *et al*., 1998; Roffey *et al*., 2000; Chang *et al*., 2003; Thangaraj *et al*., 2002; Lattanzi *et al*., 2005; Mitchell *et al*., 2006; Chang *et al*., 2007; Cadenas *et al*., 2007; Jobling *et al*., 2007; Takayama *et al*., 2009; present study). Many samples from different population groups will have to be screened for SRY or DYZ1 to determine their accuracy as gender markers. In fact, recently a combination of SRY primers and a commercial STR genotyping kit was evaluated in 97 individuals (Inturri *et al*., 2009) and the results were promising.

5.4 MODEL OF THE GENOTYPE / PHENOTYPE COMPARISON OF AMELOGENIN-NEGATIVE MALES

In order to illustrate the relationship between the sample source, e.g. forensic database, paternity testing or infertility clinic, referred to in this study, the frequency of the AMELY deletion and the deletion breakpoints, a graph was plotted for each population described. Data documented in Table 4.1 (page 53) and Figure 5.1 (page 81) was used to construct the graphs. These graphs are specified in Figure 5.2.

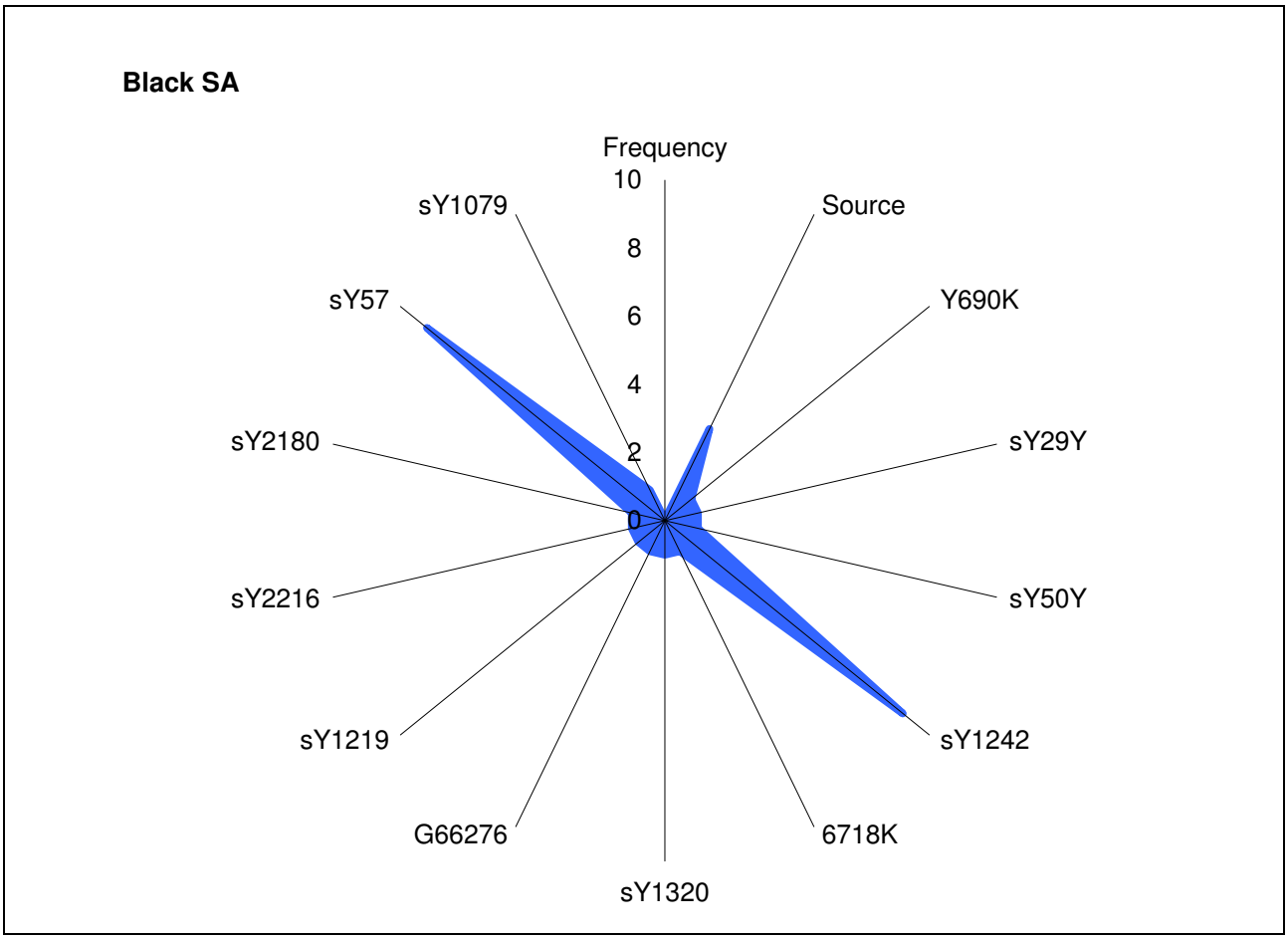
Figure 5.2: Model of the genotype/phenotype comparison of AMELY-negative males in different population groups



The population that is depicted by the graph is reflected at the top left of the graph. Axes: 1. Frequency: frequency of AMELY-negative males in the population. True values are represented for the frequency. 2. Source: three different sources are differentiated, namely paternity testing group or database group or paternity testing group + database group + infertility clinic group. The sources are represented by a value of 3 for the paternity testing group, a value of 6 for the paternity testing group + database group + infertility clinic group and a value of 9 for the database group, respectively. 3 – 14. Y-STS locus present at the breakpoint sites described for the population. The presence of a Y-STS locus at the breakpoint site is represented by a value of 9 and the absence of a Y-STS locus is represented by a value of 1. The scale is from 0 to 10 on all the axes.

The graphs indicate that the highest frequency of AMELY-negative males was observed in the Sri-Lankan population. With the exception of the Indian and Nepalese populations, the frequency of AMELY-negative males in the other populations was very low. For most of the populations, the samples were sourced from forensic databases. Samples from the Indian and European populations were obtained from various sources, i.e. forensic databases, paternity testing laboratories and infertility clinics. It is already apparent that the European population has the most variation in the breakpoint regions and also indicate shared breakpoints with most of the other populations. As would be expected from the literature, the South Asian populations share many of the breakpoints. Data generated by the present study was used to plot the graph in Figure 5.3, depicting the black South African cohort investigated in this study.

Figure 5.3: Model of the genotype/phenotype comparison of AMELY-negative males in the black South African population



The population that is depicted by the graph is reflected at the top left of the graph. Axes: 1. Frequency: frequency of AMELY-negative males in the population. True values are represented for the frequency. 2. Source: three different sources are differentiated, namely paternity testing group or database group or paternity testing group + database group + infertility clinic group. The sources are represented by a value of 3 for the paternity testing group, a value of 6 for the paternity testing group + database group + infertility clinic group and a value of 9 for the database group, respectively. 3 – 14. Y-STS locus present at the breakpoint sites described for the population. The presence of a Y-STS locus at the breakpoint site is represented by a value of 9 and the absence of a Y-STS locus is represented by a value of 1. The scale is from 0 to 10 on all the axes.

The main differences between the black South African population and the other populations can be seen from the graphs. The source was different for the black South African population. This was the only population sourced exclusively from a paternity testing database. Although the black South African population share one breakpoint with many of the other populations, a unique breakpoint was observed for this population. On the graph for the black South African population the presence of Y-STS sY57 is clearly indicated, a breakpoint site that is absent in all the other populations. The implications of the different breakpoint sites observed, are discussed in Section 4.7.

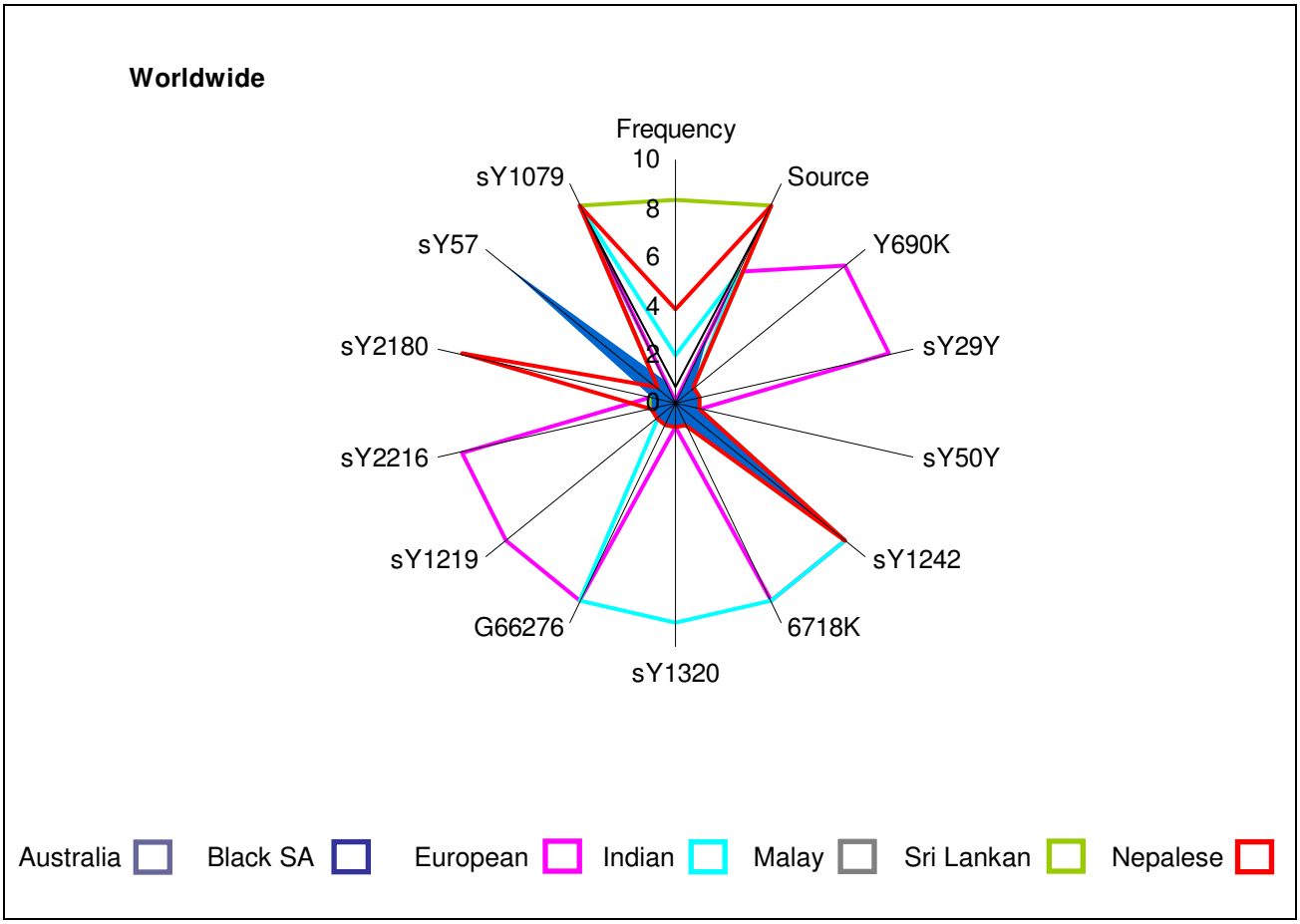
In order to appreciate the differences between the different populations at a glance, the graphs were superimposed. Figure 5.4 depicts this comparison between the black South African study population and the other populations. It is evident that the black South African cohort is unique with regard to the sY57 breakpoint.

It appears that the black South African population differs from the other populations in terms of the genetic composition of the chromosome Y deletion. As discussed in Sections 4.7 and 5.3, the deletion present in the black South African population has the same distal breakpoint as the Class I deletion described by Jobling *et al*. (2007). Proximally the breakpoint differs to that of Class I, as sY1079 is present in the South African samples, but absent in the Class I samples. The distal breakpoint is in the TSPY copy in the minor array and the proximal breakpoint of the South African samples is in the most distal copy of TSPY within the major array (Jobling *et al*., 2007). If the deletion was caused by recombination between the TSPY copy within the minor array and the most distal copy of TSPY within the major array (Jobling *et al*., 2007), as discussed in Section 4.7, this could indicate an increased susceptibility to recombination and a possible hotspot at these chromosomal areas. Vijayakumar *et al*. (2006) also suggested that the TSPY gene cluster located close to the centromere on Yp has the potential to undergo genomic rearrangement through homologous recombination. According to Vijayakumar *et al*. (2006), the genomic structure of the region possibly marks it as a hotspot for rearrangements, leading preferentially to deletions.

It would be interesting to observe whether a change in the frequency and or more similar sample sources would result in graphs that are more uniform. If the low incidence of AMELY deletions in the black South African population is due to a bias in the study population, a more varied sample source is likely to affect the frequency. The incidence of AMELY deletions is low in populations other than South Asian, with many populations

reporting only one or two deletions. As previously discussed, a greater variation was expected in the black South African population. At present the global distribution of AMELY deletions would suggest that in some populations the AMELY deletion arose independently and that in the South Asian population the frequency of the AMELY deletion increased due to inheritance by descent and subsequent expansion.

Figure 5.4: Comparative model of the genotype/phenotype composition of AMELY-negative males



The population that is depicted by the graph is reflected at the top left of the graph. Axes: 1. Frequency: frequency of AMELY-negative males in the population. True values are represented for the frequency. 2. Source: three different sources are differentiated, namely paternity testing group or database group or paternity testing group + database group + infertility clinic group. The sources are represented by a value of 3 for the paternity testing group, a value of 6 for the paternity testing group + database group + infertility clinic group and a value of 9 for the database group, respectively. 3 – 14. Y-STS locus present at the breakpoint sites described for the population. The presence of a Y-STS locus at the breakpoint site is represented by a value of 9 and the absence of a Y-STS locus is represented by a value of 1. The scale is from 0 to 10 on all the axes.

This model indicates how each interacting factor has an impact on the overall outlook and highlights the role of each component in generating data for each population. The relationship, if any, between clinical data and deletion size remains unanswered. This can only be addressed once a detailed clinical workup has been done on the AMELY-negative males.

5.5 FUTURE DEVELOPMENTS

In many publications on AMELY-negative males there is no information on whether these deletions are related by descent or arose from independent mutations. This information is necessary to consider the evolutionary history of the deletions. The data generated in this study highlights the different populations and deletion events observed for AMELY-negative males. An aspect that needs to be addressed in the future is to shed light on the possible origins of AMELY-negative males, via the determination of the Y-haplogroups of the two black South African AMELY-negative males using a panel of Y-SNP markers (The Y-chromosome consortium, 2002).

The global distribution of AMELY-negative males revealed an ethnic association with the Indian subcontinent and an apparent absence among a number of ethnic groups (Thangaraj *et al*., 2002; Chang *et al*., 2003; Kashyap *et al*., 2006; Cadenas *et al*., 2007; Jobling *et al*., 2007). This seemed true for the African population, as no AMELY-negative males have been described in a black population prior to this study. The present study has discussed the presence of four AMELY-negative males in a black population for the first time. Moreover, it confirmed that AMELY dropouts in black South Africans are due to a large deletion in the short arm of the Y-chromosome.

Wrong gender assignment has consequences in the medical, forensic, legal, paternity determination and social fields, as discussed in Section 2.7. This has serious implications for the black South African population and other African populations where this may be present. The question should be asked whether the incidence of AMELY deletions appear to be low in the African population due to lack of information or an actual low prevalence. Regardless, care should be taken in sex tests performed for detecting the presence of male DNA as this study confirms that these deletions do occur in the African population albeit at a very low level. The impact of the results generated in this study on the forensic practise of DNA testing is significant. Sex testing is often used as a major discriminating factor in criminal cases. Moreover, the South African Police Service (SAPS) is in the process of negotiating for the creation of a National Database of criminal offenders, similar to other countries. Wrong gender assignment will lead to a bias in this national Database that will be employed in the pursuit of justice. When also considering that the SAPS propose to destroy the DNA sample once the STR profile has been created, there will be no recourse available to revisit the subtle issues discovered in this study. In view of this,

the importance of this study cannot be ignored. In a country with limited resources and goals to discourage criminal offenders, the National Database should be beyond reproach.

This study is evidence that the molecular tools currently employed in sex testing, be it for application in the clinical, paternity or forensic arenas, has embedded or underlying intricacies that will only be exposed via detailed analysis of the molecular architecture of the chromosomal structure. Although, a similar detailed analytical approach may never become routine, this study highlights the importance of an investigative approach when performing routine analysis in a laboratory. Although laboratory science should be unbiased, this approach is required to consistently evaluate outlier or unexpected results. These results may have more implications than initially hypothesised. Unexpected results, as were determined in this study, are often the source of unravelling the molecular mechanisms of chromosomal plasticity.

In the medical field, gender assignment is employed for determining the success of sex-mismatched bone marrow transplants and prenatal diagnosis. In archaeology, sex typing of skeletal remains is applied for the reconstruction of historical events. In forensic science, the wrong gender assignment at a scene of a mass disaster can lead to mismatching of remains and in criminal investigations, determination of the gender of the offender is usually the initial information obtained. Socially, gender assignment has far reaching implications for participation in sporting events and any similar activity that is based on gender segregation. When considering the importance of gender determination and the existence of AMELY deletions in the black South African population, the continued use of this test should be questioned. For this reason the development of an infallible method of gender assignment is a major goal, due to the far-reaching consequences of misdiagnosis in gender testing.

CHAPTER SIX

References

6.1 GENERAL REFERENCES

Affara N.A., Ferguson-Smith M.A., Tolmie J., Kwok K., Mitchell M., Jamieson D., Cooke A. and Florentin L. Variable transfer of Y-specific sequences in XX males. *Nucleic Acids Res*, **14**, 5375-5387, 1986.

Applied Biosystems ABI PRISM® BigDye™ Terminator version 3.0 Ready Reaction Cycle Sequencing kit, protocol. Life Technologies, Carlsbad, CA, USA, 2001.

Applied Biosystems ABI PRISM® 310 Genetic Analyzer, User's Manual. Life Technologies, Carlsbad, CA, USA, 1998.

Applied Biosystems ABI PRISM® 310 Genetic Analyzer Genescan™ Software Reference Guide, User's Manual. Life Technologies, Carlsbad, CA, USA, 1997.

Applied Biosystems ABI PRISM® 310 Genetic Analyzer Genescan™ Analysis Software version 3.1, User's Manual. Life Technologies, Carlsbad, CA, USA, 1997.

Applied Biosystems ABI PRISM® Genotyper™ Software version 2.5, User's Manual. Life Technologies, Carlsbad, CA, USA, 1999.

Applied Biosystems AmpFℓSTR® SGM Plus™ PCR Amplification kit, User's Manual. Life Technologies, Carlsbad, CA, USA, 1999.

Applied Biosystems 3130/3130xl GeneMapper™ ID Software version 3.1 and 3.2, User's Manual. Life Technologies, Carlsbad, CA, USA, 2003.

Applied Biosystems 3130/3130xl Genetic Analyzers, Getting Started Guide. Life Technologies, Carlsbad, CA, USA, 2004.

Applied Biosystems 3130/3130xl Genetic Analyzers, Maintenance, Troubleshooting, and Reference guide. Life Technologies, Carlsbad, CA, USA, 2004.

Applied Biosystems 3130/3130xl Genetic Analyzers, Site Preparation Guide. Life Technologies, Carlsbad, CA, USA, 2004.

Applied Biosystems 3730/3730xl DNA Analyzers and Applied Biosystems 3130/3130xl Genetic Analyzers, AB Navigator Software Administrator Guide. Life Technologies, Carlsbad, CA, USA, 2004.

Ayub Q., Mohyuddin A., Qamar R., Mazhar K., Zergal T., Mehdi S.Q. and Tyler-Smith C. Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res*, **28**, e8, 2000.

Bachtrog D. and Charlesworth B. Towards a complete sequence of the Y chromosome. *Genome Biol*, **2**, 1016.1-1061.5, 2001.

Bailey D.M., Affara N.A. and Ferguson-Smith M.A. The X-Y homologous amelogenin maps to the short arms of both the X and Y chromosomes and is highly conserved in primates. *Genomics*, **14**, 203-205, 1992.

Barbaux S., Vilian E., Raoul O., Gilgenkrantz S., Jeandidier E., Chadenas D., Souleyreau N., Fellous M. and McElreavey K. Proximal deletions of the long arm of the Y chromosome suggest a critical region associated with a specific subset of characteristic Turner stigmata. *Hum Mol Genet*, **4**, 1565-1568, 1995.

Blanco P., Shlumukova M., Sargent C.A., Jobling M.A., Affara N. and Hurles M.E. Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J Med Genet*, **37**, 752-758, 2000.

Bosch E. and Jobling M.A. Duplications of the AZFa region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. *Hum Mol* Genet, **12**, 341-347, 2003.

Brinkman B. Is the amelogenin test valid? *Int J Legal Med,* **116**, 63, 2002.

Brown G.M., Furlong R.A., Sargent C.A., Erickson R.P., Longepied G., Mitchell M., Jones M.H., Hargreave T.B., Cooke H.J and Affara N.A. Characterisation of the coding sequence and fine mapping of the human DFFRY gene and comparative expression analysis and mapping to the Sxrb interval of the mouse Y chromosome of the Dffry gene. *Hum Mol Genet*, **7**, 97-107, 1998.

Budowle B., Koons B.W. and Errera J.D. Multiplex amplification and typing procedure for the loci D1S80 and amelogenin. *J Forenic Sci*, **41**, 660-663, 1996.

Buckland P.R. Polymorphically duplicated genes: Their relevance to phenotypic variation in humans. *Ann Med*, **35**, 308-315, 2003.

Buel E., Wang G. and Schwartz M. PCR amplification of animal DNA with human X-Y amelogenin primers used in gender determination. *J Forensic Sci*, **40**, 641-644, 1995.

Butler J.M. Forensic DNA Typing, Second Edition. Elsevier Academic Press, London, UK, 2005.

Burgoyne P.S. Genetic homology and crossing over in the X and Y chromosomes of mammals. *Hum Genet*, **61**, 85-90, 1982.

Cadenas A.M., Regueiro M., Gayden T., Singh N., Zhivotovsky L.A., Underhill P.A. and Herrera R.J. Male amelogenin dropouts: phylogenetic context, origins and implications. *Forensic Sci Int*, **166**, 155-163, 2007.

Carvalho C.M.B., Rocha J.L., Santos F.R., Kleiman S.E., Paz G. Yavetz H. and Pena S.D. Y-chromosome haplotypes in azoospermic Israeli men. *Human Biology*, **76**, 469-478, 2004.

Centre R., Lukeis R., Vrazas V. and Garson O.M. Y chromosome loss and rearrangement in non-small lung cancer. *Int J Cancer*, **55**, 390-393, 1993.

Chang Y.M., Both K. and Burgoyne L.A. Genetic variation at three DYS-STR loci and amelogenin between the Australian Caucasians and three Asian ethnic groups. *J Forensic Sci*, **47**, 918-921, 2002.

Chang Y.M., Burgoyne L.A. and Both K. Higher failures of amelogenin sex test in an Indian population group. *J Forensic Sci*, **48**, 1309-1313, 2003.

Chang Y.M., Perumal R., Keat P.Y., Yong R.Y.Y., Kuehn D.L. and Burgoyne L. A distinct Y-STR haplotype for amelogenin negative males characterised by a large Yp11.2 (DYS458-MSY1-AMEL-Y) deletion. *Forensic Sci Int*, **166**, 115-120, 2007.

Charlesworth B. The evolution of chromosomal sex determination and dosage compensation. *Curr Biol*, **6**, 149-162, 1996.

Clayton T.M., Whitaker J.P. and Maguire C.N. Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci. *Forensic Sci Int*, **76**, 7-17, 1995.

Collins P.J., Hennessy L.K., Leibelt C.S., Roby R.K., Reeder D.J. and Foxall P.A. Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFℓSTR® Identifiler® PCR amplification kit. *J Forensic Sci*, **49**, 1265-1277, 2004.

Cotton E.A., Allsop R.F., Guest J.L., Frazier R.R.E., Koumi P., Callow I.P., Seager A. and Sparkes R.L. Validation of the AmpFℓSTR® SGM Plus™ system for use in forensic casework. *Forensic Sci Int*, **112**, 151-161, 2000.

DeGraaff W.E., Van Echten J., Van der Veen A.Y., Sleijfer D.T., Timmer A., Koops H.S. and De Jong B. Loss of the Y-chromosome in the primary metastasis of a male sex cord stromal tumor: pathogenetic implications. *Cancer Genet Cytogenet*, **112**, 21-25, 1999.

Ellison J., Li X., Francke U. and Shapiro L.J. Rapid evolution of human pseudoautosomal genes and their mouse homologs. *Mamm Genome*, **7**, 25-30, 1996.

Eng B., Ainsworth P. and Wayne, J.S. Anomalous migration of PCR products using nondenaturing polyacrylamide gel electrophoresis: the amelogenin sex-typing system. *J Forensic Sci*, **39**, 1356-1359, 1994.

Evett I.W., Gill P.D. and Lambert, J.A. Taking account of peak areas when interpreting mixed DNA profiles. *J Forensic Sci*, **43**, 62-69, 1998.

Fernandes S., Paracchini S., Meyer L.H., Floridia G., Tyler-Smith C. and Vogt, P.H. A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in haplogroup N. *Am J Hum Genet*, **74**, 180-187, 2004.

Foote S., Vollrath D., Hilton A. and Page D.C. The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science*, **258**, 60-66, 1992.

Foresta C., Moro E. and Ferlin A. Y chromosome microdeletions and alterations of spermatogenesis. *Endocrine Reviews*, **22**, 226-239, 2001.

Freeman J.L., Perry G.H., Feuk L., Redon R., McCarrol S.A., Altshuler D.M., Aburatani H., Jones K.W., Tyler-Smith C., Hurles M.E., Carter N.P. Scherer S.W. and Lee C. Copy number variation: new insights in genome diversity. *Genome Res*, **16**, 949-961, 2006.

Applied Biosystems GeneAmp® PCR System 9700, User's Manual. Life Technologies, Carlsbad, CA, USA, 1997.

Graves, J.A.M. The origin and function of the mammalian Y chromosome and Y-borne genes – an evolving understanding. *Bioessays*, **17**, 311-320, 1995.

Hall T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp* Ser, **41**, 95-98, 2001.

Haas-Rochholz H. and Weiler G. Additional primer sets for an amelogenin gene PCR-based DNA-sex test. *Int J Legal Med*, **110**, 312-315, 1997.

Hayashi K., Hanaoka Y., Matsumura S., Takagi T., Kajiwara M., Tamaki N., Minaguchi K. and Sato Y. An autopsy case of Klinefelter's syndrome suspected and its DNA analysis. *Forensic Sci Int*, **113**, 119-125, 2000.

Henke J., Henke L., Chatthopadhyay P., Kayser M., Dülmer M., Cleef S., Pöche H. and Felske-Zech H. Application of Y-chromosomal STR haplotypes to forensic genetics. *Croatian Medical Journal*, **42**, 292-297, 2001.

Holt C.L., Buoncristiani M., Wallin J.M., Nguyen T., Lazaruk K.D. and Walsh P.S. TWGDAM validation of AmpFlSTR PCR amplification kits for forensic DNA casework. *J Forensic Sci*, **47**, 66-96, 2002.

Hunter S., Gramlich T., Abbott K. and Varma V. Y chromosome loss in esophageal carcinoma: an in situ hybridisation study. *Genes Chromosomes Cancer*, **8**, 172-177, 1993.

International Human Genome Consortium. Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921, 2001.

Inturri S., Robino C., Gino S., Caratti S and Torre C. Integration of the AmpFlSTR Identifiler PCR Amplification Kit with SRY-specific primers for gender identification. *Forensic Sci Int*, **S2**, 36-37, 2009.

Inoue K. and Lupski J.R. Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet*, **3**, 199-242, 2002.

Jacobs P.A. and Strong J.A. Case of human intersexuality having a possible XXY sex-determining mechanism. *Nature*, **183**, 302-303, 1959.

Jobling M.A. A survey of long-range DNA polymorphisms on the human Y chromosome. *Hum Mol Genet*, **3**, 107-114, 1994.

Jobling M.A., Samara V., Pandya A., Fretwell N., Bernasconi B., Mitchell R.J., Gerelsaikhan T., Dashnyam B., Sajantila A., Salo P.J., Nakahori Y., Disteche C.M., Thangaraj K., Singh L., Crawford M.H. and Tyler-Smith C. Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. *Hum Mol Genet*, **5**, 1767-1775, 1996.

Jobling M.A., Pandya A. and Tyler-Smith C. The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med*, **110**, 118-124, 1997.

Jobling M.A., Williams G., Schiebel K., Pandya A., McElreavey K., Salas L., Rappold G.A., Affara N.A. and Tyler-Smith C. A selective difference between human Y-chromosomal DNA haplotypes. *Curr Biol*, **8**, 1391-1394, 1998.

Jobling M.A. and Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews*, **4**, 598-612, 2003.

Jobling M.A., Lo I.C.C., Turner D.J., Bowden G.R., Lee A.C., Xue Y., Carvalho-Silva D., Hurles M.E., Adams S.M., Chang Y.M., Kraaijenbrink T., Henke J., Guanti G., McKeown B., Van Oorschot R.A.H., Mitchell R.J., De Knijff P., Tyler-Smith C. and Parkin E.J. Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing amelogenin Y. *Hum Mol Genet*, **16**, 307-316, 2007.

Jobling M.A. Copy number variation on the Y chromosome. *Cytogenet. Genome Res.*, **123**, 253-262, 2008.

Kao L., Tsai L., Lee J.C. and Hsieh H. Controversial cases of human gender identification by amelogenin test. *Forensic Science J*, **6**, 69-71, 2007.

Kashyap V.K., Sahoo S., Sitalaximi T. and Trivedi R. Deletions in the Y-derived amelogenin gene fragment in the Indian population. *BMC Med Genet*, **7**, 37-43, 2006.

Kayser M., Caglia A., Corach D., Fretwell N., Gehrig C., Graziosi G., Heidorn F., Herrmann S., Herzog B., Hidding M., Honda K., Jobling M., Krawczak M., Leim K., Meuser S., Meyer E., Oesterreich W., Pandya A., Parson W., Penacino G., Perez-Lezaun A., Piccinini A., Prinz M., Schmitt C., Roewer L., *et al*. Evaluation of Y-chromosomal STRs: a multicentre study. *Int J Legal Med*, **110**, 125-133, 1997.

Kimpton C.P., Oldroyd N.J., Watson S.K., Frazier R.R.E., Johnson P.E., Millican E.S., Urquhart A., Sparkes B.L. and Gill P. Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification. *Electrophoresis*, **17**, 1283-1293, 1996.

Krenke B.E., Tereba A., Anderson S.J., Buel E., Culhane S., Finis C.J., Tomsey C.S., Zachetti J.M., Masibay A., Rabbach D.R., Amiott E.A. and Sprecher C.J. *J Forensic Sci*, **47**, 773-785, 2002.

Kumagai R., Sasaki Y., Tokuta T., Biwasaka H. and Aoki Y. DNA analysis of family members with deletion in Yp11.2 region containing amelogenin locus. *Legal Medicine*, **10**, 39-42, 2008.

Kuroda-Kawaguchi T., Skaletsky H., Brown L.G., Minx P.J., Cordum H.S., Waterston R.H., Wilson R.K., Silber S., Oates R., Rozen S and Page D.C. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet*, **29**, 279-286, 2001.

La Fountain M., Schwartz M., Cormier J. and Buel E. Validation of capillary electrophoresis for analysis of the X-Y homologous amelogenin gene. *J Forensic Sci*, **43**, 1188-1194, 1998.

Lagerström M., Dahl N., Iselius L., Bäckman B. and Pettersson U. Mapping of the gene for X-linked amelogenesis imperfecta by linkage analysis. *Am J Hum Genet*, **46**, 120-125, 1990.

Lagerström M., Dahl N., Nakahori Y., Nakagome Y., Bäckman B., Landegren U. and Pettersson, U. A deletion in the amelogenin gene (AMG) causes X-linked amelogenesis imperfecta (AIH1). *Genomics*, **10**, 971-975, 1991.

Lahn B.T. and Page D.C. Four evolutionary strata on the human X chromosome. *Science*, **286**, 964-967, 1999.

Lattanzi W., Di Giacomo M.C., Lenato G.M., Chimienti G., Voglino G., Resta N., Pepe G. and Guanti G. A large interstitial deletion encompassing the amelogenin gene on the short arm of the Y chromosome. *Hum Genet*, **116**, 395-401, 2005.

Lau E.C., Mohandsas T., Shapiro L.J., Slavkin H.C. and Snead M.L. Human and mouse amelogenin gene loci are on the sex chromosome. *Genomics*, **4**, 162-168, 1989.

Lau Y., Chou P., Iezzoni J. and Komuves L. Expression of a candidate gene for the gonadoblastome locus in gonadoblastoma and testicular seminoma. *Cytogenet Cell Genet*, **91**, 160-164, 2000.

Luedtke N.W., Liu Q. and Tor Y. On the electronic structure of ethidium. *Chem Eur J*, **7**, 495-508, 2005.

Ma K., Inglis J.D., Sharkey A., Bickmore W.A., Hill R.E., Prosser E.J., Speed R.M., Thomson E J., Jobling M., Taylor K., Wolfe J., Cooke H.J., Hargreave T.B. and Chandley A.C. A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell*, **75**, 1287-1295, 1993.

Mannucci A., Sullivan K.M., Ivanov P.L. and Gill P. Forensic application of a rapid quantitative DNA sex test by amplification of the X-Y homologous gene amplification. *Int J Leg Med*, **106**, 109-193, 1994.

Masibay A., Mozer T.J. and Sprecher C. Promega Corporation reveals primer sequences in testing kits. *J Forensic Sci*, **45**, 1360-1362, 2000.

Mathias N., Bayes M. and Tyler-Smith C. Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet*, **3**, 115-123, 1994.

Michael A. and Brauner P. Erroneous gender identification by the amelogenin sex test. *J Forensic Sci*, **49**, 258-259, 2004.

Mitchell R.J., Kreskas M., Baxter E., Buffalino L. and Van Oorschot R.A.H. An investigation of sequence deletions of amelogenin (AMELY), a Y-chromosome locus commonly used for gender determination. *Annals Hum Biol*, **33**, 227-240, 2006.

Murphy K.M., Cohen J.S., Goodrich A., Long P.P. and Griffin C.A. Constitutional duplication of a region of chromosome Yp encoding AMELY, PRKY, and TBL1Y. *JMD*, **9**, 408-413, 2007.

Nakahori Y., Takenaka O. and Nakagome Y. A human X-Y homologous region encodes amelogenin. *Genomics*, **9**, 264-269, 1991.

Parkin E.J., Kraayenbrink T., Opgenort J.R.M.L., Van Driem G.L., Tuladhar N.M., De Knijff P. and Jobling M.A. Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: isolation and drift in the Himalayas. *Forensic Sci Int*, **166**, 176-181, 2007.

Polani P.E. Pairing of X and Y chromosomes, non-inactivation of X-linked genes, and the maleness factor. *Human Genet*, **60**, 207-211, 1982.

Pouchkarev V.P., Shved E.F. and Novikov P.I. Sex determination of forensic samples by polymerase chain reaction of the amelogenin gene and analysis by capillary electrophoresis with polymer matrix. *Electrophoresis*, **19**, 76-79, 1998.

Promega PowerPlex® Y System, Technical Manual # TMD018, Promega Corporation, Madison, USA, 2008.

Promega PowerPlex® 16 System, Technical Maual # TMD012, Promega Corporation, Madison, USA, 2008.

QIAGEN® QIAamp™ DNA Mini Kit and QIAamp™ DNA Blood Mini Kit Handbook, QIAGEN Pty. Ltd., Victoria, Australia, 2003.

QIAGEN® BioSprint™ DNA Blood Handbook, April 2004, QIAGEN Pty. Ltd., Victoria, Australia, 2003.

QIAGEN® Supplementary Protocol: Rapid purification of DNA from blood using the BioSprint15™ DNA Blood Kit (BS07 Sep-04). QIAGEN Pty. Ltd., Victoria, Australia, 2004.

QIAGEN® QIAquick™ Spin Handbook, QIAGEN Pty. Ltd., Victoria, Australia, 2002.

Quintana-Murci L. and Fellous M. The human Y chromosome: the biological role of a “functional wasteland”. *J Biomed Biotech,* **1**, 18-24, 2001.

Rao E., Weiss B., Fukami M., Rump A., Niesler B., Mertz A., Muroya K., Binder G., Kirsch S., Winkelman M., Nordsiek G., Heinrich U., Breuning M.H., Ranke M.B., Rosenthal A., Ogata T. and Rappold G.A. Pseudoautosomal deletions encompassing a novel gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat Genet*, **16**, 54-63, 1997.

Reijo R., Alagappan R.K., Patrizio P. and Page D.C. Severe oligo-zoospermia resulting from deletions of azoospermia factor gene on Y chromosome. *Lancet*, **347**, 1290-1293, 1996.

Repping S., Skaletsky H., Lange J., Silber S., Van der Veen F., Oates R.D., Page D.C. and Rozen S. Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *Am J Hum Genet*, **71**, 906-922, 2002.

Repping S., Skaletsky H., Brown L., Van Daalen S.K.M., Korver C.M., Pyntikova T., Kuroda-Kawaguchi T., De Vries J.W.A., Oates R.D., Silber S., Van der Veen F., Page D.C. and Rozen S. Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nature Genetics*, **35**, 247-251, 2003.

Roffey P.E., Eckhoff C.I. and Kuhl J.L. A rare mutation in the amelogenin gene and its potential investigative ramifications. *J Forensic Sci*, **45**, 1016-1019, 2000.

Rozen S., Skaletsky H., Marszalek J.D., Minx P.J., Cordum H.S., Waterston R.H., Wilson R.K. and Page D.C. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature*, **423**, 873-876, 2003.

Saiki R.K., Gelfand D.H., Stoffel S., Scharf S.J., Higuchi R., Horn G.T., Mullis K.B. and Erlich H.A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491, 1988.

Salido E.C., Yen P.H., Koprivnikar K.K., Yu L., and Shapiro L.J. The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes. *Am J Hum Genet*, **50**, 303-316, 1992.

Sambrook J.G. and Russel R. Molecular cloning: A laboratory manual. New York: Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY, USA, 2001.

Santos F.R., Pandya A and Tyler-Smith C. Reliability of DNA-based sex tests. *Nature Genet*, **18**, 103, 1998.

Sauter G., Moch H., Wagner U., Novotna H., Gasser T.C., Mattarelli G., Mihatsch M.J. and Waldman F.M. Y chromosome loss detected by FISH in bladder cancer. *Cancer Genet Cytogenet*, **82**, 163-169, 1995.

Schadrach B., Commane M., Hren C. and Warshawsky I. A rare mutation in the primer binding region of the amelogenin gene can interfere with gender identification. *J Mol Diag*, **6**, 401-405, 2004.

Sinclair A.H., Berta P., Palmer M.S., Hawkins J.R., Griffiths B.L., Smith M.J., Foster J.W., Frischauf A.M., Lovell-Badge R. and Goodfellow P.N. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, **346**, 240-244, 1990.

Skaletsky H., Kuroda-Kawaguchi T., Minx P.J., Cordum H.S., Hillier L., Brown L.G., Repping S., Pyntikova T., Ali J., Bieri T., Chinwalla A., Delehaunty A., Delehaunty K., Du H., Fewell G., Fulton L., Fulton R., Graves T., Hou S., Latrielle P., Leonard S., Mardis E., Maupin R., McPherson J., Miner T., Nash W., Nguyen C., Ozersky P., Pepin K., Rock S., Rohlfing T., Scott K., Schultz B., Strong C., Tin-Wollam A., Yang S., Waterston R.H., Wilson R.K., Rozen S. and Page D.C. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, **423**, 825-837, 2003.

Sparkes R., Kimpton C., Watson S., Oldroyd N., Clayton T., Barnett L., Arnold J., Thompson C., Hale R., Chapman J., Urquhart A. and Gill P. The validation of a 7-locus multiplex STR test for use in forensic casework. *Int J Legal Med*, **109**, 186-194, 1996.

Stankiewics P. and Lupski J.R. Genome architecture, rearrangements and genomic disorders. *Trends Genet*, **18**, 74-82, 2002.

Steinlechner M., Berger B., Niederstätter H. and Parson W. Rare failures in the amelogenin sex test. *Int J Legal Med*, **116**, 117-120, 2002.

Strachan T. and Read A.P. Human Molecular Genetics, Third Edition. Garland Science, Oxon, UK, 2004.

Stuppia L., Gatta V., Calabrese G., Guanciali F.P., Morizio E., Bombieri C., Mingarelli R., Sforza V., Frajese G., Tenaglia R. and Palka G. A quarter of men with idiopathic oligo-azoospermia display chromosomal abnormalities and microdeletions of different types in interval 6 of Yq11. *Hum Genet*, **102**, 566-570, 1998.

Sullivan K.M., Mannucci A., Kimpton C.P. and Gill P. A rapid and quantitative DNA sex test: Fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *Biotechniques*, **15**, 636-641, 1993.

Takayama T., Takada N., Suzuki R., Nagaoka S., Watanabe Y., Kumagai R., Aoki Y. and Butler J.M. Determination of deleted regions from Yp11.2 of an amelogenin negative male. *Legal Mededicine,* **11**, S578-S580, 2009.

Thangaraj K., Reddy A.G. and Singh L. Is the amelogenin gene reliable for gender identification in forensic casework and prenatal diagnosis? *Int J Legal Med*, **116**, 121-123, 2002.

Thangaraj K., Gupta N.J., Pavani K., Reddy A.G., Subramainan S., Rani D.S., Ghosh B., Chakravarty B. and Singh L. Y chromosome deletions in azoospermic men in India. *J Andrology*, **24**, 588-597, 2003.

The Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res*, **12**, 339-348, 2002.

Tiepolo L. and Zuffardi O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum Genet*, **34**, 119-134,1976.

Tilford C.A., Kuroda-Kawaguchi T., Skaletsky H., Rozen S., Brown L.G., Rosenberg M., McPherson J.D., Wylie K., Sekhon M., Kucaba T.A., Waterston R.H. and Page D.C. A physical map of the human Y chromosome. *Nature*, **409**, 943-945, 2001.

Tsuchiya K., Reijo R., Page D.C. and Disteche C.M. Gonadoblastoma: molecular definition of the susceptibility region on the Y chromosome. *Am J Hum Genet*, **57**, 1400-1407, 1995.

Tyler-Smith C., Taylor L. and Muller U. Structure of a hypervariable tandemly repeated DNA sequence on the short arm of the human Y chromosome. *J Mol Biol*, **203**, 837-848, 1988.

Van Deerlin V. and Leonard D. Bone marrow engraftment analysis after allogeneic bone marrow transplantation. *Clin Lab Med*, **20**, 197-225, 2000.

Vauhkonen H., Hedman M., Vauhkonen M., Sipponen P. and Sajantila A. Typing of XY (male) genotype from malignant neoplastic tissue by the amelogenin-base sex test. *J Forensic Sci*, **49**, 222-226, 2004.

Vergnaud G., Page D.C., Simmler M., Brown L., Rouyer F., Noel B., Botstein D., De La Chapelle A. and Weissenbach J. A deletion map of the human Y chromosome based on DNA hybridization. *Am J Hum Genet,* **38**, 109-124, 1986.

Vijayakumar S., Hall D.C., Reveles X.T., Troyer D.A., Thompson I.M., Garcia D., Xiang R., Leach R.J., Johnson-Pais T.L. and Naylor S.L. Detection of recurrent copy number loss at Yp11.2 involving TSPY gene cluster in prostate cancer using array-based comparative genomic hybridisation. *Cancer Res*, **66**, 4055-4064, 2006.

Vogt P.H., Edelmann A., Kirsch S., Henegariu O., Hirschmann P., Kiesewetter F., Köhn F.M., Schill W.B., Farah S., Ramos C., Hartmann M., Hartschuh W., Meschede D., Behre H.M., Castel A., Niesschlag E., Weidner W., Gröne H.J. Jung A., Engel W. and Haidl G. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum Molec Genet*, **5**, 933-943, 1996.

Vollrath D., Foote S., Hilton A., Brown L.G., Beer-Romero P., Bogan J.S. and Page D.C. The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science*, **258**, 52-59, 1992.

Von Wurmb-Schwark N., Bosinski H. and Ritz-Timme S. What do the X and the Y chromosomes tell us about sex and gender in forensic case analysis? *J of Clinical Forensic Medicine*, doi:10.1016/j.jcfm.2005.09.003., 2006.

Wallin J.M., Holt C.L., Lazaruk K.D., Nguyen T.H. and Walsh P.S. Constructing universal multiplex PCR systems for comparative genotyping. *J Forensic Sci*, **47**, 52-65, 2002.

Yen P. The fragility of fertility. *Nat Genet*, **29**, 243-244, 2001.

Young D.R., Tun Z., Honda K. and Matoba R. Identifying sex chromosome abnormalities in forensic DNA testing using amelogenin and sex chromosome short tandem repeats. *J Forensic Sci*, **46**, 346-348, 2001.

Yong R.Y.Y., Gan L.S.H., Chang Y.M. and Yap E.P.H. Molecular characterization of a polymorphic 3-Mb deletion at chromosome Yp11.2 containing the AMELY locus in Singapore and Malaysian populations. *Hum Genet*, **122**, 237-249, 2007.

Zehner R. and Bohrer U. DYS19 and amelogenin in artificial blood stains with defined amounts of male and female cells. *Int J Legal Med*, **111**, 340-342, 1998.

6.2 ELECTRONIC REFERENCES

BioEdit, Biological sequence alignment editor,
http://www.mbio.ncsu.edu/Bioedit/bioedit.html

Genbank,
http://ncbi.nlm.nih.gov/Genbank/

Position of loci on the Y-chromosome, Build 37.1, Reference assembly,
http://www.ncbi.nlm.nih.gov/mapview/

STS Primer sequences and PCR conditions,
http://www.ncbi.nlm.nih.gov/nucleotide/

Y-STR haplotype reference database,
http:www.yhrd.org

APPENDIX A

Conference at which research contained in this thesis was presented

---

Research contained in this thesis was presented at the following national meeting. The name of the presenting author is underlined.

A.1 RESEARCH PRESENTED AT NATIONAL CONFERENCE

A.1.1 **Second Annual African DNA Forensics Conference:** Pretoria, South Africa, October 2010.

Ferreira I., Towers G.W. and Olckers A. Sequence variation of the amelogenin gene on the Y-chromosome in a black South Africa cohort.