University of Venda

Isolation and characterization of microsatellite markers for human identification in the Vhembe district, Limpopo province, South Africa

By

Godfrey Azwindini Dzhivhuho

11575968

A dissertation submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Zoology, School of Mathematical and Natural Sciences

University of Venda

Private Bag X5050

Thohoyandou

0950

Supervisor : Dr. T. C. Nangammbi

: Department of Zoology

Co-supervisor : Prof. A. Samie

: Department of Microbiology

Co-supervisor : Mr. L.F. Chauke

: Department of Environmental affairs

Submission date: April 2014

UNIVERSITY OF VENDA
LIBRARY

UNIVEN LIBRARY
Library Item : 20141259

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

University of Venda
Creating Future Leaders

DECLARATION

I Mr Godfrey Azwindini Dzhivhuho declare that the dissertation (Isolation and characterization of microsatellite markers for human identification in Vhembe district, Limpopo province) submitted for Master's Degree in Zoology, of the University of Venda, has not been previously submitted for any degree or examination in this or any other university, and that all the materials contained therein have been duly acknowledged.

Student: **Surname and Initials: Dzhivhuho, G.A.**

Signature.....

Date.....07/04/2014.....

Supervisor: **Surname and Initials: Nangammbi, T.C.**

Signature.....

Date.....07/04/2014.....

Co-Supervisor: **Surname and Initials: Samie, A.**

Signature.....

Date.....07/04/14.....

Co-Supervisor: **Surname and Initials: Chauke, L.F.**

Signature.....

Date.....07/04/2014.....

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

University of Venda
Creating Future Leaders

ACKNOWLEDGEMENTS

I thank God for granting me the opportunity, ability and the strength to pursue and fulfill my dream in DNA technology research. To my parents, Mr David and Mrs Sheila Dzhivhuho, I thank you for supporting, encouraging and allowing me to continue in the field of science.

Special thanks to my supervisor Dr. Tshifhiwa Nangammbi and my co-supervisors Dr. Amidou Samie and Mr. Lucas Chauke for the wonderful support and impact they had on this project. I would like to give special thanks to Mrs. Laurie Grey for playing a critical role in my lab work experiments, not forgetting the team from microbiology department in the University of Venda and University of Virginia for their full support in this project. I would also like to thank Dr. Stephen Williams for his training in bioinformatics and technical troubleshooting during my summer research in the United States of America.

I thank my lab mates, (Matamba, E., Shihlomule, D., Mutavhatshindi, V., and Mboweni, B) for helping me in sample collections and suggestions that somehow contributed to this project.

The present study was funded by the Research and Publication Committee through a grant to Dr. T. Nangammbi and Dr. A. Samie. I also acknowledge the financial support provided by the National Research Foundation (NRF; Research and innovation program). It will be important to give thanks to all students who assisted me in this project.

Finally, special thanks to the University of Venda, particularly the Department of Zoology for allowing me do my Master's degree with them.

© University of Venda

University of Venda
Creating Future Leaders

TABLE OF CONTENTS

CONTENTS **PAGE**

© University of Venda

University of Venda
Creating Future Leaders

© University of Venda

University of Venda
Creating Future Leaders

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

University of Venda
Creating Future Leaders

© University of Venda

University of Venda
Creating Future Leaders

LIST OF ABREVIATIONS

%:	Percentage
°C:	Degrees centigrade
ABI:	Applied Biosystems
AFLP:	Amplified fragment length polymorphism
BLAST:	Basic Local Alignment Search Tool
Bp:	Base pairs
BSA:	Bovine serum album
CODIS:	Combined DNA Index System
dH2O:	Distilled water
DNA:	Deoxyribonucleic acid
dNTP:	Deoxynucleotide triphosphate
E. coli:	*Escherichia coli*
EtOH:	Ethanol
FBI:	Federal Bureau of Investigation
FSS:	Forensic Science Service
HLA:	Human leukocyte antigens
HSRC:	Human Sciences Research Council
IDT:	Integrated DNA Technology
M:	Molar

UNIVERSITY OF VENDA

© University of Venda

University of Venda
Creating Future Leaders

mA:	Milli amperes
$MgCl_2$:	Magnesium chloride
ml:	Millilitre
mM:	Milli molar
MPC:	Magnetic particle collector
NaCl:	Sodium chloride
NaOAc:	Sodium acetate
ng:	Nano gram
PCR:	Polymerase chain reaction
RAPD:	Random amplification of polymorphic DNA
RCMP:	Royal Canadian Mounted Police
RFLP:	Restriction fragment length polymorphism
rpm:	Rounds per minute
SA:	South Africa
SAPS:	South African Police Service
SGM:	Second generation multiplex
SNP:	Single nucleotide polymorphism
SOC:	Super Optimal broth (catabolite-repression)
SOP:	Standard operational procedure
SSC:	Saline-sodium citrate
SSRs:	Simple sequence repeats

© University of Venda

University of Venda
Creating Future Leaders

STR:	Short tandem repeat
TAE:	Tris-acetate EDTA
TE:	Tris EDTA
TLE:	Tris low EDTA
U:	Units
UK:	United Kingdom
US$:	American Dollar
UV:	Ultra-violet
V:	Volts
VNTR:	Variable number of tandem repeats
W:	Watts
µg:	Micro gram
µl:	Microliter

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

University of Venda
Creating Future Leaders

LIST OF TABLES

© University of Venda

University of Venda
Creating Future Leaders

LIST OF FIGURES

© University of Venda

University of Venda
Creating Future Leaders

© University of Venda

University of Venda
Creating Future Leaders

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

University of Venda
Creating Future Leaders

LIST OF APPENDICEC

© University of Venda

University of Venda
Creating Future Leaders

ABSTRACT

Forensic science is a growing discipline in South Africa. However, there are only a few forensic laboratories, mostly situated in Gauteng, Western and Eastern Cape. Due to this lack of facilities, combined with a high demand for forensic services, there is a huge backlog and turnaround time throughout South Africa; forensic testing has therefore become a major challenge. Studies have revealed that paternity disputes in rural areas are not resolved due to affordability. This contributes to an increase in single parenting among South Africans, particularly by mothers of the rejected children. The purpose of this study was to design a cost-effective Standard Operational Procedure (SOP) for isolating potentially polymorphic human microsatellite markers at a cheap and most convenient time. The designed human microsatellite markers were optimized for PCR amplification and used in genotyping a mini population of three major ethnic groups in the Limpopo province within in the Vhembe district. GENETOOLS software was used to analyze band patterns and generate data. The generated data was analyzed Using SPSS v 21. A total of 14 autosomal microsatellite markers where screened from 41 positive clones. Of the 14 screened markers, a total of 13 markers successfully amplified specific targets with an average of 3 alleles. However, only 5 markers ((GAD02; GAD06; GAD11; GAD12 and GAD13) were successfully optimized to produce a single allele per locus. A multidimensional unfolding clustering analysis grouped the three ethnic groups into 7 clusters based on band pattern similarities suggesting that the individual populations are closely related. However, three individuals clustered away from the rest of the populations suggesting that they are genetically distanced from the rest of the populations. These results need to be validated using a larger sample size with more markers. This protocol serves as brainchild for isolation of microsatellite markers for identification purposes and can be further improved for paternity test analysis.

Key words: Microsatellite markers; Paternity; Polymorphism; DNA profiling/fingerprinting; Forensics; Limpopo; Population.

© University of Venda

University of Venda
Creating Future Leaders

CHAPTER 1

1.1. GENERAL INTRODUCTION

1.1.1. Status of forensic science in developed and developing countries

One of the fast growing disciplines in the biotechnology field is forensic science, which has improved enormously and its uses continue to spread rapidly throughout the world as DNA analysis has become the fundamental of solving crimes. Presently, developed countries such as American and European countries make use of forensic science to solve criminal cases, while only a few developing countries in Africa are adopting the technology.

In the United States of America, there are established organizations and facilities working on fighting crime such as the Federal Bureau of Investigation (FBI), situated 50 miles outside Washington, D.C. in Quantico (FBI Laboratory, 2003). They have a laboratory with about 650 employees and they partner with the state and local crime laboratories throughout the country. They have developed a hierarchy of DNA profile indexes known as the Combined DNA Index System (CODIS). CODIS is recognized as a national DNA information repository maintained by the FBI that allows state and local crime laboratories to store and compare DNA profiles from crime-scene evidence and convicted offenders (FBI Laboratory, 2003).

In South Africa, forensic science has been established over the past few decades and it continues to spread with new developments and improvements in technology. South African Police Service (SAPS) forensic science laboratories make use of the AmpFISTR Profiler Plus™ PCR Amplification kit, which is a ten locus Short Tandem Repeats (STR) system which was developed and is supplied by the ABI (Applied Biosystems Corporation; United States of America) (Meintjes-van der Walt, 2008). There are few forensic laboratories in South Africa and are all based in few cities of the country,

© University of Venda

mainly, Pretoria, Johannesburg, East London and Cape Town. For the Limpopo province, all forensic tests needed are sent to the nearest forensic laboratory which is in Pretoria.

This is where all the samples are analyzed and thereafter the results are brought back to the district. The Vhembe district is one of five districts of the Limpopo province situated in the far northern part of South Africa and is mostly rural. It is further divided into four district municipalities, namely, Musina, Mutale, Thulamela and Makhado. The population groups mostly represented in this part of the country are the Vhavenda, Tsonga and Bapedi.

1.1.2. Paternity disputes in South Africa

Young people in South Africa suffer rejection of paternal responsibility due to unplanned pregnancies and a variety of studies have been undertaken to understand the reason for this (Makiwane and Kwizera, 2008; Eddy, 2009; Atuyambe et al., 2009; Jewkes et al., 2009). In 2003, a survey was conducted in Eastern Cape as a stepping stone study which reported that 26.9% of young men deny responsibility for pregnancies (Jewkes, unpublished data). Findings showed that reactions towards such denials and disputes over these pregnancies by the participants' boyfriends, varied. These reactions start from the boyfriend expressing disbelief about the news of the pregnancy, then leading to relocating and rejecting responsibility by the boyfriends were amongst the most common. Responses from the study by Jewkes et al., (2009) showed that female participants believed that denial of their pregnancy was a punishment for being careless and not taking contraceptives. Female participants suffered distress due to the constant worry from the unresolved paternity issue. It has been suggested that a reliable means of determining paternity, other than paternal resemblance, such as a DNA testing, should be made affordable and accessible in public health facilities for quicker recourse (Nduna and Jewkes, 2012).

© University of Venda

In some black South African communities, mostly in the rural areas, an indigenous traditional response to out-of-marriage pregnancy involving the woman identifying the man alleged to be responsible for the pregnancy is still being practiced (Nduna and Jewkes, 2012). This involves relatives acting on the pregnant girl's family's behalf and taking her to the man's family for them to acknowledge responsibility through *intlawulo* (a cultural practice where a man acknowledges responsibility for making a female pregnant by pledging cattle to her family) (Kaufman et al., 2001; Jewkes et al., 2009). If the male denies the pregnancy, the families wait until the baby is born to scan the infant for family resemblances (Kaufman et al., 2001; Nduna and Maseko, 2008). If the baby does not resemble any of the proposed father's family members, the woman's family may exonerate the man as being responsible and not seek any child support from him (Datta, 2007; Hunter, 2006) thus leaving the woman to be fully responsible for the child.

Presently, DNA paternity tests are available in South Africa to resolve paternity disputes. These services are unfortunately, inaccessible to a majority of young women in rural and disadvantaged areas of South Africa and elsewhere (Walker and Pohl, 1989; Durosinmi and Alabi, 1995; Christianson et al., 1999; Pearson, 2003; Ejele and Nwache, 2004; Kotze et al., 2006). Limiting factors such as cost of paternity services which is at least R2 000, makes it difficult for unemployed individuals to access private laboratories for paternity services. Reproductive health clinical services are available all over South Africa but they exclude paternity testing (Forbes et al., 2008; Ashton et al., 2009).

Unplanned pregnancies due to youngsters who have sex at a very young age leads to unwanted pregnancies and contributes to single parenting (Haskett, 2011). If the youngsters are teenagers, the father most often runs away from the responsibility leaving the mother to handle things on her own. Youth pregnancies which leaves a lot of young females as single parents, is an issue that the Vhembe district is faced with. Previous study by Maluleke (2010) in the Vhembe district involving 400 individuals of 17-25 years of age, in four villages of Vhembe district indicated that 83.8% of all the

© University of Venda

respondents were sexually active. This contributes to the tremendous increase in single parenting and paternity disputes. The percentage of children with absent but living fathers was reported by the Human Sciences Research Council to have increased from 42% in 1996 to 48% in 2009 (HSRC, 2007; Stats SA, 2008). This is a clear indication that most males deny their responsibility (Haskett, 2011) and take advantage of the fact that the mother will not be able to prove that indeed they are the fathers. Paternity disputes have become very common in the rural areas within South Africa and only those financially well-off have the privilege of getting justice because they can afford paternity tests (Nduna and Jewkes 2012).

1.1.3. The need for a forensic facility in South Africa

A feasibility study was conducted by Lithole (2011) who made use of questionnaires to determine the level of understanding of forensic science by office bearers in the Vhembe district and to identify most common tests known and requested by these individuals. The survey indicated that 79.8% of the participants were aware of forensic science and the remaining ones had a minimum idea but were willing to know more. About 98.9% indicated that there is a need for a forensic laboratory in the Limpopo province which currently does not have any. The forensic tests of interest to the participants were mainly paternity testing and crime investigations. Lithole's study showed the need for establishing a forensic science laboratory in the Vhembe district with special focus on offering the most affordable paternity and forensic test which will in turn reduce the backlog of paternity disputes and criminal cases. In addition, such a forensic lab will also offer molecular diagnostic tests which will enhance treatment for diseases such as Tuberculosis and Cancer. Not only does the above mentioned survey calls for the establishment of a forensic lab in Vhembe district, but also in other less developed areas of South Africa. Such facilities will be easily accessible as they will be closer to the locals, thus will be cost-effective and will reduce the sample process turn-around time by far.

© University of Venda

University of Venda
Creating Future Leaders

1.1.4. Microsatellites as genetic markers used in kinship and paternity studies

Microsatellites are repeated motifs of 1-6 bases arranged in tandem and are found in all prokaryotic and eukaryotic genomes. Their popularity made them to be the most preferred markers, because they are single locus, co-dominant markers for which many loci can be efficiently combined in the genotyping process, providing a fast and inexpensive technique for sampling of genomes. Microsatellite markers generally have high-mutation rates which result in high allelic diversity which makes them highly informative (Hedrick, 1999; Zane et al., 2002; Selkoe and Toonen, 2006). Their simple Mendelian mode of inheritance makes them particularly suitable for the study of fine population structure, mating systems and pedigrees (Abdelkrim et al., 2009).

However, microsatellite genotyping requires primer sequences that targets and amplify regions containing short repetitive sequences. Microsatellite markers are commonly referred to as simple sequence repeats (SSRs) or short tandem repeats (STR) (Zane et al., 2002). As with any other PCR based technique, primers are required. The primer sequence information for a species in question is however obtained through isolation of the repeat containing fragments in a process referred to as microsatellite library design.

Typically, microsatellite library design requires: isolation of pure DNA, construction of a genomic library enriched for repeated motifs, sequencing of microsatellite containing clones, primer design, optimization of PCR amplification for each primer pair, and a test for polymorphism on a few unrelated individuals (Zane et al., 2002). However, the majority of these steps are either expensive, time-consuming, or both. Furthermore, because microsatellite isolation requires the construction of genetically modified vector which gets inserted in bacteria *Escherichia coli* (*E. coli*), controlled biosecurity issues must be dealt with when molecular cloning is used, which increases the costs of marker development and project timelines. The entire process, is straightforward, but is time-consuming.

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

Scientists have explored and developed new technology that can simplify these steps. This includes a new genomic shotgun sequencing technology coupled with fast and efficient bioinformatics tools to eliminate the most intensive wet lab steps in a simple, fast, and economic way (Abdelkrim et al., 2009).

1.2. RESEARCH PROBLEM STATEMENT

The Vhembe district is faced with youth pregnancies which mostly result in single parenting because potential fathers often deny responsibility of these children. In this situation, young mothers are generally powerless as the current tool for ascertaining paternity is expensive and generally out of reach of the local communities. This is because paternity testing involves the use of genetic markers such as microsatellites. One challenge in the use of microsatellites for human identification is that microsatellite primers are not generally available for use as most commercial laboratories have these microsatellite primers patented in South Africa and other developed countries, hence there is a necessity for developing a library from scratch.

In the present study, we intended to develop a cost effective Standard Operational Procedure (SOP) for microsatellite isolation and characterization. A genomic library was enriched with trinucleotide (AAT_{12}; AGC_6; AAG_8 and tetranucleotide $CACA_6$; $AAGT_8$; $CACC_6$) microsatellite repeats, which were to be tested for their discrimination power and level of polymorphism in a mini population study of the three major ethnic groups in the Limpopo province within the Vhembe district. We anticipated a minimum of 12 autosomal polymorphic microsatellite markers which could be optimized and employed in a mini population study and possibly, paternity testing, kinship and identification of individuals in the Vhembe district, Limpopo province, South Africa.

© University of Venda

University of Venda
Creating Future Leaders

1.3. AIMS OF THE STUDY

a. To develop the first human microsatellite genomic library from a laboratory not fully equipped at the University of Venda.
b. To preliminarily characterize the isolated microsatellites loci to determine whether they are polymorphic enough for use paternity testing and forensics analysis.

1.4. OBJECTIVES OF THE STUDY

a. To construct a microsatellite genomic library for humans to be used for paternity testing and for forensic analysis.
b. To document the step by step protocol used in constructing the library in question. The documented protocol will thus serve as reference in the event where another library is to be constructed at the University of Venda.
c. To evaluate these primers on a limited number of samples to make sure that they can be used in kinship studies.

1.5. RESEARCH QUESTION

a. Which non-CODIS microsatellite loci can be used for differentiation between individuals?
b. Are these microsatellites loci polymorphic enough for precise distinction between two individuals?
c. Can these microsatellites be used efficiently for kinship analysis and paternity testing?

1.6. HYPOTHESIS

a. Isolation of microsatellites suitable for a population study and possibly paternity and kinship studies is achievable at minimum cost.

© University of Venda

University of Venda
Creating Future Leaders

1.7. LITERATURE REVIEW

1.7.1. Background of genetic and paternity testing

Genetic identity testing is a method used for identifying patterns of genetic material unique to almost every individual (Saad, 2005). It makes use of the modest DNA sequence variations in the human genome to differentiate between individuals (Butler, 2005). A variety of techniques and terms are used to test for these variations; DNA fingerprinting, DNA profiling, and DNA typing are commonly used for these techniques. These terms have been and are used interchangeably, nevertheless some technical differences do exist between them (Saad, 2005).

Paternity testing is the use of genetic fingerprinting to determine the relationship between father and child. In the early 1900s, blood types were the most common tests considered in human paternity testing (Schwarz and Dorner, 2003; Cifuentes et al., 2006). This type of paternity system was discovered in the 1901 by Karl Landsteiner and today it is known as the ABO blood-typing system. The ABO gene is located on the human chromosome 9 and consists of 7 exons (Bugert et al., 2012). In the ABO system, three alleles: co-dominant, A and B alleles, and the recessive O allele are involved. For example, if a person's blood type is A, he or she will either have two A alleles or one A allele and one O allele. Similarly, if person's blood type is B, he or she will either have two B alleles or one B allele and one O allele. However if a person's ABO blood type is O, he or she will have two O alleles. Some people have type AB blood, which means they inherited both an A allele and a B allele. In paternity testing, the ABO blood-typing can be used to exclude a man from being a child's father rather than tell who the father is. For example, a man who has type AB blood cannot father a child with type O blood, because he would pass on either the A or the B allele to all of his offspring. Despite their usefulness in this regard, ABO blood groups cannot be used to confirm whether a man is indeed a child's father. The possible antigen combinations are shown in Table 1.

© University of Venda

Because of these limitations, the use of additional blood antigens, such as those associated with the MN and Rh systems, advanced the use of blood-typing for both paternity and forensics (Cifuentes et al., 2006). Nevertheless, such blood groups systems were only about 40% effective in ruling out a man as a child's father.

Table 1: ABO blood system's possible combinations between parent and child

		Father's Blood Type				
		A	B	AB	O	
Mother's Blood Type	A	A or O	A, B, AB, or O	A, B, or AB	A or O	Child's Blood type Must Be
	B	A, B, AB or O	B or O	A, B, or AB	B or O	
	AB	A, B, or AB	A, B, or AB	A, B, or AB	A or B	
	O	A or O	B or O	A or B	O	

(http://www.canadiancrc.com/paternity_determination_bloodtype.aspx).

Another improvement in the 1970s added a distinguishing feature that made it possible to rule out men as fathers with 80% effectiveness, which tested for human leukocyte antigens (HLAs). The genes responsible for the HLA system are involved in antigen presentation to T cells. The HLA system is more polymorphic, with over 3,200 different alleles identified (Williams, 2001; Robinson et al., 2003). Despite its improvement, there were limitations to this type of testing due to the fact that: (1) tissues or blood needed to be well-fixed; (2) the HLA proteins are not always present on all cells and (3) there is a lot of mixture of the genes in gamete production and other environmental factors Robinson et al., 2003.

© University of Venda

The ABO blood group and the HLA system are phenotype-based paternity testing which either "rule in" or "rule out" possible fathers rather than confirm the presence of a father-child relationship. Because of the presented number of limitations, DNA became the main studied tool for human identification (Human genome project). Human beings share more biological similarities than differences, with 99.9% genetic similarities between individuals (Kruglyak and Nickerson, 2001).

The human genome consists of approximately 3.2 billion base pairs, of which only 3.2 million base pairs (0.1%) differ from one individual to another (Human Genome Project and Beyond, 2008). The difference in base pair configuration makes us different from each other and enables us to differentiate individuals genetically by use of genetic testing, commonly known as DNA fingerprinting.

DNA fingerprinting involves comparing DNA fragments of one individual with the other, to either match or differentiate individuals. It was developed in 1984 by Alec Jeffreys at the University of Leicester (United Kingdom) (Jeffreys et al., 1986) who later became Sir Alec Jeffreys. During his career studying humans, Jeffreys noticed that certain regions of DNA contained repeated DNA sequences distributed next to each other. He also discovered that the number of repeated sequences in a sample could differ from one individual to another. He noticed that these short repeat sequences in children were similar to those of their parents, suggesting that these sequences are inherited from both parents. These DNA repeat markers were then used as genetic markers known as variable number of tandem repeats (VNTRs) and were improved over time (Pourcel et al., 2009).

Genetic markers are regions or fragments of DNA that are found in every organism's genome which can be used to differentiate cells, individuals, populations, or species. The most popular and versatile genetic markers include various nuclear and mitochondrial DNA sequence regions, restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), random amplification of polymorphic

© University of Venda

DNA (RAPD), variable number tandem repeat (VNTR), single nucleotide polymorphism (SNP) and microsatellite or simple sequence repeat (SSR) also known as short tandem repeats (STR). For more than two decades, microsatellites have been detected throughout eukaryote genomes and new ones are continuously been discovered (Butler and Hill, 2012) and used in forensic sciences.

1.7.2. Historical perspective on microsatellite marker selection

During the early 1990s, microsatellites or STR markers became the first described effective tools for human identity testing (Edwards et al., 1991; 1992). Since then, a search for new loci was commenced by the Forensic Science Service (FSS) as well as studying of population variation with a number of STR candidates (Kimpton et al., 1993). Contributions to early efforts with STR typing were also done by the Royal Canadian Mounted Police (RCMP) (Frégeau and Fourney, 1993) along with a number of European labs (Butler, 2006). Four loci: TH01; VWA; FES/FPS and F13A1; were the first FSS multiplex applied to forensic casework (Kimpton et al., 1994) followed by the second generation multiplex (SGM) loci: TH01; VWA; FGA; D8S1179; D18S51; and D21S11 (Kimpton et al., 1996). Success of STR typing technology obtained by the UK, led to the establishment of the core STR loci (CODIS) by the FBI Laboratory which was powered through funding provided by the Congressional DNA Identification Act of 1994 (Butler, 2006; Figure 2). In 1997, the 13 STR loci (CSF1PO; FGA; TH01; TPOX; vWA; D3S1358; D5S818; D7S820; D8S1179; D13S317; D16S359; D18S51 and D21S11) were announced as the core CODIS markers (Budowle et al., 1998; Butler, 2005; Butler and Hill, 2012).

In 2011, an expanded set of core STR loci was proposed by the FBI Laboratory for the United States in order to: i) reduce the possibility of associations by chance as the number of profiles stored in the U.S. national DNA database increases; ii) increase international compatibility and contribute to the law enforcement data-sharing efforts;

© University of Venda

and iii) increase the system's discrimination power to assist missing-persons cases (Hares, 2012).

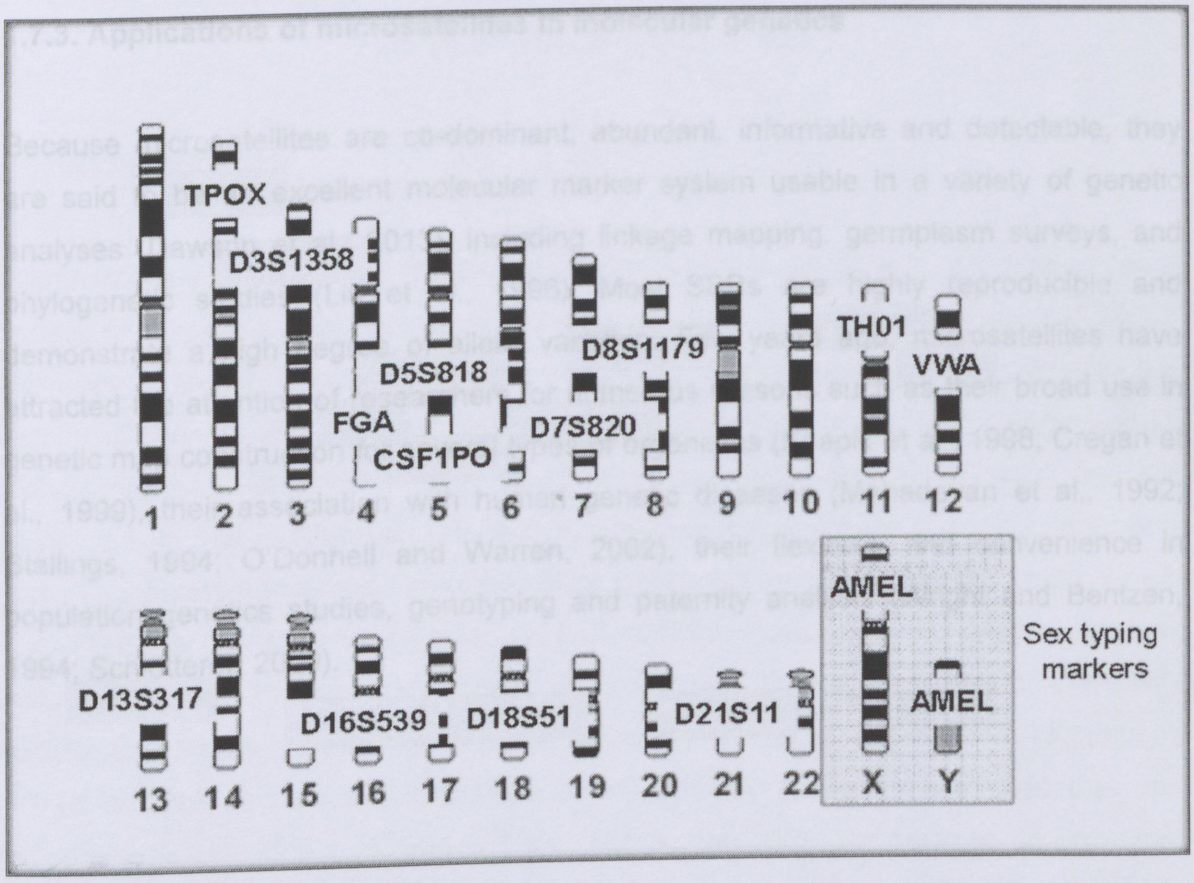


Figure 1: The 13 CODIS core STR loci with chromosomal positions.

(http://www.cstl.nist.gov/strbase/fbicore.htm).

STR markers are effective for human identification purposes because of their high variability in numbers of repeats among individuals (Butler and Hill, 2012). The widespread use of these highly polymorphic co-dominant genetic markers attracted considerable interest to marker-based kinship analysis in populations when prior lineage information is not available (Chung et al., 2004; Konovalov and Heg, 2008). Because of that, the use of STR markers for many years to come is anticipated, due to their

© University of Venda

approval and expansion in national DNA databases around the world (Butler and Hill, 2012).

1.7.3. Applications of microsatellites in molecular genetics

Because microsatellites are co-dominant, abundant, informative and detectable, they are said to be an excellent molecular marker system usable in a variety of genetic analyses (Dawson et al., 2013), including linkage mapping, germplasm surveys, and phylogenetic studies (Liu et al., 1996). Most SSRs are highly reproducible and demonstrate a high degree of allelic variation. Few years ago, microsatellites have attracted the attention of researchers for numerous reasons such as their broad use in genetic map construction for several types of organisms (Knapik et al., 1998; Cregan et al., 1999), their association with human genetic diseases (Mahadevan et al., 1992; Stallings, 1994; O'Donnell and Warren, 2002), their flexibility and convenience in population genetics studies, genotyping and paternity analysis (Wright and Bentzen, 1994; Schlötterer, 2000).

© University of Venda

University of Venda
Creating Future Leaders

CHAPTER 2

DEVELOPMENT OF A STANDARD OPERATIONAL PROCEDURE (SOP) FOR MICROSATELLITE ISOLATION

2.1. INTRODUCTION

Microsatellite markers have become important sources of genetic information for a variety of purposes (Goldstein and Schlotterer, 1999; Webster and Reichart, 2004). However, to amplify microsatellite loci using PCR, primers must be developed from the vicinity of the specific microsatellite repeats. This section of the study attempted to isolate microsatellites from human mouth wash samples using an in-house protocol modified from that of Glenn and Schable (2005) protocol to reduce the costs, labor and time. This protocol involved microsatellites isolation through eight major steps, which were (1) extraction of high molecular weight DNA from tissue (methods depending upon nature of tissue), in this study mouth wash sample, (2) fragmentation of DNA into suitable size fragments with restriction enzymes in preparation for ligation, (3) ligation of linkers to DNA fragments (Linkers are short synthetic single DNA strands that are hybridized to form double stranded DNA. The ligated ends are essential in providing binding sites for the next PCR steps at later stages. The linkers also make cloning of the fragment into the vectors easy because they are compatible with the restriction sites in the vector's multiple cloning sites (Glenn and Schable, 2005), (4) enrichment of microsatellite DNA-linker with biotin-oligos and dynabeads, (5) PCR recovery of enriched DNA fragments with microsatellites, (6) insertion of enriched DNA fragments into plasmids vector (cloning of the fragments into many copies of DNA pieces of 300-600 bp), (7) DNA recovery from the plasmid vector and (8) sequence the positive clones that make it through all the above selection steps. In this project, the above steps were essentially used for the development of polymorphic primers which were used in the population study of the three ethnic groups (Vhavenda, Tsonga and Bapedi) in the Limpopo province within the Vhembe district.

© University of Venda

University of Venda
Creating Future Leaders

2.2. MATERIALS AND METHODS

2.3. ETHICAL CLEARANCE

The study was reviewed by the Human Research Ethics committee of the University of Venda. Participants had to fill in a consent form as agreement to participation to this project. The study objectives were fully explained in local languages to ensure that the participants fully understood the project. No personal information such as name and disease status of the participants were collected with exception to participants who were willing to provide such details for all information and data generated from the study was treated with confidentiality. See Appendix A for consent form.

2.3.1. Target Population of study and collection of samples

The target population was from the Vhembe district, Limpopo province, South Africa shown in Figure 2 below. Mouth wash samples were collected from three unrelated participants for the isolation of microsatellite markers and this is enough to produce microsatellite library. Each participant represented one ethnic group labeled A for Vhavenda, B for Tsonga and C for Bapedi. The individuals were given a questionnaire about their family backround to validate their ethnic origin. Participants were given sterile industrially processed bottled water (still water) and were asked to rinse their mouth and spit into the sample collection container provided. In a 15 ml centrifuge tube, the mouth wash samples were centrifuged at 2000 rpm for 15 minutes and the pellet was transferred to a 2 ml eppendorf tube and kept in the freezer at -20°C until further analysis within 24 hours. It should be noted that DNA from oral samples contains bacterial DNA in addition to human genomic DNA (Feigelson et al., 2001; Garcia-Closas et al., 2001). However, bacterial DNA has minimal practical significance and studies have shown that DNA from oral samples gives equivalent results to DNA from blood for applications like PCR, SNP and STR genotyping (Terasaki et al., 1998; Todesco et al., 2003), hence this was not a challenge.

© University of Venda

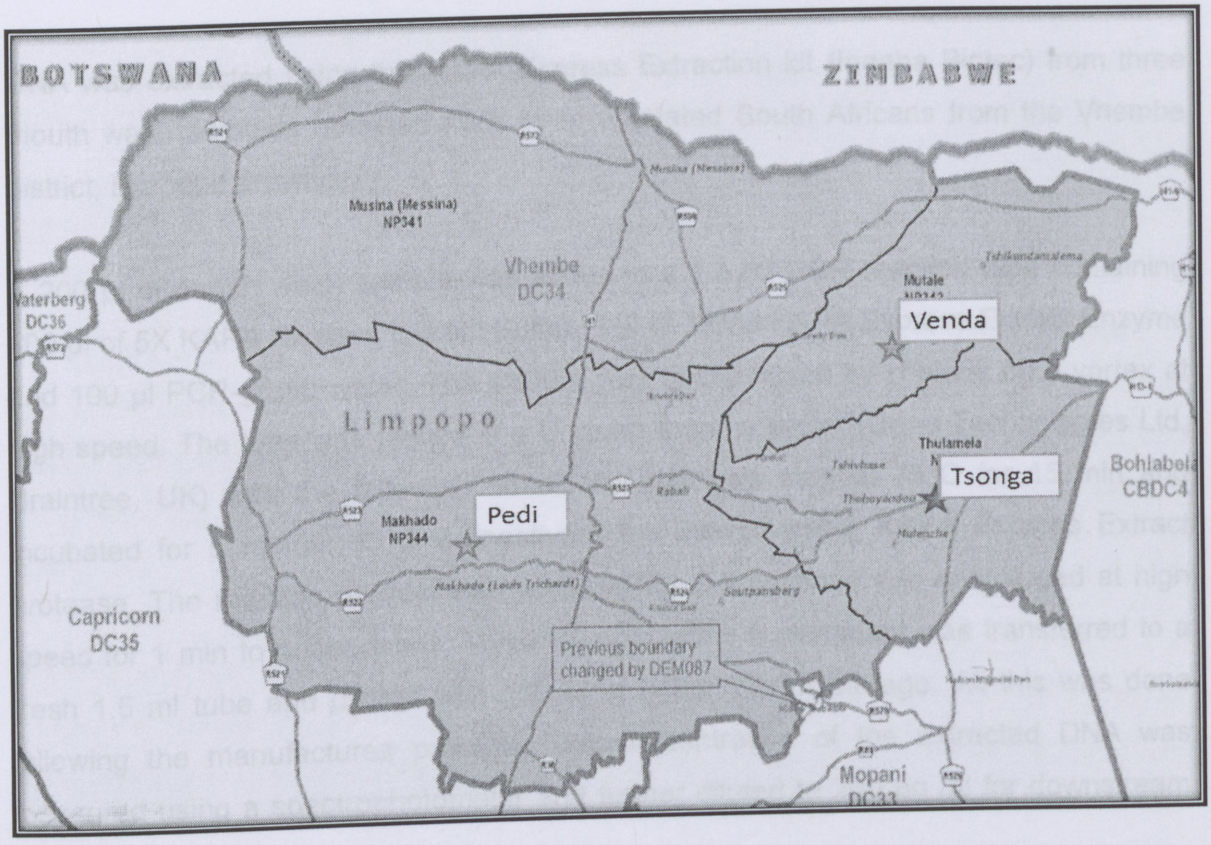


Figure 2: Study area with sites of sample collections for Venda, Tsonga and Pedi individuals.(http://commons.wikimedia.org/wiki/File:Map_No._33_of_Schedule_2_to_Notice_1998_of_2005.png).

2.3.2. Laboratory steps involved in library design

Microsatellite library construction was done through eight steps which are discussed in detail below. All recipe master mix reactions were scaled down for one reaction and multiplied by number of reaction tubes (in this case, three reactions).

© University of Venda

University of Venda
Creating Future Leaders

I. Extraction of high molecular weight DNA from tissue

DNA was extracted using the KAPA Express Extraction kit (Inqaba Biotec) from three mouth wash samples obtained from three unrelated South Africans from the Vhembe district, Limpopo province.

A 200 µl of mouth wash sample was added in a 1.5 ml PCR reaction tube containing 100 µl of 5X KAPA Express Extract Buffer, 2 µl of 1 U/µl KAPA Express Extract Enzyme and 100 µl PCR-grade water. The mixture was briefly mixed by shaking on a vortex at high speed. The tube was placed in a G-storm thermal cycler (Gene Technologies Ltd, Braintree, UK) with the following conditions: the lysis step at 75°C for 15 min and incubated for 5 min at 95°C to inactivate the thermo-stable KAPA Express Extract protease. The reaction product was vortexed for 2-3 seconds and centrifuged at high speed for 1 min to pellet debris. The DNA-containing supernatant was transferred to a fresh 1.5 ml tube and diluted with 150 µl of buffer TE for storage. All this was done following the manufactures protocol. The concentration of the extracted DNA was measured using a spectrophotometer and further diluted to 200 ng /µl for downstream applications.

Gel electrophoresis

A 1.5% agarose gel was used to visually assess the success of the DNA extractions. The gel was prepared as follows: 1.5 g of agarose was mixed with 100 ml 1X TAE buffer. The mixture was heated for approximately 3 minutes or less in a microwave depending on the dissolvability of gel particles following which the solution was cooled by placing the container under a running tap water. Following cooling, 3 µl of 0.5 µg/ml ethidium bromide staining solution was added and mixed thoroughly. The mixture was then casted into a gel-casting tray; a comb was placed in the gel and was left for 20 minutes at room temperature to solidify. The solidified gel was then placed in an electrophoresis chamber containing 1X TAE buffer (pH 8.0) wherein 4 µl of the sample's

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

University of Venda
Creating Future Leaders

genomic DNA extract was loaded. The gel was run at 100 V, 400 mA and 150 W for 40 minutes. Visualization was done under a UV trans-illuminator and pictures were taken. This procedure was repeated after extraction and all steps requiring visualization and validation of the presence of DNA. This study proceeded with only two of the best samples to later steps.

II. Restriction enzyme digestion

The purpose of this step was to fragment the long genomic DNA strand (clumsy to sequence) using restriction enzymes into fragments of approximately 300 to 500 base pairs, small enough to sequence yet possibly containing microsatellites. Two restriction enzymes (*RSa I* and *Bst U I*) were used to fragment genomic DNA of each sample briefly.

A master mix was prepared for each enzyme using the reagents and recipe below: The recipe was multiplied by the number of samples (two) plus one extra to cover for pipetting errors. Approximately 5 µl of the master mix was distributed into three 0.5 ml tubes following which 20 µl of the genomic DNA (200 ng/µL) was added to each tube containing the content of the master mix. Each cocktail was incubated at 37°C overnight. It should be noted that 10X ligase buffer was pre-heated to 70°C to ensure that the components are thoroughly dissolved not crystalized. This was done following the Glenn and Schable (2005) protocol.

III. Ligating linker to DNA fragments

During this step, double stranded DNA fragments from the previous step were ligated to double stranded linkers in both ends. All digested products from the previous step were to be processed in this step. The linkers SuperSNX24 forward and SuperSNX24+4P reverse were prepared prior to the master mix preparation.

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

The SNX oligonucleotide sequences are:

SuperSNX24 Forward: 5'GTTTAAGGCCTAGCTAGCAGAATC3'

SuperSNX24+4P Reverse: 5'pGATTCTGCTAGCTAGGCCTTAAACAAAA3'

Preparation of the Linkers

The preparation of the Super SNX linkers entailed mixing 25 µl of SuperSNX24 and SuperSNX24+4p primers with 1 µl of 5 M NaCl. The mixture was then heated at 95ºC for 2 min and slowly cooled down at room temperature.

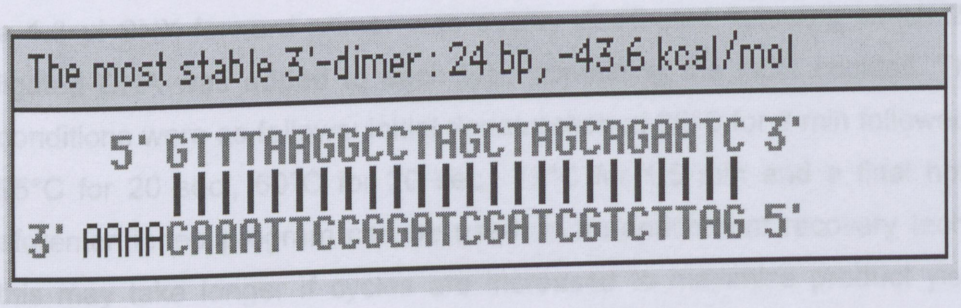


Figure 3: SuperSNX24 forward and SuperSNX24+4P reverse hybridized to form double strand DNA linker (Glenn and Schable, 2005).

Deployment of linkers to the DNA fragments

Following the Glenn and Schable (2005) protocol, 10 µl of the Linker-Ligation mix (7.0 µl ds SuperSNX linkers; 1.0 µl 10x Ligase Buffer; 2.0 µl T4 DNA ligase) was added to each of the digested DNA mixture from the previous step (restriction digests). This mixture was then incubated at 16°C overnight.With DNA ligation in progress, the success of DNA digestion for each enzymatic reaction in step (III) was assessed in 1% agarose gel electrophoresis.

© University of Venda

University of Venda
Creating Future Leaders

Verification of successful ligation

The main purpose of this step was to ensure that DNA was successfully ligated to the SNX linkers. SuperSNX24 Forward: 5'GTTTAAGGCCTAGCTAGCAGAATC3' primer was used in PCR to amplify all fragments ligated to the SNX linker to verify the success of the ligation reaction. Approximately 2 µl of the ligation products was used for this reaction. In the event wherein the original DNA amount is small (< 2 µg), PCR products were enriched using specific program to amplify the targeted fragments to make them abundant.

Following the Glenn and Schable (2005) protocol, In two 0.2 ml PCR tubes, a 23 µl cocktail containing 11 µl Kapa express ready mix taq, 2.5 µl BSA, 8.7 µl PCR water and a 1.3 µl SNX forward primer was evenly distributed following which 2 µl of the linker ligated DNA was added to each tube containing the PCR cocktail. The PCR thermal conditions were as follows: initial denaturation at 95°C for 2 min followed by 20 cycles of 95°C for 20 sec., 60°C for 20 sec., 72°C for 1.5 min and a final hold at 15°C. The aforementioned program can be used as an enrichment recovery technique, although this may take longer if cycles are increased to maximize product yield. A 4 µl PCR product was run on a 1.5% agarose gel to assess ligation success.

IV. Enrichment of microsatellite DNA-linker with biotin-oligos and dynabeads

The purpose of this step was to capture DNA fragments containing microsatellite repeats using microsatellite biotinylated probes and magnetic beads (Streptavidin M-280 dynabeads; Invitrogen; catalog No. 112-05D) at the same time eliminating those without such repeats. The principle of this step involved use magnetic beads to attach to biotinylated probes which were hybridized to the complimentary microsatellite sequence on the DNA fragments. Because magnetic beads are attached to the fragments with microsatellites, magnet was used to separate fragments containing microsatellites from those without. At this stage, the type of repeat to be isolated was decided upon; in this

© University of Venda

case trinucleotide and tetranucleotides were slected. Sequence information of the type of repeats which were intended to be isolated in this project can be found in Table 1. Note: Tri and tetranucleotides were selected because their reliability compared to mono and di nucleotide repeats. This is because trinucleotide and tetranucleotides repeats are most likely to appear in coding regions because they do not cause a frame shift (Tóth et al., 2000). Mononucleotide repeats are less reliable because of problems with amplification. Hence, the longer the microsatellites repeat, the more reliable it is.

Microsatellite probe cocktail preparation

A probe cocktail was prepared by mixing 1.7 µl of each of the probes in Table 1 in a 0.2 ml PCR reaction tube. Probes were synthesized by the Integrated DNA Technology (IDT) through White Head Scientific (Whitehead Scientific, Cape Town, South Africa).

Table 2: Oligonucleotides selected for hybridization with microsatellite DNA

Tri's	Tetra's
AAT (12 bp)	CACA (6 bp)
AGC (6 bp)	AAGT (8 bp)
AAG (8 bp)	CACC (6 bp)

Enriching linker legated DNA

In line with the Glenn and Schable (2005) protocol, the linker ligated DNA from step (IV) was enriched in a 0.2 ml PCR tube reaction mix comprising of 25 µl of 2X hybridization solution (warmed to ensure that the solution is completely dissolved), 10 µl biotinylated microsatellite probes, 10 µl linker ligated DNA from previous step and 5 µl ddH2O. See Appendix B for preparations of hybridization and washing solutions.

© University of Venda

Microsatellite repeats contained in the linker ligated DNA was allowed to hybridize with their complimentary probes (see Table 1) using the thermal cycling conditions by Glenn and Schable (2005) below:

Denaturation at 95°C for 5 minutes, a quick ramp to 70°C, followed by a step down at 0.2°C every 5 seconds for 99 cycles (i.e., 70°C for 5 sec., 68.8°C for 5 sec., 68.6°C for 5 sec. down to 50.2°C), and at 50°C for 10 minutes. Further ramp down which involved 0.5°C which decreased in every 5 seconds for 20 cycles (i.e., 50°C for 5 sec., 49.5°C for 5 sec., 49°C for 5 sec,. down to 40°C), and finally cascade to 15°C. this was as described by Glenn and Schable (2005)

Preparation of the dynabeads

While the ligated DNA and probe mixture is in the thermal cycler, 50 µl of the dynabeads was washed as described below as:

Fifty microliters of dynabeads was transferred into a 1.5 ml tube after which 250 µl of buffer TE was added. The mixture was shaken and spun down. The beads were captured using the Magnetic Particle Collecting (MPC) unit and buffer TE was removed using a pipette without touching the beads. These steps were repeated twice to ensure ultimate purity of the beads. Finally, the beads were re-suspended in 150 µl of 1X Hybridization Solution.

DNA probe mixture wash-up

In this step, microsatellite containing fragments were captured using a magnet while those without were filtered through.

According to Glenn and Schable (2005), the procedure involved the addition of DNA probe mix from step III to the washed beads in step IV. The mixture was incubated in a slowly rotating orbital set at room temperature (22°C). The beads were captured using the MPC unit following which the supernatant were removed with a P 200 pipette. The

© University of Venda

dynabeads were then re-suspended in 400 µl of 2X SSC, 0.1%, vortex and spun down to wash them. The beads were captured with the MPC unit and the supernatant was removed using P 200 pipette tips with wide base. The washing step was repeated for the second wash followed by washing twice with 400 µl of 1X SSC, 0.1% SDS at 45 or 50°C. Two more final washes were done with 1X SSC, 0.1% SDS while heating the solution at 50°C. Two hundred microliter of buffer TLE was added and the mixture was vortexed and incubated at 95°C for 5 minutes. The beads were then captured and the supernatant was transferred into a new tube immediately after the incubation. At this stage, the supernatant contained the enriched fragments. A total of 23 µl of 3 M NaOAc was added to the supernatant and mixed by flicking the tube. This was followed by adding 444 µl of 95% EtOH and mixing, by inverting the tube and placing it on ice for 15+ min. (or -20°C freezer for as long as necessary).

The tube was centrifuged at full speed for 10 minutes and the supernatant was discarded followed by adding about 0.5 ml of 70% EtOH and centrifuging for 1 min. The supernatant was carefully pipetted out and the pellet was air dried. The remainder EtOH was spun and dried until there was no trace (smell) of it left. The pellet was then re-suspended in 25 µl of buffer TLE. This was the "pure-gold" microsatellite containing DNA which was hydrated before use in PCR.

V. PCR recovery of enriched DNA fragments with microsatellites

The purpose of this step was to increase the amount of "pure-gold" DNA. This was achieved by using 2 µl of "pure-gold" DNA from the previous step as PCR template.

A PCR master mix containing 2.5 µl 10X PCR buffer, 2.5 µl BSA, 2 µl 25 mM $MgCl_2$, 1.5 µl 10 nM dNTP's, 1.30 µl 10 µM Super SNX-Forward primer, 0.20 µl *Taq* polymerase (5u/ µl) and 13.0 µl dH_2O was prepared since we had tri and tretra nucleotide 'pure gold' enriched DNA product to process further. PCR thermal cycling parameters were as follows: 95°C for 2 min initial denaturation; followed by 25 cycles of 95°C for 20 sec., 60°C for 20 sec., 72°C for 1.5 min and 72°C for 30 min, then held at 15°C.

© University of Venda

A 1% agarose gel was prepared to asses successfully DNA recovery by running 4 µl of the PCR product against a 100 bp ladder. Note: we anticipated a smear of fragments centered at the 500 bp region.

VI. Attachment of enriched DNA fragments into plasmids vector

The PCR products from the previous step were cloned in bacterial cell in order to recover microsatellite repeat sequences. Cloning was done using Clone JET PCR Cloning Kit (Thermo Fisher Scientific Inc. Gauteng, South Africa). It involved preparation of blunting reaction mixture, ligation of the "pure gold" PCR product, inserting the ligated product into bacterial cells and growing bacterial cell which was used as DNA source for subsequent DNA recovery PCR.

A master mix containing 10 µl 2X reaction buffer, 1 µl PCR product/other sticky-end DNA fragment, 6 µl ddH_2O and 1 µl DNA blunting enzyme was prepared on ice, vortexed briefly and centrifuged for 3-5 seconds, incubated at 70°C for 5 min and chilled on ice.

Ligation

The pJET1.2 /blunt Cloning Vector (50 ng/µl) and T4 DNA Ligase (Thermo Fisher Scientific Inc. Gauteng, South Africa) was added to the above mixture. The mixture was vortexed briefly and centrifuged for 3-5 seconds. Finally, the ligation mixture was incubated at room temperature (22°C) for 5 min. Note: cloning was repeated using TOPO cloning kit (Invitrogen, USA) following the manufactures protocol to produce high output clones.

© University of Venda

University of Venda
Creating Future Leaders

Transformation (cloning)

Approximately 2 µl of the ligated DNA was added to a vial (pink screw cap vial that comes with the kit) containing 50 µl DHα competent bacterial cells. The mixture was incubated on ice for 30 minutes, heat shocked in a water bath set at 42°C for 30 seconds and put back on ice immediately. A total of 250 µl of SOC media was added and then incubated at 37°C in a rotating oven for 1 hour.

Growing bacterial cells

Sixty microliters of the transformed bacterial cells was grown on an ampicillin selective media containing 40 µl X-Gal. A flaming hockey stick was cooled down and used to evenly spread the bacterial cells in 4 separate plates where 2 contained tri and 2 contained tetranucleotides. The plates were incubated overnight at 37 °C. The remaining bacterial cells were also plated on media for each enriched DNA.

VII. DNA recovery from the plasmid vector by colony PCR

In this step, cloned DNA was recovered by using colonies from the media as DNA templates for PCR.

A master mix containing 10 µl of 2X PCR ready mix Taq buffer containing all components except primer and DNA, 0.5 µl of 10 µM pJET1.2 forward sequencing primer, 0.5 µl of 10 µM pJET1.2 reverse sequencing primer and 9 µl ddH_2O was prepared.

In a 96-wells reaction plate, 25 µl of the mix was evenly distributed in 45 wells. A single colony was used as DNA template in each well containing the content of the mix and was amplified with the following cycle conditions: 95 °C for 2 min initial denaturation, followed by 25 cycles of 95 °C for 20 sec., 60 °C for 20 sec., 72 °C for 1.5 min; then 72 °C for 30 min and then hold at 15 °C.

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

The PCR products were run on a 1% agarose gel to assess clones with inserts.

VIII. Cycle sequencing of positive clones

The purpose of this step was to obtain sequence information of the positive clones which contain microsatellites. Primers were designed from the flanking regions of the microsatellites. This procedure was carried out using a thermal cycler machine.

A sequence reaction containing 1 µl big dye, 3 µl big dye buffer, 0.6 µl M13 forward primer and 5.6 µl PCR product was prepared. The reaction was carried out using the following conditions: initial denaturation of 96 °C for 1 min 25 cycles of (96 °C for 10 sec., 50 °C for 5 sec., and 60 °C for 4 min).

Cycle sequence products were purified using EtOH/EDTA precipitation method. Sequences were generated by the ABI 3730xl® DNA Analyzers (Applied Biosystems) for capillary electrophoresis and fluorescent dye terminator detection at GENEWIZ (GENEWIZ, Inc., USA).

2.4. PRIMER DESIGN

2.4.1. Introduction

Studies have shown that degraded DNA is not easy to analyze using the CODIS system (Abrahams and Benjeddou, 2012). Shorter STR markers have been developed and used for amplification and genotyping of degraded or compromised DNA samples (Butler et al., 2003; Coble and Butler, 2005). However these markers have to be designed from the flanking regions of the studied genomic DNA. The Powerplex 16 commercial kit (CODIS system) (Asamura et al., 2007) is an STR genotyping kit that shows partial profiles for a majority of samples (Coble and Butler, 2005). In the present study, short microsatellites ranging from 200 to 300 bp were designed from the flanking

© University of Venda

regions of microsatellites marker detected from the sequences generated from positive clones using the web based Primer3 software (Rozen and Skaletsky, 2000).

2.4.2. Primer designing

The purpose of this step was to identify microsatellite primers from the sequences that were recovered in previous section, which were designed from the sequences that surround the repeat motifs known as 'flanking regions'. Factors such as: Dimer formation, self-complementarity, melting temperature, unique primers, internal stability was a focus point in development of successful microsatellite primers.

Selection of primers was dependent on the following conditions:

1) Primers are free from complementarity as far as possible at their 3' prime ends to reduce significant primer-dimer artifact formation which reduces greatly the efficiency and specificity of the PCR reaction.

2) Self-Complementarity: Individual primers should not exhibit self-complementarity especially at the 3' end as this will result in the formation of hairpins and internal primer extension which can impact the yield severely.

3) Melting Temperature: The G-C bond is a higher energy bond than the A-T bond. A higher % GC leads to a higher melting temperature in general. The GC/AT ratio of a PCR primer should be similar or higher than that of the amplified template. The difference in melting temperature between the template and primer should be minimized for higher efficiency. For a PCR product of ≤ 500 base pairs short (16-18 nucleotides) primers are preferred.

4) Unique Primers: It is important for the primer to uniquely hybridize at a single location.

5) Sequence will be selected at least, at 25-80 base pairs from the target sequence.

© University of Venda

University of Venda
Creating Future Leaders

2.4.3. Primer designing software for microsatellites

Primers were designed using Primer 3 software (Rozen and Skaletsky, 1998). Primer3 is the most commonly used, freely available software for designing primers and probe selection. It is less complicated and has a flexible input-output file format that makes it suitable to be used with ease for advanced primer design projects. Primer3 is freely available on several websites and is suitable for use in small-scale projects. Features supported by Primer3 are:

1) Flexibility in setting weights for various scores and penalties calculated in primer design; 2) Evaluation of an overall score for a primer pair and returns pairs in descending order of desirability; 3) Supports all major primer design constraints including excluded regions, included regions, melting temperature range, primer self-complementarity, primer-primer complementarity, mispriming against a mispriming library, PCR product length, primer length, GC content etc; 4) Selection of reaction parameters like salt concentration is allowed; 5) User-definable number of primer pairs to be returned; 5) Statistics for accepted and rejected sequences.

The best primer sequences were selected based on their specific complementarity to the target sequence, balanced GC content of 40%, average base length of 20 bp and PCR products. They were sent to IDT Whitehead scientific (Whitehead Scientific, Cape Town, South Africa) laboratory for synthesis.

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

2.5. RESULTS

2.5.1. Extraction of high molecular weight DNA from mouth wash samples

Both samples extracted yielded high molecular weight DNA at an average of 300-700 ng/µl concentration as measured by the spectrophotometer. The samples are visually represented on the gel image in Figure 4 below.

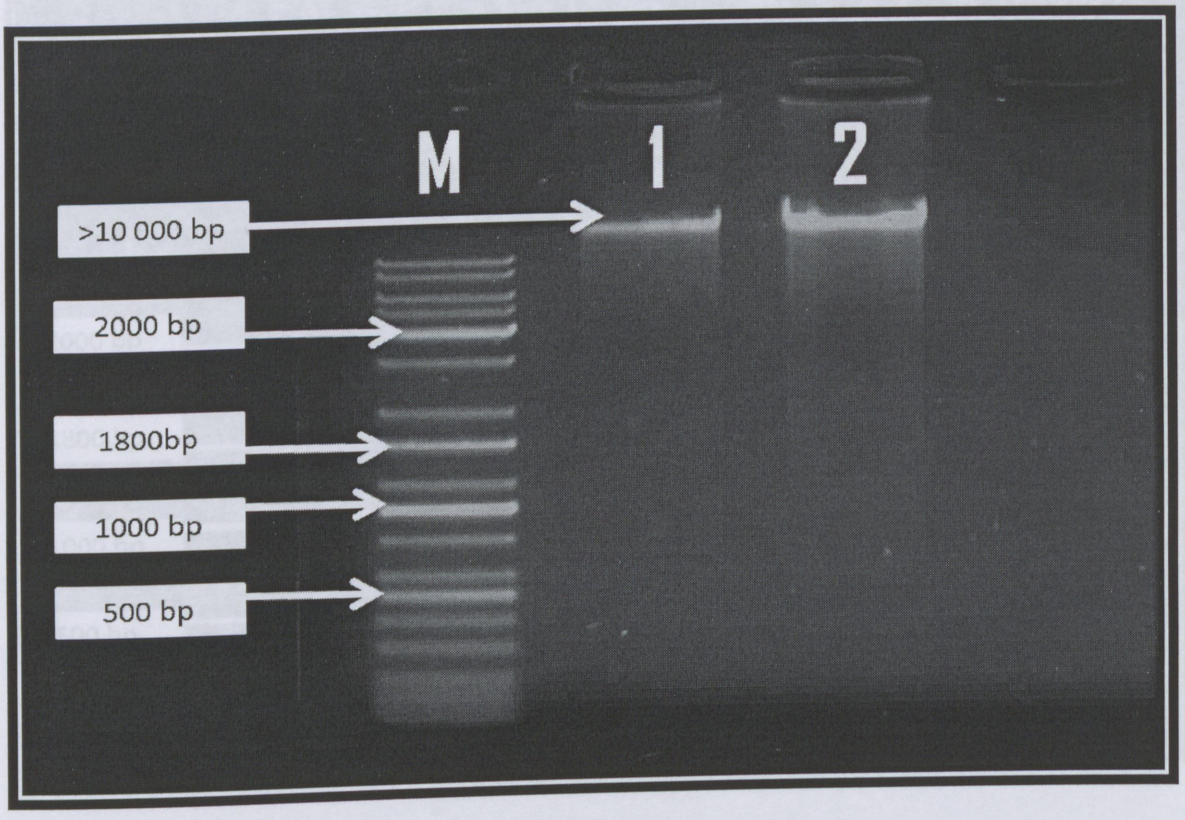


Figure 4: Representation of successful genomic DNA extraction from two unrelated individuals. Lane M represents a 1 kb molecular ladder; lane 1 and 2 represents genomic DNA from two different individuals respectively.

© University of Venda

2.5.2. Restriction enzyme digestion

Two restriction enzymes, *RSa I* and *Bst U I*, were used to digest the total genomic DNA of the two samples. Of the two enzymes, *RSa I* successfully digested the two genomic DNA samples whereas *Bst U I* proved unsuccessful. The cut samples are visually represented on the gel image in Figure 5 below.

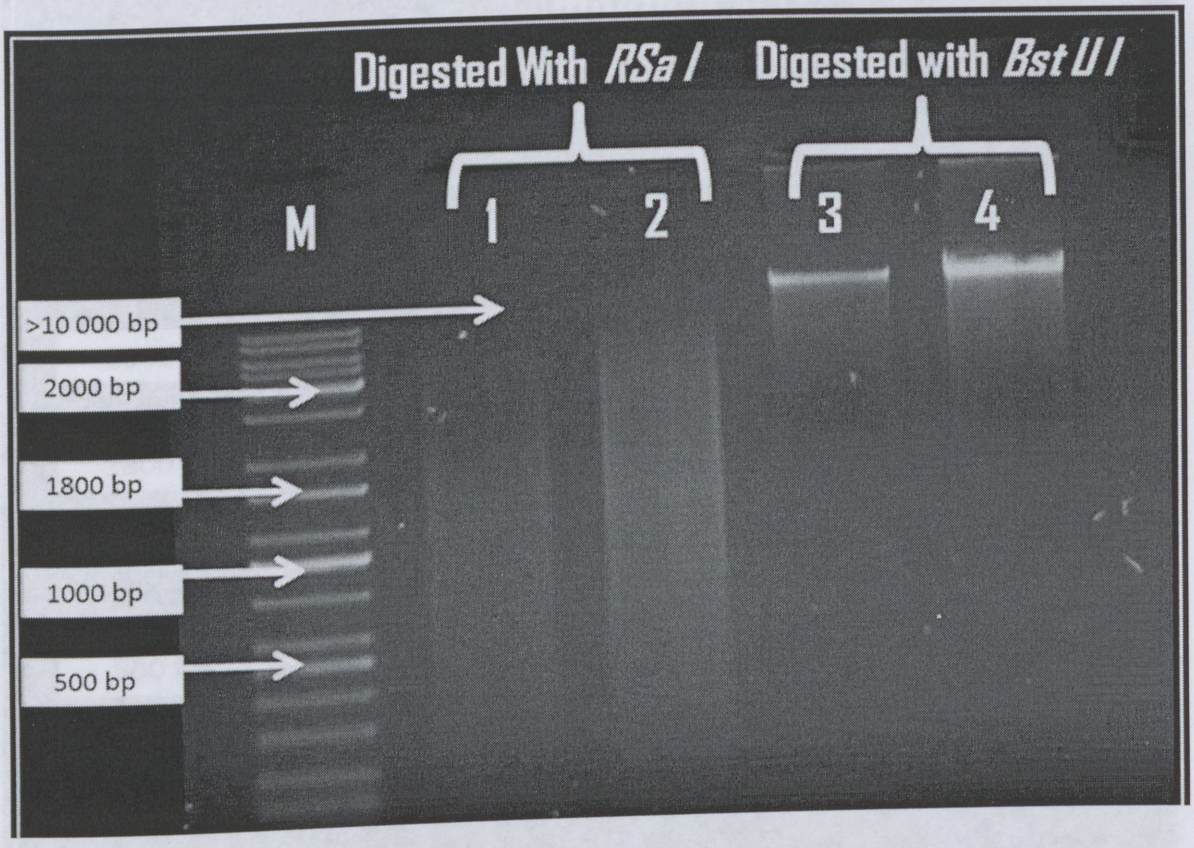


Figure 5: Representation of successful fragmentation of genomic DNA of both samples 1 and 2. Lane M represents a 1 kb molecular ladder; lane 1 and 2 represents two individuals digested with restriction enzyme *RSa I* respectively and lane 3 and 4 represents two individuals digested with *Bst U I* enzyme respectively.

© University of Venda

2.5.3. Linkers ligation to successfully fragmented DNA

Linkers were successfully attached to the fragmented DNA. They were quantified by PCR amplification for downstream steps. The samples are visually represented on the gel image in Figure 6 below.

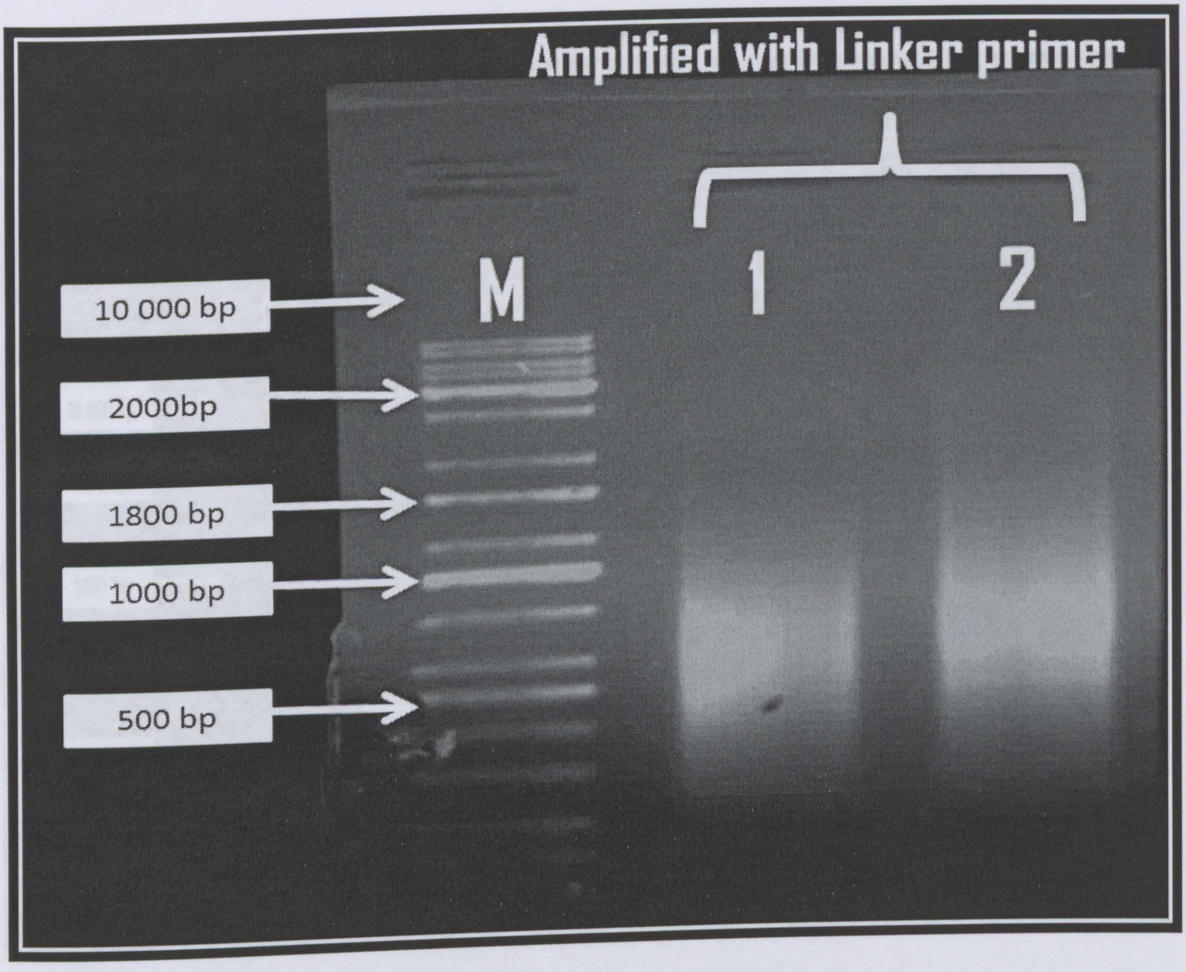


Figure 6: Representation of successfully amplified DNA linker ligation products with a continuous smear from ~300 to ≥1000 bp. Lane M represents a 1 kb molecular ladder; lane 1 and 2 represents two DNA linker ligation products.

© University of Venda

2.5.4. Enrichment of microsatellite DNA-linker with biotin-oligos and dynabeads

The DNA-linker ligated products were enriched successfully with tri nucleotides and tetranucleotides biotinylated probes. Tetranucleotide reaction yielded brighter bands compared to trinucleotide on a 1% agarose gel electrophoresis. The samples are visually represented on gel image in Figure 7.

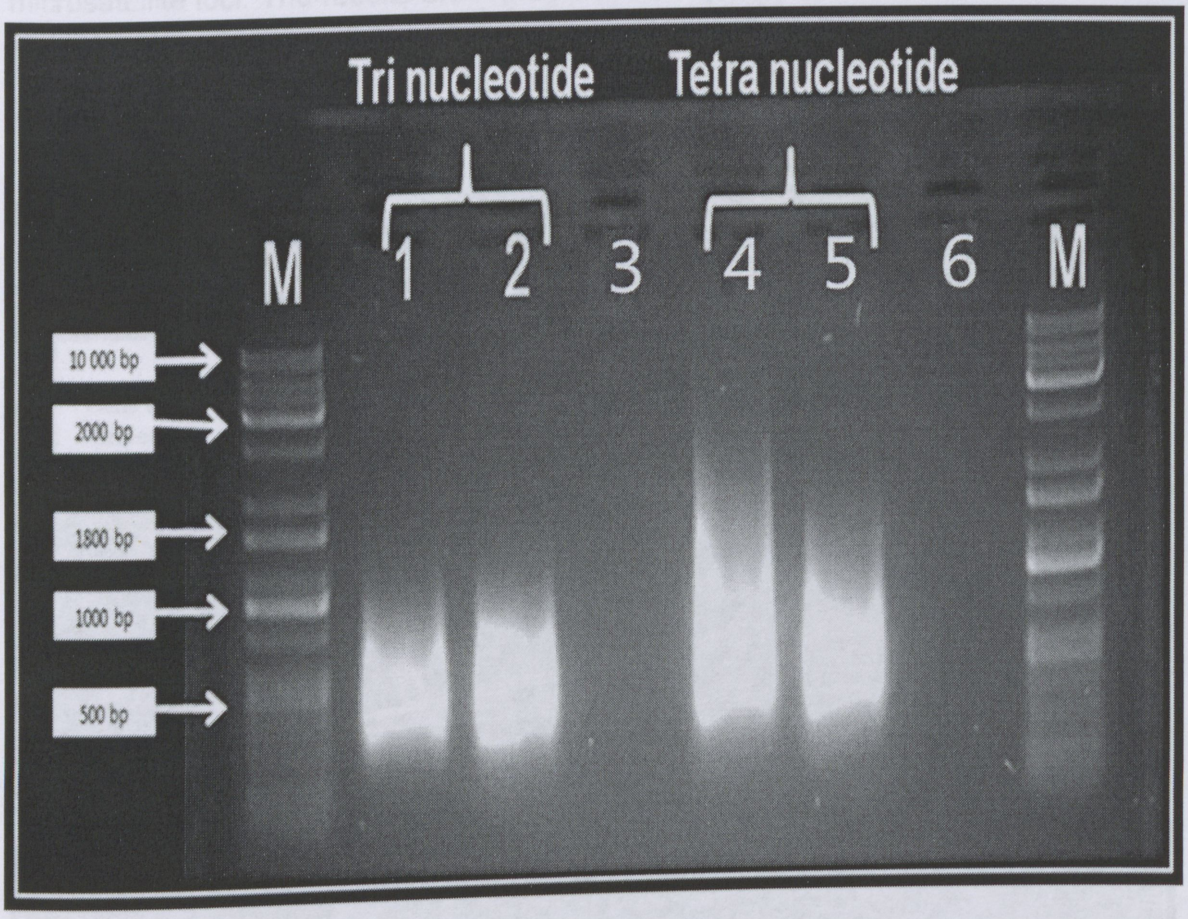


Figure 7: Representation of successful enrichments with tri and tetranucleotide probes. Lane M represents a 1 kb molecular ladder; lane 1 and 2 represents sample enriched for trinucleotides and lane 4 and 5 represents samples enriched for tetranucleotides; lane 3 and 6 represents negative control containing no DNA.

© University of Venda

2.5.5. Cloning and enriched DNA fragments to a plasmid vector and growing of transformed bacterial cells

Enriched microsatellite containing DNA was successfully cloned with TOPO plasmid vector (Invitrogen, USA) and grown in *E. coli* bacteria. The clones grew on a selective media containing X-Gal and ampicillin. About 80% of colonies contained the desired clones as they were white in color whilst 20% were blue and contained clones without microsatellite loci. The results are represented in Figure 8 below. Note: the Clone Jet kit resulted in low yield of clones.

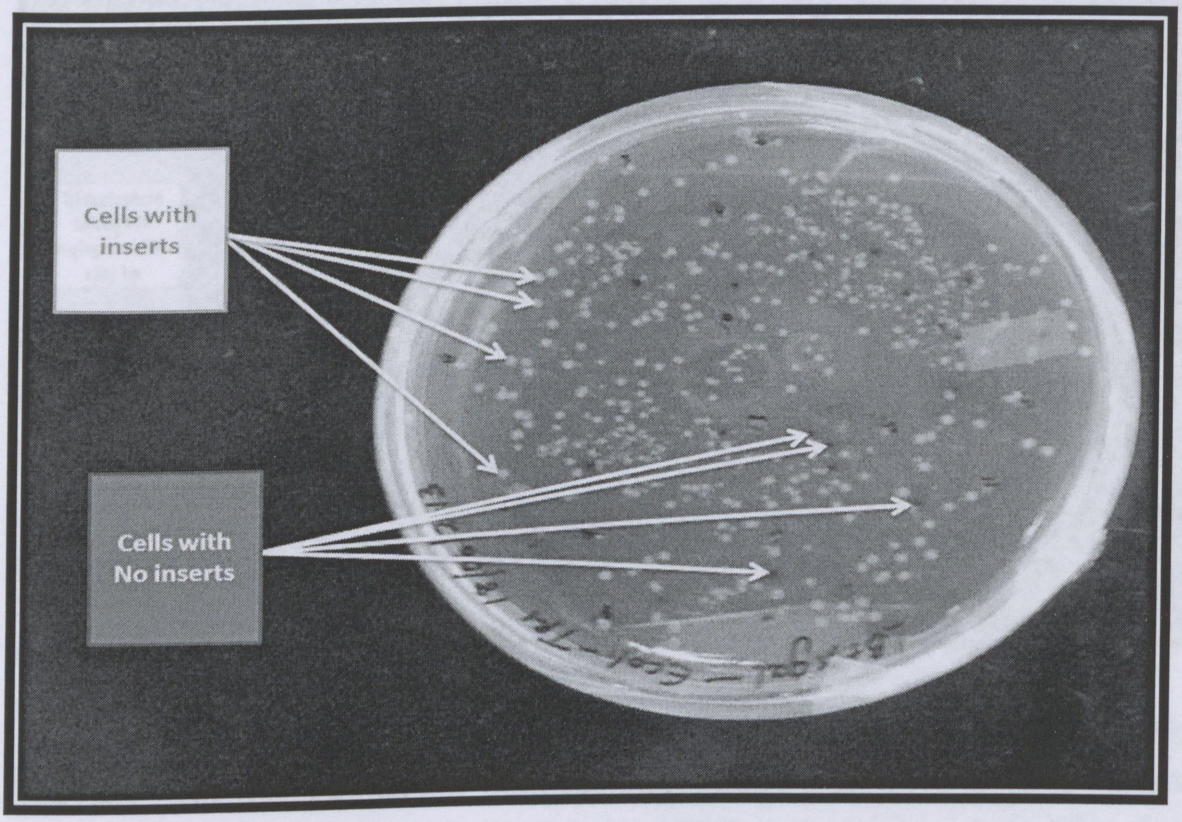


Figure 8: Image of *E. coli* grown on selective media containing ampicillin and X Gal. White spots represent colonies with inserts and blue colonies represent colonies with no true insert.

© University of Venda

2.5.6. DNA recovery from the plasmid vector using polymerase chain reaction

A total of 46 clones were amplified with PCR to recover both tri (lane 1-24) and tetranucleotides repeats (lane 25-46). Out of 46 amplified colonies, only 41 (89%) yielded positive results (Figure 9). Three (6.5%) clones: 10, 13 and 43 contained a double band which was due to picking of two clones at once (contamination of clones). The 5 clones: 31, 34, 38, 39 and 46 did not amplify adding up to a total of 8 unsuccessful clones out of 46 resulting in 83% success. The results are represented on the gel image below in Figure 9.

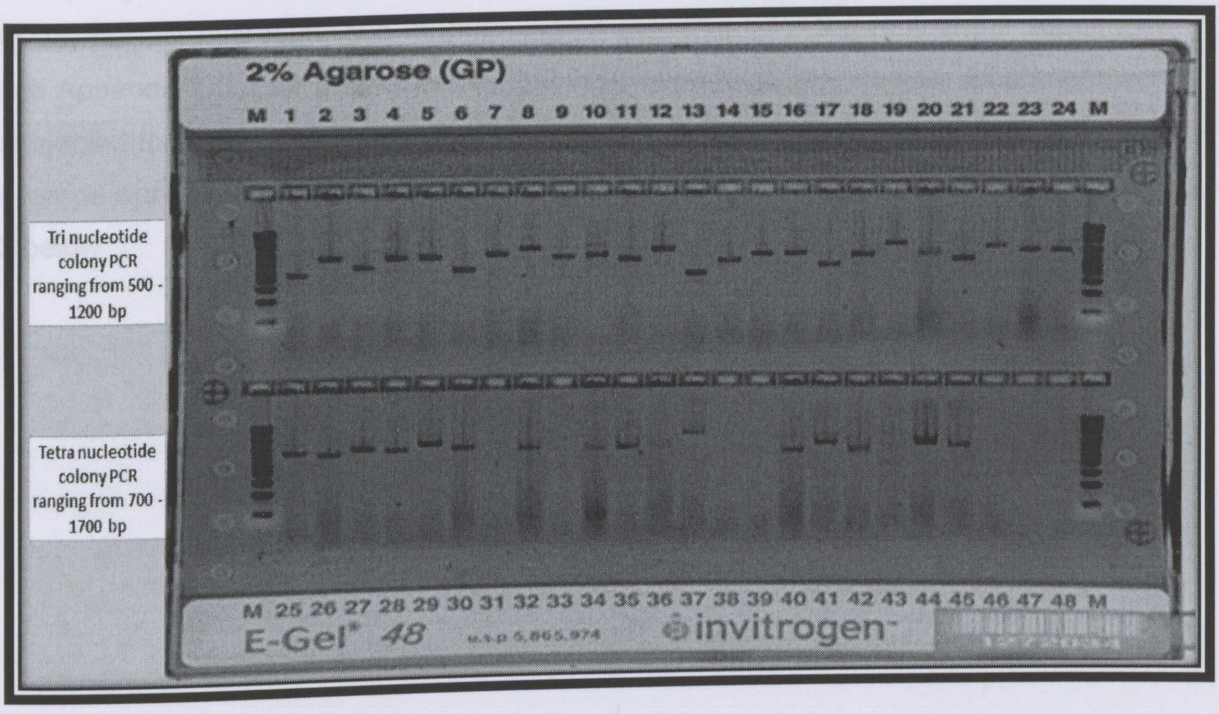


Figure 9: Representation of PCR results for 46 bacterial clones randomly selected. Top row represents trinucleotides enriched clones and bottom row represents tetranucleotides enriched clones. PCR products are run against a 1 kb molecular ladder on an E-gel.

© University of Venda

2.5.7. Primer design from sequences containing microsatellites

Sequences of the 38 successful amplified clones were obtained from GENEWIZ (GENEWIZ, Inc., USA) company and were aligned in BLAST (Basic Local Alignment Search Tool) to verify if they were human. All sequences were found be of human samples with 98% identity match and the sequences ranged from 470 bp to 1700 bp in length. Sequences containing a minimum of four repeat motifs were selected for primer design. However only one sequence containing a minimum of three was selected to assess its success. Some sequences contained microsatellites that where not enriched for and where selected if they meet the requirements (amplifying human DNA, containing minimum of three repeat motifs and a primer binding site). Only 39% (14 out of 39) perfect microsatellites markers where selected from the rest of the sequences, see Appendix D for all 14 sequences that were used for primer design. All primers were subjected to BLAST for specificity. Average size of the primers was 22 base pairs and average annealing temperature was at 59°C. The primers are represented in the Table 3 below.

© University of Venda

Table 3: Microsatellite primers selected for use in genetic testing

Locus	Repeat Motif	Chromosome	Primer sequence	Expected band size range	Annealing Temp in °C
GAD01	GGAA(17)	4	F-cctgggcaacagagagagac R-gacaccggggctgtttcta	240bp	60.5
GAD02	AAG (7)	3	F-cttcccctcaggacagtcag R-tggcctcattcttctcatcc	161-221bp	60.0
GAD03	TGG(5)	17	F-gttgttggccttcacctgag R-cctccattacgtccaccatc	223-324bp	60.5
GAD04	AAC(6)	7	F-ggcgaaaatgacacacacac R-atgcaggtcccacaaaactc	209-273bp	60.0
GAD05	AGC(17)	X	F-ttctctgattgcttagtggacaa R-tcattccgatggtgaaatcc	115-245bp	59.9
GAD06	TTGT (4)	1	F-ttgtgagcagggagtgtcag R-ctctgcctcccagagagacc	152-208bp	60.8
GDA07	GCT(10)	5	F-ctatctggccccaaatcaga R-tatggcaacctcacgtggta	209-273bp	60.0
GDA08	GT(13)	2	F-gggaaacttcccttccaaag R-acctgcagccagagtgatg	181-261bp	59.9
GDA09	AC(11)	X	F-ggcccgtaatccaatatgac R-atttgctgaagggatgtgagg	223-236bp	59.6
GDA10	AAT(4)	8	F-agaaggaggttgcagtgagc R-acacacccaggaaaagcact	207-223bp	59.6
GDA11	TTGT (4)	1	F-gacggaattcagcaggaact R-cctcccagagagaccctgta	182-247bp	59.3
GDA12	ATTG(5)	8	F-agcacagcagagatgctcaa R-ggaggaatttggtaaggcaga	159-223bp	60.0
GDA13	GAA(3)	6	F-tagctctccagccctctctg R-cttccaagcaggagctgact	229-309bp	59.8
GDA14	TATT (4)	3	F-aaggtgagggtccagtttca R-ttttctgaacaccaagcctgt	175-247bp	59.7

© University of Venda

University of Venda
Creating Future Leaders

2.6. DISCUSSION AND CONCLUSIONS

Results of this study confirm that designing a protocol for the isolation of human microsatellite at a low cost is achievable with less effort involved just as existing protocols; however, there is need to compare in details the expenditure compared to existing protocols. A total of 46 clones were screened from a library enriched with tri and tetranucleotides. A total of 14 potentially polymorphic microsatellite loci were recovered with annealing temperature ranging between 59-61°C, making them ideal for multiplexing. This technique is derived and modified from version of several published protocols, such as the one by Glenn and Schable, (2005) with an exception that it was optimized for a small library built and easy to be performed by genetics students in the University of Venda. Several steps were modified, this include incubation steps involved in PRC, restriction digest, colony PCR waiting periods. Because of this, the protocol does well by cutting time spent without compromising results. This only required six days to get the library designed and ready for sequencing.

As previously demonstrated in other studies, a library can be designed in one week (Glenn and Schable, 2005). This makes it ideal for rapid isolation of microsatellites across a variety of taxa for different applications. Our modified protocol also proved efficient in terms of time, with only five days for a library design, hence suitable for rapid isolation of microsatellites for different taxa for different applications. With components such as the *Xmn I* enzyme reported to be crucial for microsatellite library design (Glenn and Schable, 2005), we demonstrate that the library can be built without it and counteracted by amplification of the library to increase probability of hybridization to DNA template complementary to probes. This enzyme (*Xmn I*) is believed to prevent dimerization (self-ligation) of linkers, thus vital for success of the experiment (Glenn and Schable, 2005). In this project, this was bypassed by amplifying the linkers attached to microsatellite containing DNA. However, this can result in reduction of diversity of microsatellites if the DNA fragments are not numerous in the reaction, because amplification will be biased and amplify one type of microsatellite containing DNA

© University of Venda

University of Venda
Creating Future Leaders

attached to a linker. This nevertheless reduces the costs without compromising the efficiency of the protocol.

Being the first microsatellite isolation protocol conducted at the University of Venda, there were few technical challenges encountered which affected the percentage yield of potential polymorphic loci. Minor challenges were mainly due to a lack of equipment, such as a belly dancer for proper mixing of substances. Particularly in the hybridization of probes to the complementary DNA, a vortex machine was used to create low vibrations in a nearby surface were tubes containing probe-DNA mixer were placed. Though this seemed to have work, it may have reduced the binding power of the probes to the DNA.

Another limitation was encountered in cloning procedure. Cheaper cloning kit (Clone JET) cloning kit from Thermo Scientific in contrast to the commonly and widely used TOPO TA from Invitrogen which is a bit expensive, however affordable. This however resulted in low colony yield (~10%); hence we resorted to the TOPO cloning kit which improved our clone number significantly (~30%). It is suspected that linear vectors such as the Clone JET, do not work well with highly concentrated DNA. We recommend that diluted DNA from PCR products should be used with such kits to improve efficiency of the vector. It was also observed that the screening of the number of positive clones gets decreased when a larger volume of transformed bacteria are plated. This is due to the fact that, clones grow too close to each other, making it difficult to pick, and hence, reduces the number of colonies to be picked. Low amount of transformed bacteria improves the screening technique by producing less but distantly arranged bacterial clones that make them easy to pick, hence avoiding picking of more than one colony that will not be considered for cloning. This limitation however, does not affect our protocol in a drastic way; since other protocols reported only ~2% yield (Zane et al., 2002).

Out of 39 sequences containing microsatellite inserts, only 36% of the sequences contained perfect microsatellites and the remaining 64% of sequences was ignored to avoid incorrect primer designing. A total of 14 potentially polymorphic microsatellite loci

© University of Venda

were recovered with annealing temperature ranging between 59-61°C, making them suitable for multiplexing (Henegariu et al., 1997). The distribution of these markers was not across all chromosomes; however, 9 chromosomes were represented. No markers represented the male Y chromosome sequenced library; hence this was a major limitation in the use of these markers for paternity testing. Some studies say the "Y chromosome is such a superb tool for investigating recent human evolution from a male perspective and has specialized but important roles in medical and forensic genetics" (Jobling and Tyler-Smith, 2003). This makes the Y chromosome a critical marker to have in a microsatellite library intended for paternity and kinship analysis (Rolf et al., 2001). With little resources, the markers were tested for PCR amplification and their level of polymorphism within the three ethnic groups (Vhavenda, Tsonga and Bapedi). Testing and optimization is detailed in chapter 3.

© University of Venda

University of Venda
Creating Future Leaders

CHAPTER 3

OPTIMIZATION AND VALIDATION OF 14 MICROSATELLITE PRIMER PAIRS

3.1. INTRODUCTION

An important limitation regarding the use of microsatellites for polymorphism or genetic diversity studies is the prior need for optimization of PCR conditions for each microsatellite marker (Ogliari et al., 2000). Desired amplification may not be obtained using the recommended protocol due a variety of factors such as difference in types/brands of thermal cyclers, reaction components, or even minor differences in thickness of walls of PCR tubes (Dograr and Akkaya, 2001). It is of great importance that the primers be optimized for PCR. However, this takes a considerable amount of time and effort. Even after optimization of cycling conditions, some microsatellite markers may still amplify null alleles (amplifies more than a single locus or non specific) (Meglécz et al., 2004). Difficulties in optimization and characterization of microsatellite markers can result in of only a few well-resolved markers obtained (Nève and Meglécz, 2000).

Primers have to be optimized and tested for the following criteria: 1) Specificity: The primer must be specific and amplify one and only one product, meaning that the primers must anneal only at the location that is chosen for amplification. 2) Efficiency: The primer must have higher product yield efficiency, meaning that more quantity of product is produced in fewer cycles. 3) Fidelity: High fidelity means that there are a very low number of errors introduced by the DNA polymerase. Mostly this refers to the correct elongation of primers in the third phase of the PCR cycle.

© University of Venda

University of Venda
Creating Future Leaders

3.2. MATERIALS AND METHODS

3.2.1. PCR optimizations of 14 microsatellite loci

Fourteen microsatellite loci were optimized using the total genomic DNA from three individuals representing the three tribes (see Table 2). Details of how the genomic DNA was extracted from each sample and how they were collected can be found in Chapter 2. In order to avoid confusion, tribes were labelled as follows when conducting PCR reactions: A for Vhavenda, B for Tsonga and C for Bapedi. The PCR reactions for each locus were performed using KAPA 2G ready mix Taq polymerase kit from Inqaba Biotec.

The reactions were carried out in a 20 µl reaction volume containing 10 µl of KAPA 2G ready mix solution consisting of Taq polymerase (0.5 U/ µl), dNTP (0.2 mM), MgCl2 (1.5 mM), buffer (1X), 2 µl BSA (10X), 1 µl (20 pM) of each primer, 1 µl of 50 ng/µl DNA template, and 5 µl of nuclease-free H2O. A negative control containing no DNA was also prepared following the manufactures protocol.

The thermal cycling conditions were as follows: initial denaturation step at 96°C for 10 min followed by 35 cycles of 95°C for 30 seconds, annealing at 55 – 60°C (depending on primer pairs) for 45 seconds, extension at 72°C for 1 minute and final elongation at 72°C for 10 minutes. PCR reactions were then taken out of the thermal cycler immediately for further agarose gel assessment or stored at 4°C overnight.

Another reaction was carried out using 2X Taq readymix containing 2.0 mM $MgCl_2$ and annealing temperature was adjusted to 58°C for 45 seconds.

The 14 loci were also tested to confirm robust amplification in a multiplex reaction, using gradient PCR at annealing temperatures ranging from 55-60°C. Four multiplex reactions were designed and amplified based on the calculated respective annealing temperatures seen in Table 4.

© University of Venda

University of Venda
Creating Future Leaders

Table 4: Multiplex scheme for 14 microsatellite loci

Multiplex Combination		Tm in °C	Multiplex Combination		Tm in °C	Multiplex Combination		Tm in °C	Multiplex Combination		Tm in °C
Loci	Size		Loci	Size		Loci	Size		Loci	Size	
GAD05	245	59.0	GAD03	292	60.0	GAD09	236	59.3	GAD13	229	59.5
	245	60.7		292	61.1		236	59.2		229	59.8
GAD06	152	60.0	GAD01	240	59.8	GAD10	207	59.1	GAD14	175	59.8
	152	61.5		240	60.2		207	60.1		175	59.7
GAD07	209	60.0	GAD04	209	60.7	GAD11	182	59.6	-	-	-
	209	60.0		209	60.2		182	59.6	-	-	-
GAD08	181	59.9	GAD02	161	60.0	GAD12	159	59.9	-	-	-
	181	60.0		161	60.0		159	60.4	-	-	-

All PCR mixtures were incubated with the following cycling conditions: Pre-denaturation at 96°C for 10 minutes followed by 35 cycles of (denaturation at 95°C for 30 seconds, annealing at 55-60°C for 45 seconds and extension at 72°C for 1 minute), final elongation step at 72°C for 10 minutes and stored at 10°C until visualization.

All PCR products were visualized in a 2% agarose gel to assess their quality and specificity. GENETOOLS software Version 4.01 (Syngene systems) was used to analyze the gel, calculating the number and molecular weight of alleles. This software accurately counts the number of alleles based on number of bands, molecular weight of the bands and the quantity of the band (intensity of the band) on the gel. Data in a form of peaks was generated with each peak representing a band, hence an allele with reference to the molecular ladder used. Null alleles were eliminated using GENETOOLS (Syngene systems).

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

University of Venda
Creating Future Leaders

3.3. RESULTS

3.3.1 PCR products for 14 microsatellite markers at 1.5 mM $MgCl_2$ concentrations cycled at 58°C annealing temperature

Fourteen loci screened for specificity of primer binding and polymorphism at 1.5 mM $MgCl_2$ concentration and annealing temperature of 58°C. Of the tested loci, 13 yielded thick brighter bands of a range of specific band sizes (115 - 323 bp) whilst one locus (GAD01) proved unsuccessful (Figure 10). However, nonspecific faint bands were also observed in all 13 samples which yielded possible PCR product. The product size of non-specific band sizes ranged between 350 and1982 bp.

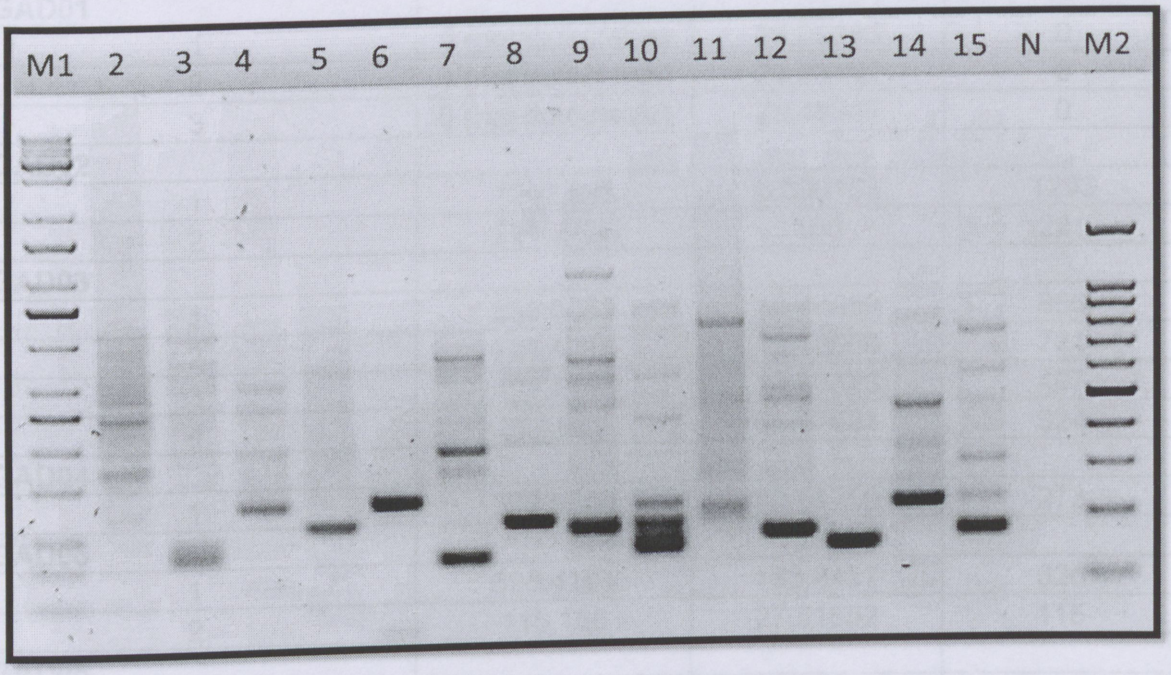


Figure 10: Representation of PCR products amplified with 14 microsatellite loci on a 2% agarose gel. M1 represents a 1 kb marker M2 represents a 100 bp marker, lane 2-15 represents loci GAD1-GAD14 PCR products. N represents a negative control without DNA.

© University of Venda

University of Venda
Creating Future Leaders

GENETOOLS (Syngene systems) gel analysis software detected one allele for locus 4 and locus 12 and two alleles for locus 2, locus 5, locus 7, locus 10 and locus 13. Conversely, three alleles were detected for locus 11 whereas four were observed for locus 3, locus 6, and locus 9. A total of 5 alleles where detected for locus 14. Allelic sizes detected for bands between loci ranged from 115 to 1982,271 bp as observed in Table 5.

Table 5: Gel analysis generated using GENETOOLS for 14 potential microsatellite markers

Number of alleles per Locus	Molecular weight	Quantity	Base pairs
GAD01			
1	0 (not detectable)	25,06355	0
2	0 (not detectable)	22,63192	0
3	0 (not detectable)	25,49349	0
GAD02			
1	1292,538	6,030153	1293
2	221,2725	100	221
GAD03			
1	856,0068	11,46659	856
2	721,4218	12,18069	721
3	496,7143	10,73383	497
4	323,882	66,97681	324
GAD04			
1	272,9599	80,94759	273
GAD05			
1	326,4101	183,4457	326
2	115,155	27,51552	115
GAD06			
1	1047,784	28,95814	1048
2	492,8672	65,66614	493
3	415,3765	20,47779	415
4	207,9283	140,8445	208
GAD07			
1	524,4977	7,874555	524

© University of Venda

2	272,9599	209,3297	273
GAD08			
1	1982,271	19,95802	1982
2	1000,027	20,09826	1000
3	869,4219	16,77671	869
4	693,9139	13,01677	694
5	260,5186	189,6679	261
GAD09			
1	876,2083	15,23852	876
2	622,3453	23,36061	622
3	311,5324	15,39007	312
4	222,9997	234,7512	223
GAD10			
1	1323,042	17,44135	1323
2	299,6536	55,48595	300
GAD11			
1	1168,278	19,67482	1168
2	710,29	32,62003	710
3	246,7183	220,6743	247
GAD12			
1	222,9997	207,9396	223
GAD13			
1	667,4549	40,53461	667
2	309,1195	172,729	309
GAD14			
1	1224,07	20,04821	1224
2	883,0475	12,47941	883
3	431,8427	24,27102	432
4	318,8844	25,46624	319
5	246,7183	161,7385	247

© University of Venda

GENETOOLS (Syngene systems) generated peaks based on band intensity and molecular weight with reference to a 1 kb molecular ladder. This is shown for all band patterns that appeared to be clear on the gel (Figure 11). For all 14 markers see Appendix C.

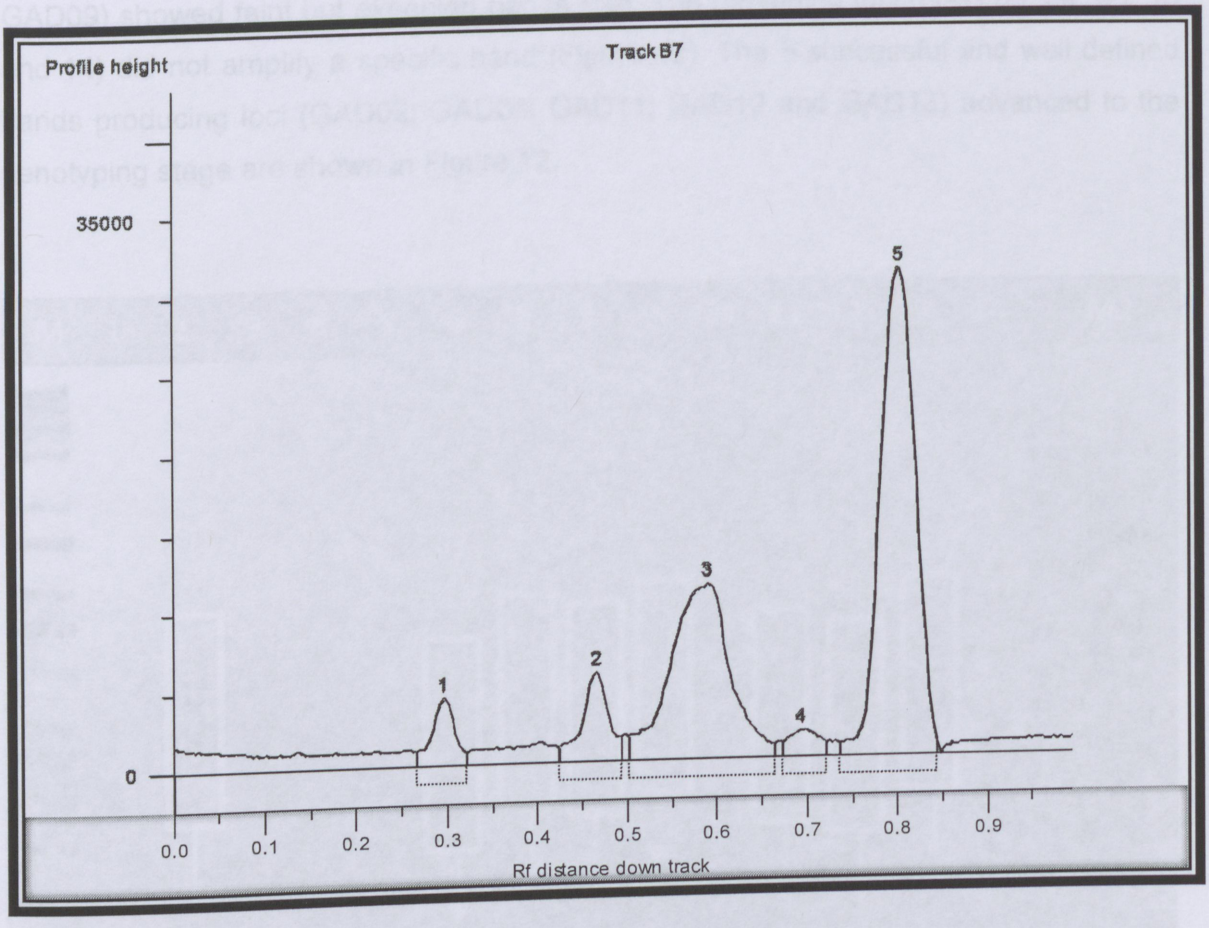


Figure 11: Representation of peaks detected by GENETOOLS software, peak numbers 1-5 represent the number of alleles produced by the respective marker, the highest peak 5 represents the brightest band on the gel hence the expected main band.

© University of Venda

3.3.2. PCR products for 14 microsatellite markers at 2 mM $MgCl_2$ concentration cycled at 58°C annealing temperature

At 2 mM $MgCl_2$ concentration with annealing temperature of 58°C, only 8/13 (62%) loci (GAD02; GAD06; GAD07; GAD08; GAD09; GAD11; GAD12 and GAD13) successfully amplified a specific target with one visible band. Three loci (GAD07; GAD08 and GAD09) showed faint but expected bands size. The remaining loci (GAD 03; 04; 05; 10 and 14) did not amplify a specific band (Figure 12). The 5 successful and well defined bands producing loci (GAD02; GAD06; GAD11; GAD12 and GAD13) advanced to the genotyping stage are shown in Figure 12.

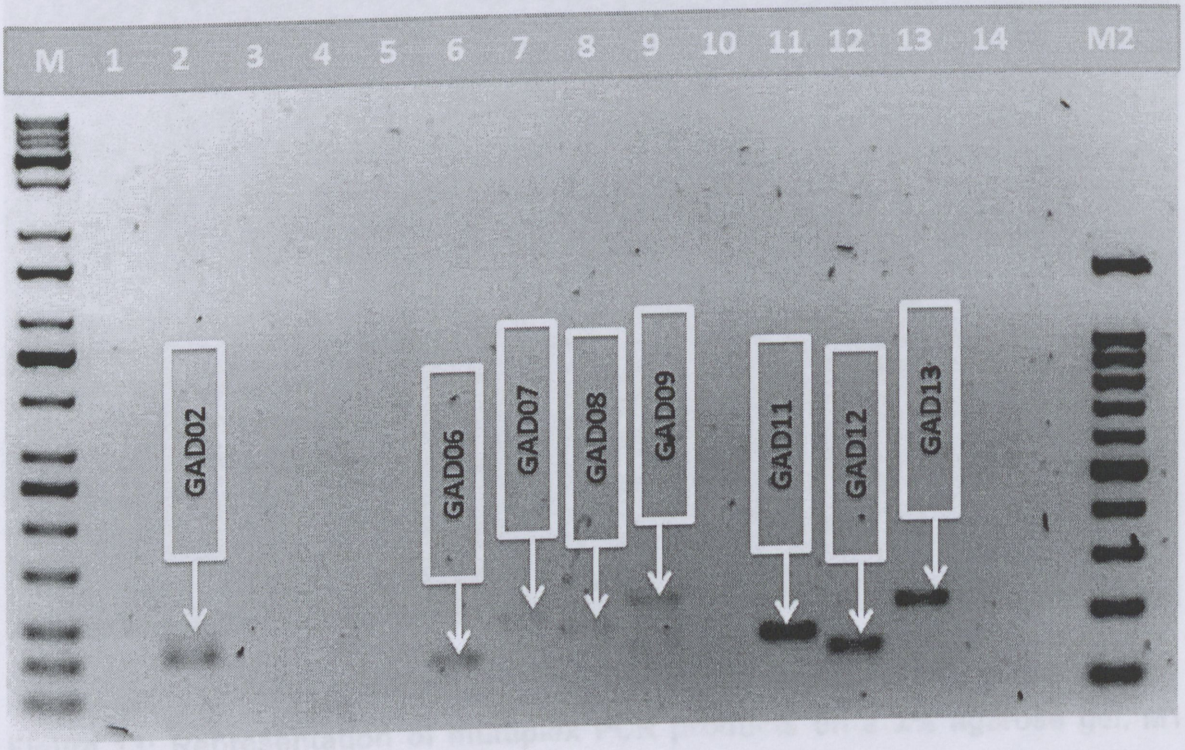


Figure 12: Representation of PCR products showing single band producing markers on a 2% agarose gel; M1 represents a 1 kb and M2 represents a 100 bp molecular ladder, lanes having a band represents PCR products and the assigned locus; N represents a negative control containing no DNA.

© University of Venda

3.3.3. Multiplex PCR products for 14 microsatellite markers at 2 mM $MgCl_2$ concentrations cycled at 58°C annealing temperature

Of the 4 multiplex reactions, only 2 yielded positive PCR results. However, not all alleles were amplified for multiplex. This is seen in Figure 13 below.

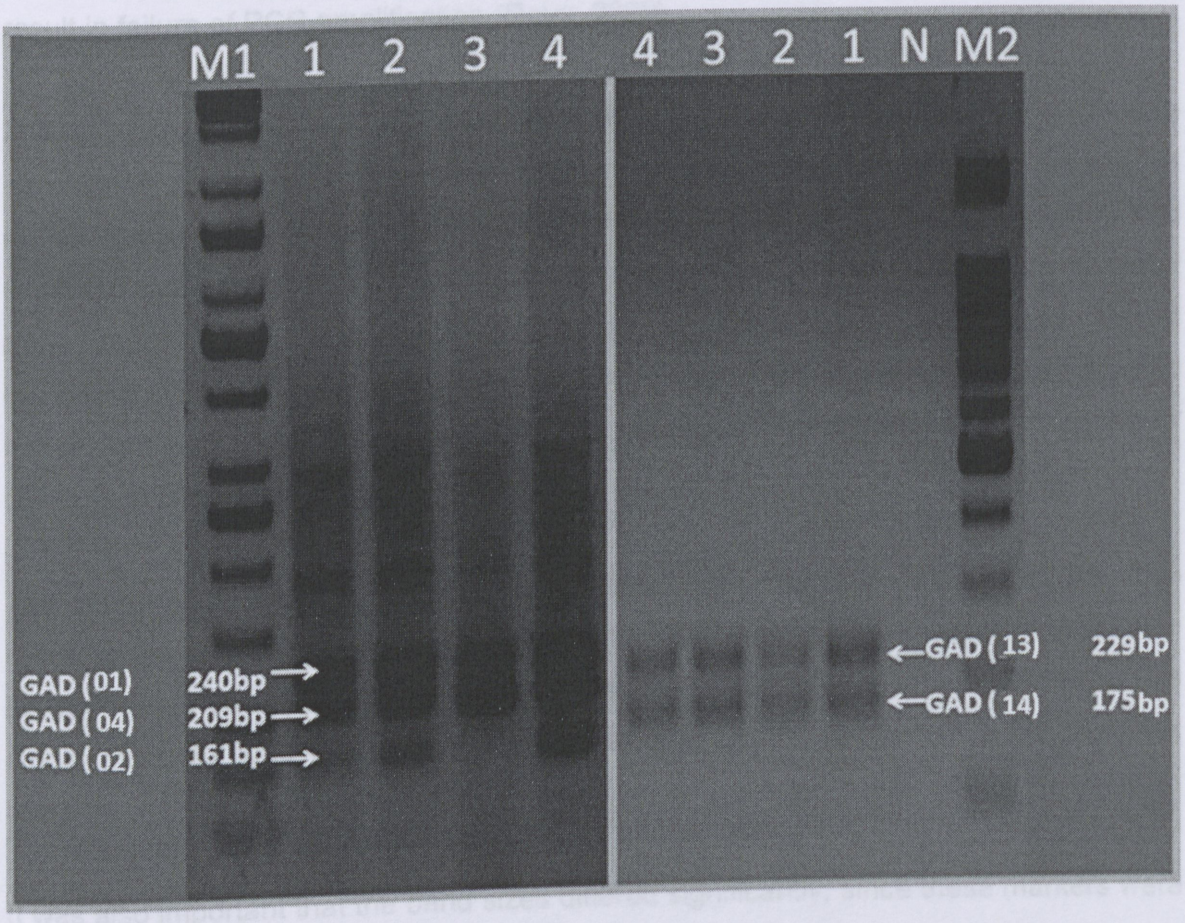


Figure 13: Representation of multiplex PCR products on a 2% agarose gel. M1 represents a 1 kb and M2 represents a 100 bp molecular ladder, N represent negative control with no DNA, Lane 1-4 represents PCR products amplified with indicated loci.

© University of Venda

3.4. DISCUSSION AND CONCLUSIONS

Microsatellite detection can become a challenging task when it comes to optimizing the primers for PCR amplification (Ogliari et al., 2000). This requires modifications and adjustments of few but critical parameters such as annealing temperatures, $MgCl_2$ concentration and DNA template concentration. Failure to proper adjustments may result in failure of PCR amplification (Roux, 2009)

In the present study, 14 potentially polymorphic microsatellite markers were successfully designed and optimized for PCR. All the markers amplified specific regions in PCR amplification except for marker GAD01 which did not show any specificity and expected size band. The 13 markers amplified at average, 3 alleles (Table 4) with major alleles ranging from 115 to 323 base pair. The results of this study confirm that the 14 loci designed for humans using the designed cost effective protocol, amplify specific regions in the genome, hence the protocol is working. This is similar to studies by Butler (2006) and Glenn and Schable (2005).

This study also managed to optimize to amplify all loci in a single reaction, thus reducing the cost. This is because it is important that there is no overlap in product size between the individual loci in the multiplex reaction (Masi et al., 2003). However in this study, multiplex was not achieved as by smeary bands encounter in the multiplex reaction.

It was also important that the band sizes differed significantly, since these markers were not fluorescently labeled. This however is not a major problem, because the difference in band size obtained were more than 20 bases apart, making it easy to differentiate using standard agarose gel electrophoresis.

© University of Venda

Limitations to these markers includes the fact that they can only be analyzed using a gel, hence the accuracy of band size relies on GENETOOLS software (Syngene systems), which relies on how far the band has travelled on the gel with reference to a molecular ladder. This is because they were not fluorescently labeled for analysis using an ABI genetic analyzer. In a gel were band patterns are closely similar due to the size of the STR, it would be difficult to tell if the band is correct or not. However, few markers can be selected for a multiplex reaction to eliminate this in genotyping. These markers can be used in genotyping of populations to understand the genetic structure of the studied population. Furthermore, the markers developed can be used I conjunction with already existing microsatellite markers in the market and from GenBank. This will further validate these markers for robust genetic analysis.

© University of Venda

University of Venda
Creating Future Leaders

CHAPTER 4

EVALUATION OF MICROSATELLITE PRIMERS ON A LIMITED COMMUNITY IN VHEMBE DISTRICT, LIMPOPO SOUTH AFRICA

4.1. INTRODUCTION

South Africa consists of a diversified human population mix which is a rich mosaic of distinctive minorities (Rammala, 2003). There are eleven official languages, with a majority of them (seven) spoken in the Limpopo Province: Northern Sotho (Sepedi), Xitsonga, Tshivenda, Setswana, Isindebele, Afrikaans and English. Population groups speaking these languages are scatted throughout Limpopo province with Northern Sotho (Sepedi) (52.7%), Xitsonga (22.6%), Tshivenda (15.5%) being the largest population groups respectively and only (9.2%) being the Setswana, Isindebele, Afrikaans and English population groups. The major (Sepedi, Xitsonga and Tshivenda population groups are believed to be related to the Lesotho, Mozambique and Zimbabwean populations respectively (Nxumalo, 2000; Rammala, 2003). This is possibly because of Botswana, Mozambique and Zimbabwe are the closest neighboring countries to the Limpopo province, hence migration is an important factor allowing gene flow between these population groups.

Population genetic data from countries of great ethnic and cultural diversity like South Africa are important world-wide (Nxumalo, 2000). The availability of such data from both indigenous and admixed populations is important for a number of human scientific disciplines which include: forensic genetics, genetic association studies, and anthropological studies (Cloete et al., 2010). In this study, a total of 13 autosomal microsatellite markers were characterized in three ethnic groups of the Limpopo province (Vhavenda, Bapedi and Tsonga). Under the hypothesis: individuals from the same province should be closely related, hence cluster together genetically.

© University of Venda

University of Venda
Creating Future Leaders

4.2. MATERIALS AND METHODS

4.2.1. Sample collection

Consent forms were provided to participants and samples were collected as previously described in chapter 2. A total of 14 mouth wash samples were collected for each ethnic group (Vhavenda, Tsongas and Bapedi); in total= 42 mouth wash samples. Samples were processed as previously described in chapter 2 to get the genomic DNA. There was no need for eliminating bacterial DNA since it had little or no effect in this analysis as previously described in chapter 2.

4.2.2. Genomic DNA extraction

DNA was extracted with KAPA Express Extraction kit following the manufacture's protocol as previously described in chapter 2, followed by 1% agarose gel electrophoresis to visually assess the DNA quality of the extracts.

4.2.3. Microsatellite screening among three ethnic groups (Vhavenda, Tsonga, Bapedi)

A total of 14 individuals per ethnic group were screened for polymorphism at five well defined autosomal sites. Polymerase chain reaction was performed using five autosomal microsatellite loci (GAD02; GAD06; GAD11; GAD12 and GAD13) that were successful in producing well defined bands during optimization from the library enrichments described in chapter 3. Screening was performed using the thermal cycling conditions described following conditions: A 20 µl PCR reaction mixture consisting of 5 µl of 20 ng genomic DNA and 10 µl of 2X KAPA ready mix (Inqaba Biotec) containing Taq polymerase (0.5 U/ µl), dNTP (0.2 mM), MgCl2 (2.0 mM), buffer (1X), with 1 µl of forward primer and 1 µl reverse primer (selected in chapter 3) for five loci and balanced with nuclease-free water. The PCR mixture was incubated with the following cycling conditions: Pre-denaturation at 96°C for 10 minutes, 35 cycles of (denaturation at 98°C

© University of Venda

University of Venda
Creating Future Leaders

for 25 seconds, annealing at 55-60°C for 45 seconds and extension at 72°C for 1 minute), final elongation step at 72°C for 10 minutes and stored at 10°C. All PCR products we run against a 1 kb molecular ladder on a 2% agarose gel.

4.2.4. Data analysis

Gel analysis was performed using GENETOOLS software (Syngene systems) and data was analyzed using Statistical Package for Social Science (SPSS, Inc., 2009, Chicago, IL, www.spss.com) version 21.0. A multidimensional unfolding analysis was performed to cluster individuals based on the number of alleles, size, and location on the gel. This was done under a 2 dimensional and linearized model for similarities across the studied population. Transformation of data was done for all sources simultaneously.

4.3. RESULTS

4.3.1. Gel electrophoresis for microsatellite genotyping across 3 ethnic groups

Majority of individuals at two loci GAD02 and GAD06 showed high similarity, with close (98%) similarities across all ethnic groups. Loci GAD11; GAD12 and GAD13 amplified different band patterns between the three ethnic groups represented by A, B and C which were Venda, Tsonga and Pedi respectively on the Figure 14 below. Some loci including GAD11, amplified heterozygous allele in some individuals, also shown in individual B (Tsonga) on Figure 14 below. Loci GAD12 and GAD13 were the most variable across all ethnic groups. Only major alleles were considered in the analysis of the gel.

© University of Venda

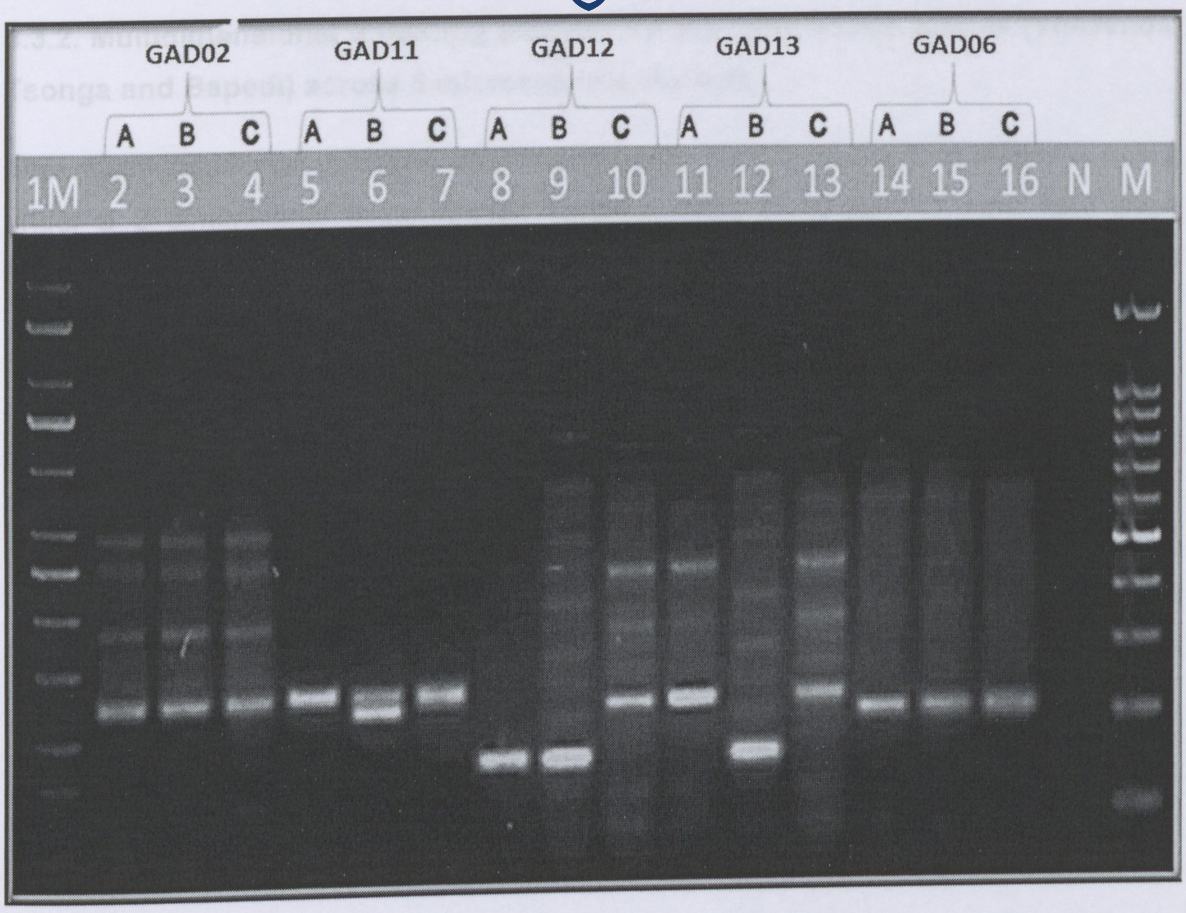


Figure 14: Representation of the 5 selected loci across three ethnic groups A, B and C (Vhavenda, Tsonga and Bapedi respectively) on a 2% agarose gel. 1M represents a 1 kb and M represents a 100 bp molecular ladder, N represents negative control; A, B and C represent Venda, Tsonga and Pedi representatives respectively. Loci GAD02; 11; 12; 13 and 06 are shown on the top of the gel, lane 2-16 are the number of individuals genotyped.

© University of Venda

University of Venda
Creating Future Leaders

4.3.2. Multidimensional unfolding analysis for the three ethnic groups (Vhavenda, Tsonga and Bapedi) across 5 microsatellite markers

The multidimensional unfolding analysis showed 7 (1-7) clusters at different points under a 2 dimensional linear model. Three clusters (5, 6 and 7) were very close together for all the ethnic groups. Cluster (1) showed that 3 individuals from a population were distanced far away from the rest of the clusters. The 5 microsatellite loci grouped all the individuals in one common cluster shown in the Figure 15 below.

© University of Venda

University of Venda
Creating Future Leaders

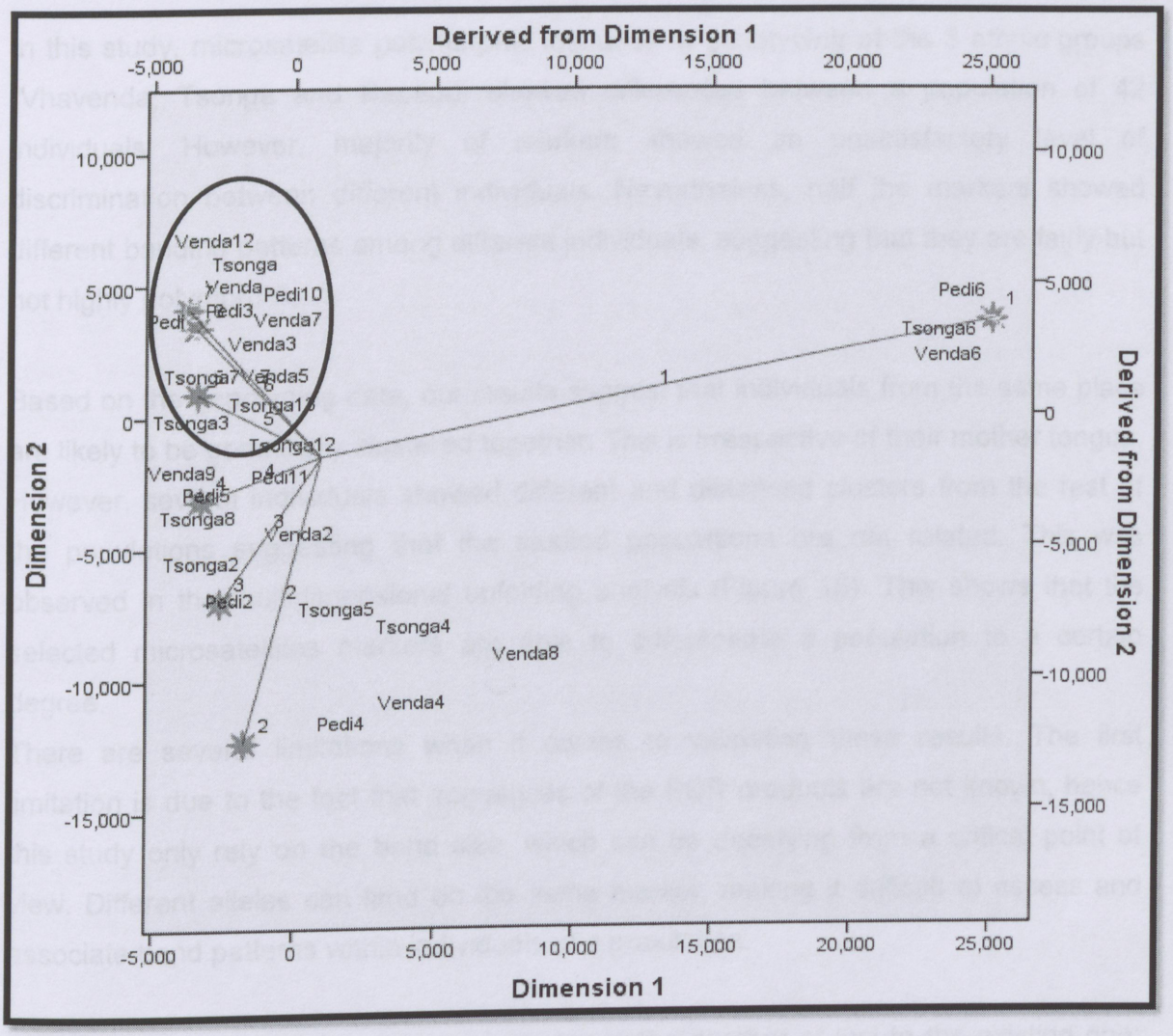


Figure 15: Multidimensional unfolding joint plot for 42 individuals (14 Venda, 14 Tsonga and 14 Pedi) across 5 microsatellites. Red circle represents point of major clustering of the individuals due to similarities of variables (band size, number of allele with reference to molecular marker.

© University of Venda

4.4. DISCUSSION AND CONCLUSIONS

In this study, microsatellite polymorphic loci used in genotyping of the 3 ethnic groups (Vhavenda, Tsonga and Bapedi), showed differences between a population of 42 individuals. However, majority of markers showed an unsatisfactory level of discrimination between different individuals. Nevertheless, half the markers showed different banding patterns among different individuals, suggesting that they are fairly but not highly polymorphic.

Based on the genotyping data, our results suggest that individuals from the same place are likely to be genetically clustered together. This is irrespective of their mother tongue. However, several individuals showed different and distanced clusters from the rest of the populations suggesting that the studied populations are not related. This was observed in the multidimensional unfolding analysis (Figure 15). This shows that the selected microsatellites markers are able to differentiate a population to a certain degree.

There are several limitations when it comes to validating these results. The first limitation is due to the fact that sequences of the PCR products are not known, hence this study only rely on the band size, which can be deceiving from a critical point of view. Different alleles can land on the same marker, making it difficult to assess and associate band patterns within individuals of a population.

This study however encourages an increase in the number of loci to the existing one: (GAD02; GAD06; GAD11; GAD12 and GAD13). This will significantly increase the discrimination power across individuals. Including the Y chromosome will also enable the markers to be used in a paternity testing. These markers can act as a stepping stone to continuous isolation of microsatellites markers for human genetic studies. The hypothesis of this study is therefore accepted since markers were isolated with low resources. However, only as few as 5 markers could be isolated.

© University of Venda

University of Venda
Creating Future Leaders

CHAPTER 5

RECOMMENDATIONS AND FUTURE APPROACHES

Microsatellites are categorised as one of the most used molecular markers, with wide range of applications (Dawson et al., 2013). However, isolating them involves two main challenges. They can either be expensive and/or time consuming based on equipment and other lab technicalities. With insufficient basic equipment, a library can be built but, there is a high probability of failure, especially if the species of interest is not well documented. However, in this study, we improvised to achieve a limited but working set of microsatellites primers.

The existing microsatellite markers designed can be further improved and optimized for higher resolution and discrimination power. Their use in a population study is more feasible at this point than their use in a paternity test. However, with more enrichments made, the discrimination power and level of polymorphism increase (Butler, 2006).

Technical challenges that include low harvest of polymorphic microsatellites can be resolved. In this study, blasting of sequences before designing the primers flanking the microsatellites is recommended. This is because; primer binding sites may have mutations that lead to non-specificity, hence failure of amplification (Butler, 2007). Reference sequences are also good for comparing primer binding sites.

For future studies, the designed protocol can be further modified to increase its success rate in microsatellites isolation. A larger library construction can be attempted following the existing protocol. Testing of the designed markers in a larger population can give a clear and conclusive description about the three ethnic groups in the Vhembe district, Limpopo province, South Africa. An attempted of this protocol in other different taxa is recommended to further explore its capabilities.

© University of Venda

University of Venda
Creating Future Leaders

REFERENCES

Abdelkrim, J., Robertson, B.C., Stanton, J.L. and Gemmell, N.J. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Journal of Biological Techniques* **46**(3):185-192.

Abrahams, Z. and Benjeddou, M. 2012. Population genetic data for six non-combined DNA index system (non-CODIS) miniSTR loci from the Xhosa and Cape Muslim populations of South Africa. *African Journal of Biotechnology* **11**(65):12798-12802.

Asamura, H., Fujimori, S., Ota, M. and Fukushima, H. 2007. MiniSTR multiplex systems based on non-CODIS loci for analysis of degraded DNA samples. *Forensic Science International* **173**(1):7-15.

Ashton, J., Dickson, K. and Pleaner, M. 2009. Evolution of the national adolescent-friendly clinic initiative in South Africa. In: *World Health Organization and Pathfinder International*. Geneva: World Health Organization.

Atuyambe, L., Mirembe, F., Annika, J., Kirumira, E.K. and Faxelid, E. 2009. Seeking safety and empathy: adolescent health seeking behavior during pregnancy and early motherhood in central Uganda. *Journal of Adolescence* **32**(4):781-796.

Budowle, B., Moretti, T.R., Niezgoda, S.J. and Brown, B.L. 1998. CODIS and PCR-based short tandem repeat loci: Law enforcement tools, P.73-88. In proceedings of the second European Symposium on Human Identification (June 1998, Innsbruck, Austria) Promega corporation, Madison,WI.

Bugert, P., Rink, G., Kemp, K. and Klüter, H. 2012. Blood group ABO genotyping in paternity testing. *Transfusion Medicine and Hemotherapy* **39**(3):182-186.

Butler, J.M. 2005. Forensic DNA Typing: biology, technology, and genetics of STR markers (2nd Edition). *Elsevier Academic Press*, New York, 688 pages.

© University of Venda

University of Venda
Creating Future Leaders

Butler, J.M. 2006. Genetics and genomics of core STR loci used in human identity testing. *Journal of Forensic Science* **51**(2):253-265.

Butler, J.M. 2007. STR and Molecular Biology Artifacts. Houston DNA training workshop.

Butler, J.M. and Hill, C.R. 2012. Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. *Forensic Science Review* **24**(1):15-26.

Butler, J.M., Shen, Y. and McCord, B.R 2003. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *Journal of Forensic Science* **48**(5):1054-1064.

Christianson, A.L., Jenkins, T. and Madolo, N. 1999. Report of the Subcommittee on Genetic Laboratory Services to the National Task Team on the Transformation of Laboratory Services. Department of Health.

Chung, D.T., Drábek, J., Kerry, L., Opel, K.L., Butler, J.M. and McCord, B.R. 2004. A study on the effects of degradation and template concentration on the amplification efficiency of the STR miniplex primer sets. *Journal of Forensic Sciences* **49**(4):733-740.

Cifuentes, L.O., Martinez, E.H., Acuna, M.P. and Jonquera, H.G. 2006. Probability of exclusion in paternity testing: time to reassess. *Journal of Forensic Science* **51**(2):349-350.

Coble, M.D. and Butler, J.M. 2005. Characterization of new miniSTR loci to aid analysis of degraded DNA. *Journal of Forensic Sciences* **50**(1):43-53.

Cloete, K., Ehrenreich, L., D'Amato, M.E., Leat, N., Davison, S. and Benjeddou, M. 2010. Analysis of seventeen Y-chromosome STR loci in the Cape Muslim population of South Africa. *Legal Medicine* **12**(1):42-45.

© University of Venda

University of Venda
Creating Future Leaders

Cregan, B., Mudge, J., Fickus, E.W., Marek, L.F., Danesh, D., Denny, R., Shoemaker, R.C., Matthews, B.F., Jarvik, T. and Young, N.D. 1999. Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. *Theoretical and Applied Genetics* **98**:919-928.

Datta, K. 2007. 'In the eyes of a child, a father is everything': changing constructions of fatherhood in urban Botswana? *Women's Studies International Forum* **30**(2):97-113.

Dawson, D.A., Ball, A.D., Spurgin, L.G., Martín-Gálvez, D., Stewart, I.R.K., Horsburgh, G.J., Potter, J., Molina-Morales, M., Bicknell, A.W.J., Preston, S.A.J., Ekblom, R., Slate, J. and Burke, T. 2013. High-utility conserved avian microsatellite markers enable parentage and population studies across a wide range of species. *BMC Genomics* **14**:176.

Dograr, N. and Akkaya, M.S. 2001. Optimization of PCR amplification of wheat simple sequence repeat DNA markers. *Turkish Journal Biology* **25**:153-158.

Durosinmi, M.A. and Alabi, O.A. 1995. Disputed paternity: experience in Ile-Ife, Nigeria. *West Africa Medical Journal* **14**(2):88-90.

Eddy, G., 2009. "Unhappy families". Fast Facts: *South African Institute of Race Relations* **7**:2-12.

Edwards, A., Civitello, A., Hammond, H.A. and Caskey, C.T. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics* **49**(4):746-756.

Edwards, A., Hammond, H.A., Jin, L., Caskey, C.T. and Chakraborty, R. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* **12**(2):241-253.

Ejele, O.A. and Nwache, C.A. 2004. Determinants of paternity disputes in the Niger delta region of Nigeria. *Nigeria Post-grad Medical Journal* **11**:187-189.

© University of Venda

University of Venda
Creating Future Leaders

FBI Laboratory. 2003. Federal Bureau of Investigation, Quantico, Virginia, 2004.

Feigelson, H., Rodriguez, C., Robertson, A., Jacobs, E., Calle, E., Reid, Y. and Thun M.J. 2001. Determinants of DNA yield and quality from buccal cell samples collected with mouthwash. *Cancer Epidemiology, Biomarkers & Prevention* **10**(9):1005-1008.

Forbes, K.M., Rahman, N., McCrae, S. and Reeves, I. 2008. Integrated community-based sexual health services for young people in urban areas: are we meeting the needs of the local community? *International Journal of STD & AIDS* **19**(10):713-714.

Frégeau, C.J. and Fourney, R.M. 1993. DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *Journal of Biological Techniques* **15**(1):100-119.

Garcia-Closas, M., Egan, K., Abruzzo, J., Newcomb, P., Titus- Ernstoff, L., Franklin, T. Bender P.K., Beck J.C., Le Marchand, L., Lum, A., Alavanja, M., Hayes, R.B., Rutter, J., Buetow, K., Brinton, L.A. and Rothman, N. 2001. Collection of Genomic DNA from Adults in Epidemiological Studies by Buccal Cytobrush and Mouthwash. *Cancer Epidemiology, Biomarkers & Prevention* **10**(6):687-696.

Glenn, T.C. and Schable, N.A. 2005. Isolating microsatellite DNA loci. *Methods in Enzymology* **395**:202-222.

Goldstein, D.B. and Schlotterer, C. 1999. "Microsatellites: Evolution and Applications". Oxford University Press, Oxford. UK.

Hares, D.R. 2012. Expanding the CODIS Core Loci in the United States. *Forensic Science International: Genetics* **6**(1):e52-e54.

Haskett, A. 2011. Why Are Single Parent Families On The Increase?. Retrieved October 30th 2013, from http://ezinearticles.com/?Why-Are-Single-Parent--Families-On-The-Increase?&id=6756875.

© University of Venda

University of Venda
Creating Future Leaders

Hedrick, P.W. 1999. Perspective: highly variable loci and their interpretation in evolution and conservation. *International Journal of Organic Evolution* **53**(2):313-318.

Henegariu, O., Heerema, N.A., Dlouhy, S.R., Vance, G.H. and Vogt, P.H. 1997. Multiplex PCR: Critical Parameters and Step-by-Step Protocol. *Bio-Techniques* **23**:504-511.

Human Genome Project and Beyond. 2008. http://www.ornl.gov/hgmis/publicat/primer U.S. Department of Energy Genome Research Programs, visited 09.10.2013.

Human Sciences Research Council. 2007. Youth Policy Initiative Roundtable 4: Learner Retention. Reserve Bank, Pretoria.

Hunter, M. 2006. Fathers without amandla: Zulu-speaking men and fatherhood. In: L. Richter and R. Morrell, (eds). *Baba: men and fatherhood in South Africa. Cape Town*: HSRC Press, 99-107.

Jeffreys, A.J., Wilson, V., Thein, S.L., Weatherall, D.J. and Ponder, B.A.J. 1986. DNA "Fingerprints" and Segregation Analysis of Multiple Markers in Human Pedigrees. *American Journal of Human Genetics* **39**:11-24.

Jewkes, R., Morrell, R. and Christofides, N. 2009. Empowering teenagers to prevent pregnancy: lessons from South Africa. *Culture, Health & Sexuality* **11**:675-688.

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

Kaufman, C.E., Dewet, T. and Stadler, J. 2001. Adolescent pregnancy and parenthood in South Africa. *Studies in Family Planning* **32**:147-160.

Kimpton, C.P., Fisher, D., Watson, S., Adams, M., Urquhart, A., Lygo, J. and Gill, P. 1994. Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. *International Journal of Legal Medicine* **106**:302-311.

© University of Venda

University of Venda
Creating Future Leaders

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

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. 1996. Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification. *Electrophoresis* **17**:1283-1293.

Knapik, E.W., Goodman, A., Ekker, M., Chevrette, M., Delgado, J., Neuhauss, S., Shimoda, N., Driever, W., Fishman, M.C. and Jacob, H.J. 1998. A microsatellite genetic linkage map for zebrafish (*Danio rerio*). *Nature Genetics* **18**:338-343.

Konovalov, D.A. and Heg, D. 2008. Technical advances: A maximum-likelihood relatedness estimator allowing for negative relatedness values. *Molecular Ecology Resources* **8**(2):256-263.

Kotze, M.J., Scholtz, C.L. and Opperman, P. 2006. Genetics in Family Practice: Health implications and counseling for paternity testing. *South African Family Practice* **48**(1):34.

Kruglyak, L. and Nickerson, D.A. 2001. Variation is the spice of life. *Nature genetics* **27**(3):234-236.

Lithole, A. 2011. The feasibility of forensic science in the Vhembe district of Limpopo province in South Africa. Honours thesis, University of Venda.

Liu, Z.W., Biyashev, R.M. and Saghai-Maroof, M.A. 1996. Development of simple sequence repeat markers and their integration into a barley linkage map. *Theory of Applied Genetics* **93**:869-876.

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

University of Venda
Creating Future Leaders

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., Leblond, S., Earle Macdonald, J., De Jong, J. and Wieringa, B. 1992. Myotonic dystrophy mutation: An unstable CTG repeat in the 38 untranslated region of the gene. *Science* **255**:1253-1258.

Makiwane, M. and Kwizera, S. 2008. Youth and well-being: a South African case study. *Social Indicators Research* **91**:223-242.

Maluleke, T.X. 2010. Sexual risk behavior amongst young people in the Vhembe district of the Limpopo province. *South Africa 'health SA Gesondheid* **15**(1):505-507.

Masi, P., Spagnoletti Zeuli, P.L. and Donini, P. 2003. Development and analysis of multiplex microsatellite markers sets in common bean (*Phaseolus vulgaris L.*). *Molecular Breeding* **11**:303-313.

Meglécz, E., Petenian, F., Danchin, E., Coeur D'Acier, A., Rasplus, J-Y. and Faure, E. 2004. High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Molecular Ecology* **13**:1693-1700.

Meintjes-Van der Walt, L. 2008. An overview of the use of DNA evidence in South African criminal courts. *South African Journal of Criminal Justice* **21**(1):22-62.

Nduna, M. and Jewkes, R. 2012. Denied and disputed paternity in teenage pregnancy: topical structural analysis of case studies of young women from the Eastern Cape Province, Social Dynamics. *A Journal of African studies* **38**(2):314-330.

Nduna, M. and Maseko, V. 2008. Discontinued intimacy, denied paternity. *Sexuality in Africa Magazine* **4**:11-14.

Nève, G. and Meglécz, E. 2000. Microsatellite frequency in different taxa. *Trends in Ecology and Evolution* **15**:376-378.

Nxumalo, N.E. 2000. *The status and role of minority African Languages in South Africa's new and democratic language policy*. University of the North.

© University of Venda

University of Venda
Creating Future Leaders

O'Donnell, W.T. and Warren, S.T. 2002. A decade of molecular studies of fragile X syndrome. *Annual Review of Neurosciences* **25**:315-338.

Ogliari, J.B., Boscariol, R.L. and Camargo, L.E.A. 2000. Optimization of PCR amplification of maize microsatellite loci. *Genetics and Molecular Biology* **23**(2):393-398.

Pearson, B. 2003. 'A matter of opinion': unofficial paternity tests and the impacts on children. Child Support Analysis. UK: submission to the Human Genetics Commission.

Pourcel, C., Hormigos, K., Onteniente, L., Sakwinska, O., Deurenburge, R.H. and Vergnaud, G. 2009. Improved Multiple-Locus Variable-Number Tandem-Repeat Assay for *Staphylococcus aureus* Genotyping, Providing a Highly informative Technique Together with Strong Phylogenetic Value. *Journal of Clinical Microbiology* **47**(10):3121-3128

Rammala, J.R. 2003. Draft language policy for the University of the North. Unpublished Work.

Robinson, J., Waller, M.J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L.J., Stoehr, P. and Marsh, S.G.E. 2003. "IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex". *Nucleic Acids Research* **31**(1):311-314.

Rolf, B., Keil, W., Brinkmann, B., Roewer, L. and Fimmers, R. 2001. Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations. *International Journal of Legal Medicine* **115**(1):12-15.

Roux, K.H. 2009. Optimization and Troubleshooting in PCR. *Cold Spring Harbor Protocols* **4**(4):1-6.

© University of Venda

University of Venda
Creating Future Leaders

Rozen, S. and Skaletsky, H.J. 2000. Primer3 on the www for general users and for biologists programmers. In: Bioinformatics Methods and protocols: Methods in Molecular Biology (eds krawetz S, Miseners S), pp. 365-386. Human Press, Totowa, New Jersey.

Saad, R. 2005. Discovery, development and current applications of DNA testing. *BUMC Proceedings* **18**(2):130-133.

Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**:365-371.

Schwarz, H.P. and Dorner, F. 2003. Historical review. *British Journal of Haematology* **121**(4):556-565.

Selkoe, K.A. and Toonen, R.J. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**(5):615-629.

Stallings, R.L. 1994. Distribution of trinucleotide microsatellites in different categories of mammalian genomic sequence: Implications for human genetic diseases. *Genomics* **21**(1):116-121.

Statistics South Africa. 2008. Statistical Release Community Survey 2007., Pretoria: Statistics SA.

Terasaki, P., Chia, D. and Sugich, L. 1998. Saliva as DNA source for HLA typing. *Human Immunology* **59**:597-598.

Todesco, L., Torok, M., Krahenbuhl, S., and Wenk, M. 2003. Determination of 3858G –› A and 164C –› A genetic polymorphisms of CYP1A2 in blood and saliva by rapid allelic discrimination: large difference in the prevalence of the 3858G-› A mutation between Caucasians and Asians. *European Journal of Clinical Pharmacology* **59**:343-346.

Tóth, G., Gáspári, Z. and Jurka, J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research* **10**(7):967-981.

© University of Venda

University of Venda
Creating Future Leaders

Walker, R.H. and Pohl, B.A. 1989. Paternity testing with an absent mother: the probability of exclusion of red cell surface antigen, Gm, Hp, and HLA systems in North American whites and blacks. *Transfusion* **29**(1):31-35.

Webster, M. and Reichart, L. 2004. Use of microsatellites for parentage and kinship analyses in animals. In: *Methods in enzymology 395, part B, molecular evolution*. (ed. E. A. Zimmer and E. Roalson), pp. 222-238. *Elsevier Academic Press*, San Diego, USA.

Williams, T.M. 2001. Human leukocyte antigen gene polymorphism and the histocompatibility laboratory. *Journal of Molecular Diagnostics* **3**(3):98-104.

Wright, J.M. and Bentzen, P. 1994. Microsatellites: genetic markers for the future. *Reviews in Fish Biology and Fisheries* **4**:384-388.

Zane, L., Bargelloni, L. and Patarnello, T. 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**:1-16.

UNIVEN LIBRARY
Library Item : 20141259

© University of Venda

University of Venda
Creating Future Leaders

APPENDIX

Appendix A

UNIVERSITY OF VENDA

CONSENT TO PARTICIPATE IN RESEARCH

Research Topic: Isolation and characterization of microsatellite markers for human identification in Vhembe district, Limpopo province

---

Principal investigators: Dr T. C Nangammbi, Prof A. Samie, Mr. L.F. Chauke and Dzhivhuho G.A

You are invited to participate in a research study conducted by Godfrey Dzhivhuho, who is a Masters student from the Zoology Department in the University of Venda. Mr. Dzhihvuho is conducting a study on microsatellites isolations for human identification under the supervision of Dr. Nangammbi T.C, Prof Samie A and Mr. Chauke F.L. This study is funded by the National Research Foundation (NRF) and UNIVEN RPC.

Your participation in this study is entirely voluntary. You should read the information below and ask questions about anything you do not understand, before deciding whether or not to participate.

- **PURPOSE OF THE STUDY**

The purpose of this study is to develop a standard operation procedure for microsatellites isolation that can be used for human identification purposes. These markers will be useful in DNA fingerprinting studies, desirably paternity and kinship analysis. They will also be used to define the population structure of three ethnic groups in Limpopo province, Vhavenda, Tsonga and Bapedi.

UNIVERSITY OF VENDA
LIBRARY

© University of Venda

- **PROCEDURES**

If you volunteer to participate in this study, we will ask you to do the following:

1. Fill a questionnaire that will give us your general physical description (Age, gender, height, weight, nationality, mother tongue, area of residence and brief family history.
2. Provide a DNA sample in form of a mouth wash sample.

- **POTENTIAL RISKS AND DISCOMFORTS**

We expect that any risks, discomforts, or inconveniences will be minor and we believe that they are not likely to happen. If discomforts become a problem, you may discontinue your participation.

- **POTENTIAL BENEFITS TO SUBJECTS AND/OR TO SOCIETY**

It is not likely that you will benefit directly from participation in this study, but the research should help us learn how to develop a microsatellite genomic library that is essential for DNA fingerprinting. This will allow us to provide forensic services in the Limpopo province in the near future. The services will be provided in the University of Venda for local hospitals and local community at large.

- **COMPENSATION FOR PARTICIPATION**

You will not receive any payment or other compensation for participation in this study. There is also no cost to you for participation.

- **CONFIDENTIALITY**

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. Confidentiality will be maintained by means of a code number known by the researcher. We will not use your name in any of the information we get from this study or in any of the research reports.

© University of Venda

- **PARTICIPATION AND WITHDRAWAL**

You can choose whether or not to be in this study. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you do not want to answer. There is no penalty if you withdraw from the study and you will not lose any benefits to which you are otherwise entitled. The investigator may withdraw you from this research if your physician tells us that continued participation may injure your health.

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

_____

Printed Name (optional) or ID code of Subject

Signature of Subject _____

Date _____

© University of Venda

University of Venda
Creating Future Leaders

Appendix B

Solution preparations

20 X SSC solution recipe for 1 litre volume:

Component	Amount	20X Stock Concentration	Final 1X Concentration
NaCl	175.3 g	3 M	150 mM
Na3CitrateX2H2O	88.2 g	300 mM	15 mM

Dissolve 175.3 g of NaCl and 88.2 g of Sodium Citrate in 800 ml water. Adjust the pH to 7.0 with a few drops of 14 N solution of HCl. Adjust the volume to 1 litre with ultrapure water. Dispense into aliquots. Sterilize by autoclaving.

Dilute to working solutions (12 X SSC using distilled H_2O) for preparation of hybridization and washing solutions

2 X Hybridization solution: 12 X SSC; 0.2% SDS

1 X Hybridization solution: 6 X SSC; 0.1% SDS

Wash solutions: 2 X SSC, 0.1% SDS and 1 X SSC, 0.1% SDS

UNIVERSITY OF VENDA
LIBRARY

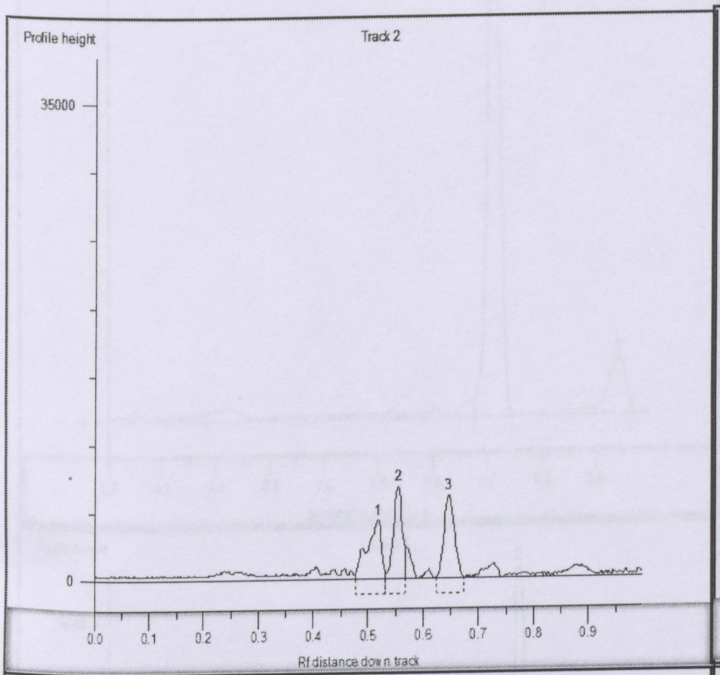
© University of Venda

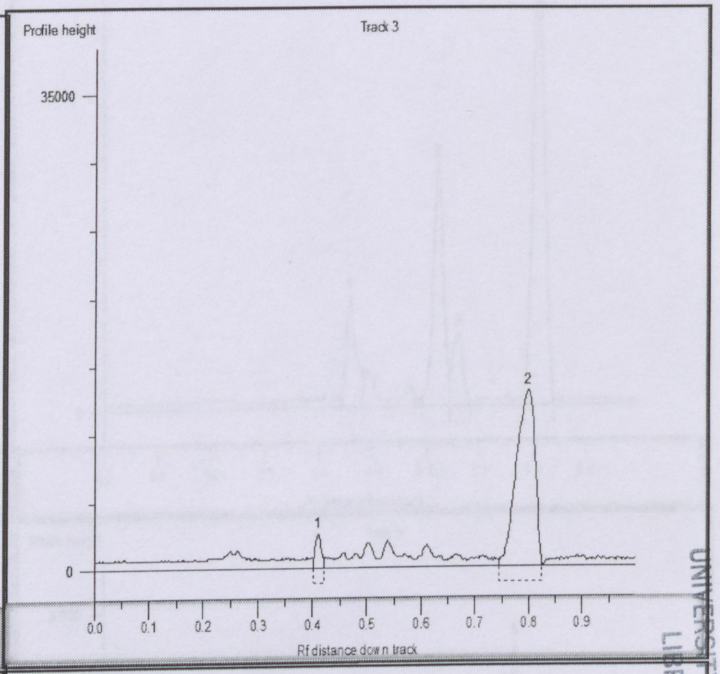
University of Venda
Creating Future Leaders

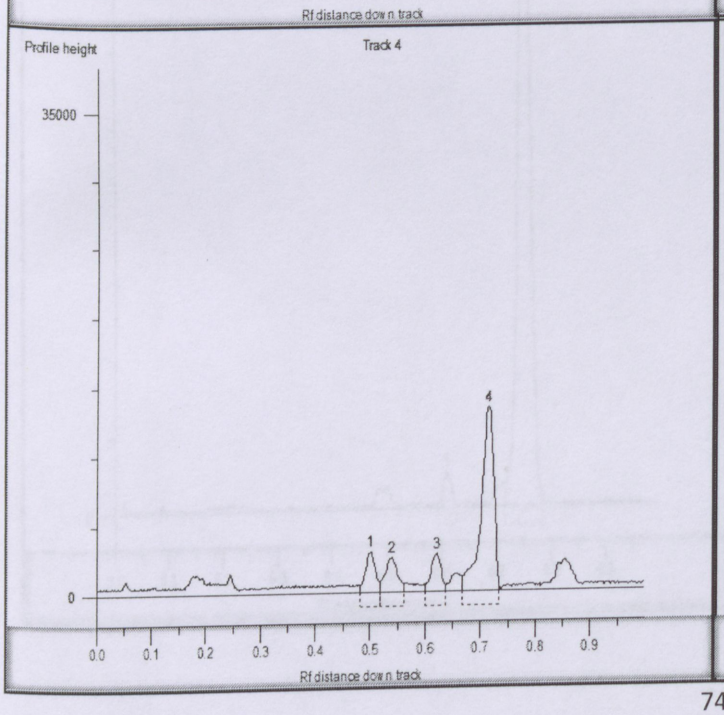
Appendix C

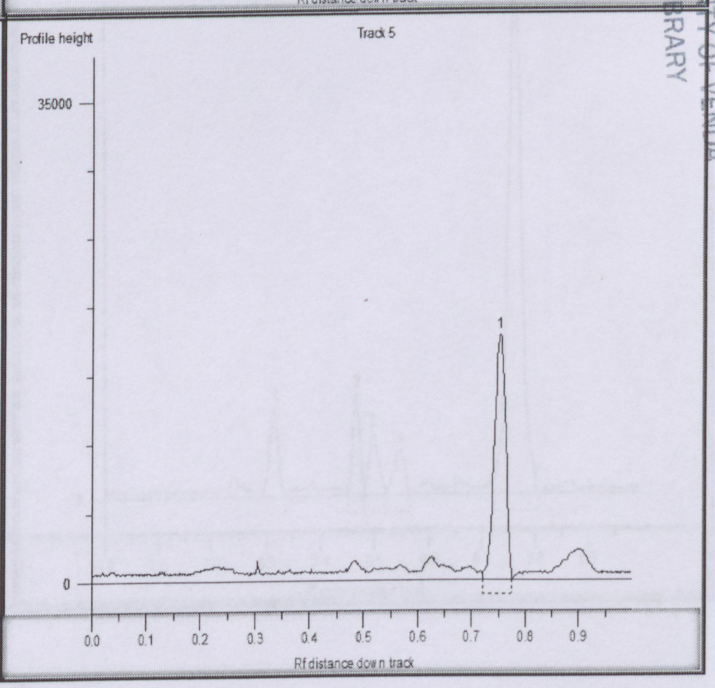
Peaks generated in GENETOOLS for 14 potential microsatellite markers

Track 2 through 15 represents Locus 1 through 14 respectively





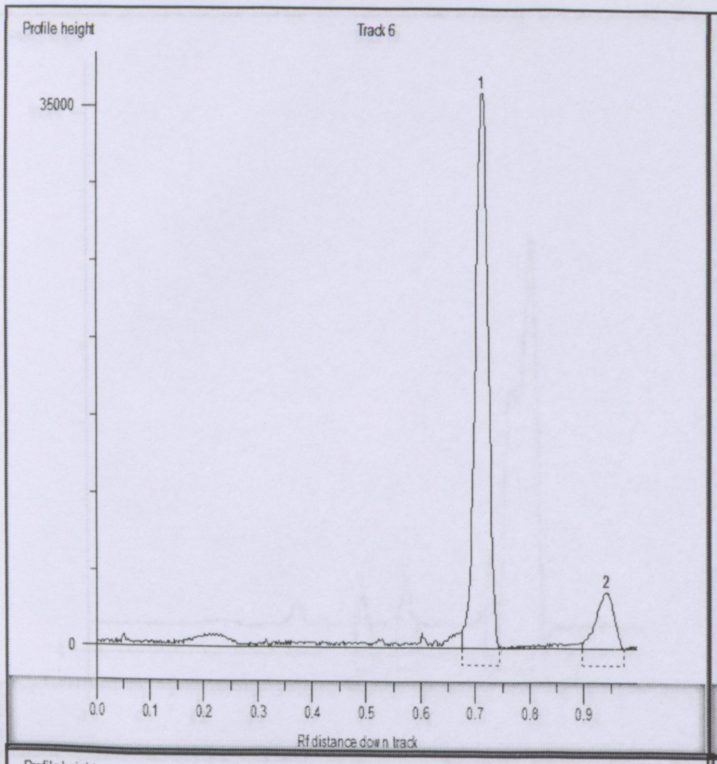


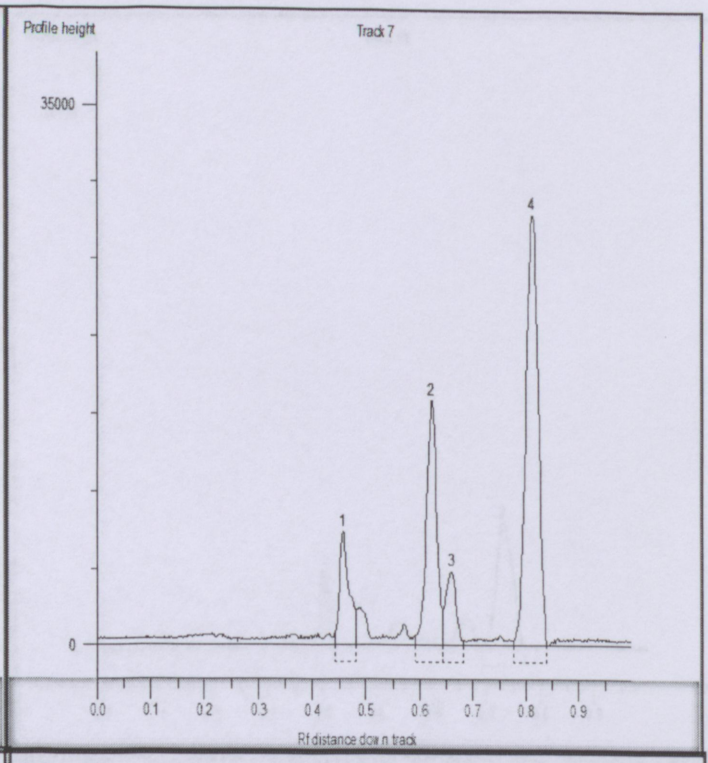


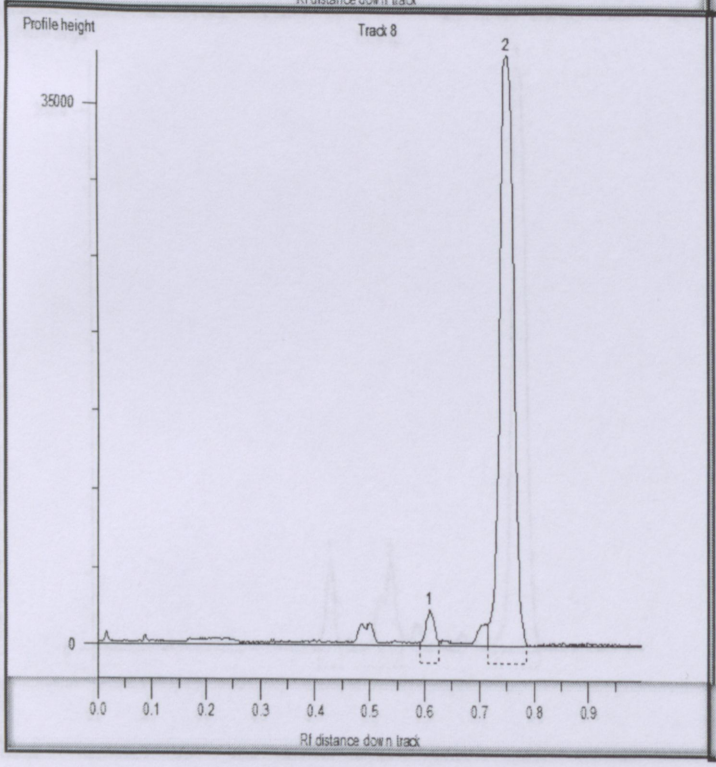
UNIVERSITY OF VENDA
LIBRARY

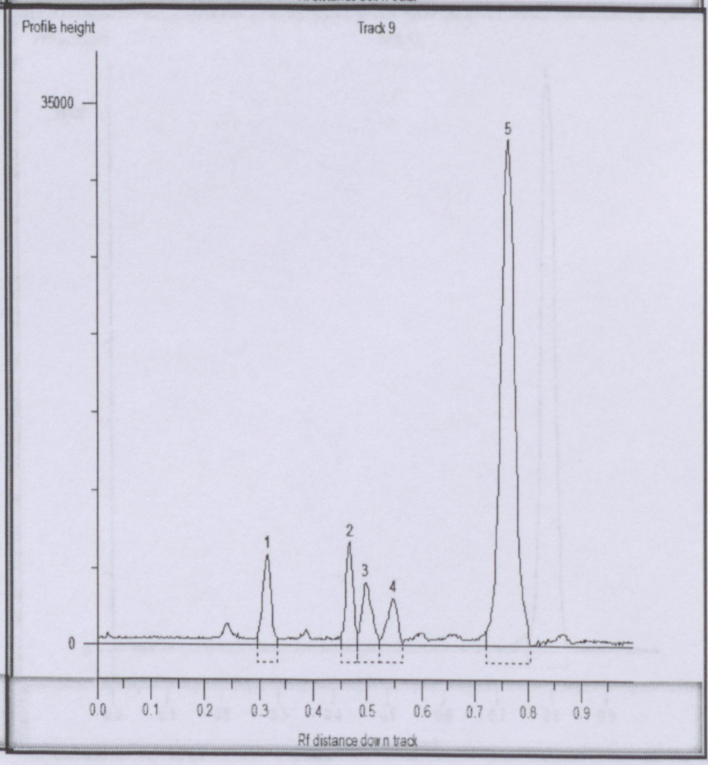
© University of Venda

University of Venda
Creating Future Leaders



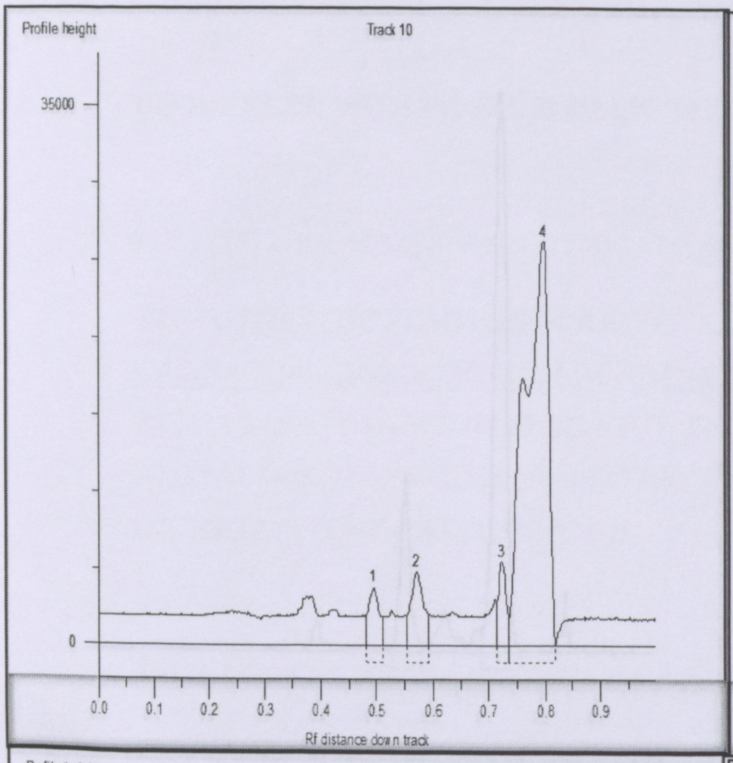


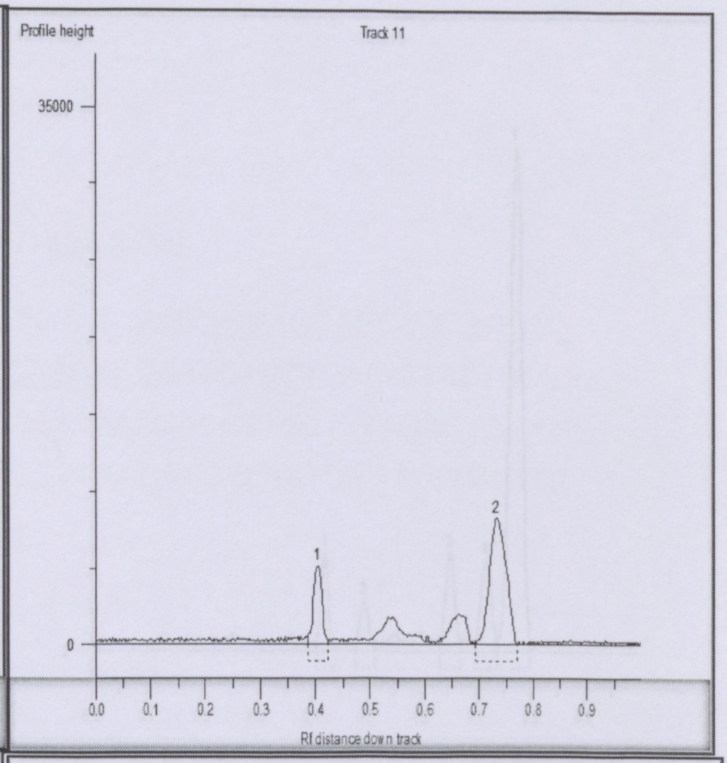


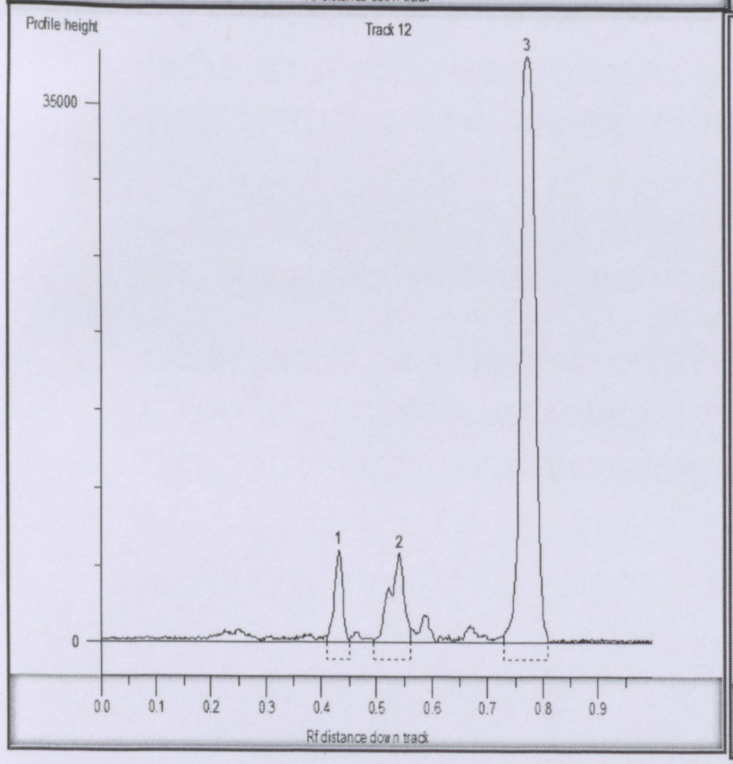


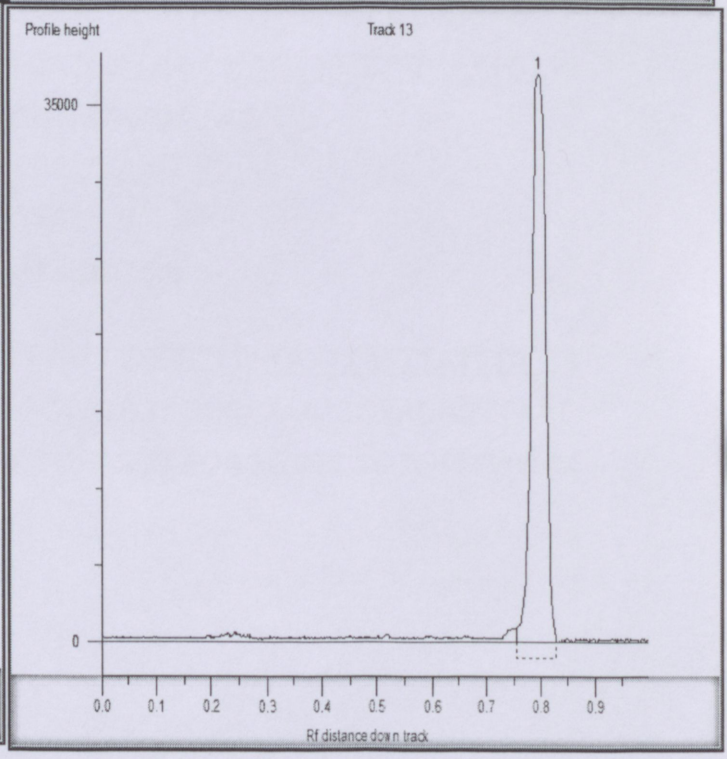
© University of Venda

University of Venda
Creating Future Leaders

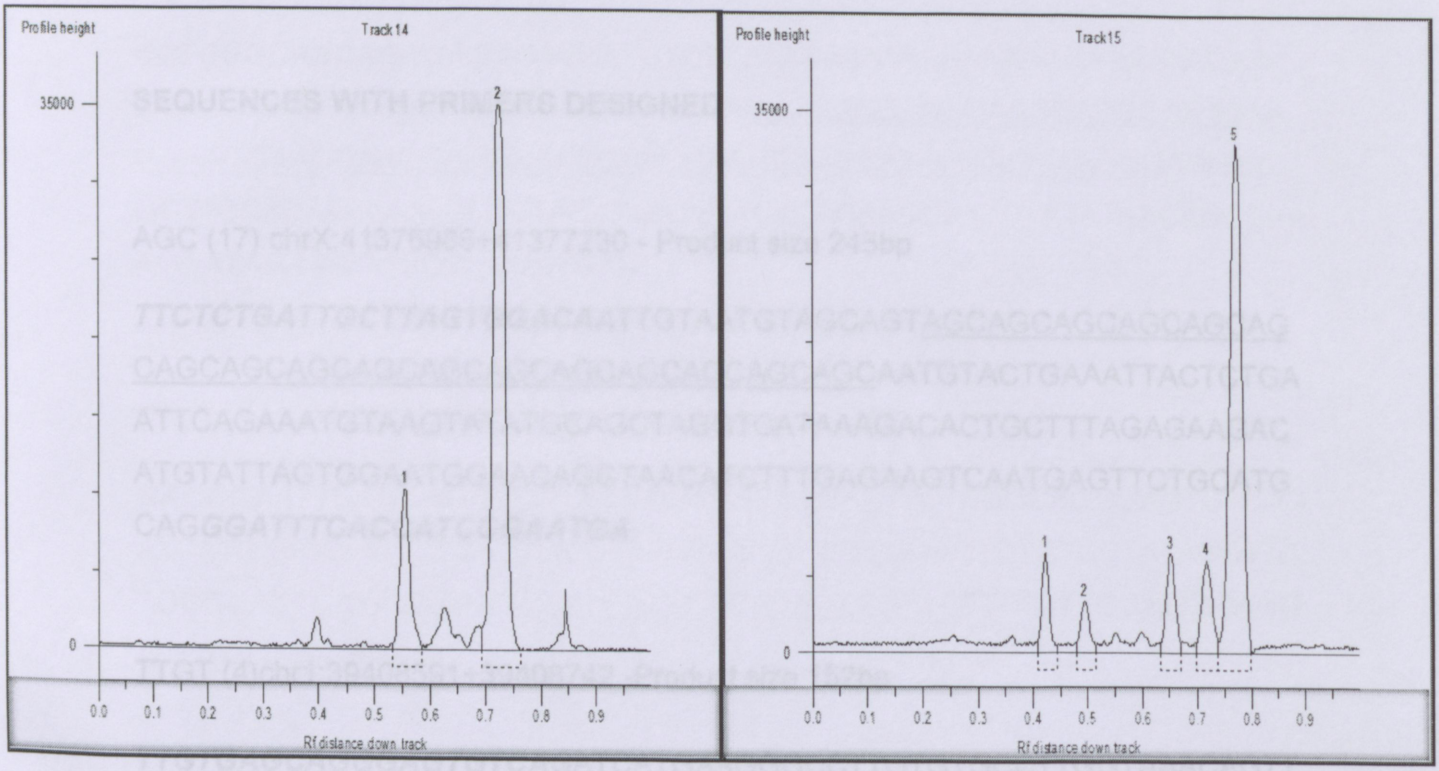








© University of Venda



© University of Venda

University of Venda
Creating Future Leaders

Appendix D

SEQUENCES WITH PRIMERS DESIGNED

AGC (17) chrX:41376986+41377230 - Product size 245bp

TTCTCTGATTGCTTAGTGGACAATTGTAATGTAGCAGTAGCAGCAGCAGCAGCAG
CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAATGTACTGAAATTACTCTGA
ATTCAGAAATGTAAGTATATGCAGCTAGGTCATAAAGACACTGCTTTAGAGAAGAC
ATGTATTAGTGGAATGGAACAGGTAACATCTTTGAGAAGTCAATGAGTTCTGCATG
CAG***GGATTTCACCATCGGAATGA***

TTGT (4)chr1:39408591+39408742 -Product size 152bp

TTGTGAGCAGGGAGTGTCAGATCATGAAGGGCCTTGTGTGCCTTGGTAAACAGTT
TGGACTGTGTCCTATAGGTTCTGGGGAGTCATAGAAAGATTGTTTGTTTGTTTGTTC
GTTTGTTTGTTGCAATACAG***GGTCTCTCTGGGAGGCAGAG***

TATT (4)chr3:132874869+132875043 -Product size175bp

AAGGTGAGGGTCCAGTTTCAATCTTGTTTTTATTTTTTCTTTTATTTATTTATTTATTT
ATTTATTCTTTTTGGTAGGGATAAGGTTTTGCTATGTTGCCCAGGCCTACACTTAAT
TTTTCTAGTTAACTTCTTTCTCCTGAGATACTCAAAGT***ACAGGCTTGGTGTTCAGAA
AA***

© University of Venda

GGAA(17.5)chr4:8068005+8068244 - Product size240bp

CCTGGGCAACAGAGAGAGACTATGTCTCAAAAAAAAAAAAAAAAGAAAAAAGAAA
GGAGGGAGGGAGAGAGGGAGGGAGGGAGGGAAGGAAGGAAGGAAGGAAGGAA
GGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGGAGGGGT
GGTTCCACATAAAAACCTGTACATGATGTTCACAGAAGCATTCTATATAACAGCGAA
AGGA***TAGAAACAGCCCCGGTGTC***

TTGT (4.25)chr1:39408556+39408737 - Product size182bp

GACGGAATTCAGCAGGAACTACTCAAAGTTCAGAATTGTGAGCAGGGAGTGTCAG
ATCATGAAGGGCCTTGTGTGCCTTGGTAAACAGTTTGGACTGTGTCCTATAGGTTC
TGGGGAGTCATAGAAAGATTGTTTGTTTGTTTGTTCGTTTGTTTGTTGCaa***TACAGG***
GTCTCTCTGGGAGG

AAG (7)chr3:46750995+46751155 - product size 161bp

CTTCCCCTCAGGACAGTCAGACCCCAGGACCTTGTCTCACCACTATCACATGGTC
TCTTCCCCCTGCCCCACAGAGGATAAGAAGAAGAAGAAGAAGAAGAAGAAGGACA
GTGTGGACACAGTGGCCATCAAAGTAGAGGA***GGATGAGAAGAATGAGGCCA***

GAA(3)chr6:44636291-44636519 -Product size229bp

TAGCTCTCCAGCCCTCTCTGCCACTCCCTCCCCAGCAGGTCTCATCTGGGCACAG
GAGACCGCTTGGCCTCAAGCTAAGATGTAGTTTCAAGACTGAGCCCGTGGGCCAG
GACTCTGAGAGCAGTGGGCCGTGGATGCTGCCAGTTCTGTCTTCCGGAAGAAGAA
AAATGAGTTAGAGGTTGGGAAGATGTTTCTGGGGCCCAGGAGGC***AGTCAGCTCCT***
GCTTGGAAG

© University of Venda

GCT(10)chr5:153265744-153265952 -Product size209bp

CTATCTGGCCCCAAATCAGAAACTTTGTAAAGGTTGGAGGTGGGATTCAAACCAG
GCAGTCGGTTCCCTTAAGCTCCAGTAGCCTTGAGAAGCACTTGTCTCAGCTTCTGA
CAAATAGTAAAACTTGATGAAACTGGGTTGGGCAGATGCTGCTGCTGCTGCTGCTG
CTGCTGCTGCTACTGCTGCTATT***TACCACGTGAGGTTGCCATA***

TGG(5)chr17:78287063+78287354-Product size 292bp

GTTGTTGGCCTTCACCTGAGCCTCAGTTCTTCATCTACAAAATGGAGTTGCTCCTG
TCTACCTGGCAGGGTGGCTCTGAGGCAGACAGATGCCGGGTGCAAAGTGCTGCT
GCACTGCCTCGTGTCCCCTGGCCGGGCACAGTCCCTGTGGTGATGGAGGTGTGG
AGGTGATGGTGATGGTTGGTGGTGGTGGTGGTAGAGGTGATGGTGGTGGAGGTG
ATGGTGATGGTTGGTGGTGGTGGTGGTAGAGGTGATGGTGGTGAAGGTGATGGT
GATGGTGGACGTAATGGAGG

AAC(6.6)chr7:50907160+50907368 -Product size209bp

GGCGAAAATGACACACACACAAAAACCAGATGACAACAACAACAACAACAAATAA
GACAACTCTTTCGCCGTAAGTCAAGCAGAACATTCCAGCACCATGGTGTGGGGAA
GATGAATCCCATGGAAGAACGAGCAGGTCAGGGAGAACCTTCCTCCAAAGCAGCT
GCTCTCAGGATGCATCCTTGTGAT***GAGTTTTGTGGGACCTGCAT***

AC(11)chrX:63037800+63038035 -Product size236bp

GGCCCGTAATCCAATATGACTGGTGTCCTTATAAAAAGGGGAAACTTAAACACAGA
CACAGATATACACACACACACATACACACACACACACACACACGAAGAACACCATG
TGAAGATGAAGGCAGAGATCTAGGTGATGCAGCAGAATACAAGGAATGCCCAAGA
TTTCCAGCAAACCACCAGAAGCTAGGAGAAAGCCATGTAACAGATACTC***CCTCACA***
TCCCTCAGCAAAT

© University of Venda

GT(13.5)chr2:73870810+73870990 -Product size181bp

GGGAAACTTCCCTTCCAAAGATTAGCTAACCTATTGATGATCAGCCACGCTGGGC
CCGTCATCTAGGTTAGGTAGTTTGGACTCTATTCCAAGAGCAAGGGGAGGACATTG
ATATGCACGGAGGGGAGGAGGTGTGTGTGTGTGTGTGTGTGCGCACGCAAG***CAT***
CACTCTGGCTGCAGGT

AAT(4)chr8:6723383+6723589 Product size207bp

AGAAGGAGGTTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGTGACA
GAGCAAGACTCCATCTCAAAAAACAAAAACAGAAAAAAACAAAAATAATAATAATCA
AACACACAGCAGGACCCCTGGCCAGCTGGGGAGGGGGCCAGTGTGGGTGCGTCT
TCCTGGGGCCAGCCTAGGAGAA***AGTGCTTTTCCTGGGTGTGT***

AATG(5)chr8:6768702+6768860- Product size159bp

AGCACAGCAGAGATGCTCAAGACATATTTGTTGAAGAAATGAATGAATGAATGAAT
GGATGCAGATTCACTTACGCAGCAATCCCCTGTCAGGTGAGAAGGTAGAGAGTGC
TGCTCTAATCACTTCTCTCTTTACCTC***TCTGCCTTACCAAATTCCTCC***

© University of Venda